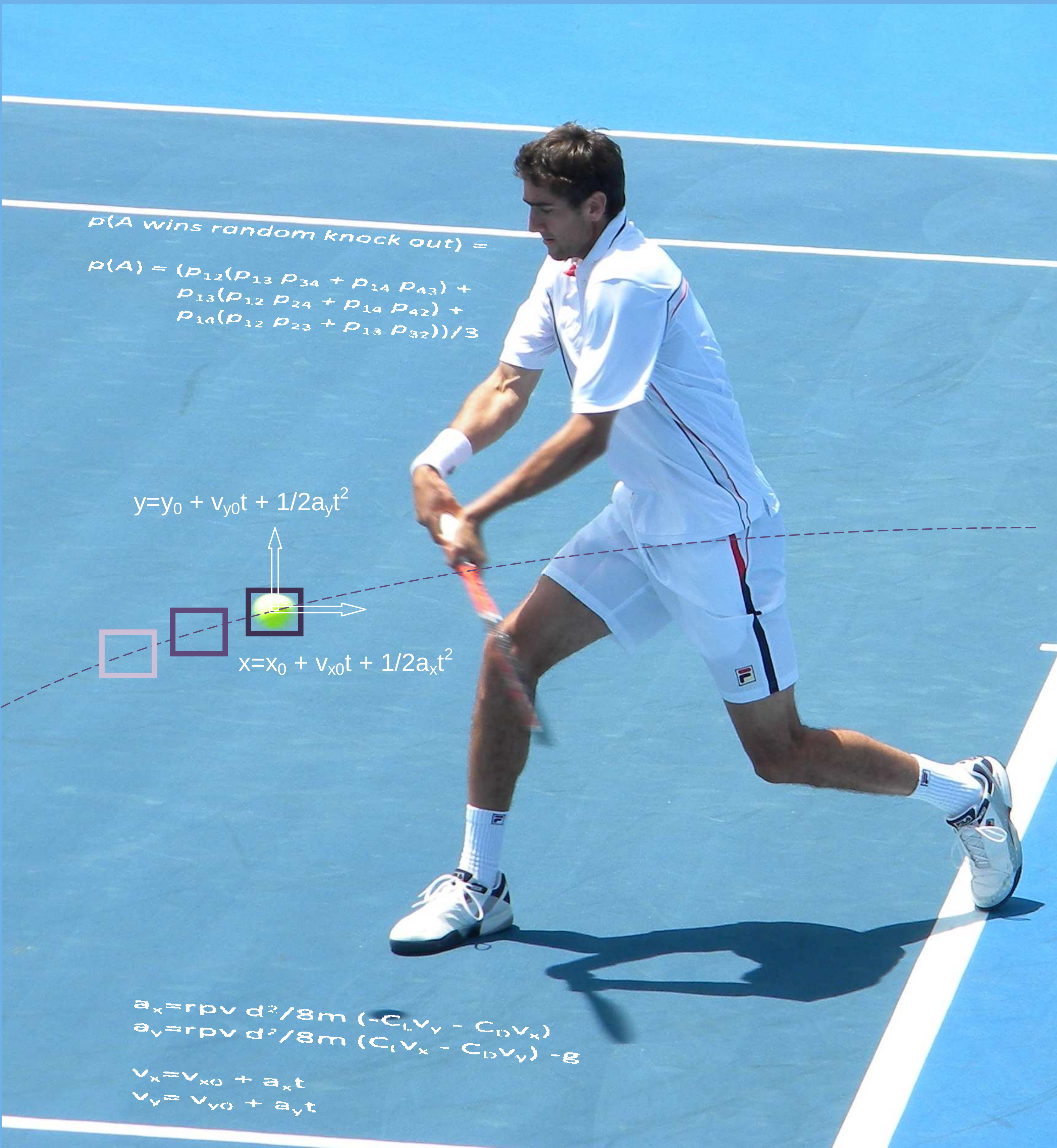
PROCEEDINGS OF THE TENTH AUSTRALASIAN CONFERENCE ON

MATHEMATICS AND COMPUTERS IN SPORT



July 5 – 7, 2010. Darwin, Northern Territory, Australia

Edited by A Bedford and M Ovens

PROCEEDINGS OF THE TENTH AUSTRALASIAN CONFERENCE

ON

MATHEMATICS AND COMPUTERS IN SPORT

edited by
Anthony Bedford and Matthew Ovens
RMIT University

10M&CS

July 5th to 7th 2010

Darwin, Northern Territory, Australia

Published by MathSport (ANZIAM)

ISBN: 978-0-9578623-5-7

The full papers in these proceedings have gone through a full peer review refereeing process.

Cover image: Marin Čilić in action at the 2010 Australian Open
Original photo taken by Matthew Ovens, additional graphics by Anthony Bedford

TABLE OF CONTENTS

Conference Director's Report vi

Keynote Address: **Ray Stefani** *A World of Sports and Rating Systems* .... 1 – 12

Timothy Chartier, Erich Kreutzer, Amy Langville, **Kathryn Pedings** and Yoshitsugu Yamamoto *Minimum Violations Sports Ranking Using Evolutionary Optimization and Binary Integer Linear Program Approaches* .... 13 – 20

Anthony Bedford, **Tristan Barnett** and Monique Ladds *Risk Taking in Badminton to Optimize In-The-Run Performance* .... 21 – 26

Monique Ladds and Anthony Bedford *Using Importance to Determine a Service Strategy in Badminton* .... 27 – 34

Graham Pollard and Geoff Pollard *The Efficiency Of Scoring Systems With Outcome Dependent Points* .... 35 – 42

Stephen R. Clarke and John M. Norman *An Introductory Analysis of Challenges in Tennis* .... 43 – 48

Graham Pollard, Geoff Pollard and Denny Meyer *Some Aspects of Ordering, Ranking and Seeding 1* .... 49 – 56

Denny Meyer, Graham Pollard and Geoff Pollard *Some Aspects of Ordering, Ranking and Seeding 2* .... 57 – 64

Allison O'Riordan and **Chris Barnes** *A Level Playing Field for AWD Competition: Choosing the Paralympian of the Decade*
(Abstract Only) .... 65 – 66

Peter Christ, Jörg Mielebacher, Martin Haag, and Ulrich Rückert *Detection of Body Movement and Measurement of Physiological Stress with a Mobile Chest Module in Obesity Prevention* .... 67 – 74

T. Churchill, D. Sharma, and B. M. Balachandran *Identifying Taper Shape and its Effect on Performance* .... 75 – 82

M. H. Korayem, S. A. Mustafa, A. H. Korayem, and A. Amanati *Modeling of the Snatch Weightlifting* .... 83 – 88

Keynote Address: **John Hammond** *A 'Potted History' of Mathematics and Computers in Sport: in Australasia (1992-2008)* .... 89 – 96

Jonathan Sargent and Anthony Bedford *Long-Distance Relationships: Positional Classification of Australian Football League Players*.. 97 – 102

Adrian Schembri and Anthony Bedford *Revisiting the Draft: A Revised Probability Based Approach for Allocating Draft Selections in Australian Rules Football* ... 103 – 112

Darren O'Shaughnessy *On the Value of AFL Player Draft Picks*.................... 113 – 120

Richard Ryall and Anthony Bedford *Fitting Probability Distributions to Real-Time AFL Data for Match Prediction* .... 121 – 128

Kelly Foreman *A Conceptual Framework for Global Positioning Systems Within the Australian Football League: A Discussion of Two Models*................................ 129 – 136

Per Wilhelm, Patrick Thomas, Emad Monier, Robert Timmermann, Michael Dellnitz, Felix Werner and Ulrich Rückert *An Integrated Monitoring and Analysis System for Performance Data of Indoor Sport Activities* 137– 144

Joachim Gudmundsson and **Thomas Wolle** *Towards Automated Football Analysis: Algorithms and Data Structures* .... 145 – 152

Reza Kiani Mavi and Neda Kiani Mavi *Measuring Efficiency of Football Teams by Common Set of Weights in DEA* .... 153 – 162

Darren O'Shaughnessy *An Improved World Cup Tournament Design* (Abstract Only) .... 163 – 164

Adrian Schembri, Shahin Golkhandan and Anthony Bedford *Re-scheduling to Minimise the Potential for Match Fixing in the FIFA World Cup* .... 165 – 172

Michael Bailey, Stephen Clarke and Donald Forbes *Home Continent Advantage on the American and European Professional Golf Tours* .... 173 – 180

Bob Buckley and Conny Draper *Stroker: Software for Presenting Rowing Data* (Abstract Only) .... 181 – 182

Minh Vu Huynh and Anthony Bedford *Skills Acquisition in Badminton: A Visual Based Approach to Training*.. 183 – 188

Ian Tim Heazlewood *Talent ID in 100M Sprinting: How the World's Best Sprinters Solve the Problem Based on Mathematical Models and Multivariate Statistics* .... 189 – 196

Ian Tim Heazlewood and Hovik Keshishian *A Comparison of Classification Accuracy for Karate Ability Using Neural Networks and Discriminant Function Analysis Based on Physiological and Biomechanical Measures of Karate Athletes* .... 197 – 204

Joe Walsh *An Investigation into the Possibility of Theoretical Modelling for the Purposes of Examining the Non-Associative Nature of Probabilities of Game Based Skill Tasks Within Competitive Sports Matches* .... 205 – 212

Rebecca Shepherd, **Anthony Bedford** and Adrian Schembri *A Two-Stage Simulation to Predict Medalists in Pistol Shooting* .... 213 – 220

Graham Pollard, Geoff Pollard, Ian Lisle and Rod Cross *Bias in Sporting Match Statistics* .... 221 – 228

Sohail Akhtar and Philip Scarf *Modelling Match Outcomes in Test Cricket* (Abstract Only) .... 229 – 230

Steven Stern *What Will They Score? In-Game Projection of First Innings Final Scores in Limited-Overs Cricket* .... 231 – 238

Abid Ghaffar and Ehtesham Taha *Data Mining in Professional Cricket for Performance Enhancement* .... 239 – 244

M. H. Korayem, M. Bamdad and A. H. Korayem *Sport Simulation by Using Six Degree of Freedom Cable Suspended Robot* .... 245 – 254

Neville De Mestre *Tournament Roulette* .... 255 – 258

Joe Walsh *A Markov States Model for Touch Football* .... 259 – 267

CONFERENCE DIRECTOR'S REPORT

Welcome to the Northern Territory, Australia, for the tenth Australasian conference on Mathematics and Computers in Sport (10M&CS). This is the first time the conference has been hosted in Darwin, Australia's northern most capital city. Being winter, Darwin is the perfect setting for Australia's southern state delegates to escape the depths of winter, and experience some tropical warmth!

Darwin is one of Australia's fastest growing cities. A vibrant, multicultural and cosmopolitan city, Darwin is home to people from over 70 different ethnic backgrounds and 60 nationalities. Located on edge of the Timor Sea, it is renowned for its picturesque sunsets and tropical nights. The Northern Territory has a number of national parks and attractions, including Litchfield, Uluru, Kakadu and Katherine Gorge.

This year we welcome Professor Ray Stefani (California State University) as the day one keynote speaker, speaking on Ratings and Ranking. His topic introduces a strong theme throughout the conference, and we look forward to his presentation.

We also welcome Professor John Hammond (Southern Cross University) as the day two keynote speaker, who will share with us the history of MathSport. His timely presentation should spark much discussion on the future direction of the group, especially as many delegates have connections with sporting bodies such as state and national institutions of sport.

In this the tenth conference, we find no shortage of papers on a number of diverse topics. For example, we have papers from cricket, tennis, rugby, athletics, AFL, and football, to golf, roulette, badminton, weightlifting and karate; there is much to pique our interest, crossing many domains of sport science, mathematics, statistics and computing. As well as the strong Australasian presence, we welcome contributions from the USA, Iran, India, and England.

It is also notable that, for the first time, the Australasian nations of Australia and New Zealand are both represented at the FIFA World Cup, and like 8M&CS, we look forward to early morning semi-finals contests during the conference.

To end our conference, this year we introduce the (Emeritus Professor) Neville de Mestre (Bond University) Awards for Best Student Presentation and Best Student Publication.

All full papers in these proceedings have been peer refereed, and I thank all the reviewers for their swift returns given the tight time frame. For assistance in the organisation of 10M&CS, we are very grateful to co-editor Matthew Ovens (RMIT University) who has contributed a significant amount of effort in both compiling and reviewing papers, along with our outstanding scientific committee members. I also thank Monique Ladds (RMIT University) for her significant contribution in preparing those little things like emails and reviewers.

On behalf of Co-Director Associate Professor Tim Heazlewood (Charles Darwin University), and the MathSport executive, I welcome you to Darwin and hope you will enjoy the presentations, social outings, networking opportunities and wonderful sights of Darwin and the Northern Territory.

Dr Anthony Bedford
RMIT University
Conference Co-Director, 10M&CS.

A WORLD OF SPORTS AND RATING SYSTEMS

Raymond T. Stefani

California State University, Long Beach, USA
Corresponding author: Stefani@csulb.edu

Abstract

As the result of a three-year effort, a comprehensive, definitive and comparative survey is elucidated for internationally recognized federations, for the sports organized by those federations and for the rating systems employed by those federations to rate those sports. Mind sports and physical sports are both included. As of December 2009, 106 Olympic and non-Olympic sports are recognized by the IOC, an additional 25 sports are recognized by AGFIS and an additional 25 sports are organized by additional federations listed by Wikipedia under "List of International Sports Federations". Of the 156 sports, 17 are combat sports, in which opponents are in direct physical contact (as in boxing), 73 are independent sports in which contact is not allowed (as in swimming and archery) and 66 are object sports in which indirect contact is allowed while opponents attempt to control an object (as in basketball and chess). A few of the lesser known sports are the combat sports Wushu and Kabbadi; the independent sports Apnea and Dragon Boats and the object sports Underwater Rugby and Sepak Takraw. Of the 156 sports, 59 sports have no rating system, two sports have subjective rating systems, 82 sports have accumulative systems in which points accrue monotonically over some window of time, and 13 sports have adjustive systems in which the a rating adjusts in any direction by a factor times the difference between some result and a predictions of that result. For accumulative rating systems, features discussed including converting results to points, ageing results more than one year old, and adjusting points for performance quality. The adjustive systems include Elo, probit and averaging methods. This study provides reference material of general interest and trade-offs for rating system developers.

Keywords: Sports, rating systems, Olympics, international sports federations, ELO, probit

1. INTRODUCTION

This paper intends to offer a comprehensive overview of officially recognized international sports and the various rating systems for those same sports, as published by the sports federations which organize competition. It is well beyond the scope of this paper to survey previous contributions spanning the agenda. There are many excellent reference books and almanacs listing various types of sports, the history of each, rules and past results. Similarly, many specific types of rating systems have been offered in the literature from simple to complex, operating only on game score or applying regression to many game statistics. A selected list of important international sports ratings systems appears in Stefani (1997). In Stefani (1998), 83 widely played sports are selected based on International Olympic Committee recognition and on additional personally-selected sports that appeared to be widely played. A taxonomy of some of those sports is offered in Stefani (1999), using systems methods for categorization. In this paper, the number of analysed sports will be almost doubled. The 156 selected sports are those for which competition is organized by international federations, which are recognized by three definitive agencies.

The rest of this paper is organized as follows. Section 2 covers 156 sports of the world, providing a unifying taxonomy. It will be noted that many of the categorizations will be in sets of three. Section 3 provides a critical survey of the three basic types of ratings systems, provided by governing sports federations. This section is intended to be a resource for those wanting to create a rating system. The reader will have guidance into the basics of creating each type of system. Section 4 will provide conclusions.

2. SPORTS OF THE WORLD

To begin, what is a sport? I will generalize the usual dictionary definition; so as to be inclusive, not having to disqualify some widely recognized competitions. I define a sport as a competition using established rules for determining the winner. Sports so defined fall into two classes. In a mind sport, a surrogate (human, mechanical or computer) can make a play for the competitor as in chess and bridge. No physical action is need. For example, chess has been contested with moves sent via computer or by mail. In a physical sport, the competitor must make each play, requiring physical prowess as in running and soccer. What I call a physical sport follows the more widely used dictionary definition of a sport. It should be noted that the first Mind Sports Games were contested in Beijing following the Beijing Olympics for chess, bridge, draughts (checkers), Go and Xianggi (Chinese Checkers). I suggest that the world of sports has plenty of room for both mind and physical sports.

Categories of Sports

The *Official World Encyclopedia of Sports and Games* (1979) lists 15 categories of what are therein called games (mind sports here) and 13 categories of what are called sports therein (physical sports here). For example, some physical sport categories are court sports, team sports, water sports and stick and ball sports. Note that those categories are hardly mutually exclusive. For compactness henceforth, physical sports will not be distinguished from mind sports. I suggest with the certainty of physics, that there are three categories of sports following the three ways that two objects (competitors) can interact in three dimensional space without merging. First, in a combat sport competitors are in direct contact: the goal of a competitor is to control the opponent as in wrestling and boxing. Second, in an independent sport no significant contact between competitors is allowed as in running swimming and shooting: the goal is for the competitor to control his or her own self. Third, in an object sport the competitors interact indirectly as in soccer, chess and rugby: the goal is to control an object.

Each activity within each sport also falls into one of those three categories. Consider the corner kick in soccer. As the ball is being readied for play, players may jostle for position (a combat activity). As the ball is centred, some players try to control the ball (an object activity). The goalkeeper may not be interfered with in the six yard box (an independent activity).

Determining the Outcome

The outcome of each sport is determined by one of three methods: by subjective decision, by measurement or by scoring. The outcomes of most combat sports are determined subjectively; the outcomes of most independent sports are determined by scoring or direct measurement whilst the outcomes of most object sports are determined by scoring. The rating systems that follow are best selected taking into account the method by which the outcome of that sport is determined.

Recognized International Sports

As mentioned earlier, the purpose of this paper is to select sports with wide international recognition and to survey the rating systems published by the relevant sports federations. Three sources of recognition are chosen herein. It was necessary to visit the website of each recognized federation to locate all sports organized by each federation. It was necessary to distinguish a sport (swimming, athletics) from a discipline within that sport (butterfly, pole vault) and only count sports. First, the International Olympic Committee, IOC, is obviously a world leader in sport; however, the IOC uses non-standard terminology as to the term “sport” in that “sport” and “international sports federation” are used interchangeably. (Due to the very large number of federation abbreviations to follow, the reader is directed to the appropriate website for each definition.) For example, aquatics is considered to be one sport due to the one federation, FINA, which organizes competition in what the IOC calls four “disciplines”, swimming, diving, water polo and synchronized swimming, which in common terminology would be considered to be four sports. Henceforth, this paper will use the term “sports federation” for what the IOC calls a sport and will use the term “sport” for what the IOC calls a discipline. As of December 2009, the IOC website www.olympic.org recognized the numbers of sports federations and sports for Summer Olympic, Winter Olympic non-Olympic competitions as shown in

Recognition	Sports Federations	Sports	Combat	Independent	Object
IOC Summer	26	37	6	21	10
IOC Winter	7	15	0	13	2
IOC Recognized	34	54	3	23	28
Total IOC	67	106	9	57	40
AGFIS (additional)	21	25	6	10	9
Other (Wikipedia)	24	25	2	6	17
Total Additional	45	50	8	16	26
Total	112	156	17	73	66

Table 1: International Sports Federations and Sports

	Type of Sports Rating System				
Sport	Number	Accumulative	Adjustive	Subjective	None
Table 4 Combat Sports	17	3	1	2	11
Table 5 Independent Sports	73	51	3	0	19
Table 6 Object Sports	66	28	9	0	29
Total	156	82	13	2	59

Table 2: Types of Sports Rating Systems

Sport	Int. Fed.	Recognition	Type of Rating System	Years for Accumulative System
Aikido	IAF	AGFIS Recognized	None	
Boxing	AIBA	IOC Summer	Accumulative	4
Fencing	FIE	IOC Summer	Accumulative	1
Judo	IJF	IOC Summer	Accumulative	2
Ju-Jitsu	JJIF	AGFIS Recognized	None	
Kabbadi	WFK	Other	None	
Karate	WKF	IOC Recognized	None	
Kendo	FIK	AGFIS Recognized	None	
Kickboxing	WIKF, WAKO	AGFIS Recognized	Subjective	
Mixed Martial Arts	ISCF	Other	Subjective	
Muay Thai	IFMA	AGFIS Recognized	None	
Sambo	FIAS	AGFIS Recognized	None	
Sumo Wrestling	ISF	IOC Recognized	Adjustive ELO	
Taekwondo	WTF	IOC Summer	None	
Wrestling-Freestyle	FILA	IOC Summer	None	
Wresting- Greco Roman	FILA	IOC Summer	None	
Wushu	IWUF, IKF	IOC Recognized	None	

Table 3: International Rating Systems for 17 Combat Sports

Table 1. The sports are broken down into combat, independent and object categories. The Vancouver Olympic website provided 15 icons, one each for the 15 recognized sports whilst it would be expected that a website for the upcoming London Games of 2012 would have 37 icons. In addition to the 52 Olympic sports, another 54 non-Olympic sports are recognized. Table 2 will be discussed in relation to rating systems. Tables 3-5 contain the combat, independent and object sports respectively, showing the source of recognition for each sport and details to follow about rating systems.

The second source of recognition is the international organization AGFIS, also known under the English abbreviation GAISF, which recognizes all IOC-recognized federations as well as an additional 21 federations organizing 25 additional sports, See the References section for the web link.

The third source or recognition is Wikipedia. As of December 2009, Wikipedia had organized coverage of international sports federations (as in this paper) into IOC-recognized federations, additional AGFIS/GAISF recognized federations and another 24 federations organizing 25 sports. The Wikipedia website provides convenient links to all sports included here, although some links do not work and some sports are not in the correct branch of the IOC-AGFIS-Wikipedia trilogy. See the References section for the web link. Table 1 includes 156 sports, 2/3 of which are IOC recognized. About 10% (17) of these are combat sports whilst the remaining 90% of the sports are about evenly distributed among 73 independent sports and 66 object sports. AGFIS and Wikipedia also include a number of federations that organize social meetings, organize international competitions for the other sports and organize sports for separately-abled athletes. Such sports are not included here.

Among the 156 sports in Tables 3-5, are well known combat sports such as boxing and fencing, independent sport such as archery and swimming and object sport such as basketball and soccer. Many other less well-known sports are well worth exploring.

Some Lesser-Known Sports

Wushu is a combat sport with an odd name to the Western ear. Better known as Kung Fu, the Wushsu athlete must make breathtaking in-air movement and work with a variety of martial-arts weapons. Muay Thai (boxing, Thai-style) requires skill at hitting with the feet from various positions. Sambo is an Eastern-European combination of wrestling and boxing, requiring significant upper-body strength. Kabbadi is an Asian Indian sport, involving a raiding run onto the opponent’s half of a field. The raider must contact an opponent and return to the raider’s field half in one breath, proven by repeatedly saying “Kabbadi-kabbadi” or ‘Kit-kit”.

Among lesser-known independent sports is Apnea, an underwater breath-holding sport. Events include Free-dive Apnea and Dynamic Apnea, where the athlete swims horizontally with long fins, going as far as possible in one breath. In the 2009 World Championships, Goran Colak of Croatia broke the world record in Dynamic Apnea with a distance of 244 m, almost five lengths of a 50 m pool on one breath. The female winner, Lidija Lijic of Croatia, completed 182 m on one breath. Other independent sports include rowing huge ornamental boats (Dragon Boats), Frisbee Golf and a simulated combat shooting sport called Practical Shooting.

Underwater Rugby is an unusual object sport with a water-filled ball which must be stuffed into an underwater basket to score a “try”. Underwater hockey involves flipping a lead puck along the pool bottom with a small stick. Bandy (on ice) and Shinty (on grass) are Gaelic stick and ball sports with similar rules and equipment. Hurling is another Gaelic stick and ball sport wherein the ball can be carried on the end of the stick and then hit in midair toward the goal with remarkable accuracy. Gaelic football has the same rules as for Hurling. Gaelic Football and its cousin Australian Rules Football preserve football as it was played in the early 1800s. Although there is a World Cup of American Football, that fact was apparently unknown to the Americans who did not field a team in the first two competitions (1999 and 2003), both won by Japan. The Americans redeemed themselves somewhat by defeating Japan in the 2007 final. Rugby Fives (a form of court handball), Guts Frisbee (where the Frisbee is made purposely hard to catch) and a foot-volleyball sport called Sepak Takrow (literally kick ball) are other lesser-known object sports.

Sport	Int. Fed.	Recognition	Type of Rating System	Years for Accumulative System
Aerobatics	FAI	IOC Recognized	None	
Airsoft Practical Shooting	IAPS	Other	Accumulative	2
Alpine Skiing	FIS	IOC Winter	Accumulative	1
Apnea	CMAS	IOC Recognized	None	
Archery	FITA	IOC Summer	Adjustive	
Artistic Roller Skating	FIRS	IOC Recognized	None	
Athletics(Track and Field)	IAAF	IOC Summer	Accumulative	1
Auto Racing	FIA	Other	Accumulative	1
Ballooning	FIA	IOC Recognized	None	
Biathlon	IBU	IOC Winter	Accumulative	1
BMX cycling	ICU	IOC Summer	Accumulative	1
Bobsled	FIBT	IOC Winter	Accumulative	1
Bodybuilding	IFBB	AGFIS Recognized	None	
Bowling	FIQ	IOC Recognized	Accumulative	1
Canoe	ICF	IOC Summer	Accumulative	1
Casting	ICSF	AGFIS Recognized	None	
Cross County Skiing	FIS	IOC Winter	Accumulative	1
Cycling-road	ICU	IOC Summer	Accumulative	1
Cycling-Track	ICU	IOC Summer	Accumulative	1
Dance Sport	IDSF	IOC Recognized	Accumulative	1
Darts	WDF	AGFIS Recognized	Accumulative	2
Diving	FINA	IOC Summer	Accumulative	1
Dragon Boats	IBSF	AGFIS Recognized	None	
Equestrian	FEI	IOC Summer	Accumulative	1
Figure Skating	ISU	IOC Winter	Accumulative	2
Finswimming	CMAS	IOC Recognized	None	
Fishing	CIPS	AGFIS Recognized	None	
Freestyle Frisebee	WFDF	AGFIS Recognized	Accumulative	1
Freestyle Skiing	FIS	IOC Winter	Accumulative	1
Frisbee Golf	WFDF	AGFIS Recognized	Adjustive	
Glider Racing	FAI	IOC Recognized	Accumulative	3
Golf	IGF	IOC Recognized	Adjustive	
Gymnastics	FIG	IOC Summer	Accumulative	1
Hang Gliding	FAI	IOC Recognized	Accumulative	3
Horseshoes	NHPA	Other	None	
Ice Climbing	UIAA	IOC Recognized	Accumulative	1
Kayak	ICF	IOC Summer	Accumulative	1
Life Saving	ILSF	IOC Recognized	None	
Luge	FIL	IOC Winter	Accumulative	1
Minigolf	WMF	AGFIS Recognized	Accumulative	3
Modern Pentathlon	UIPM	IOC Summer	Accumulative	1
Motorcycle Racing	FIM	IOC Recognized	Accumulative	1
Mountain Bike Cycling	ICU	IOC Summer	Accumulative	1
Mountain Running	WMRA	Other	None	
Nordic Combined	FIS	IOC Winter	Accumulative	1
Orienteering	IOF	IOC Recognized	Accumulative	1
Power Boating	UIM	IOC Recognized	Accumulative	1

Powerlifting	IPF	AGFIS Recognized	Accumulative	1
Practical Shooting	IPSC	Other	None	
Rhythmic Gymnastics	FIG	IOC Summer	Accumulative	4
Rowing	FISA	IOC Summer	Accumulative	1
Sailing	ISAF	IOC Summer	Accumulative	2
Shooting	ISSF	IOC Summer	Accumulative	2
Short Track Speed Skating	ISU	IOC Winter	Accumulative	1
Skeleton Sled	FIBT	IOC Winter	Accumulative	1
Ski Jumping	FIS	IOC Winter	Accumulative	1
Ski mountaineering	UIAA	IOC Recognized	Accumulative	2
Skibobbing	FISB	Other	None	
Skydiving	FAI	IOC Recognized	None	
Sled Dog Racing	IFSS	AGFIS Recognized	Accumulative	1
Snowboarding	FIS	IOC Winter	Accumulative	1
Speed Roller Skating	FIRS	IOC Recognized	None	
Speed Skating	ISU	IOC Winter	Accumulative	1
Sport Climbing	UIAA	IOC Recognized	Accumulative	1
Surfing	ISA	IOC Recognized	Accumulative	1
Swimming	FINA	IOC Summer	Accumulative	1
Synchronized Swimming	FINA	IOC Summer	None	
Trampoline	FIG	IOC Summer	Accumulative	4
Triathlon	ITU	IOC Summer	Accumulative	2
Ultra-light Aircraft	FAI	IOC Recognized	None	
Underwater Orienteering	CMAS	IOC Recognized	None	
Water Skiing	IWSF	IOC Recognized	Accumulative	1
Weightlifting	IWF	IOC Summer	Accumulative	1

Table 4: International Rating Systems for 73 Independent Sports

3. SPORTS RATING SYSTEMS FOR THE WORLD SPORTS

It is important to distinguish a rating from a ranking. A rating is a numerical value assigned to a competitor, based on results and other factors whilst a ranking is the ordinal placement based on the ratings. The federation websites for all 156 sports were carefully searched for rating systems. In many cases, an existing system was not easily located by information on the home page; however, a subsequent search via Google did locate an existing system. In each case, data over some fixed period are analysed sequentially to establish the ratings. As has been true of other taxonomies, sports rating systems may be separated into three mutually exclusive types, depending on how new ratings are arrived at for each update over the data window. Ratings are either subjective; objective-non-decreasing (called accumulative here) or objective-able to increase, decrease or remain the same (called adjustive here). The adjustive rating systems usually tend to be the best predictors of future performance since each adjustment follows from a predictor-corrector action; hence predictability is built in. Accumulative systems are preferred by many tournament-rich sports because the accumulation of points requires top athletes to enter as many tournaments as possible, which encourages ticket sales and TV revenue.

Table 2 contains the result of the survey of sports rating systems. Federations for 59 sports do not publish ratings. Of the 97 published rating systems, only two are subjective, 82 are accumulative and 13 are adjustive. It is clear that accumulative systems are favoured.

Table 3 (for the 17 combat sports), Table 4 (for the 73 independent sports) and Table 5 (for the 66 object sports) show the federation for each sport, the source of recognition, the type of rating system if any and the number of years in the data window used by each accumulative system. Table 5 identifies true team sports among the object sports. Each type of rating system is now covered in detail.

Sport	Int. Fed.	Recognition	Team Sport	Type of Rating System	Years for Accumulative System
American Football	IFAF	AGFIS Recognized	Team	None	
Badminton	BWF	IOC Summer		Accumulative	1
Bandy	FIB	IOC Recognized	Team	None	
Baseball	IBAF	IOC Recognized	Team	Accumulative	4
Basketball	FIBA	IOC Summer	Team	Accumulative	8
Beach Volleyball	FIVB	IOC Summer	Team	Accumulative	1
Bocci	CBI	IOC Recognized		None	
Bridge	WBF	IOC Recognized		Accumulative	8
Broomball	IFBA	Other	Team	None	
Carom Billiards	WCBS,UMB	IOC Recognized		Accumulative	2
Chess	FIDE	IOC Recognized		Adjustive ELO	
Court Handball	USHA	Other		Accumulative	1
Cricket	ICC	IOC Recognized	Team	Adjustive	
Croquet	WCF	Other		Adjustive ELO	
Curling	WCF	IOC Winter		Accumulative	7
Double Disc Court Frisbee	WFDF	AGFIS Recognized		Accumulative	1
Draughts	FMJD	AGFIS Recognized		Adjustive ELO	
English Billiards	WCBS IBSF	IOC Recognized		None	
Field Hockey	FIH	IOC Summer	Team	Accumulative	4
Fistball	IFA	AGFIS Recognized	Team	None	
Floorball	IFF	IOC Recognized	Team	None	
Gaelic Football	GAA	Other	Team	None	
Go	IGF, EGF	AGFIS Recognized		Adjustive ELO	
Guts Frisbee	WFDF	AGFIS Recognized	Team	Accumulative	1
Handball	IHF	IOC Summer		None (Ind & All-Time Best)	
Hurling	GAA	Other	Team	None	
Ice Hockey	IIHF	IOC Winter	Team	Accumulative	4
Inline Roller Hockey	FIRS	IOC Recognized	Team	None	
Korfball	IKF	IOC Recognized	Team	Accumulative	4
Lacrosse	FIL, IFWLA, ELF	Other	Team	None	
Lawn Bowls	CMSB, World Bowls LTD	IOC Recognized		Accumulative	4
Netball	IFNA	IOC Recognized	Team	Adjustive, similar to ICC	
Pelota Vasca	FIPV	IOC Recognized		None	
Pesapallo	PESIS	Other	Team	None	
Petanque	CMSB, FIFJP	IOC Recognized		None	
Polo	FIP	IOC Recognized	Team	None	
Pool	WCBS,WPW	IOC Recognized		Accumulative	1
Racketlon	FIT	Other		Accumulative	2
Racquetball	IRF, IRT	IOC Recognized		Accumulative	1
Real Tennis	IRTPA	Other		Accumulative	1
Rock it Ball	IRIBF	Other	Team	None	

Roller Hockey	FIRS	IOC Recognized	Team	None	
Rounders	NRA	Other	Team	None	
Rugby Fives	RFA	Other		Accumulative	1
Rugby League	RLIF, RLEF	Other	Team	None	
Rugby Sevens	IRB	IOC Recognized	Team	Accumulative	1
Rugby Union	IRB	IOC Recognized	Team	Adjustive	
Sepak Takrow	ISTAF	AGFIS Recognized		None	
Shinty	CA	Other	Team	None	
Snooker	WCBS WPSBA	IOC Recognized		Accumulative	2
Soccer	FIFA	IOC Summer	Team	Adjustive (M) Adjustive ELO (W)	
Soft Tennis	ISTF	AGFIS Recognized		Accumulative	1
Softball	ISF	IOC Recognized	Team	None	
Sport Bowls	CMSB, FIB	IOC Recognized		None	
Squash	WSF PSA(M) WISPA(W)	IOC Recognized		 Adjustive	
Table Hockey	ITHF	Other		Accumulative	2
Table Soccer	ITSF	Other		Accumulative	1
Table Tennis	ITTF	IOC Summer		Accumulative	4
Tennis	ITF, ATP(M) WTA(W)	IOC Summer		Accumulative	4 1 1
Throwball	ITF	Other	Team	None	
Tug of War	TWIF	IOC Recognized	Team	None	
Ultimate Fisbee	WFDF	AGFIS Recognized	Team	Accumulative	1
Underwater Hockey	CMAS	IOC Recognized	Team	None	
Underwater Rugby	CMAS	IOC Recognized	Team	None	
Volleyball	FIVB	IOC Summer	Team	Accumulative	4
Water Polo	FINA	IOC Summer	Team	None (World League Only)	

Table 5: International Rating Systems for 66 Object Sports
Year = the number of years included for an accumulative system

Subjective Rating Systems

The only two subjective systems are for WACO Kickboxing and the ICSF Mixed Martial Arts. A panel of experts ranks competitors and those individual rankings are combined for the overall ranking. There are also non-internationally recognized systems in boxing published by the WBC, WBA, IBF and WBO organizations and in UFC mixed martial arts, each of which employs a champion-challenger system rather than international tournaments.
There is one anomaly of this accounting system in that the word “None” describes the fact that the IFNA Muay Thai federation does not publish Muay Thai ratings. The WIKF Kickboxing federation publishes ratings for the Kickboxing events that WIKF organizes and WIKF also publishes Muay Thai ratings as a service to Muay Thai, although Muay Thai competition is organized by IFNA.

Accumulative Rating Systems

An accumulative rating system for competitor *i* follows the form shown in (1), where summation is over a window of past results for competitor *i*.

$$\text{New rating for } i = \Sigma f_i \text{(results, weights, ageing, old ratings, other factors)} \quad (1)$$

The function f_i in (1) includes converting results for competitor i to points, weighting points by importance, ageing data from previous years, including old ratings and other factors. The term “accumulative” follows from the fact that all $f_i (..) \geq 0$, hence ratings for each i are non-decreasing as summation moves over the data window. A particularly simple accumulative system employs the “best” operator for the function f_i. For example, the current year’s best result is used by IAAF athletics and FINA swimming whilst the three best results of the current year are used by FINA diving.

Two distinctly different methods are employed to convert results to points weighted by importance. In FIVB Basketball, the various championships held over the eight-year window are each given a vector of weights varying from 5 to 0.1. Placement in each championship is given a common set of points from 50-1. It is necessary to multiply the weight by the placement points for a given competition. Conversely, for ATP (men’s professional) Tennis and WTA (women’s professional) Tennis, a matrix is published where each row contain result-scores for a given championship (say for Wimbledon) whilst each column contains the placement scores (for say, being eliminated in the quarter-finals). The matrix approach is recommended since both athlete and sports follower can easily denote the points to be accumulated for a given placement in a given championship.

Table 6 shows the number of data-window years used by the 83 accumulative systems. Notice that a one year window is the most frequent data window, clearly favoured by the independent sports in Table 4 and by the individual (non-team) object sports in Table 5, where the skill of an individual can change rather dramatically from year to year. The number of systems drops for two and three year windows and increases for four year windows, which usually include one world cup and/or Olympic cycle. Multiple-year windows are favoured by the team object sports in Table 5, where a team may play a limited number of international matches in a given year requiring a number of years for valid comparisons.

Systems	Number of Years
53	1
12	2
3	3
11	4
1 (Curling)	7
2 (Basketball, Bridge)	8

Table 6: Number of Years for the 83 Accumulative Systems.

When there is a multiple-year window, the previous year’s results (relative to the current year) may be “aged” by multiplying the current year’s results by 100% and by multiplying each previous year’s results by a lower value.

Year	Uniform Ageing	Non Uniform Ageing FIVB Volleyball	
	All Points	4 year events	2 year events
1	100%	100%	100%
2	75%	75%	50%
3	50%	50%	0
4	25%	25%	
5	0	0	

Table 7: Ageing of Data
Uniform ageing: IIHF Ice Hockey, IKF Korfball, ITF Tennis, FIH Field Hockey, ITFF Table Tennis.
Non uniform ageing: FIVB Volleyball

Table 7 shows two methods used to age data. Ageing occurs over a four-year window for five sports with all events in a given year being aged exactly the same way, with a graduated ageing that drops by the same fraction each year. Conversely, FIVB volleyball employs non-uniform ageing in that some event results are aged differently than others contested in the same year, depending on the frequency of the event.

IDF Darts is the only accumulative system using money won. In some sports, the final position value is multiplied by values dependent on other factors for that sport. In FAI Hang Gliding and Paragliding, other factors include the quality of entrants, the number of entrants, the relative time and the number of skills used. In ISAF Sailing, other factors include the importance of the race and the quality of entrants. In ISSF Shooting, other factors include the importance of the match, the score relative to the world record and the score relative to minimum standards. In IFSS Sled Dog racing, points are earned based on the length of race, the importance of the race and relative time. In ITU Triathlon, a bonus is given based on the number of top 20 entrants; however, no points are earned if the final time is worse than a cut off time.

Adjustive Rating Systems

An adjustive rating system for competitor *i* has the form shown in (2).

$$\text{new rating for } i = \text{old rating for } i + K\,[\text{ new result - prediction (old results, weights, old ratings)}] \quad (2)$$

This type of rating system follows the format of a predictor-corrector in which a rating for *i* can increase, stay the same or decrease as each new result is compared to each prediction based on information available prior to the competition. The value of *K* must be chosen carefully. Too large of a value for *K* would make the ratings respond too forcefully to the error term in the square brackets *[...]*, probably making ratings oscillate thereafter whilst too small of a value for K would make the ratings unresponsive to *[...]*.

Type	New Result	Prediction
Elo	(1, .5, 0)	*P(d)*
Probit	(1, 0, -1)	*k d* with limiting
Averaging	Chosen values	Past average of chosen values

Table 8: Types of Adjustive Systems

According to Table 2, there are 13 adjustive systems published by the federations listed in Tables 3-5. In Table 8, these adjustive systems fall into another trilogy of categories, based on how the new result is used to make adjustments vis-a-vis the method used to predict each result. Let team *i*'s most recent opponent be designated *j*. Let *d* represent the rating difference between the ratings of competitors *i* and *j* prior to the most recent competition, as given by (3).

$$d = \text{old rating for } i - \text{old rating for } j \quad (3)$$

Six sports are rated using the Elo system, three are rated using Probit systems whilst four sports employ averaging methods. Each type is now covered in order.

Elo Rating Systems

A strong point of the Elo system is simplicity, in that the system depends more on theory than on ad-hoc parameter selection. For the Elo System

$$P(d) = 1 / (1 + 10^{-d/[2\sigma_t]}) \quad (4)$$

Here σ_t denotes the standard deviation of team performance, usually set at 200 arbitrarily whilst the mean is also set arbitrarily. The adjustive mechanism causes the rating distribution to follow whatever mean and standard deviation are selected. Only *K* must be selected purposely. It may seem illogical to use a base *10* exponent rather than a base *e* exponent. Actually, (4) is an approximation to a base *e* function.
The "new result" for the Elo system follows the scale (1 = a win, .5 = a draw and 0 = a loss) for competitor *i* against opponent *j*, which is compared to the a prior probability *P(d)*. The rightmost adjustment term in (2) becomes *K [w – P(d)]* where *w* is the new result. The maximum positive adjustment is *K [1-0]* or *K*. The ELO system was developed by Chess Master Arpad Elo. That system is used for six very diverse sports: FIDE Chess, FMJD Draughts (Checkers), WCF Croquet, IGF Go, FIFA Women's Soccer and ISF Sumo Wrestling. Three applications are to mind sports and three are to physical sports.
For each of the six Elo applications, Table 9 shows the value of assumed team standard deviation and the range of *K* values. In order to compare the systems, the rightmost column shows the maximum rating change divided by the team standard deviation, which equals *K* divided by σ_t. The last column of Table 9 thus measures the sensitivity of the rating system. Where *K* has a range of values, the maximum rating change also has a range of values.

Sport	σ_t	K	[Max. rating change K] / σ_t
FIDE Chess, FMJD Draughts	200	10-25	.05 - .12
IGF Go	70-200	10 - 116	.14-.58
ISF Sumo Wrestling	≥ 200	≥ 50	.25 - .45
FIFA Women's Soccer	200	10 – 40	.05 - .20
WCF Croquet	250	50	.20

Table 9: Six Elo Rating Systems

For the mind sports of Chess, Draughts and Go as well as for the physical sport of Sumo Wrestling, the smallest maximum rating change is for a grand master whilst the largest maximum rating change is for a beginner. In such sports, it is logical that a newcomer wants to move up quickly whilst a grand master wants to remain highly rated; hence such rating sensitivities appeal to both competitors and meet directors. Conversely for Women's Soccer, the smallest maximum rating change is for a friendly whilst the largest maximum rating change is for a World Cup final, as is reasonable for that sport. For croquet all competitors are subject to the same maximum rating adjustment. The most sophisticated selection of *K* is done for arguably the least sophisticated sport, Sumo Wrestling, in that the other sports employ ad-hoc values of *K* whilst Sumo Wrestling employs a theoretically-derived formula. Note that a rating system designed to facilitate accurate predictions whilst not being beholding to competitors or meet directors would most likely treat highly-rated and lower-rated competitors with the same adjustment mechanism.

Probit Rating Systems

For the three Probit adjustive systems in Table 10, the predicted probability of a win is a linear function *k d*, limited to a range such as +1 to -1. The new result is then evaluated for competitor *i* as (1 = a win, 0 = a draw and -1 = a loss). The adjustment term in Equation 2 become *K [w – k d]* where *w* is the new result. The maximum positive adjustment is *K [1- (-1)]* or *2 K*. For each Probit system, Table 10 summarizes the value of the match prediction *k d*, the ratings smoother *K*, home advantage and the maximum change in terms of team standard deviations so as to compare with the Elo systems of Table 9.

The ICC Cricket and IFNA Netball systems are identical, both having been developed by David Kendix. Here prediction is given by *d/50* and K depends on the number of games played.

For IRB Rugby, prediction is given by *d/10* and 3 rating points are added to *d* for a home advantage if *i* plays at home. As compared to FIFA Elo Women's Soccer which adds uses 100 rating points to *d* as home advantage, the assumed fraction of home wins for *d* = 0 is about the same (65% for Rugby and 64% for soccer). The maximum change for Cricket is the same as the maximum rating change for Rugby using a window *n* of about 10.

The IRB Rugby system is at the upper range of sensitivities compare to the Elo systems of Table 9. Note that the IRB Rugby system is the only one using score difference as a factor in determing *K*, which also depends on match importance. Since the inclusion of score provides more information than just winning, this IRB Rugby is recommended as an excellent example of a Probit system.

Averaging Systems

In an averaging system, a new result is compared with a past average of that same result, causing rating adjustment in (2). Five sports use averaging systems. In FIJA Archery, each score is adjusted for the cohort tournament average compared with the last cohort World Cup average and the scores are averaged over one year (minimum of 3) matches. In WFDF Frisbee Golf, a score more than two standard deviations from the mean is not used. The last eight scores are averaged. In IGF Golf, points are accrued depending on tournament placement and a two-year average is used (minimum of 40 tournaments). In WSF Squash tournament points are averaged over one year (minimum of 10 matches). For a Men's Soccer match, FIFA multiplies factors which depend on importance, continent, outcome and opponent rating, Ratings follow using a four-year weighted average.

Sport	Prediction $k\ d$	K	Home Adv.	Home Wins	[Max rating change $2K$] / σ_t
ICC Cricket IFNA Netball (David Kendix)	d/50 with limiting	50/n n = games played	0		3.4 / n
IRB Rugby	d/10 With limiting	1-3 (win margin, importance)	3	65%	.13 - .38
(FIFA Elo Wom. Soccer)			100	64%	

Table 10: Three Probit Rating Systems (FIFA Elo is shown for comparative purposes)

4. CONCLUSIONS

Based on three sources of recognition, 156 widely played international sports were identified, 17 of which are combat sports, 73 are independent sports and 66 are object sports. The organizing sports federations published rating systems for 97 of those sports, only two of which are subjective. Among the 83 accumulative systems, several convert performance points via a very compact and useful placement-importance matrix. A one year window is commonly used for individual independent and object sports; a four-year window is commonly used for team object sports and ageing is may be done uniformly in equal steps for all points earned in a given year. Of the adjustive systems, the Elo system is noted for its simplicity whilst the IRB Rugby Probit system is particularly well structured. The nine Elo and Probit adjustive systems had a maximum rating adjustment that varied from about 0.1 to 0.6 team standard deviations.

References

AGFIS Recognized International Sports Federations: http://www.agfisonline.com/vsite/vnavsite/page/directory/0,10853,5148-176060-193278-nav-list,00.html

IOC Recognized International Sports Federations: http://www.olympic.org/en/content/The-IOC/

Stefani, R. (1997) Survey of the Major World Sports Rating Systems, *Journal of Applied Statistics, 24*, 635-646.

Stefani, R. (1998) Predicting Outcomes. *Statistics in Sport*, Arnold Press.

Stefani, R. (1999) A Taxonomy of Sports Rating Systems. *IEEE Trans. On Systems, Man and Cybernetics, Part A, 29*, 116-120.

The Official Encyclopedia of Sports and Games, New York, London: Paddington Press, 1979.

Wikipedia List of International Sports Federations http://en.wikipedia.org/wiki/List_of_international_sport_federations:

MINIMUM VIOLATIONS SPORTS RANKING USING EVOLUTIONARY OPTIMIZATION AND BINARY INTEGER LINEAR PROGRAM APPROACHES

Timothy Chartier[a], Erich Kreutzer[a], Amy Langville[b], Kathryn Pedings[b,d], Yoshitsugu Yamamoto[c]

[a] *Davidson College, Davidson, NC, USA*
[b] *College of Charleston, Charleston, SC, USA*
[c] *University of Tsukuba, Tsukuba, Japan*
[d] *Corresponding author: kepeding@edisto.cofc.edu*

Abstract

Typically the binary integer linear program (BILP) formulation of the minimum violations ranking (MVR) problem and related rank aggregation problem is the preferred way to find a ranking that minimizes the number of violations to hillside form. However, for very large ranking problems, the BILP formulation is limited by the $O(n^3)$ number of constraints. Even when constraint relaxation techniques are employed, there are practical limits on the size of n, the number of items being ranked. One goal of this paper is to demonstrate these limits on several ranking problems drawn from a wide range of application areas. Another goal is to overcome these limitations by using a evolutionary optimization (EO) algorithm to solve large MVR ranking problems. Our EO algorithm uses many features of the BILP formulation to improve its speed and convergence. Though EO, unlike BILP, is not guaranteed to produce the global optimum, its speed, scalability, and flexibility make it the method of choice for solving very large-scale linear ordering problems.

Key words: evolutionary optimization, binary integer linear program, minimum violations, ranking, hillside form, March Madness

1. Introduction

In this paper, we present a rating method that, given information on the pairwise comparisons of n items, minimizes the number of inconsistencies in the ranking of those items. Though Minimum Violations Ranking (MVR) methods have many applications (Reinelt et al., 1984), we use examples from sports to explain our new MVR methods. There are two algorithms discussed in this paper that both use MVR. The binary integer linear program (BILP) shown in this paper always gives the optimal solution to the optimization problem. The evolutionary optimization (EO) gives a heuristic solution. In order to understand the methods, we must define terms that will be used throughout the paper.

The matrix **D** below, which we call a point differential matrix, contains pairwise comparison data and is commonly and easily produced for many sports.

$$\mathbf{D} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 9 & 0 & 4 & 0 & 2 \\ 5 & 0 & 0 & 0 & 0 \\ 15 & 3 & 8 & 0 & 5 \\ 6 & 0 & 3 & 0 & 0 \end{pmatrix} \end{array}$$

The $(2, 3)$-entry means that team 2 beat team 3 by 4 points in their matchup. We will analyze this point differential matrix in order to produce a ranking of these five teams. At this point we introduce a definition.

A matrix $\mathbf{D}$ is in *hillside form* if

$$d_{ij} \leq d_{ik} \quad , \qquad \forall\ i \text{ and } \forall\ j \leq k$$
$$d_{ij} \geq d_{kj} \quad , \qquad \forall\ j \text{ and } \forall\ i \leq k.$$

The name is suggestive as a cityplot of a matrix in hillside form looks like a sloping hillside as in figure 2. As an illustrative example, consider two point differential matrices from two different seasons. Since this represents seasonal data, it is possible that some of the teams played multiple matchups. For instance, one possible scenario for matrix **B** is that teams 1 and 3 played two times during the season, the first time team 1 beat team 3 by 5 points (**B**(1,3)=5) and the second time team 3 beat team 1 by 7 points (**B**(3,1)=7). The matrix given in **A** is in hillside form and the season represented in **B** is not.

$$\mathbf{A} = \begin{pmatrix} 0 & 3 & 5 & 8 & 15 \\ 0 & 0 & 2 & 4 & 9 \\ 0 & 0 & 0 & 3 & 6 \\ 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$\mathbf{B} = \begin{pmatrix} 0 & 3 & 5 & 8 & 15 \\ 0 & 0 & 2 & 4 & 9 \\ 7 & 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

For $n \times n$ matrices in hillside form, the ranking $\mathbf{r}$ of the items is clear: $\mathbf{r} = (1 \quad 2 \quad \cdots \quad n)$. For non-hillside matrices, we can count the number of violations of the hillside conditions. In the above example, $\mathbf{B}$ has 7 violations. Often a matrix that appears to be non-hillside can be symmetrically reordered so that it is in hillside or near hillside form. In fact, the non-hillside matrix $\mathbf{D}$ when reordered according to the vector $(5 \quad 2 \quad 4 \quad 1 \quad 3)$ forms the hillside matrix $\mathbf{A}$. *Finding such a hidden hillside structure is exactly the aim of both the EO and BILP methods.* Our MVR methods find a reordering of the items that when applied to the item-item matrix of differential data forms a matrix that is as close to *hillside form* as possible. We will discuss our measure of closeness to hillside form as violations which are defined in the next paragraph.

Hillside form gives a great deal of information about the difference in the strengths of teams. For example, matrix $\mathbf{A}$ says that not only is team 1 ranked above teams 2, 3, 4, and 5, but we expect team 1 to beat team 2 by some margin of victory, then team 3 by an even greater margin, and so on. Sometimes a data matrix has been reordered to be as close to hillside form as possible, yet violations remain. These violations are of two types: *upsets* and *weak wins*. Nonzero entries in the lower triangular part of the reordered matrix correspond to upsets, i.e., when a lower ranked team beat a higher ranked team. Weak wins manifest as violations of the hillside conditions that occur in the upper triangular part of the matrix. This is when a high ranked team beats a low ranked team but by a smaller margin of victory than expected. Our MVR paper inherently weights upsets more than weak wins. The example matrix $\mathbf{B}$ above demonstrates this well. Notice that the presence of the 7 in the lower triangular part of the matrix accounted for 6 of the 7 violations. Looking across the third row, the 7 is to the left of four number which are smaller in magnitude, giving 4 of the violations. Down the first column, 7 is below two zeroes giving the final 2 of the violations. The last of the 7 violations can be seen in the fifth column since the 5 falls below the 4.

Although we have not experimented with alternate ways to weight the data, it is possible. The user could weight the seriousness of the upset. For example, a 12^{th} ranked team beating a 4^{th} ranked team would be weighted more heavily than a 9^{th} ranked team beating an 8^{th} ranked team. This would be implemented by simply doing a Hadamard product of the data matrix with a weight matrix where the values in the lower, left-hand corner would be larger than those around the rest of the matrix. A second idea is to weight the input data by date. If the ranking is used as a predictive method for a tournament, then games closer to the tournament would be weighted more heavily. Many different weights could be used here. Other methods such as Colley and Massey use linear, exponential, logarithmic, or step functions to weight the games.

There are many types of data that can be used as input to create MVR rankings. Below are four common data matrices that can be used for sports teams.

1. Point Matrices:
   (a) $\text{Psum}_{i,j}$ = sum of all the points scored by team i against team j
   (b) $\text{Pavg}_{i,j}$ = average of all points scored by team i against team j
2. Point Differential Matrices:
   (a) $\text{Dsum}_{i,j}$ = sum of all the positive point differences scored by team i against team j
   (b) $\text{Davg}_{i,j}$ = average of all positive point differences scored by team i against team j
3. Difference Matrices:
   (a) $\text{Diffsum}_{i,j}$ = sum of the points scored by team i minus points scored by team j
   (b) $\text{Diffavg}_{i,j}$ = average of the points scored by team i minus points scored by team j
4. Rank Aggregation Matrices:
   (a) For rank aggregation, we use rankings from other models and combine the data. We combine the data in the following way:
   (b) $\text{RankAgg}_{i,j}$ = # lists having i above j

It is important to note that our algorithms have been tested with data from entire seasons to use as prediction models for tournaments. However, this does not necessarily mean that we have information on every head to head match-up between the teams. The rankings can still be computed using the indirect relationships in the data. Future work needs to be done to determine whether there is a warm up period needed for the algorithms to run successfully.

This paper is outlined as follows. First, in Section 2, we summarize the major findings from our prior MVR solution technique, which uses mathematical programming to find MVR rankings. Then in Section 3, we propose our new MVR solution technique, which uses the very intuitive method of EO. Section 4 gives results on data from the Southern Conference region of NCAA Division I basketball in the United States. The paper ends with some experiments from NCAA basketball and thoughts on future work for this topic.

2. Findings from Prior Work in Mathematical Programming

Other researchers have proposed various methods for solving the MVR problem (Ali et al., 1986; Cassady et al., 2005; Coleman, 2005; Park, 2005) and in another paper (Langville et al., 2009), we formulated a binary integer linear program (BILP) to solve the MVR problem described above. In that paper, our MVR methods used ideas from mathematical programming. While specialized knowledge of that field is required to appreciate and implement those methods, in this section we summarize the findings from that paper that pertain to this work. Our goal in this paper is to solve the MVR problem using a more intuitive technique that requires no specialized knowledge or software.

The BILP that we formulated and explained in detail in (Langville et al., 2009) is below.

$$\begin{aligned} \min \sum_{i=1}^{n}\sum_{j=1}^{n} & \quad c_{ij}\, x_{ij} \\ x_{ij} + x_{ji} &= 1 \quad \text{for all } i \neq j \\ x_{ij} + x_{jk} + x_{ki} &\leq 2 \quad \text{for all } i \neq j \neq k \\ x_{ij} &\in \{0,1\}, \end{aligned}$$

where

$$x_{ij} = \begin{cases} 1, & \text{if item } i \text{ is ranked above item } j \\ 0, & \text{otherwise.} \end{cases}$$

and $\mathbf{C}$ is a matrix of constants formed from the data matrix $\mathbf{D}$. One definition assigns

$$c_{ij} := \#\{\, k \mid d_{ik} < d_{jk} \,\} + \#\{\, k \mid d_{ki} > d_{kj} \,\}, \quad (1)$$

where # denotes the cardinality of the corresponding set. Thus, $\#\{\, k \mid d_{ik} < d_{jk} \,\}$ is the number of items ranking item j above item i. Similarly, $\#\{\, k \mid d_{ki} > d_{kj} \,\}$ is the items ranking item i above item j.

Industrial software such as Xpress-MP finds the globally optimal MVR ranking. When the optimization algorithm concludes, these x_{ij} variables can be assembled into a binary matrix $\mathbf{X}$, which is then used to create a ranking of the n items. The item with the greatest number of 1s in its row is the highest ranked team. In fact, this item has a row sum of $n-1$, meaning that it is ranked above every other item. The second place item will have a row sum of $n - 2$, the third place item will have a row sum of $n - 3$, and so on down to the last place item, which has a row sum of 0, meaning that it is ranked above no items. In addition, in (Langville et al., 2009) we described a linear time algorithm for scanning the optimal ranking to identify multiple optimal solutions. In other words, multiple optimal solutions correspond to an optimal MVR ranking with ties in some rank positions.

We discovered that the $O(n^3)$ inequality constraints dramatically limit the size of ranking problems that can be solved with the BILP method. Consequently, we used classical relaxation techniques from the field of mathematical programming. One relaxation solves the linear program (LP) relaxation of the original BILP. The LP results were very good. We were able to prove that the LP solution was optimal for the BILP under certain conditions. In the few cases where the LP produces a suboptimal solution, we used bounding techniques to produce a measure ϵ, indicating that the LP solution is within ϵ% of the optimal BILP solution. Further, the LP relaxation requires slightly less computation, immediately identifies multiple optimal solutions, and enables sensitivity analysis, which enables measures of confidence in the assignments of items to rank positions. While we solved a 347-team example in under a minute using Xpress-MP software, we were still unsatisfied with the size of ranking (also known as linear ordering) problems we could solve. As a result, in this paper, we present a very intuitive solution technique called EO that requires no specialized knowledge of mathematical programming or its software and enables the solution of even bigger ranking problems.

3. EO

3.1. Overview of EO

EO, as the name suggests, takes it modus operandi from natural evolution, and every EO algorithm uses the basic evolutionary ideas of mating, mutation, and survival of the fittest to solve an optimization problem. The trick is to tailor these basic ideas to fit the specific problem at hand. The idea is to start with some initial population of p possible candidate solutions for the problem of interest. Each member of the population is evaluated for its fitness. The fittest members of the population are mated to create offspring that contain the best properties of their parents. Continuing with the evolutionary analogies, the less fit members are mutated in asexual fashion while the least fit members are dropped and replaced with immigrants. This new population of p members is evaluated for its fitness and the process continued. As the iterations proceed, it is fascinating to watch Darwin's principle of survival of the fittest. The populations march toward more evolved, fitter collections of members. Perhaps more fascinating is the fact that there are theorems proving that evolutionary algorithms converge to optimal solutions in many cases and near optimal solutions under certain conditions (Fogel and Michalewicz, 2004). Unfortunately, evolutionary algorithms can be slow-converging, which is why careful tailoring to the application is so important (Fogel and Michalewicz, 2004).

3.2. Tailoring EO to the MVR problem

In this section, we tailor the general EO ideas above to our specific MVR problem.

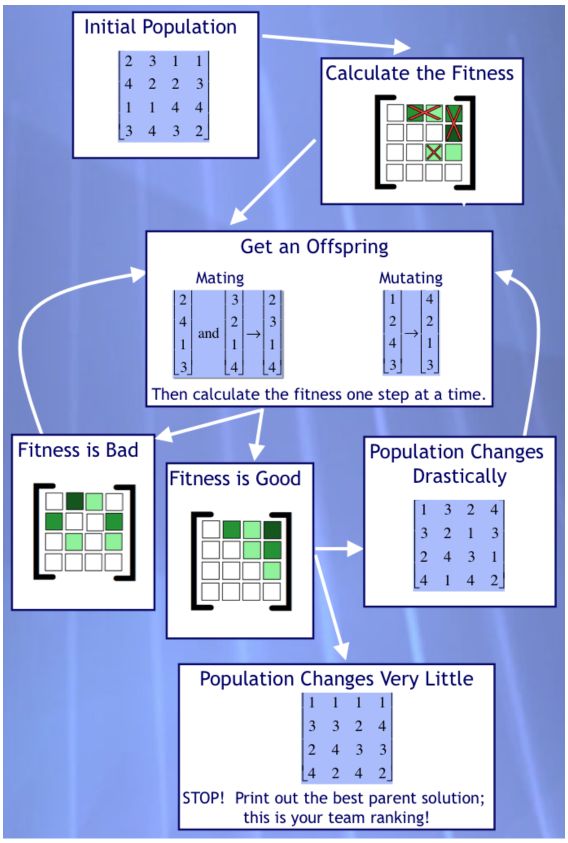


Figure 1: Overview of steps of EO for ranking problem

3.2.1. Members of the Population

For our MVR problem, each member of the EO population is a ranking vector, i.e., a permutation vector of the integers 1 through n.

3.2.2. Initialization

The EO algorithm always gets increasingly closer to hillside form as it progresses. This is motivation to find a good initial population. There are many established methods of ranking that have done well in predictive settings such as the NCAA March Madness.[1] We use the output from these methods as our initial parent population. Some methods used are Massey, Colley, and mHits (Colley, 2002; Govan et al., 2009; Massey, 1997). As seen in table 1, the initialization truly makes a difference. These experiments were run using all 347 NCAA Division 1 basketball teams.

Init.	time (sec)	violations
Random	19.6	1,404,783
Best 10	10.3	1,163,143
Worst 10	30.8	1,182,274

Table 1: Runtimes (in seconds) and number of violations for EO with different initializations

Starting solutions from the Colley, Massey, and mHITS methods were considered. The number of violations to hillside form was calculated for each solution to determine the best and worst 10 solutions. It can be seen that the runtime and number of violations were both lowest for the best initialization. Although the worst initialization had the highest runtime, it had better results for the number of violations than the random initialization. These results are just preliminary, but suggest that further work should be done on the sensitivity of the initialization.

3.2.3. Fitness

The fitness function for EO is the number of violations to hillside form for the reordered data matrix. This can be calculated using the same **C** matrix defined for the BILP in equation 1.

[1] Information about the NCAA March Madness Tournament can be found at http://www.ncaa.com/sports/m-baskbl/ncaa-m-baskbl-body.html

3.2.4. Offspring

There are two ways to create offspring: mating and mutating. When considering Darwinian ideas, mating should take preference over mutating. Mutating is used as a means to break the population out of a local optimum. Our algorithm allows the user to choose both the percentage of time to mate versus mutate and set the probability density function (pdf) to determine which mating or mutating algorithm is used most often. All of our experiments were run with mating set at 85% and a uniform pdf over each mating and mutating algorithm. Future work should be performed to determine how these percentages affect the outcome.

This section presents the mating and mutating algorithms we used for our MVR problem.

1. Mating: These use two or more parent solutions to create one offspring.
    (a) Borda Count: To compute the Borda Count for a particular team, for each parent list and for each team in that list, count the number of teams that it ranks above. Sum this for each list.
    (b) Average Rank: This method uses exactly two parent solutions. It averages the corresponding entries then ranks these averages.
2. Mutating: These use one parent solution to create one offspring.
    (a) Flip: Randomly choose two teams and flip their positions.
    (b) Insert: Randomly choose a team and put it in a different location.
    (c) Displace: Choose a group of teams and put them in a different location.
    (d) Reverse: Choose a group of teams and reverse their order.

After an offspring is formed, the fitness of that offspring is checked. If the fitness is better than the fitness of the worst parent solution, the offspring becomes part of the parent population and the worst parent is kicked out. This allows for the population to always be moving toward improving hillside form.

3.2.5. Stopping Criteria

Texts suggest many different ways to select a stopping criteria (Fogel and Michalewicz, 2004). Our

EO algorithm uses the average change of the average fitness of the parent population. Every time a new offspring is added to the population, the average fitness is calculated. We then calculate the average change of these fitness values. The user sets a tolerance level for the average change to fall below. To allow the algorithm ample chance to escape a local minimum, the average fitness change has to fall below this tolerance level five times in a row in order for the algorithm to stop. Upon termination, the best parent solution in the population is the ranking.

Figure 1 gives a global view of the EO algorithm applied to the MVR problem.

3.3. *Summary of EO for the MVR problem*

We pause to consider the pros and cons of EO when compared to the alternative solution techniques from mathematical programming, described in Section 2. **Pros:**

- EO is easy to understand and requires no specialized knowledge of mathematical programming.
- EO is easy to code and requires no specialized industrial software.[2]
- EO is flexible and adaptable. For instance, a user can easily implement multi-objective functions or secondary or tertiary objectives by changing the fitness function.
- EO can handle big datasets because it scales up well and can be parallelized.
- Early termination of an EO algorithm gives meaningful results that are at least locally optimal.

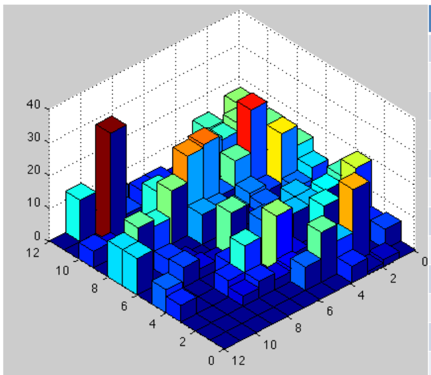
Cons:

- The EO solution is usually only locally optimal. Unlike the mathematical programming methods of Section 2, there is no guarantee that the solution is globally optimal or within some percentage of globally optimal.
- No sensitivity analysis is available with the EO method.

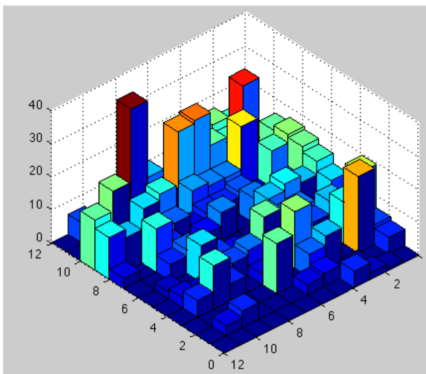
However, it is possible to use the EO solution to initialize or provide good bounds for the mathematical programming methods, thereby greatly reducing their runtimes, and enabling optimality guarantees and sensitivity analysis.

4. Small Example of EO vs. BILP

For this section, we will use the 2009-2010 Southern Conference (SoCon) Men's Basketball data. For this dataset, a total of 119 games were played with each team playing every other team approximately twice. The results shown use the Davg data matrix formulation. The heights of each of the bars in the pictures represent the magnitude of the entry in the matrix. The top plot shows the matrix reordered with the SoCon actual standings, the middle plot shows the matrix reordered with the results from EO and the bottom plot shows the matrix reordered with the results from the BILP. For the EO, only one initialization was used with results from Massey, Colley, and mHits, and all types of mating and mutating described in Section 3 were used. Accompanying these plots are tables with the results.



SoCon Standings
Wofford
College of Charleston
Appalachian State
Western Carolina
Davidson
Citadel
Furman
Chattanooga
GA Southern
UNC – Greensboro
Samford
Elon



EO Results
Wofford
Appalachian State
College of Charleston
Western Carolina
Citadel
Furman
Davidson
UNC – Greensboro
Elon
Chattanooga
GA Southern
Samford

[2]Our EO code for the MVR problem is available upon email request.

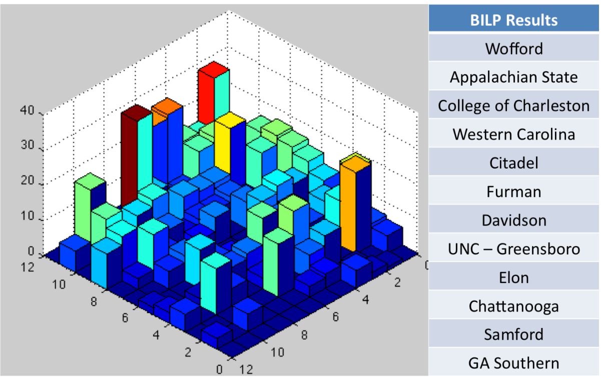


Figure 2: SoCon Standings, EO Results, and BILP Results on 2009-2010 SoCon data.

You will notice that the EO and BILP differ only in the last two teams. What is interesting about this is that the number of violations found for both the EO and BILP is 436. Since both of the optimal values were the same, but the rankings were different, we know there must be multiple optimal solutions for this SoCon example.

5. Large Experiments

This section explores some larger examples for both EO and BILP.

5.1. *Datasets of Varying Size*

We tested the runtime and number of violations for data sets of different sizes for each of the algorithms. The first dataset used is all of the data from the 2009 - 2010 NCAA Division I basketball season with 347 total teams. The second dataset is all of the data from the 2009 NCAA football season with 634 total teams. The third dataset is all of the data from the 2009 - 2010 NCAA basketball season with 1041 total teams. The fourth dataset is all of the data from the 2009 - 2010 college basketball season with 2034 total teams. The runtimes in the table are in minutes.

	EO		BILP/LP	
n	time	violations	time	violations
347	.35	1,161,919	.66	1,147,912
634	4.2	2,054,127	106.8	1,538,490
1041	14.2	11,412,879	1343.8	9,311,502
2034	21.3	42,109,102	-	-

Table 2: Runtimes (in minutes) and number of violations for EO vs. BILP for problems of increasing size

Here we note some observations from Table 2.

- The EO number of violations for the $n = 347$ example is within 1.22% of the optimal BILP number of violations.
- The BILP took almost 10 times the amount of minutes to run for the $n = 1041$ example. The results are much better for the BILP, but one must consider the time factor in determining which algorithm to use.
- The BILP was allowed to run for 24 hours on the $n = 2034$ example and did not obtain a solution. We stopped the algorithm as the time was so much greater than that for the EO.

5.2. *March Madness*

Both of these methods have been used to predict winners in each game of the Division I NCAA Men's basketball tournament, which is often called March Madness, in American College basketball for the past two years. Before the tournament begins, many fans complete brackets predicting winners of each tournament game. Once the tournament begins, each correct prediction in a bracket accrues points. As such, a pool of brackets can be formed and compete for the highest score. A popular online pool is the ESPN Tournament Challenge where over 4 million submissions were submitted for the 2009-2010 tournament. In this online pool, the brackets are scored so that correct guesses in later rounds earn the fan more points out of the total 1920 points possible. The following table shows results of using the various data matrices and methods to predict March Madness Games. The final column gives the percentile ranking for the corresponding ESPN score. For example, the EO Pavg method with ESPN score of 930 scored above 93.3% of the over 4 million brackets submitted by fans.

Method	**ESPN Score**	**Percentile**
EO Pavg	930	93.3
EO Psum	930	93.3
EO Rankagg	910	92.9
BILP Rankagg	900	92.7
EO Diffavg	850	90.9
EO Davg	780	86.7
EO Diffsum	750	84.1
EO Dsum	740	83.2
BILP Pavg	710	80.1
BILP Dsum	700	78.9
BILP Diffsum	660	72.2
BILP Diffavg	650	69.9
BILP Davg	570	44.2
BILP Psum	480	15.8

Table 3: Table of March Madness Results

Let us note some observations from the table:

- Overall, the EO brackets did much better than the BILP brackets. The 2009 - 2010 season was filled with many upsets. This may provide a reason why the local solutions of the EO better predicted the tournament. However, this is a very intriguing result that will need further investigation.
- The P and Rankagg data matrices performed the best.
- Almost all of the submissions were above the 50^{th} percentile of all brackets submitted.

6. Conclusions

This paper presents two equivalent, but different methods to solve the MVR problem. EO has a simple structure and can handle large datasets, but it gives a heuristic, often locally optimal value. The binary integer linear program gives a global optimum and can be converted to an LP which helps with multiple optimal solutions and sensitivity analysis, but has a lot of constraints that significantly affect the run time.

Our EO algorithm has become much more sophisticated since we began this research; however, there is more work that can be done. We need to update our code using the definition of the **C** matrix to count the number of violations to hillside form. This should decrease the runtime. We can also do analysis for each of the mating and mutating algorithms to determine which lead to quicker convergence. Finally, there is a lot of user choice involved in initializing the algorithm. There are multiple ways to do this using rankings from other methods or preprocessing the data itself. We believe that the EO algorithm can become a competitive method for ranking sports teams.

Acknowledgements

Dr. Chartier's research was supported in part by a research fellowship from the Alfred P. Sloan Foundation. Dr. Langville's research was supported in part by NSF grant CAREER-CCF-0546622. Dr. Yamamoto's research was supported in part by the Grant-in-Aid for Scientific Research (B) 18310101 of the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

We would also like to thank DASH for giving us use of the Xpress-MP Optimization Software through the Academic Partner Program.

References

Ali, I., Cook, W. D., and Kress, M. (1986). On the minimum violations ranking of a tournament. *Management Science*, 32(6):660–672.

Cassady, C. R., Maillart, L. M., and Salman, S. (2005). Ranking sports teams: A customizable quadratic assignment approach. *INFORMS: Interfaces*, 35(6):497–510.

Coleman, B. J. (2005). Minimizing game score violations in college football rankings. *INFORMS: Interfaces*, 35(6):483–496.

Colley, W. N. (2002). Colley's bias free college football ranking method: The colley matrix explained.

Fogel, D. B. and Michalewicz, Z. (2004). *How to Solve It: Modern Heuristics*. Springer.

Govan, A. Y., Langville, A. N., and Meyer, C. D. (2009). Offense-defense approach to ranking team sports. *Journal of Quantitative Analysis in Sports*, 5(1):1–17.

Langville, A. N., Pedings, K., and Yamamoto, Y. (2009). A minimum violations ranking method. preprint.

Massey, K. (1997). Statistical models applied to the rating of sports teams. Bachelor's thesis, Bluefield College.

Park, J. (2005). On minimum violations ranking in paired comparisons.

Reinelt, G., Grotschel, M., and Junger, M. (1984). A cutting plane algorithm for the linear ordering problem. *Operations Research*, 32(6):1195–1220.

RISK TAKING IN BADMINTON TO OPTIMIZE IN-THE-RUN PERFORMANCE

Anthony Bedford [a,c] , Tristan Barnett [b] , Monique Ladds [a]

[a] *RMIT University*
[b] *Victoria University*
[c] *Corresponding author:* *anthony.bedford@rmit.edu.au*

Abstract

In tennis the serve can be a most powerful weapon. However in badminton, the serve holds a much lower advantage in comparison to tennis, and for many players, yields a net disadvantage. Badminton's most common service used is a short serve requiring accuracy, as opposed to a long serve requiring power. This is because badminton does not allow for the advantage of a second serve on fault of the first, someway explaining the conservative nature of serving, and low success probabilities. The short serve allows the receiver to gain the advantage, putting the server under pressure on the third shot. In this work, we develop a model to ascertain whether a player should be taking a high or low risk serve. Using Bayesian models, we hypothesize how a player's performance could be optimized conditional on the state of the match in progress. Practical implications for players are discussed, given that the rules of badminton allow for coach intervention during a match in progress.

Keywords: Badminton, risk taking, coach intervention, Markov Chain

1. INTRODUCTION

There are currently three racket sports played at the Olympic Games: tennis and table tennis (which have been played since 1988) and badminton (since 1992). Tennis enjoys world wide popularity, whilst table tennis and badminton are predominately popular in Asia and Europe. Olympic Sports have the highest recognition amongst the vast number of competitive sports that now exist, and as a consequence Badminton Australia are interested in many areas of science (including mathematics) to potentially improve player performance in badminton.

It has been shown in tennis (Barnett et al., 2008) that against certain opponents on specific surfaces, players could possibly increase their chances of winning a point by taking more risk on the second serve. Pollard (2008) considered this problem and found that the range of risk in men's singles tennis serves fall into a quadratic relationship rather than a linear one. Therefore it may not always be optimal to use a 'hard' first serve and a 'soft' second serve. Some players may benefit from taking risky first and second serves, while others may improve results by playing both serves safely. Unlike tennis, only one fault is allowed in badminton, and players generally rely on a low risk short serve, which requires accuracy, rather than a long serve that requires power (Edwards et al., 2005). Badminton players will generally perform better by constantly using a low risk serve, rather than a high risk serve. However, a common strategy amongst players is to occasionally use a high risk serve to catch the opponent off-guard and possibly force them in to making a poor return. This type of strategy has been analysed in tennis using a game theory approach, in which the expectations of the opponent are taken into consideration (Hannan, 1976). In this research we analyse the strategy of when to use an occasional high risk serve throughout a match in progress to potentially enhance player performance. The concept of importance (Morris, 1977) and Bayesian models (Carlin and Louis, 2000) are used in this analysis. Given the rules of badminton allow for coaching intervention during play, this creates great opportunities for live data collection, computer analysis, and intervention which could greatly assist in improving player performance.

2. THE GAME

A badminton match is decided through the best-of-three games. A game can be won only once the score reaches 21 points. If a player reaches 21 and is two points or more ahead, they win the game. If the score reaches 20-20, play continues until one player has obtained a two point lead and is the winner. If the score reaches 29-29, the winner of the next point wins the game. A toss of the coin allows a player to choose the end that they wish to play and whether to serve or receive. The player who wins the point takes serve and thus continues into the following game, with points capable of being won on either a players serve or return of serve. Badminton rules allow for only one service; if a fault is served, the point and service is immediately won by the opponent.

3. MARKOV CHAIN MODEL

Let us consider the possible outcomes of both the game (known as a set in tennis) and winning a match.

3.1 Probability of winning a game

Firstly, to the probability of winning a game. Let p_A and p_B represent the constant probabilities of player A and player B winning a point on serve. Let $S_A(a,b)$ and $S_B(a,b)$ represent the conditional probabilities of player A winning a game, conditional on the point score (a,b), for player A serving and player B serving respectively. These probabilities can be obtained recursively as follows:

$S_A(a,b) = p_A S_A(a+1,b) + (1- p_A)S_B(a,b+1)$
$S_B(a,b) = p_B S_B(a,b+1) + (1- p_B)S_A(a+1,b)$
The boundary values are:

$S_A(a,b) = S_B(a,b) = 1$ if a=21 and b ≤19 or (a,b) = (22,20), (23,21), (24,22), (25,23), (26,24), (27,25), (28,26), (29,27), (30,28), (30,29)

$S_A(a,b) = S_B(a,b) = 0$, if b=21 and a ≤19 or (a,b) = (20,22), (21,23), (22,24), (23,25), (24,26), (25,27), (26,28), (27,29), (28,30), (29,30)
$S_A(29,29) = p_A$
$S_B(29,29) = 1-p_B$

3.2 Probability of winning a match

Given the match of badminton is the best-of-three games, the outcomes are simple to ascertain. Let m_A and m_B represent the probabilities of player A winning a match when player A and player B are serving first in the match respectively. For notational simplicity, let $S_A(0,0)=s_A$ and $S_B(0,0)=s_B$. Table 1 exhibits the probabilities of player A winning a match for the different game outcomes for when player A and player B are serving first in the match.

Game outcome	Player A serving first	Player B serving first
WW	$s_A s_A$	$s_B s_A$
WLW	$s_A (1-s_A)s_B$	$s_B(1-s_A)s_B$
LWW	$(1-s_A)s_B s_A$	$(1- s_B)s_B s_A$

Table 1: Probabilities of player A winning a match for the different set outcomes

It follows from the table that

$m_A = s_A^2 + 2s_A s_B - 2s_A^2 s_B$
$m_B = s_B^2 + 2s_A s_B - 2s_A s_B^2$

Table 2 exhibits the probabilities of player A winning a game and match for different values of p_A and p_B, and for both player A and player B serving first. It can be observed from the table that when p_A and p_B are greater than 0.5, it is an advantage to serve first in the match. Similarly, when p_A and p_B are less than 0.5, it is an advantage to receive first in the match.

p_A	p_B	s_A	s_B	m_A	m_B
0.46	0.46	0.495	0.505	0.497	0.503
0.48	0.46	0.549	0.556	0.577	0.581
0.50	0.46	0.602	0.607	0.653	0.656
0.52	0.46	0.653	0.655	0.723	0.725

0.48	0.48	0.497	0.503	0.499	0.501
0.50	0.48	0.551	0.554	0.578	0.579
0.52	0.48	0.604	0.604	0.654	0.654
0.50	0.50	0.500	0.500	0.500	0.500
0.52	0.50	0.554	0.551	0.579	0.578
0.52	0.52	0.503	0.497	0.501	0.499

Table 2: Probabilities of player A winning a game and match for different values of p_A and p_B, and for both player A and player B serving first.

3.3 Mean number of points in a game

Let $E_A(a,b)$ and $E_B(a,b)$ represent the mean number of points remaining in a game conditional on the point score (a,b) for player A and player B serving respectively. These probabilities can be obtained recursively as follows:

$$E_A(a,b) = 1+p_A\, E_A(a+1,b) + (1- p_A)\, E_B(a,b+1)$$
$$E_B(a,b) = 1+p_B\, E_B(a,b+1) + (1- p_B)\, E_A(a+1,b)$$

The boundary values are:
$E_A(a,b) = E_B(a,b) = 0$, if a=21 and b $\leq$19 or b=21 and a $\leq$19 or (a,b) = (22,20), (23,21), (24,22), (25,23), (26,24), (27,25), (28,26), (29,27), (30,28), (20,22), (21,23), (22,24), (23,25), (24,26), (25,27), (26,28), (27,29), (28,30)
$E_A(29,29) = E_B(29,29) = 1$

These derivations provide additional interest for both players and coaches during a match in progress.

3.4 Importance of a point to winning a game

Let $I_A(a,b)$ and $I_B(a,b)$ represent the importance of a point to winning a game when player A and player B are serving respectively. $I_A(a,b)$ and $I_B(a,b)$ are defined by Morris (1977) and can be obtained as follows:

$$I_A(a,b) = I_B(a,b) = S_A(a+1,b) – S_B(a,b+1)$$

Note that $I_A(a,b) = I_B(a,b)$ as a result of the rotation of serve in badminton. This result does not occur in a set of tennis since player's alternate serve at the end of each game. It can be shown that $I_A(29,29) = I_B(29,29) = 1$ since the next point decides the winner of the game, and therefore represents the highest level of importance to winning the game.

4. STRATEGIES

We consider a series of strategies to employ whilst a match is in progress.

4.1 Strategy 1 – game theory approach

As stated in the introduction, badminton players will generally perform better by constantly using a low risk serve, rather than a high risk serve. However, a common strategy amongst players is to occasionally use a high risk serve to catch the opponent off-guard and possibly catch them in to making a poor return.

Consider the following between player A (server) and player B (receiver)
a - player A serves low risk and player B is expecting a low risk serve
b - player A serves low risk and player B is expecting a high risk serve
c - player A serves high risk and player B is expecting a low risk serve
d - player A serves high risk and player B is expecting a high risk serve
Based on the above it is reasonable to assign probabilities to the outcomes of a, b, c and d with the condition that d < a < c < b. An example is given in table 3 below in a game theory matrix.

		Player B	
		expected low risk serve	expected high risk serve
Player A	low risk serve	0.55	0.60
	high risk serve	0.57	0.45

Table 3: Game theory matrix of how much risk to take on serve in badminton

Solving this two-person zero-sum game with the Minimax theorem gives mixed strategies for player A of 0.706 low risk serve, 0.294 high risk serve and for player B of 0.882 expecting a low risk serve, 0.118 expecting a high risk serve.

Let p_{iL} represent the proportion of time that player i serves a low risk serve for the match and let p_{iH} represent the proportion of time that player i serves a high risk serve for the match. It follows that $p_{iL} + p_{iH} = 1$, given they cover all scenarios. From the example above $p_{iL}=0.706$ and $p_{iH}=0.294$.

According to game theory analysis, this proportion of high risk serves should be randomized. Before the player serves, the coach could signal to the player what type of serve they should use, through a computer based selection.

This form of game theory can also be applied to tennis. For example, even though a tennis player may perform better overall by serving high risk on the first and low risk on the second serve, rather than a high risk on both the first and second, a player may further improve their performance by serving a high risk second serve a proportion of the time as a 'surprise' factor. Game theory analysis could determine the proportion of time that a player should serve out wide, down the line and into the body. Game theory analysis could also determine what proportion of the time that a player should be serve-and-volleying.

4.2 Strategy 2 – risk importance

It is convenient to analyse tennis for devising risk strategies that depend on the level of importance, and then establish the connection to badminton.

It has been established in tennis that the more important the point, the lower the probability that the server wins the point (Klaassen and Magnus, 2001). The model developed in Barnett et al. (2008) and Pollard et al. (2009), is used to determine if the server can increase their chances of winning a point by serving high risk on the second serve.

The following definitions are given for each type of serve:
A high risk serve is a typical first serve by each player
A low risk serve is a typical second serve by each player

Let:
a_{hi} = percentage of high risk serves in play for player i
b_{hi} = percentage of points won on high risk serves (conditional on them being 'in') for player i
b_{li} = percentage of points won on low risk serves (unconditional) for player i
c_{hi} = percentage of points won on return of high risk serves (conditional) for player i
c_{li} = percentage of points won on return of low risk serves (unconditional) for player i
d_{hij} = percentage of points won on high risk serves (conditional) for player i, for when player i meets player j
d_{lij} = percentage of points won on low risk serves (unconditional) for player i, for when player i meets player j
c_{ha}=average percentage (all players) of points won on return of high risk serves (conditional)
c_{la}=average percentage (all players) of points won on return of low risk serves (unconditional)

The following assumptions are given: $a_{hi} < a_{li,}$ $b_{hi} > b_{li}$, $c_{hi} < c_{li}$, $d_{hij} > d_{lij}$, $d_{hij} = b_{hi} - c_{hj} + c_{ha}$ and
$d_{lij} = b_{li} – c_{lj} + c_{la}$

The following two serving strategies are defined:
Strategy 1 – high risk serve followed by a high risk serve
Strategy 2 – high risk serve followed by a low risk serve

The percentage of points won on serve by player i by using each strategy is:
Strategy 1 – $a_{hi} * d_{hij} + (1- a_{hi}) * a_{hi} * d_{hij}$
Strategy 2 – $a_{hi} * d_{hij} + (1- a_{hi}) * d_{lij}$
Thus, player i should use Strategy 1 (two high risk serves) rather than Strategy 2 if
$a_{hi} * d_{hij} + (1- a_{hi}) * a_{hi} * d_{hij} > a_{hi} * d_{hij} + (1- a_{hi}) * d_{lij,}$
and this inequality simplifies to $a_{hi} * d_{hij} > d_{lij}$

The following is developed to determine if the server can increase their chances of winning a point by serving high risk on the second serve, conditional on the "importance" of the point. We will assume that the server's probability of winning a point on serve is affected only by serving low or high risk on the second serve.

The percentage of points won on serve by player i by using a high risk first serve and a low risk second serve is given by:
$a_{hi} * d_{hij} + (1- a_{hi}) * d^\wedge_{lij}$

The superscript ^ is used as the server's probability of winning a point on a low risk serve is now conditional on the "importance" of the point.
From above, $d^\wedge_{lij} = b^\wedge_{li} – c_{lj} + c_{la}$

The following result follows from Klaassen and Magnus (2001), where it was established that a server's probability of winning a point decreases with the more "important points". Given two score lines in tennis x_1 and x_2, if the "importance" at score line x_1 is greater than the "importance" at score line

x_2, then $b\hat{}_{li}$ and consequently $d\hat{}_{lij}$ is lesser at score line x_1 than score line x_2.

Player i should use Strategy 1 (two high risk serves) rather than Strategy 2 if $a_{hi} * d_{hij} > d\hat{}_{lij}$. This is evidence to suggest that the server would be encouraged to take more risk on the more "important" points.

Using the above information, we make the following suggestion as a strategy that could be used in men's badminton: *More high risk serves should occur on the more "important" points, and less often on the lesser "important" points.*

Let $p_{iL}(a,b)$ be the proportion of time that player i serves a low risk serve at point score (a,b) and let $p_{iH}(a,b)$ be the proportion of time that player i serves a high risk serve at point score (a,b). The following is discussed based on a game of tennis, to gain insight to obtaining values for $p_{iL}(a,b)$ and $p_{iH}(a,b)$. The average importance of a point to winning a game in tennis (I_{AV}), as defined by Morris (1977), is given as $I_{AV}=N/L$ where $N=dS(0,0)/dp$, L = mean number of points in a game of tennis, $S(0,0)$ represents the probability that the server will win the game at point score (0,0) and p represents the probability that the server wins a point. The following formula is intuitive $p_{iH}(a,b)=[I(a,b)/I_{AV}]p_{iH}$. Note that a game of badminton requires two parameters and therefore a generalized expression for average importance is required.

4.2 Strategy 3 – Bayesian

The proportion of time that a player takes a high and low risk serve can be updated based on the initial proportions and what has occurred during the match using Bayesian analysis. Consider the binomial distribution, with Y the number of events in n independent trials, and μ the event probability. The sampling distribution is defined as $P(Y = y \mid \mu) = {}^nC_y\, \theta^y(1-\theta)^{n-y}$. The posterior distribution of μ given Y is calculated in Carlin and Louis (2000), and is Beta(a,b) with mean $\theta\hat{} = M / (M+n)\ \mu + n / (M+n)\ (Y/n)$, where $M=a+b$, $\mu = a/(a+b)$. Let n_i represent the number of points served by player i, $\mu_i = p_{iH}(a,b)$, Y_i/n_i = the actual percentage of points won on a high risk serve by player i, and M = weighting parameter. Table 4 represents the updated proportion of time that player i should be using a high risk serve for different values of Y_i/n_i and M, given $p_{iH}(a,b) = 0.2$ and $n_i = 5$. Due to the "small" sample size which occurs in the first game of the match, it would appear logical that more weighting should be given towards $p_{iH}(a,b)$ initially, whereas towards the end of the match more weighting should be given to the actual values that occurred during the match. For this reason M=75 appears to be a more reasonable weighting parameter at the start of the match (compared with M=40), whereas M=40 could possibly be a more reasonable weighting parameter towards the end of the match. One method to overcome this problem is to let M_i represent the expected number of serves remaining in the match for player i, where M_i can be obtained from the Markov Chain model in section 3.

	θ^	
Y_i/n_i	M=40	M=75
0	0.18	0.19
1/5	0.20	0.20
2/5	0.22	0.21
3/5	0.24	0.23
4/5	0.27	0.24
1	0.29	0.25

Table 4: The updated proportion of time that player i should be using a high risk serve for different values of Y_i/n_i and M, given $p_{iH}(a,b) = 0.2$ and $n_i = 5$.

5. CONCLUSIONS

In this paper, we outline the simple method of evaluation of success probabilities recursively. From this, game and match success probabilities are evaluated, thereby assisting in the basic merits of differing types of serving strengths. From this, three levels of strategies have been identified in the paper as potential ways to increase player performance. Strategy 1 is a relatively simple game theory approach to determine when a player should use a high risk serve, with a clear element of surprise approach. Strategy 2 is an extension of Strategy 1. By identifying that more high risk serves should occur on the more "important" points and less often on the lesser "important" points, the adjustment allows for a more specific approach to success. Strategy 3 uses Bayesian analysis to update the initial estimates based on what has occurred during the match, and attempts to optimize in-the-run performance based on current information. All of these strategies can be implemented in a live match, since the rules of badminton allow for coaching intervention during play, and thereby offer technical approaches to success not normally possible in other racquet sports.

References

Barnett, T., Meyer, D., & Pollard, G.H. (2008). Applying match statistics to increase serving performance. *Medicine and Science in Tennis* **13**, 2, 24-27.

Carlin B.P. and Louis T.A. (2000). *Bayes and empirical* B*ayes methods for data analysis*, 2nd ed., London: Chapman & Hall/CRC.

Edwards, B., Lindsay, K. and Waterhouse, J. (2005). Effect of time of day on the accuracy and consistency of the badminton serve. *Ergonomics* **48**, 11, 1488-1498.

Hannan E.L. (1976). An analysis of different serving strategies in tennis. In *Management Science in Sports*, S.P. Ladany, R.E. Machol and D.G. Morrison eds., Amsterdam: North–Holland Publishing Company, 125–136.

Klaassen F.J.G.M. and Magnus J.R. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association,* **96,** 500–509.

Klaassen F.J.G.M. and Magnus J.R. (2007). Myths in Tennis. In *Statistical Thinking in Sports*, J. Albert and R.H. Koning eds., CRC Press, 217-237.

Morris C (1977). The most important points in tennis. In *Optimal Strategies in Sports,* S.P. Ladany and R.E. Machol eds., Amsterdam: North– Holland, 131–140.

Pollard G (2008). What is the best serving strategy? *Medicine and Science in Tennis* **13**, 2, 34-38.

Pollard, G.N., Pollard, G.H., Barnett, T., & Zeleznikow, J. (2009). Applying tennis match statistics to increase serving performance during a match in progress. *Medicine and Science in Tennis* **14**, 3, 16-19

USING IMPORTANCE TO DETERMINE A SERVICE STRATEGY IN BADMINTON

Monique Ladds[a] and Anthony Bedford[a,d]

[a] *School of Mathematical and Geospatial Sciences, RMIT University*
[d] *Corresponding author: anthony.bedford@rmit.edu.au*

Abstract

In this work we look at the optimal criteria for taking a long (risky) serve in badminton. A simulation model is built and describes the process in determining when to execute a serve based upon point importance. Importance is the likelihood of winning a game given you succeed in the next point, against winning a game given you do not win the next point. By considering different types of player serve and return combinations, we explore the variations in game probabilities to ascertain the ideal time to execute a risky serve for each approach.
Whilst many racquet sports are able to be dominated by players with the best serve, this is untrue for badminton, where the serve is not perceived as an advantage as in tennis, especially as badminton rules do not allow for a second serve in case of fault on the first. Badminton players opt for a 'safe' short serve over a 'risky' long serve. In most cases, the need for the server to get the shuttle in play immediately results in the advantage of the receiver, and pressure for the server on the third shot. The rule developed in this paper identifies when 'high risk', or long serves, and 'low risk', or short serves, should be played during a match. Practical implications for this work are possible, given that the rules of badminton allow for coach intervention during a match in progress.

Keywords: Badminton, serve, strategy, importance, simulation.

1. INTRODUCTION

Badminton is the world's fastest racquet sport. With smashes in-play reaching over 300 km/h, it easily surpasses tennis' fastest recorded shot (a serve) currently held by Andy Roddick, at 246 km/h. Badminton is both a popular recreational game and professional sport. Unlike tennis, badminton is especially strong in Asia. The Asian nations of China, South Korea, Indonesia and Malaysia dominate the Olympic medal count, and, at time of publishing, held 23 of 24 gold medals; 21 of 24 silver medals and 25 of 28 bronze medals (69 of 76 in total). Of the 63 nations to have competed for medals, only seven have medalled.

Badminton competition comprises five forms; men's and women's singles, men's and women's doubles and mixed doubles. It is played on a rectangular court marked with both singles and doubles lines. A 1.52 metre high net separates the players and a shuttle is hit back and forth between players with the aim of a player to make the shuttle fall to ground within the bounds of the opposition's court. A badminton match is decided through the best-of-three games. If a player reaches 21 and is two points or more ahead, they win the game. If the score reaches 20-20, play continues until one player has obtained a two point lead and is then declared the winner. If the score reaches 29-29, the winner of next point wins the game. A toss of the coin allows a player to choose the end that they wish to play and whether to serve or receive. The player who wins the point takes serve and thus continues into the following point, with points capable of being won on either a players serve or return of serve. Badminton rules allow for only one service; if a fault is served, the point and service is immediately awarded to the opponent.

Two types of serve are utilised in badminton. These are the short serve and the long serve. Edwards, Lindsay and Waterhouse (2005) describe the short serve as requiring "minimal strength and speed, in order to get the shuttle to 'float over the net' into the service box", and the long serve as requiring "speed and strength, to be able to put the shuttle high and down the centre of the line and to fall vertically at the back of the court". Unlike tennis, no faults are allowed in badminton, and the point is immediately lost. Players generally rely on a low risk short serve, as seen in Figure 1A, rather than a long serve as shown in Figure 1B.

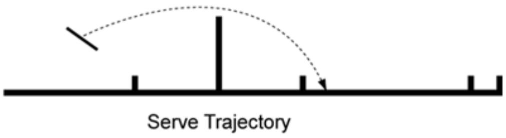


Figure 1A. 'Safe' Short serve trajectory

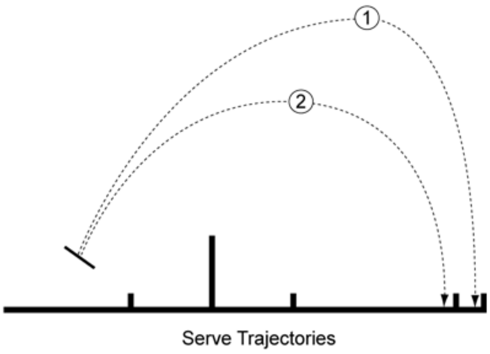


Figure 1B. 'Risky' Long serve trajectories (1) used in doubles and (2) in singles.

Current trends in men's singles badminton players reveal that a long serve is used for only 5% of all serves, although in the women's game it is used with greater frequency. These serves have different uses in matches. In singles, the long serve is used when the desire is to push your player into the back of the court. This will open up the court, however if that player is a strong attacker, this provides them with the opportunity to obtain an advantage. In doubles, this serve is used even less. The short serve is more commonly used at the highest level as players are typically strong attackers. Badminton players will generally perform better by constantly using a low risk serve, and strategically using a high risk serve sparingly. The dilemma remains as to when to choose to execute a high risk serve. Pollard (2008) is one of many papers in tennis looking at service strategy.

In this paper we utilise simultaneous simulation at fours levels of point importance by four different serve-return levels to determine the best time to execute a risky serve.

2. METHOD

In order to compare the success rate of different levels of importance, we first need to define the methods used to ascertain the probabilities of winning a game. This process was detailed in previous work Bedford, Barnett and Ladds (2010), and is covered here in brief.

Let p_A and p_B represent the constant probabilities of player A and player B winning a point on serve, and we assume that these probabilities are independent and identically distributed.

Let $S_A(a,b) = P(A\,win\,game \,|\, (a,b) \cup A\,serve)$ and $S_B(a,b) = P(A\,win\,game \,|\, (a,b) \cup B\,serve)$ represent the conditional probabilities of player A winning a game, conditional on the point score (a,b), for player A serving and player B serving respectively. These probabilities can be obtained recursively using (1) and (2) as follows:

$$S_A(a,b) = p_A S_A(a+1,b) + (1-p_A) S_B(a,b+1) \quad (1)$$

$$S_B(a,b) = p_B S_B(a,b+1) + (1-p_B) S_A(a+1,b) \quad (2)$$

The boundary values require definition, specific on the possible states of the match prior to game termination. The first boundary condition (game won by *A*) is given by

$$S_A(a,b) = S_B(a,b) = 1 \quad (3)$$

where $(a,b) \in (21,y), and\ y \le 19$ or $(a,b) \in (x,y)$ such that $(22 \le x < 30) \cap (x-y=2) or\ x=30, y=29$

The second boundary condition (game won by *B*) is given by

$$S_A(a,b) = S_B(a,b) = 0 \quad (4)$$

where $(a,b) \in (x,21), where\ x \le 19$ or $(a,b) \in (x,y)$ such that

$22 \le y < 30 \cap y-x=2\ or\ y=30, x=29$.

The third boundary state is given by

$$S_A(29,29) = p_A \quad (5)$$

The final boundary state is given by

$$S_B(29,29) = 1 - p_B \quad (6)$$

Utilising equations (1) and (2) from (3) through (6), we can essentially fill-in the probability of winning a game conditional on the current server and point score of the game at any point in the game.

2.1 Importance

It has been established in tennis that the more important the point, the lower the probability that the server wins the point (Klaassen and Magnus, 2001). It has also been shown that in tennis, at game and break point, men take extra risks on their serve and more important points occur at break points.

Following on from the definition of importance in Morris (1977), an important point is determined by

$$I(a,b) = S_A(a+1,b) - S_B(a,b+1) \qquad (7)$$

Every combination of points has an importance associated with it, and typically the closer the game the greater the importance.

Probability of winning on serve

Probabilities of winning on serve, with the exception of top level players, are typically lower than when receiving. In this work, we assume the probability that player A will win on serve is also dependent on the probability that player B will win on return.

For each point, we need to determine who is serving, and use a standardised probability set for when A is serving and B is receiving, or when B is serving and A is receiving.

We used four scenarios in our approach. The following probabilities were used:

	Scen. 1		Scen. 2		Scen. 3		Scen. 4	
	Sv	Rt	Sv	Rt	Sv	Rt	Sv	Rt
A	0.5	0.5	0.5	0.65	0.35	0.8	0.3	0.65
B	0.5	0.5	0.45	0.7	0.45	0.65	0.6	0.8

Table 1. Probabilities of winning on serve (Sv) and return (Rt) player A and B

From Table 1, we note that scenario 1 is a balanced match, whereby there is no advantage in serving or receiving. This is not a true representation of a typical player's performance, as most players are slightly better at one or the other. In Scenario 2, we have the strong return cases, with player A slightly better server than B, and the reverse in terms of receiving. Scenario 3 is the case with a weak serve but strong return for player A, and scenario 4 the dominant player B.

From Table 1 we arrive at the following inputs for the simulation as detailed in Table 2.

	Scen. 1	Scen. 2	Scen. 3	Scen. 4
p_A	0.50	0.42	0.35	0.27
p_B	0.50	0.41	0.36	0.48

Table 2. Simulation inputs.

Risk

A risk here is defined as executing a long serve, also referred to as a high risk shot. We assume that it is only completed by player A, as this is the player, in practise, who will be coached to take a risk, based on this research. In our simulation, taking a risk increases a player's probability of winning a point. In this case, a risk improvement of .45 is added when the player first plays a long serve. After this, the improvement reduces each time a long serve is played by .05, making the shot redundant after using it nine times.

The increase of .45 typically guarantees that Player A will win their serve when playing a risky serve for the first time. The time to execute the risky serve is governed by the importance of the next point. However one must consider what that level of importance is before executing such a serve, hence our use of simulations.

The values in this work are based on evidence collected by the Australian coach, however long term player-based analysis is critical to a successful implementation, and this work forms a basis for more detailed future modelling.

Simulation

We built the simulation in Excel using @Risk. For every single simulation run, we generate $X_i \sim U(0,1)$ values for all states of the game (i.e. up to 59 possible points for 3 games, rarely all needed). If the seed generated for each point *i* is less than the probability of Player A winning that point (taking into account if a risky serve is taken) then player A wins. The simulator was specified so that a single value from X_i was used for both Risk and No Risk serves, thereby allowing the best possible direct comparison for every point. So we set up a simultaneous simulation to attempt to best determine the net effect. Nonetheless, once a risky serve is taken, the result of the match is more than likely to change in some way that negates this approach. For example, if the first game is at 17-18, and we execute a risk, the outcome may change to a win for player A so that the score becomes 18-18. However the no risk simulation may yield a 17-19 score. Thereon in the match is not perfectly matched, as we do not then simulate the remaining no risk versus risk from 18 – 18. Furthermore, there is no conditionality on the success of risks. That is, a coach would be reluctant to continually encourage a player to execute more risk serves if they all fail!

A screen shot of the summary window is given in Figure 2. This illustrates the outcome of one iteration of a simulation. In this example, the no risk match (our pseudo-control) yielded a loss to Player A in three games, whereas in the risk match (where a risk was taken when $I > 0.25$) the match was won. In this match, the final game was won with the execution of two risky serves.

In this work, we look at setting up the problem and exploring some solutions, and leave the overall general solution for further work.

The Importance values used are as follows:

Simulation	1	2	3	4
K	0.1	0.25	0.4	1

Table 3. Simulations using $I(a,b)>K$ to determine Risky serve execution for Player A

In this work, we ran 16 simulations of 10,000 simultaneous matches (risk v no-risk). Simulation 1 is a low importance trigger level of 0.1; this can be viewed as executing a serve very early in a game in order to save a match before it gets away. Simulation 3 might be used if only a match is dire.

One could argue that simulation 3 could be a suitable criterion for good players against poor opponents; and simulation 2 for a more conservative player. Simulation 4 was our testing value as Importance can never exceed 1.

As a summary, the method is as follows:

At (0,0), $R = 0.45$.

For all (a,b)

If $I(a,b) > K$ then $p_{A,win} = max(p_A, p_A + R)$; and $R=R-0.05$ else $p_{A,win} = p_A$

If $p_{A,win} < X_i$ then player A wins point else B wins.

Figure 2 provides a case where two risks are taken in the final game to win the match.

All simulations were conducted 10,000 times for each of the four levels of K. Simulation 4, where K = 1, was used to validate the accuracy of the no risk case. That is, the results were compared to the theoretical values of winning the match (see Bedford et. al.(2010) for these calculations).

There are six possible outcomes for each paired match simulated in terms of player A. We define the possible outcome of a match within a simulation as follows:

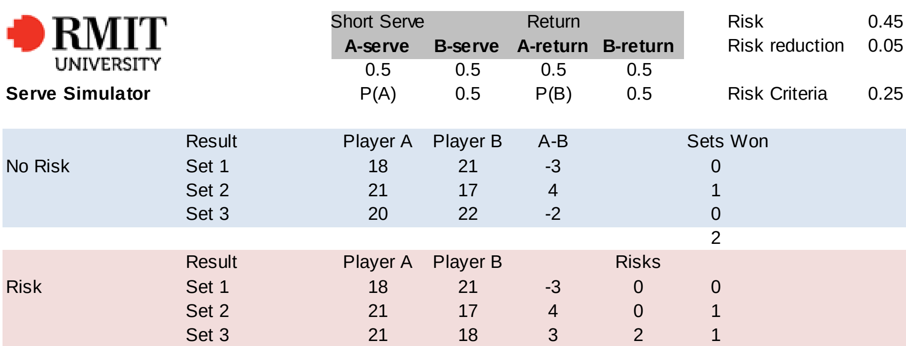
Let $\left(m_i^{NR}, m_i^{R}\right)$ describe the outcome of match i, where $i = 1,..,10000$; and NR = no risks taken and R = opportunity to take a risk when $I > K$.

So from this we have the following possible outcomes:

$$m_i^{NR} \in \{Win, Loss\} and$$

$$m_i^{R} \in \{Win \mid R, Win \mid NR, Loss \mid R, Loss \mid NR\}.$$

The matching of these events is of interest, and we can obtain the following six feasible pairings:



RMIT UNIVERSITY		Short Serve		Return			
		A-serve	B-serve	A-return	B-return	Risk	0.45
		0.5	0.5	0.5	0.5	Risk reduction	0.05
Serve Simulator		P(A)	0.5	P(B)	0.5	Risk Criteria	0.25
	Result	Player A	Player B	A-B		Sets Won	
No Risk	Set 1	18	21	-3		0	
	Set 2	21	17	4		1	
	Set 3	20	22	-2		0	
						2	
	Result	Player A	Player B		Risks		
Risk	Set 1	18	21	-3	0	0	
	Set 2	21	17	4	0	1	
	Set 3	21	18	3	2	1	

Figure 2: Simulation 2 screenshot of an iteration for $K = 0.25$, Scenario 1

P1. $(Loss, Loss \mid R)$ = Risk but no change Loss

P2. $(Win, Win \mid R)$ = Risk but no change Win

P3. $(Loss, Win \mid R)$ = Risk changes Loss to Win

P4. $(Win, Loss \mid R)$ = Risk changes Win to Loss

P5. $(Loss, Loss \mid NR)$ = Loss

P6. $(Win, Win \mid NR)$ = Win

The comparisons of interest are those between the first four pairings and the final two. The proportional sum of all Win and Loss states used in each simulation should converge to the stationery probabilities obtained utilising (1) – (6), and for all outcomes when K =1.

3. RESULTS

We first consider the outcomes of the matches for Scenario 1, when the players are equal in quality. In the following tables, Win = P2 + P3 + P6, and Loss = P1 + P2 + P5.

3.1 Scenario 1

As seen in Table 4, great improvement is yielded (against K = 1) when K = 0.25 and 0.4, with smaller improvement for K = 0.1.

	K			
Outcome	1	0.4	0.25	0.1
Win	0.500	0.624	0.623	0.565
Loss	0.500	0.376	0.377	0.435

Table 4. Probability of win/loss by Simulation for Scenario 1

Within these simulations, we exhibit the proportion of outcomes by the six pairings graphically. We only display the 0.25 and 0.4 graphs, as 0.1 yielded a large amount of undesirable outcomes. Firstly, every match saw the execution of a risky serve, which still improved the outcomes of player A winning, however yielded a large number of P4 (1,670 of 10,000). A basic analysis of the costs for levels of K is given in Table 5. We first define the following statistics:

Let RiskCost = P4/(P4+P1), so the proportion of Losing matches when a risk was taken paired to a match where a win resulted without risk / all matches where a loss occurred when a risk was taken; RiskBenefit = P3/(P3 + P2), the proportion of Winning matches when a risk was taken paired to a match where a loss resulted without risk/ all matches where a win occurred when a risk was taken; and Ratio = RiskBenfit / RiskCost.

We obtain the following:

	K			
Outcome	1	0.4	0.25	0.1
RiskCost	0	0.090	0.149	0.384
RiskBenefit	0	0.232	0.269	0.417
Ratio	-	2.588	1.802	1.085

Table 5. Costs by Simulation for Scenario 1

So in Table 5, K=0.4 yields the best result, with a lower RiskCost than K=0.25 and overall higher ratio. Looking at Figures 3A and 3B we note that when number of risks = 0, the risk changes lines are 0 as they cannot occur, and No Change L and No Change W denote the proportion of P5 and P6.

From table 4 we note that player A is able to increase their chances of winning by taking risks. Figures 3.1A – 3.1B display the proportion of not only the No Change (P1 + P5 and P2 + P6) but Risk Changes L to W (P3) and Risk Changes W to L (P4) for scenario 1 with K = .25

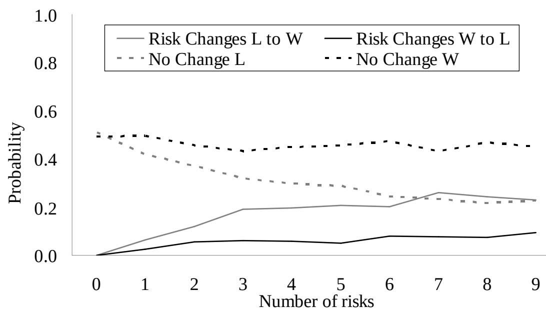


Figure 3.1A. Probability of winning given number of risks taken for Scenario 1, K = .25

Some similarity is seen in figures 3.1A and 3.1B in terms of the Changes lines.

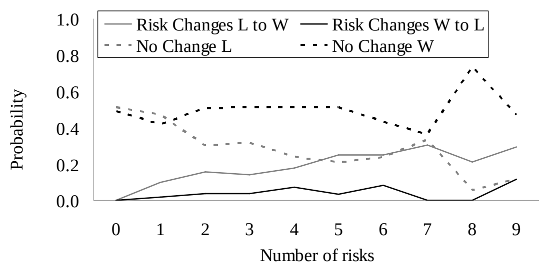


Figure 3.1B. Probability of winning given number of risks taken for Scenario 1, K = .4

The notable similarity is between the Changes lines, with matches requiring a high number of risks yielding a higher proportion of changed results to the paired match. In a balanced match, the use of *K*=0.4 appears the optimal result.

3.2 Scenario 2

In Scenario 2, Player A has a slight advantage over player B.

As seen in Table 6, player A has a nice advantage without utilising any risks. This does improve substantially when any risk is taken, with again great improvement when *K* = 0.25 and 0.4, with smaller improvement for *K* = 0.1.

	K			
Outcome	1	0.4	0.25	0.1
Win	0.539	0.644	0.655	0.602
Loss	0.461	0.356	0.345	0.398

Table 6. Probability of win/loss by Simulation for Scenario 2

In terms of the specifics of the benefits, we exhibit table 7.

	K			
Outcome	1	0.4	0.25	0.1
RiskCost	0	0.072	0.151	0.345
RiskBenefit	0	0.243	0.285	0.351
Ratio	-	3.389	1.884	1.016

Table 7. Costs by Simulation for Scenario 2

Notably, *K* = 0.4 returns the highest ratio by some margin. A look at the specifics for *K* = 0.25 and 0.4 are detailed in Figures 3.2A and 3.2B.

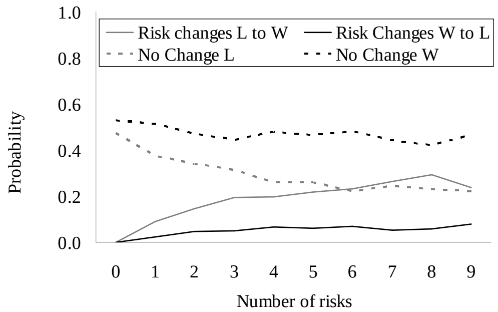


Figure 3.2A. Probability of winning given number of risks taken for Scenario 2, *K* = .25

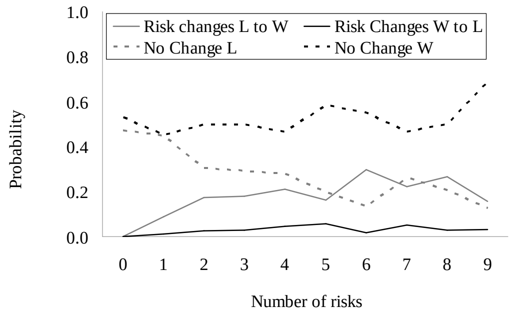


Figure 3.2B. Probability of winning given number of risks taken for Scenario 2, *K* = .4

Similarly to Figures 3.1, we see *K* = .4 yield the higher Change to win. The peak occurs around six risks.

3.3 Scenario 3

Beginning with a Player A disadvantage, it is interesting to see that by using risk greatly enhances a players chances.

As seen in Table 8, taking a large number of risks improves the chances of player A dramatically.

	K			
Outcome	1	0.4	0.25	0.1
Win	0.469	0.602	0.604	0.549
Loss	0.531	0.398	0.396	0.451

Table 8. Probability of win/loss by Simulation for Scenario 3

In terms of the specifics of the benefits, we exhibit table 9.

	K			
Outcome	1	0.4	0.25	0.1
RiskCost	0	0.070	0.175	0.305
RiskBenefit	0	0.272	0.335	0.404
Ratio	-	3.898	1.916	1.324

Table 9. Costs by Simulation for Scenario 3.

The largest ratio yet is shown in Table 9 (3.898). Notably the risk cost is near the same as in Scenario 2, with slightly higher benefit. We provide the detailed plots in figure 3.3A and 3.3B.

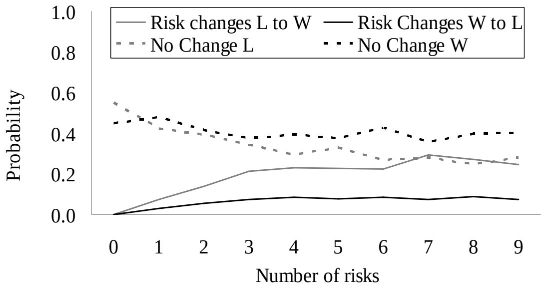


Figure 3.3A. Probability of winning given number of risks taken for simulation 3, K = .25

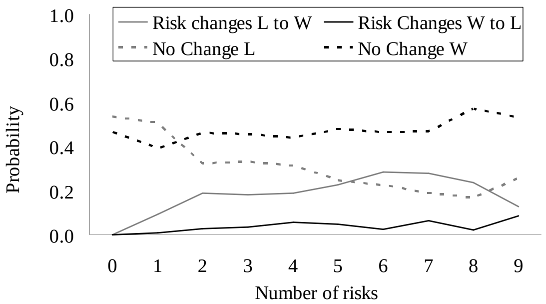


Figure 3.3B. Probability of winning given number of risks taken for simulation 3, K = .4

Once again, we see in figure 3.3A and 3.3B that K = .4 delivers better results, with the exception of nine risks. Overall is performs much better than K=.25.

We now look at the final scenario.

3.4 Scenario 4

This scenario is of great interest as player A has a major disadvantage.

As seen in Table 10, taking any risk improves the chances of player A. This result is not unexpected as the chances of

	K			
Outcome	1	0.4	0.25	0.1
Win	0.039	0.162	0.145	0.081
Loss	0.961	0.838	0.855	0.919

Table 10. Probability of win/loss by Simulation for Scenario 4

In terms of the specifics of the benefits, we exhibit table 11.

	K			
Outcome	1	0.4	0.25	0.1
RiskCost	0	0.004	0.014	0.008
RiskBenefit	0	0.320	0.518	0.531
Ratio	-	79.937	37.472	63.742

Table 11. Costs by Simulation for Scenario 4.

The ratios in Table 11 are very large. This is expected given the very low likelihood of winning.

We provide the detailed plots in Figure 3.4A and 3.4B.

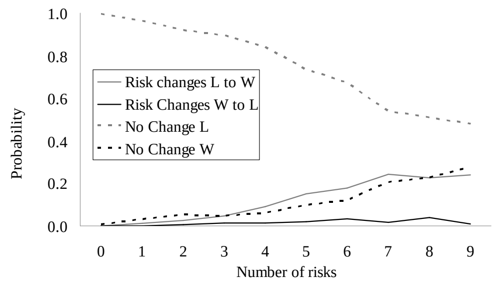


Figure 3.4A. Probability of winning given number of risks taken for Scenario 4, K = .25

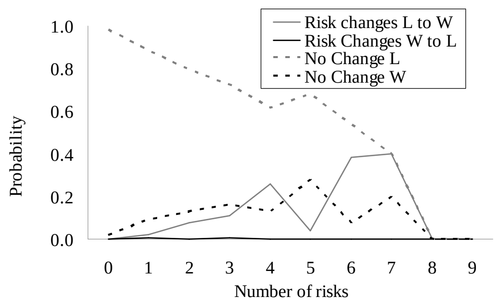


Figure 3.4B. Probability of winning given number of risks taken for Scenario 4, K = .4

3.5 Comparison

The level of the trigger for a long serve, K, was checked at several levels as detailed in Tables 3-10. Each of these levels determined when Player A would take a risk for all scenarios. Given the 0.25 and 0.4 values clearly the two optimal values, we plot the number of risks by each of these triggers for Scenario 1.

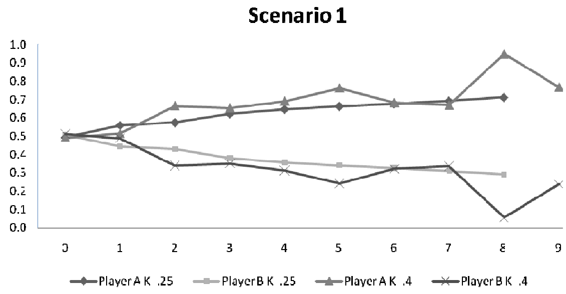


Figure 4. Winning probabilities for number of risks for scenario 1 by K = .25 and .4.

We note that the results for .4 are at or higher than .25 for player A, with the exception of 1 risk.

4. DISCUSSION

By considering four scenarios for four levels of K, we identified that .4 has the most consistent results to be deemed the better criteria for risky serve execution. However, this finding is clearly restricted.

The implementation of this trial involves three values of a trigger for long serve based on importance; and a check value. Ideally we would need to optimise this value given the results.

So, we have not varied K; we have simply selected values to hone in on a local maximum.

Further work is needed to vary K for differing values of p_A and p_B. These probabilities are assumed independent identically distributed, which, in all likelihood, may not turn out to be true, especially in light of the work of Klaassen and Magnus (2001). We believe the use of a fixed probability is a good proof concept; however a Bayesian approach seems the appropriate step forward.

Also, the premise of this model is to execute a high risk serve when points are important – what is to say that unimportance could also be used? That is, we could fit a simulation to improve winning based on risky serves taken on an unimportant point threshold.

5. CONCLUSION

In this research, we demonstrated components of our simulation model that calculates the likelihood of winning a badminton game based on importance levels. In further work, we aim to compare this method to a simpler both points behind model. This simply involves counting to see if a player falls x points behind before executing a risky serve. It is simpler to use points behind given there is no need for technology, although we presume this would result in taking a large number of risks earlier in a match.

Nonetheless, we hope to trial this model with some of the athletes, in conjunction with Badminton Australia, to determine the on-court practicalities of such an idea.

Acknowledgments

The authors wish to thank the ongoing support of Lasse Bundgaard, head coach of Badminton Australia, and the atheletes.

References

Bedford A, Barnett T and Ladds M (2010). *Risk taking in badminton to optimize in-the-run performance*. In proceedings of the Tenth Australian Conference on Mathematics and Computers in Sport, Bedford, A. and Ovens, M. (eds.), Darwin, 5-7 July, 2010, p. 21-26.

Edwards, B.J., K. Lindsay, and J. Waterhouse (2005), Effect of time of day on the accuracy and consistency of the badminton serve. Ergonomics, 48: p. 1488-1498.

Klaassen, F.J.G.M. and J.R. Magnus (2001), Are Points in Tennis Independent and Identically Distributed? Evidence from a Dynamic Binary Panel Data Model, Journal of the American Statistical Association, 96, 500-509.

Morris, C. (1977). The most important points in tennis. In Ladany, S.P and Machol, R.E. (Eds.), Optimal strategies in sport (pp. 131-140), North-Holland Publishing Company: Amsterdam.

Pollard, G. (2008) What is the best serving strategy? Medicine and Science in Tennis, 13(2): p. 34-38.

THE EFFICIENCY OF SCORING SYSTEMS WITH OUTCOME DEPENDENT POINTS

Pollard, Graham [a, c] and Pollard, Geoff [b]
[a] *Faculty of Information Sciences and Engineering, University of Canberra, Australia*
[b] *Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia*
[c] *Corresponding author: graham@foulsham.com.au*

Abstract

In almost all modelling of sports such as tennis, squash, badminton, volleyball, …, it is assumed that point probabilities are constant, and points are independent. Mathematics has been applied to find such things as the mean and variance of the number of points in the match, the probability that each player wins the match, and other characteristics. Indeed, Miles (1984) developed a very elegant theory of the relative efficiency of different scoring systems at correctly identifying the better player, assuming that points were independent.
However, there is some evidence that in tennis for example this assumption may not be *strictly* true (Klaassen & Magnus, 2001; Pollard, Cross & Meyer, 2006). In squash when the 'server only scores' rule was used, it was generally accepted that the better of two players would often 'lift his/her p-value' when he/she had lost the previous point. The major objective of this research was to investigate whether Miles' work on the efficiency of scoring systems could possibly be extended to the more general situation in which points are dependent in some manner. Thus, the focus in this paper is on the efficiency of scoring systems with one-step dependent points.

Keywords: Correlated points, 1-step dependent points, importance of a point, play-the-loser, play-the-winner.

1. INTRODUCTION

In almost all mathematical modelling of sports scoring systems for various sports, it is assumed that point probabilities are constant, and points are independent. For example, Pollard (1983), using a methodology applicable to many sports, found the mean and variance of the number of points in a best-of-three (advantage or tiebreak) sets match of tennis, and evaluated the probability that each player wins. Miles (1984) developed a very elegant theory of the relative efficiency of different scoring systems, assuming points were independent.
There is some evidence that points in tennis are not *strictly* independent (Klaassen & Magnus, 2001; Pollard, Cross & Meyer, 2006). It would appear however that the evidence in tennis is probably not strong enough to seriously question the general conclusions reached when using the independence assumptions. When the 'server only scores' rule was used in squash, it was generally accepted that the better of two players 'lifted' his/her p-value when receiving (i.e. having lost the previous point and was at risk of losing a point, rather than just the right to serve).
Whilst in some particular sports most players may not have the capacity to play in a non-independent manner, it may be that some players do have that capacity (Pollard, 2004). Further, a scoring system may be used for (say) two sports, and independence may be a reasonable assumption for one sport but not for the other one. For these reasons the analysis in this paper is of a general nature, and considers families of scoring systems.
Miles (1984) considered the efficiencies of scoring systems when points were independent. For 'win-by-n' (W_n) scoring systems in which player A has a constant probability p of winning every point (unipoints), it can be shown that the probability player A wins, P, and Q (= 1 - P) and μ satisfy the equations

$$(P-Q)/\mu = (p-q)/n$$

$$(P/Q) = (p/q)^n$$

where q = 1-p. Also, W_n has the constant probability ratio (cpr) property (Pollard, 1992). That is, the ratio of the probability that player A wins in n + 2m points divided by the probability that he loses in n + 2m points (m = 0, 1, 2, …) is constant, and is equal to P/Q. Using Wn (points) as the family of scoring systems with unit efficiency, Miles (1984) showed that the efficiency of a general 'unipoints' scoring system with key characteristics P and μ was

$$\rho = \frac{(P-Q)\ln(P/Q)}{\mu(p-q)\ln(p/q)} \quad (1)$$

Thus, ignoring the factors involving p (and q), ((P-Q)/μ)*ln(P/Q) is the measure of relative efficiency for a unipoints scoring system *given underlying independent points*.

Miles (1984) also considered systems relevant to tennis (and volleyball), which he called 'bipoints' scoring systems. He assumed that the probability player A (B) wins a point on service is p_a (p_b), and points are independent. Noting the work of Wald (1947) and using Wn (point-pairs) as the standard family of scoring systems with unit efficiency, he showed that the efficiency of a general bipoints scoring system with key characteristics P and μ was given by

$$\rho = \frac{2(P-Q)\ln(P/Q)}{\mu(p_a - p_b)\ln(p_a q_b / p_b q_a)} \quad (2)$$

where q_a = 1 - p_a and q_b = 1 - p_b. Thus, the measure of relative efficiency is also ((P-Q)/μ)*ln(P/Q) for a bipoints system *with underlying independent points*. Pollard and Pollard (2008) showed that this is also the measure of relative efficiency for independent quad-points (tennis doubles).

The major objective of this research was to consider extending this efficiency measure for unipoints and bipoints to the more general situation in which points are one-step dependent.

2. METHODS

Unipoints with dependencies, and the Wn system

Under unipoints with dependencies, player A's probability of winning a point is equal to p_w = p + ε after winning the previous point, and is equal to p_l = p – ε after losing the previous point. Here ε is the incremental (decremental) effect on player A's probability of winning a point by having won (lost) the previous point.

Example 1. Consider the W_2 scoring system, which is the system used in a game of tennis once 'deuce' has been reached. With two equal players and ε = 0.1, the initial point probability is p = 0.5, and thereafter it is either p_w = 0.6 after a point has been won by player A, or it is p_l = 0.4 after a point has been lost by player A. This is a case of 'positive' dependency from one point to the next (ε > 0). There are 5 possible 'live' states for this scoring system (the starting state, +1win, -1loss, 0win and 0loss using an obvious notation with the win or loss referring to the previous point outcome for player A). The importance of a point (Morris, 1977) is defined as the probability that player A wins under the scoring system given he wins that point minus the probability that he wins given he loses that point. Using recurrence methods, it turns out that the importance of each of the above five states, I, is equal to 1/(2(1 - ε)) which equals 5/9 when ε = 0.1 (and p = 0.5). Correspondingly, when ε = 0.1 (and p = 0.5), the importance of every point in W_3 is 5/13, and the importance of every point in W_4 is 5/17,….
In general, for this 'unipoints with 1-step dependency' structure, the importances of every state in Wn are equal when p = 0.5.

A fundamental equation for scoring systems

Suppose I_i is the importance of the i^{th} non-absorbing state within a scoring system when the players are equal. Now suppose there is a shift (from the 'equal-players' situation) in favour of player A of δ in the 'p-value' for every point. Then, after this shift, P, μ, the expected number of times state i is visited in one realization of the scoring system, n_i , and, the proportion of time spent in state i, π_i, satisfy

$$P = 0.5 + \sum_i n_i I_i \delta$$

$$(P-Q)/\mu = 2\sum_i \pi_i I_i \delta$$

for a fair scoring system, where a fair scoring system has the property that P = 0.5 for two equal players. Also, if every point in the scoring system is equally important when the players are equal, we have

$$(P-Q)/\mu = 2I\delta$$

a useful equation (Pollard, 1992).

The 'module' method for W_n systems (n large)

Some complex W_n systems can be decomposed into smaller independent components called 'modules', which can in turn be analysed to produce values from which the asymptotic values of P, μ and ρ can be derived (Pollard, 1990). This method is useful when analysing scoring systems with outcome dependent probabilities.

Within a W_n system, these modules are steps in a general one-dimensional walk in discrete time. Using the approach and notation of Cox and Miller (1965, p. 46-58), the steps in the random walk, Z_i, are mutually independent random variables on the integers …, -2, -1, 0, 1, 2,… and the moment generating function (mgf) of Z_i is defined by

$$f^*(\theta) = \sum_{j=-\infty}^{\infty} \exp(-j\theta) P(Z_i = j)$$

If P(Q) represents the probability of absorption in states [a, infinity) ((-infinity, -b]), then, neglecting any excess over the barriers,

$$P/Q = (1 - \exp(\theta_0 b))/(\exp(-\theta_0 a) - 1)$$

when $E(Z_i) \neq 0$, where θ_0 is the non-zero solution of the equation $f^*(\theta) = 1$. For the W_n scoring system, we set a and b equal to n. Using this module approach, the efficiency of scoring system 1, SS_1 relative to SS_2 is given by the ratio

$$(\theta_{01} E(S_1)/E(D_1))/(\theta_{02} E(S_2)/E(D_2)) \qquad (3)$$

where E(S) is the expected shift (in points) in favour of player A in the playing of one module, and E(D) is the expected number of points played in one module (Pollard, 1990). Note that the expression ((P-Q)/μ)*ln(P/Q) converges to $(\theta_0 E(S))/E(D)$ as n tends to infinity.

The efficiency of unipoints (with dependencies) scoring systems

We firstly set up a family of scoring systems against which the efficiency of any unipoints (with dependencies) general scoring system can be measured. Noting that the first point with probability p occurs only once, and that the remainder of the time player A's probability of winning a point is either p_w or p_l, we can consider the scoring system family $W_1(W_n^{pw}, W_n^{pl})$, where the superscript refers to player A's probability of winning the first point in that section of the scoring system. We consider this system not only because it is a natural system against which to compare others, but also because it has the cpr property, and this allows a general formula for efficiency to be constructed. This system has the key characteristics

$$P/Q = (p_w^{\,2n-1} p_l)/(q_l^{\,2n-1} q_w)$$

$$\mu = 2(P-Q)(1 + (n-1)(q_w + p_l))/(p_w - q_l)$$

Noting that the efficiency of the scoring system SS_1 with key characteristics P_1 and μ_1 relative to the scoring system SS_2 with key characteristics P_2 and μ_2 is given by μ_2 / μ_1 when $P_1 = P_2$, we have

$$2n - 1 = (\ln(P/Q) - \ln(p_l/q_w))/\ln(p_w/q_l)$$

$$\rho = ((P-Q)/\mu)(2 + (q_w + p_l)F/(p_w - q_l)$$

$$\text{(where } F = \frac{\ln(P/Q) - \ln(p_l/q_w)}{\ln(p_w/q_l)} - 1) \qquad (4)$$

where ρ is an expression for the efficiency of a general unipoints (with dependencies) scoring system with key characteristics P and μ. Note firstly that this expression reduces to the above expression for the efficiency of unipoints systems (without dependencies) when $p_w = p_l = p$. Secondly, for large n, (P - Q) tends to unity and μ and ln(P/Q) are large, and ρ tends to ((P – Q)/μ)*ln(P/Q) multiplied by a function of the parameters p_w and p_l. Thus, when comparing the efficiencies of two unipoints (with dependencies) scoring systems with large means, we simply need to compare their values of ((P – Q)/μ)*ln(P/Q) or their values of $(\theta_0 E(S))/E(D)$.

Example 2. Consider W_2 with ε = 0.1. For two unequal players with p = 0.6, p_w = 0.7, and p_l = 0.5, recurrence relations can be used to show that player A's probability of winning is equal to P = 91/134 and the mean duration is equal to μ = 216/67 =3.2239 points. Recalling from Example 1 that the importances I of all the points/states were 5/9 when ε = 0.1, it can be seen that (P - Q)/ μ = (p - q)*I = 1/9. Also, ρ = 0.9821 (a little less than unity as the first point is a p point rather than a p_w or p_l point), and ((P-Q)/μ)*ln(P/Q) = 0.0833. (When p = p_w = p_l = 0.6 (no dependency), ((P-Q)/μ)*ln(P/Q) = 0.0811, and ρ = 1.)

Example 3. Consider the scoring system W_3 with unequal players (and p = 0.6, p_w = 0.7, and p_l = 0.5). Recurrence relations or absorbing Markov chain methods can be used to show that player A's probability of winning is equal to P = 0.7482 and the mean duration is equal to μ = 6.4530 points. Recalling from Example 1 that the importances were equal to 5/13 when ε = 0.1, it can be seen that (P - Q)/ μ = 0.0769 = 1/13 = (p - q)*I. Also, ρ = 0.9903 and ((P-Q)/μ)*ln(P/Q) = 0.0838, both being slightly greater than their values in Example 2. It follows that W_3 is slightly more efficient than W_2. In fact, it can be shown that ρ and ((P-Q)/μ)*ln(P/Q) both increase as the n in W_n increases. This is related to the fact that the proportion of the time spent in the initial state gets smaller as n increases. (When p = p_w = p_l = 0.6 (no dependency) ((P-Q)/μ)*ln(P/Q) = 0.0811, and ρ = 1, as above.)

Example 4. For the scoring system W_n (with n large), consider the module beginning with a p_w point and finishing when the next p_w point is about to be played. This has a 'geometric' structure, and $E(D) = p_w + (q_w(1 - p_l^2))/(p_l q_l)$ which equals 1.6 when $p_w = 0.7$ and $p_l = 0.5$. Further, $E(S) = p_w - (q_w q_l)/p_l = 0.4$, and so $E(S))/E(D) = 1/4$. The mgf of this module is equal to $p_w*Exp(-\theta) + (q_w p_l)/(1 - q_l*Exp(\theta))$, which when equated to unity gives $\theta_0 = \ln(p_w/q_l) = 0.3365$. It follows that $(\theta_0*E(S))/E(D)$ [the asymptotic version of $((P-Q)/\mu)*\ln(P/Q)$] is equal to 0.0841, and this is slightly bigger than the values in Examples 2 and 3, as expected. The efficiency ρ is obviously equal to unity.

Example 5. Consider W_n (with n large) for the case in which the outcome dependency is negative ($\varepsilon = -0.1$, $p = 0.6$, $p_w = 0.5$, and $p_l = 0.7$). Here $E(S)/E(D) = 1/6$ (smaller than above), whilst $\theta_0 = \ln(p_w/q_l) = \ln(0.5/0.3) = 0.5108$ (larger than above). The efficiency ρ is obviously unity and $(\theta_0*E(S))/E(D) = 0.0851$, which is larger than its value in Example 4. An interpretation of this slightly higher value (0.0851 Vs 0.0841) is, per point played, the negative dependency situation is slightly more effective than the positive dependency one at identifying the better player. When there is no outcome dependency and $p_w = p_l = 0.6$, $(\theta_0*E(S))/E(D) = 0.0811$, which is less than both of the above cases. Thus, for the W_n system, both the positive and negative dependency cases are, per point played, slightly more effective at identifying the better player than is the independence case. That is, the underlying 'structural variation' or dependency that has been considered in our model actually increases the probability that the better player wins (per point played).

Unipoints with dependencies, and the B_{2n-1} system

Table 1 gives the probability that player A wins, the unconditional and conditional means and variances of duration, and the efficiency of the scoring system B_{2n-1}. Several things are observable, including

(i) P increases as the dependency (either positive or negative) increases

(ii) as expected, μ decreases as the dependency becomes more positive, and it increases as the dependency becomes more negative

(iii) for B_5, the variance increases as the dependency becomes more positive, and it decreases as the dependency becomes more negative

(iv) $Exp = (P-Q)/\mu)*\ln(P/Q)$ and ρ increase as the dependency becomes more positive, as in the W_n system (see examples 2, 3 and 4)

(v) $Exp = (P-Q)/\mu)*\ln(P/Q)$ and ρ decrease as the dependency becomes more negative, unlike in the W_n system.

We note that if the family of scoring systems against which the efficiency of unipoints (with dependencies) is measured was $W_n(W_2^{pw}, W_2^{pl})$, then the efficiency ρ' is $((P-Q)/\mu)*\ln(P/Q)$ times a function of p_w and p_l. Thus, given two scoring systems SS_1 and SS_2, and (p_w, p_l), it is clear that $\rho(SS_1) < \rho(SS_2)$ if and only if $(P_1 - Q_1)/\mu_1)*\ln(P_1/Q_1) < (P_2 - Q_2)/\mu_2)*\ln(P_2/Q_2)$.

Efficiency of bipoints (with dependencies)

Suppose player A's probability of winning a point when serving is equal to $p_{a+} = p_a + \varepsilon$ after winning the previous point on service and $p_{a-} = p_a - \varepsilon$ after losing the previous point on service, and player B's probability of winning a point when serving is equal to $p_{b+} = p_b + \varepsilon$ after winning the previous point on service and is equal to $p_{b-} = p_b - \varepsilon$ after losing the previous point on service.

We firstly set up a scoring system against which the efficiency of any bipoints with dependencies general scoring system can be measured. Noting that the first a point and the first b point occur only once each, and the remainder of the time player A's probability

SS	P, p_w, p_l	P	μ_w	σ_w^2	μ_l	σ_l^2	μ	σ^2	Exp	Eff, ρ
B_3	0.6, 0.6, 0.6	0.648	2.4444	0.2469	2.5454	0.2479	2.48	0.2496	0.0728	0.8982
B_3	0.6, 0.61, 0.59	0.6480	2.4352	0.2458	2.5341	0.2488	2.47	0.2491	0.0732	0.9015
B_3	0.6, 0.7, 0.5	0.65	2.3538	0.2286	2.4286	0.2449	2.38	0.2356	0.0780	0.9185
B_3	0.6, 0.59, 0.61	0.6480	2.4537	0.2479	2.5568	0.2468	2.49	0.2499	0.0726	0.8946
B_3	0.6, 0.5, 0.7	0.65	2.5386	0.2485	2.6571	0.2253	2.58	0.2436	0.0720	0.8510
B_5	0.6, 0.6, 0.6	0.6826	3.9873	0.6201	4.2339	0.5824	4.0656	0.6213	0.0688	0.8478
B_5	0.6, 0.61, 0.59	0.6826	3.9704	0.6237	4.2146	0.5922	4.0479	0.6266	0.0691	0.8514
B_5	0.6, 0.7, 0.5	0.6855	3.8155	0.6392	4.0302	0.6652	3.883	0.6573	0.0744	0.8781
B_5	0.6, 0.59, 0.61	0.6826	4.0042	0.6161	4.2529	0.5723	4.0831	0.6156	0.0685	0.8441
B_5	0.6, 0.5, 0.7	0.6855	4.1539	0.5679	4.4181	0.4722	4.237	0.5528	0.0682	0.8055

Table 1: Some characteristics of the B_3 and B_5 systems with dependent points

of winning an a point is either p_{a+} or p_{a-}, and player B's probability of winning a b point is either p_{b+} or p_{b-}, we can consider the scoring system $W_n(W_1(W_2^{a+}, W_2^{a-}), W_1(W_2^{b+}, W_2^{b-}))$, (n = 1, 2, 3, …). We consider this system (even though it is not a particularly efficient one) as the basis against which to compare various other systems because this system has the cpr property and this allows a general formula for efficiency to be evaluated. This scoring system consists of two main components. The first has only a+ and a- points and has the cpr property. It can be seen that

$P_1/Q_1 = (p_{a+}{}^3 p_{a-})/(q_{a-}{}^3 q_{a+})$, and

$$\mu_1 = 2(P_1 - Q_1)(1 + (q_{a+} + p_{a-}))/(p_{a+} - q_{a-})$$

where P_1 is player A's probability of winning this first component, and $Q_1 = 1 - P_1$. The second component has only b+ and b- points, has the cpr property, and has corresponding equations

$$P_2/Q_2 = (p_{bw}{}^3 p_{bl})/(q_{bl}{}^3 q_{bw})$$

$$\mu_2 = 2(P_2 - Q_2)(1 + (q_{bw} + p_{bl}))/(p_{bw} - q_{bl})$$

where, reversing the roles of P and Q, P_2 is player B's probability of winning the second component and $Q_2 = 1 - P_2$. The probability player A wins under the full scoring system when n = 1, $P_{1,2}$, is equal to $(P_1Q_2)/(P_1Q_2 + Q_1P_2)$, and the expected duration $\mu_{1,2}$ is equal to $(\mu_1 + \mu_2)/(1-R_{1,2})$ where $R_{1,2}$ is the probability each player wins exactly one of the two components, i.e. $R_{1,2} = P_1P_2 + Q_1Q_2$. The full scoring system (with general n) has the cpr property, and

$$P_n/Q_n = (P_{1,2}/Q_{1,2})^n$$

$$\mu_n = n(P_n - Q_n)\mu_{1,2}/(P_{1,2} - Q_{1,2})$$

where P_n is player A's probability of winning (for general n), and μ_n is the mean duration. Thus, the efficiency of a general bipoints (with dependencies) scoring system (with key characteristics P and μ) is

$$\rho = \frac{(P-Q)\ln(P/Q)\mu_{1,2}}{\mu(P_{1,2} - Q_{1,2})\ln(p_{1,2}/Q_{1,2})} \quad (5)$$

Noting that $\mu_{1,2}$, $(P_{1,2} - Q_{1,2})$, and $\ln(P_{1,2}/Q_{1,2})$ are simply functions of the underlying parameters, it can be seen that *the relative efficiencies of two scoring systems with this same dependent bipoints structure can be noted by simply comparing their values for the expression ((P–Q)/μ)ln(P/Q).*

The efficiency of play-the-loser bipoints, PL, (with dependencies) for Wn systems (n large)

The module approach can be used to analyse the W_nPL system for n large. Consider a module starting with an a+ point and using the PL rule. Player B might win 0, or 1, or 2, or…points until finally player A wins on an a- point (it is actually on the a+ point if it is the first point played in the module). This completes the first part of the module, and the last part of the module commences with a b+ point, and the module is completed when B wins a b- point (or possibly the very first b+ point). This is shown diagrammatically below, where the points played are shown on the first line, and the outcomes of those points are shown directly below on the second line.

// a+ a- a- a-…..a- / b+ b- b- b-…..b- // a+
B// B B B B…..A / A A A A…..B //

The geometric distribution is used to analyse this module. Suppose $S_1(S_2)$ is the shift (or gain) in points by player A during the first (second) part of this module. Then the distribution of S_1 is given by $P(S_1{=}1) = p_{a+}$, $P(S_1{=}0) = q_{a+}p_{a-}$, $P(S_1{=}{-i}) = q_{a+}p_{a-} {*} (q_{a-})^i$ (i = 1, 2, 3, …), and the mgfs of S_1 S_2 are

$M_{S_1}(\theta) = p_{a+}e^{-\theta} + (p_{a-}q_{a+})/(1 - q_{a-}e^{\theta})$, and

$$M_{S_2}(\theta) = p_{b+}e^{\theta} + (p_{b-}q_{b+})/(1 - q_{b-}e^{-\theta}) .$$

The mgf of $S = S_1 + S_2$, $M_S(\theta)$, is equal to the product of these two mgfs, and the expected number of points played in a module, E(D), is easily evaluated, as is E(S).

Example 6: When $p_{a+} = 0.9$, $p_{a-} = 0.7$, $p_{b+} = 0.7$, and $p_{b-} = 0.5$, the equation $M_S(\theta) = 1$ gives $\theta = 0.6931$, E(S)/E(D) = 1/6 and (θ*E(S))/E(D) is 0.1155.

The efficiency of play-the-loser, PL, bipoints (with dependencies) for Wn systems (n small)

Example 7: Consider the system $W_1(W_2(PL^a), W_2(PL^b))$ with the 3 types of points (initially) for each player. For two equal players, the points are not equally important, and so (P-Q)/μ does not have a 'nice' value. In fact, when $p_a = 0.8$, $p_{a+} = 0.9$, $p_{a-} = 0.7$, $p_b = 0.6$, $p_{b+} = 0.7$, and $p_{b-} = 0.5$, (P - Q)/μ = 0.0421, μ = 20.8514, and P = 0.9388, and so ((P-Q)/ μ)*ln(P/Q) = 0.1149, a little bit smaller than the value in Example 6 for large n.

The efficiency of play-the-winner, PW, bipoints (with dependencies) for Wn systems (n large)

Consider the module commencing with an a- point in the first part of the module, and the second part of the module commencing with a b- point. Then, for this PW module, the two relevant mgfs are

$M_1(\theta) = q_{a-}e^{\theta} + (p_{a-}q_{a+})/(1 - p_{a+}e^{-\theta})$ and

$M_2(\theta) = q_{b-}e^{-\theta} + (p_{b-}q_{b+})/(1 - p_{b+}e^{\theta})$,

and the mgf for the module is their product.

Example 8: When $p_{a+} = 0.9$, $p_{a-} = 0.7$, $p_{b+} = 0.7$, and $p_{b-} = 0.5$, the equation $M_1(\theta)^* M_2(\theta) = 1$ gives $\theta = 0.2231$, E(S)/E(D) = 1/2, and $(\theta^*E(S))/E(D) = 0.1116$, which is a bit smaller than for the PL case. This is expected since we know from published work on bipoints without dependencies that, when $p_a + p_b > 1$, play-the-winner W_n systems have slightly smaller efficiencies than play-the-loser W_n systems.

The efficiency of play-the-winner, PW, bipoints (with dependencies) for Wn system (n small)

Example 9: Consider the system $W_1(W_2(PW^a), W_2(PW^b))$ with the 3 types of points (initially) for each player. For two equal players, the points are not equally important, and so (P-Q)/μ does not have a 'nice' value. In fact, when $p_a = 0.8$, $p_{a+} = 0.9$, $p_{a-} = 0.7$, $p_b = 0.6$, $p_{b+} = 0.7$, and $p_{b-} = 0.5$, (P-Q)/μ = 0.0683, μ =9.6912, and P = 0.8310, and so ((P-Q)/μ)*ln(P/Q) = 0.1088, a bit smaller than 0.1116.

The efficiency of alternating bipoints, AL, (with dependencies) for Wn systems (n large)

Consider the module approach to AL point-pairs. Suppose the module starts with an (a+, b+) point (state 1), and continues until the next (a+, b+) is about to be played. It is clear that the next state after the start of the module is (a+, b-) (state 2) with probability $p_{a+}q_{b+}$, (a+, b+) with probability $p_{a+}p_{b+}$ (and so the module has finished), (a-, b-) (state 3) with probability $q_{a+}q_{b+}$, and (a-, b+) (state 4) with probability $q_{a+}p_{b+}$. Now, if $M_S(\theta)$ is the mgf of the shift in favour of player A during a module, and if $M_i(\theta)$ is the mgf of the shift in favour of player A from state i to the end of the module, we have

$$M_S(\theta) = p_{a+}q_{b+}e^{-\theta}M_2(\theta) + p_{a+}p_{b+} + q_{a+}q_{b+}M_3(\theta) + q_{a+}p_{b+}e^{\theta}M_4(\theta),$$

$$M_2(\theta) = p_{a+}q_{b-}e^{-\theta}M_2(\theta) + p_{a+}p_{b-} + q_{a+}q_{b-}M_3(\theta) + q_{a+}p_{b-}e^{\theta}M_4(\theta),$$

$$M_3(\theta) = p_{a-}q_{b-}e^{-\theta}M_2(\theta) + p_{a-}p_{b-} + q_{a-}q_{b-}M_3(\theta) + q_{a-}p_{b-}e^{\theta}M_4(\theta),$$

$$M_4(\theta) = p_{a-}q_{b+}e^{-\theta}M_2(\theta) + p_{a-}p_{b+} + q_{a-}q_{b+}M_3(\theta) + q_{a-}p_{b+}e^{\theta}M_4(\theta).$$

Omitting θ, the last three equations can be written as

$a_iM_2 + b_iM_3 + c_iM_4 = d_i$ (i=1, 2, 3) where

$a_1 = 1 - p_{a+}q_{b-}e^{-\theta}$, $b_1 = -q_{a+}q_{b-}$, $c_1 = -q_{a+}p_{b-}e^{\theta}$, $d_1 = p_{a+}p_{b-}$,
$a_2 = -p_{a-}q_{b-}e^{-\theta}$, $b_2 = 1 - q_{a-}q_{b-}$, $c_2 = -q_{a-}p_{b-}e^{\theta}$, $d_2 = p_{a-}p_{b-}$,
$a_3 = -p_{a-}q_{b+}e^{-\theta}$, $b_3 = -q_{a-}q_{b+}$, $c_3 = 1 - q_{a-}p_{b+}e^{\theta}$, $d_3 = p_{a-}p_{b+}$

It can be shown that

$$M_3(\theta) = (E_2E_7 + E_3E_8)/(E_1E_9 + E_3E_{10})$$

$$M_2(\theta) = (E_1M_3 + E_2)/E_3$$

$$M_4(\theta) = ((a_1E_1 + b_1E_3)M_3)/(-c_1E_3) + (a_1E_2 - d_1E_3)/(-c_1E_3)$$

where $E_1 = c_3b_2 - b_3c_2$, $E_2 = c_2d_3 - c_3d_2$, $E_3 = c_2a_3 - c_3a_2$, $E_4 = c_1a_2 - c_2a_1$, $E_5 = c_1b_2 - b_1c_2$, $E_6 = c_1d_2 - d_1c_2$, $E_7 = c_1a_3 - c_3a_1$, $E_8 = c_3d_1 - c_1d_3$, $E_9 = a_1c_3 - c_1a_3$, and $E_{10} = b_1c_3 - c_1b_3$.

Substituting for $M_2(\theta)$, $M_3(\theta)$ and $M_4(\theta)$ in the first equation, and putting $M_S(\theta) = 1$, we can solve for θ.

Example 10: When $p_{a+} = 0.9$, $p_{a-} = 0.7$, $p_{b+} = 0.7$, and $p_{b-} = 0.5$, $\theta = 0.4581$, E(S)/E(D) = 0.25 and $(\theta^*E(S))/E(D) = 0.1145$, which is a value between 0.1116 (in Example 8) and 0.1155 (Example 6). This is not surprising as it is known that for bipoints without dependencies when $p_a + p_b > 1$, play-the-winner W_n systems have slightly smaller efficiencies than alternating W_n systems, and play-the loser W_n systems have slightly greater efficiencies than alternating W_n systems.

A relationship between the solutions for the PW and PL cases (n large)

The solution θ_{PL} for the PL module is the solution of $M_{S1}(\theta_{PL})M_{S2}(\theta_{PL}) = 1$, and the solution θ_{PW} for the PW module is the solution of $M_1(\theta_{PW})M_2(\theta_{PW}) = 1$. Surprisingly, it turns out that these θ solutions are related, and can be obtained by putting $M_{S1}(\theta_{PL})$ equal to $M_1(\theta_{PW})$ and $M_{S2}(\theta_{PL})$ equal to $M_2(\theta_{PW})$, and solving simultaneously for θ_{PL} and θ_{PW}. Putting $x_1 = e^{\theta PL}$, $y_1 = 1/x_1$, $x_2 = e^{\theta PW}$ and $y_2 = 1/x_2$, we have

$$x_2 = \frac{(p_{a-} + \delta_a p_{b-})}{(p_{b-} + \delta_a p_{a-})}, \text{ and}$$

$$y_1 = (1 - p_{b+}x_2)/(q_{b-} - \delta_b x_2)$$

where $\delta_a = p_{a+} - p_{a-}$ and $\delta_b = p_{b+} - p_{b-}$.

Example 11: When $p_{a+} = 0.9$, $p_{a-} = 0.7$, $p_{b+} = 0.7$, and $p_{b-} = 0.5$, $x_2 = 1.25$ and $y_1 = 0.5$, giving $\theta_{PL} = 0.6931$ as in Example 6, and $\theta_{PW} = 0.2231$ as in Example 8. These results for x_2 and y_1 were verified for several other parameter values. When $\delta_a = \delta_b = 0$, $p_{a+} = p_{a-} = p_a$, $p_{b+} = p_{b-} = p_b$ these expressions for x_2 and y_1 give $\theta_{PW} = \ln(p_a/p_b)$ and $\theta_{PL} = \ln(q_b/q_a)$ in agreement with earlier work (Miles, 1984 and Pollard, 1992).

An explicit solution for the alternating module

The value of θ for the AL module, although difficult to evaluate, turns out to be the average of θ_{PL} and θ_{PW}. In Table 2, and in line with the paper by Miles (1984), W_nAL systems are given unit efficiency. It

Pa-, pa+, pb-, pb+		Exp(θ)	theta, θ	E(S)/E(D)	θ *E(S)/E(D)	Efficiency
0.8,0.8,0.6,0.6	AL	Sqrt(8/3)	0.4904	1/5	0.0981	1.0
0.8,0.8,0.6,0.6	PL	2.0	0.6931	1/7	0.0990	1.0096
0.8,0.8,0.6,0.6	PW	4/3	0.2877	1/3	0.0959	0.9777
0.7,0.9,0.5,0.7	AL	Sqrt(2.5)	0.4581	¼	0.1145	1.0
0.7,0.9,0.5,0.7	PL	2.0	0.6931	1/6	0.1155	1.0086
0.7,0.9,0.5,0.7	PW	1.25	0.2231	½	0.1116	0.9741
0.9,0.7,0.7,0.5	AL	Sqrt(3.215385)	0.5840	1/6	0.0973	1.0
0.9,0.7,0.7,0.5	PL	2.2	0.7885	1/8	0.0986	1.0126
0.9,0.7,0.7,0.5	PW	1.461538	0.3795	¼	0.0949	0.9748

Table 2: Some characteristics of the AL, PL and PW modules

can be seen that the efficiency (5) is greater than unity for the PL and less than unity for PW when ε is positive or negative.

An extension of the fundamental equation

We give an extension of the earlier equation from individual points to point-pairs, which can result in a win, a draw or a loss to player A. The importance of winning a point-pair rather than drawing it is equal to the probability of winning given the point-pair is won minus the probability of winning given the point-pair is drawn. Also, the importance of drawing a point-pair rather than losing it is the probability of winning given the point-pair is drawn minus the probability of winning given the point-pair is lost. Thus, using an obvious notation, when the score is (i, j), the importance of winning a point-pair rather than drawing it, $I_{i, j, W}$, is equal to P(i+2, j) - P(i+1, j+1), and the importance of drawing a point-pair rather than losing it, $I_{i, j, D}$, is equal to P(i+1, j+1) - P(i, j+2). The earlier equation becomes

$$P = 0.5 + \sum n_{i,j}(I_{i,j,W}\delta_{i,j,W} + I_{i,j,D}\delta_{i,j,L})$$

where, as before, the importances are evaluated for the case in which the two players are equal, and the $n_{i, j}$ are evaluated when player A has an increase in probability of $\delta_{i, j, W}$ of winning the point-pair and a decrease in probability of $\delta_{i, j, L}$ of losing a point-pair.

Example 12. Consider playing just two point-pairs, and if the final score is 2-2, a coin is tossed to determine the winner. The first point-pair is an (a, b) point-pair. The second is (a+, b-) if A wins the first point-pair; it is (a+, b+) if A wins the first a point and B wins the first b point; it is (a-, b-) if player A loses the first a point and B loses the first b point, etc. If the two players are equal, P equals 0.5, $I_{2,0,W}$ = 0, $I_{2,0,D}$ = 0.5, $I_{-2,0,W}$ = 0.5, $I_{-2,0,D}$ = 0, $I_{1,1,W}$ = 0.5 and $I_{1,1,D}$ = 0.5. Also, if $p_a = p_b = 0.7$, $p_{a+} = p_{b+} = 0.8$ and $p_{a-} = p_{b-} = 0.6$, $I_{0,0,W}$ = 0.44 and $I_{0,0,D}$ = 0.44. If A is better than B and p_a = 0.8, p_{a+} = 0.9, p_{a-} = 0.7, p_b = 0.6, p_{b+} = 0.7 and p_{b-} = 0.5, P equals 0.6606. Also, for example, $n_{0,0}$ = 1, $n_{2,0}$ = 0.32 and $n_{-2,0}$ = 0 12. It can be seen, for example, that $\delta_{0,0,W}$ = 0.8*(1 - 0.6) - 0.7*0.3 = 0.11 and $\delta_{0,0,L}$ = 0.3*0.7 - 0.6*(1 - 0.8)= 0.09. The right hand side of the above equation is 0.5 + 1*0.44*0.2 + 0.32*0.5*0.07 + 0.48*0.5*0.2 + 0.08*0.5*0.2 + 0.12*0.5*0.09 = 0.6606, verifying the above equation for this example.

Example 13. Consider $W_{2point\text{-}pairs}$ab using dependent bipoints, starting with an (a, b) point-pair. The possible states are (0, a, b), (2, a+, b-), (2, a+, b+), (2, a-, b-), (0, a+, b-), (0, a+, b+), (0, a-, b-), (0, a-, b+), (-2, a+, b+), (-2, a-, b-), and (-2, a-, b+). All of the importances for this scoring system are equal, and are equal to 5/18 when $p_a = p_b = 0.7$, $p_{a+} = p_{b+}$ = 0.8 and $p_{a-} = p_{b-}$ = 0.6. As the importances are equal, (P –Q)/μ has a 'nice' value (1/18 when p_a = 0.8, p_{a+} = 0.9, p_{a-} = 0.7, p_b = 0.6, p_{b+} = 0.7 and p_{b-} = 0.5).

Example 14. A stochastically identical 'single-point' version of the scoring system in example 13 is $W_4(AL^a)$. It can be seen that there are 3 'initial' states, 0a, followed by 1b(+) or -1b(-), where, for example, 1b(+) represents a score of 1 with a b point to be played and the next a point being an a+ point. There are 12 b states when the scores are 3, 1, -1 or -3, and 10 a states when the scores are 2, 0, or -2. The importances of all these states are equal, and equal to 5/18 for the above parameter values.

Example 15. The system $W_1(W_2PL^{a+(b+)}, W_2PL^{b+(a+)})$ (where $PL^{a+(b+)}$ indicates that the first point is a+ and the first b is b+) has equally important points for two equal players (and equal to 5/24 for the above parameters). Further, $W_1(W_2PW^{a-(b-)}, W_2PW^{b-(a-)})$ has equally important points for two equal players.

Relevance to golf

The extension of the above equation to the situation where draws are possible has relevance to other sports such as golf. For example, in match play golf, each hole is won, drawn or lost. The equation can be extended even further to where there are more than just 3 possible outcomes. For example, when two golfers play each other at stroke play and we assume each golfer has an eagle, a birdie, a par, a bogie or a double bogie on each hole, the two players' scores on each hole can differ by +4, +3, +2, …, -4. The importance of having an eagle rather than a birdie, etc or having a +4 rather than a +3, etc, can be evaluated. This approach is taken in a forthcoming paper focussing on golf (Pollard & Pollard, 2011).

3. RESULTS

Various results for unipoints and bipoints scoring systems without dependencies have been seen to carry over to the case where one-step dependencies exist. Efficient scoring systems and systems with equally important points for two equal players have been identified. The detailed results are given in the methods section above, and are summarised below.

4. CONCLUSIONS

Formulae for the efficiency of unipoints and bipoints scoring systems with one-step dependencies have been derived. In both cases the relative efficiency of two scoring systems can be measured by the same function, namely $((P-Q)/\mu)*\ln(P/Q)$. Thus, the formula for relative efficiency has been extended to the case of one-step dependent points.

The very efficient scoring systems for unipoints and bipoints without dependencies have points that are equally important points for two equal players. This is also the case when one-step dependencies exist.

In the tennis context, the play-the-loser service exchange mechanism is seen to be slightly more efficient than the alternating mechanism, which in turn is more efficient than the play-the-winner mechanism, whether dependencies exist or not.

A relationship between solutions to play-the-loser and play-the-winner scoring systems with one-step dependencies and large mean durations has been derived. Also, a relationship between these systems and the alternating system has been observed.

Some results for the best-of-2n-1 unipoints scoring systems with one-step dependencies are given.

By generalizing the definition of the importance of a point, the formula relating the increased probability of winning under a scoring system to the increased probability of winning the points within that system has been extended from the two outcome win/loss to the three outcome win/draw/loss structure. Indeed, there is a further extension to the situation where there is a four or more outcome structure. This can be useful for other sports such as golf.

References

Cox, D. R. & Miller, H. D. (1965) The theory of Stochastic Processes. London: Chapman and Hall.

Klaassen, F. J. G. M. & Magnus, J. R. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96, 500-509.

Miles, R. (1984). Symmetric sequential analysis: the efficiencies of sports scoring systems (with particular reference to those of tennis). *Journal of the Royal Statistical Society B*. 46(1), 93-108.

Morris, C. (1977). The most important points in tennis. In *Optimal Strategies in Sports*, edited by S. P. Ladany and R. E. Machol, 131-140. Amsterdam:North-Holland. (Vol 5 in Studies in Management Science and Systems).

Pollard, G. H. (1983). An analysis of classical and tie-breaker tennis. *Australian Journal Statistics, 25(3)*, 496-505.

Pollard, G. H. (1990). A method for determining the asymptotic efficiency of some sequential probability ratio tests, *The Australian Journal of Statistics*, 32, 191-204.

Pollard, G. H. (1992). The optimal test for selecting the greater of two binomial probabilities, *The Australian Journal of Statistics*, 34, 273-284.

Pollard, G. H. (2004). Can a tennis player increase the probability of winning a point when it is important? In *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport*, edited by R. H. Morton & S. Ganesalingam, Massey University, Massey, New Zealand, 253-256.

Pollard, G. H., Cross, R. & Meyer, D. (2006). An analysis of ten years of the four grand slam men's singles data for lack of independence of set outcomes, *Journal of Sports Science and Medicine*, 5, 561-566.

Pollard, G. N. and Pollard, G. H. (2008). The efficiency of doubles scoring systems. *Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport*, edited by J. Hammond, Mathsport (ANZIAM), 45-51.

Pollard, G. H. & Pollard. G. N. (2011). An analysis of individual matchplay and stroke play golf. (draft)

Wald, A. (1947). Sequential Analysis. New York: Wiley.

AN INTRODUCTORY ANALYSIS OF CHALLENGES IN TENNIS

Clarke, Stephen R [1], Norman, John M [2]

[1] *Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia:* sclarke@swin.edu.au
[2] *Sheffield University, Sheffield, UK*

Abstract

The use of technology in sport to assist umpires has been gradually introduced into several sports. This has now been extended to allow players to call upon technology to arbitrate when they disagree with the umpire's decision. Both tennis and cricket now allow the players to challenge a doubtful decision, which is reversed if the evidence shows it to be incorrect. However the number of challenges is limited, and players must balance any possible immediate gain with the loss of a future right to challenge. With similar challenge rules expected to be introduced in other sports, this situation has been a motivation to consider challenges more widely. We use Dynamic Programming to investigate the optimal challenge strategy and obtain some general rules.

Keywords: Sports, Dynamic Programming, Tennis, Challenge

1. INTRODUCTION

Technology is being used increasingly in sport to assist, and in some cases replace, judges and referees. The photo finish has been used to help the steward decide close horse races since the camera was invented. In swimming, electronic timing has largely replaced the judges in determining finishing order. In tennis we have had the electronic service line machine for many years, and now also a device for picking up net cords. More recently, some forms of Rugby have used a video referee to decide on the legality of tries. Cricket followed suit, allowing umpires to refer decisions on run outs, stumpings and catches to a third umpire with access to video footage. In many other sports the media use video replays or other technology (such as the hot spot in cricket) to provide evidence for or against the umpire's decision.

In 2008 the Australian Tennis Open saw an interesting development with the introduction of Hawkeye. This system, relying on several video cameras and some mathematical modelling, was originally used in cricket, where it claims to show where the ball actually went or would have gone had it not hit the batsman. In tennis, it displays a schema of the court lines along with a mark where the ball is believed to have bounced, along with a decision on whether it was in or out. (Interestingly, the path of the ball is never shown with any error bounds. The public and players appear to accept that it is exact and infallible.) The interesting development in tennis was that the players, not the umpires, under certain conditions were allowed to challenge the umpire's decision by referring to Hawkeye. If Hawkeye's view was consistent with the appeal, the umpire's decision was reversed. The International Cricket Committee has now introduced a similar rule into the playing conditions of some cricket series, and it seems inevitable that allowing players to challenge umpires' decisions will play an increasing part in many sports. But some poor decision making by players shows they do not always make good use of their right to challenge. The number of challenges is limited, and players must balance any possible immediate gain with the loss of a future right to challenge. This provides the motivation to investigate the optimal strategy in the use of challenges in a wider context.

Norman (1995) gives several examples of the use of Dynamic programming to find optimal strategies in sport. It can be used here to find the optimal challenge strategy under simplified rules, and thus formulate some general rules for the real situation.

2. A SIMPLE MODEL TO DETERMINE CHALLENGE POLICY

In a game of tennis, at the end of a point, a player may challenge a line call. For example, if his opponent's ball is called in when it may have been out, then a successful challenge will win the point (or if the miscall occurs on his opponent's first serve, the opponent will be required to serve again). Again, if his own ball is called out when it may have been in, a successful challenge will earn a replay of the point, or even win the point if it is deemed that his opponent would have been unable to return the ball. A player may make up to three unsuccessful challenges in a set, up to four if a tie break is reached.

The scoring system in tennis makes the game difficult to model and we consider here a simpler game in which two players compete to be the first to gain 20 points. It could be thought of as tennis with different scoring. It is not very different from the game of table tennis played under the old rule of first to 21.

Suppose X (our man) is playing Y. With probability p_c a challenge opportunity may occur for X, and if X makes a challenge the probability of success is s_c. With probability $(1 - p_c)$ no challenge opportunity occurs, and X wins the point outright with probability p. The point has just been played and one of three states occurs:

W: X is about to be given the point outright;

L: Y is about to be given the point and X thinks this is right;

C: A call has been made such that if it stands Y will win the point, but X thinks there is a good chance he would be awarded it if he challenges the call.

The probabilities of the three states W, L and C occurring are respectively $(1\text{-}p_c)*p$, $(1\text{-}p_c)*(1\text{-}p)$ and p_c

Before the umpire says anything, we take the score to be i-j. X has m challenges left.

We take the state of the system to be (i, j, m, θ) where θ can take the values W, L and C.

Define $f(i, j, m, \theta)$ as the maximum probability of X winning the game, with the score i–j and X having m challenges left, with θ the state of play. Then

$f(i, j, m, W) = (1\text{-}p_c)*p*f(i+1,j, m ,W) + (1\text{-}p)*f(i+1, j, m, L) \} + p_c*f(i+1, j, m, C)$

$f(i, j, m, L) = (1\text{-}p_c)*\{p*f(i, j+1, m, W) + (1\text{-}p)*f(i, j+1, m, L) + p_c*f(i, j+1, m, C)$

$f(i, j, m, C) = \max$

don't challenge: $(1\text{-}p_c)*\{p*f(i, j+1, m, W) + (1\text{-}p)*f(i, j+1, m, L) + p_c*f(i, j+1, m, C)$

challenge: $s_c*[(1\text{-}p_c)*\{p*f(i+1, j, m, W) + (1\text{-}p)* f(i+1, j, m, L)\} + p_c* f(i+1, j, m, C)]$

$+(1\text{-} s_c)*[(1\text{-}p_c)*\{p*f(i, j+1, m\text{-}1, W) + (1\text{-}p)*f(i, j+1, m\text{-}1, L)\} + p_c*f(i, j+1, m\text{-}1, C)]$

Since having an extra challenge can never decrease X's chance of winning ($f(i, j, m, \theta) >= f(i, j, m\text{-}1, \theta)$), it is easily shown that the challenge test quantity is monotone increasing in s_c. Suppose the two test quantities are equal when $s_c = \pi$. Then for $s_c > \pi$ it is better to challenge, and for $s_c < \pi$ it is better not to challenge. Thus the form of the optimal policy is 'Challenge only if the probability of success is greater than some probability π'.

3. A COMPUTABLE MODEL FOR TENNIS

One (big) disadvantage of this formulation is the number of variables in the state description. For computational simplicity, we may reduce the number of variables by one by taking the time at which a decision is made to be when player X is about to serve (or receive a serve). We suppose that he then asks himself whether if an opportunity to challenge occurs, he will take it.

The state of the system is (i, j, m) where i - j is the score: i and j are the points each player (X and Y, respectively) has earned so far in the game; m is the number of challenges left. We consider X as the player who decides whether or not to challenge, and who has three challenges available at the start of the game. We define $f(i, j, m)$ as the probability of X winning the set using an optimal policy.

We suppose that challenge possibilities are of two types, occurring with probabilities p_1 and p_2. If player X makes a challenge, his probability of success is s_1 and s_2 respectively ($s_1>s_2$). If his challenge is successful, the state of the system becomes $(i+1, j, m)$ but if unsuccessful $(i, j+1, m\text{-}1)$.

Just before a point is played, player X may consider three possibilities:

A: With probability 1-p_1-p_2, the point proceeds without any question of the ball being in or out of play and X wins with probability p and loses with probability 1-p

B: With probability p_1, the ball is called out when player X thinks it is in and unless he makes a successful challenge, he will lose the point. With probability s_1 his challenge is successful and he will gain the point; if it is unsuccessful; he will lose the point and lose one right to challenge.

C: With probability p_2 the ball is called out when player X thinks it is in and unless he makes a successful challenge, he will lose the point. With probability $s_2 < s_1$ his challenge is successful and he will gain the point; if it is unsuccessful, he will lose the point and lose one right to challenge.

The decision problem faced by X before the point is played is thus to choose one of three alternatives:

I: Not to challenge, even if a possibility occurs

II: To challenge if and only if possibility B occurs

III: To challenge if either possibility B or possibility C occurs

(since s_1>s_2 it is never optimal to challenge only if possibility C occurs)

The functional equation is thus

$f(i, j, m)$=(1-p_1-p_2)*{p*$f(i$+1,j, m)+(1-p)*$f(i, j$+1, m)} + max

I: (p_1+p_2)* $f(i, j$+1, m)

II: p_1*{ s_1* $f(i$+1, j, m) +(1- s_1)*$f(i, j$+1, m-1)} + p_2*$f(i, j$+1, m)

III: p_1*{ s_1*$f(i$+1, j, m) +(1-s_1)*$f(i, j$+1, m-1)}
+ p_2*{ s_2*$f(i$+1, j, m) +(1-s_2)*$f(i, j$+1, m-1)}

for i, j < 20, m= 1, 2, 3.

$f(i, j$, 0) =(1-p_1-p_2)*{p*$f(i$+1 ,j, 0)+ (1-p)*$f(i, j$+1, 0)} + (p_1+p_2)*$f(i, j$+1, 0) for i, j < 20

f(20, j, m) =1 and $f(i$, 20, m)=0,for all i, j < 20, m = 0, 1, 2, 3.

4. MODEL CALIBRATION

Table 1 summarises data on the success rate of challenges during the 2009 Wimbledon championship, obtained from http://www.wimbledon.org/en_GB/scores/challenge/index.html.

The "total challenges" in Table 1 relate to matches played with Hawkeye, a subset of 47 out of 127. Assuming the presence of Hawkeye has no effect on the number of sets in a match, we can calculate the number of challenges per set. We would really like to have average challenges per set, so we looked at the men's singles results. 474 sets were played in 125 matches played to completion, giving the average sets per match as 473/125 = 3.8. The average number of challenges per set is thus 6.7/3.8 = 1.8.

	Men	Women
Total challenges	314	130
Successful challenges	93	38
Unsuccessful challenges	221	92
Percentage successful	29.6	29.2
Average challenges per match	6.7	3.8

Table 1: Statistics on Challenges, Wimbledon 2009

In the above model challenges are of two types, with different probabilities of success. We could suppose these probabilities to be 0.4 and 0.2. (They can't be close to 1, as the line judges rarely make bad mistakes, and they can't be close to 0 for then a challenge would not be worthwhile).

How often do challenge opportunities occur? More guesswork is needed. Suppose, on average, that one type 1 opportunity and two type 2 opportunities occur per set. We might suppose that players take up all type 1 opportunities and half of type 2, giving an average of two challenges per set and an average success rate of (0.4 + 0.2)/2 =0.3 or 30%.

It's easy to juggle with these figures. If only a quarter of type 2 opportunities are taken up, then there will be an average of 1.5 challenge per set and an average success rate of (0.4 + 0.1)/1.5 =0.33 or 33%. The point of all this is to suggest a set of credible values. Let's take the ones in the preceding paragraph and suppose that, on average, a player has 0.5 type 1 opportunity and 1.0 type 2 opportunity per set.

		X score																			
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
	0																				
	1												OPTION 2								
	2																				
Y	**3**												CHALLENGE ONLY IF HIGH PROBABILITY OF SUCCESS								
	4																				
s	**5**																				
c	**6**																				
o	**7**																				
r	**8**																				
e	**9**																				
	10												OPTION 3								
	11																				
	12												CHALLENGE AT EVERY OPPORTUNITY								
	13																				
	14																				
	15																				
	16																				
	17																				
	18																				
	19																				

Table 2: Optimal strategy in first to 20 game, with one challenge left, $p = .6$

These values seem reasonable. A player is allowed up to three unsuccessful challenges in a set. This is presumably thought to be a reasonable maximum. If a player makes one challenge in a set, on average, then if the number of challenges follows a Poisson distribution, the probability that he makes three or fewer challenges is 0.98, a 2% chance that he cannot make as many challenges as he would like.

A BASIC computer program (verified with an Excel spreadsheet) has been written for this game of first to 20, and with a slight modification a first to 30 game. As any number of points between 20 and 39 could be played in the first and any number between 30 and 59 in the second, we take an average of 30 points played in the first and 45 in the second. Thus we arrive at values of $p_1 = 0.5/30 = 0.0167$ and $p_2 = 1/30 = 0.0333$ for playing up to 20 and $p_1 = 0.5/45 = 0.0111$ and $p_2 = 1/45 = 0.0222$ for playing up to 30. $s_1 = 0.4$ and $s_2 = 0.2$ as suggested earlier. The probability of winning a point outright for player X was taken to be 0.6.

5. RESULTS

In both games, with more than one challenge left, in almost every situation, the optimal decision is Option III to take up every challenge opportunity. The cases where this choice is not optimal are when the game is virtually over, when Player X leads by many points and is very close to winning. For example, when X leads 19 - 4 in the first to 20 game and 29 - 10 in the first to 30 game, the optimal decision is Option I, not to challenge if an opportunity occurs. In such situations, Player X has a probability of winning very close to one and it matters little which decision he makes.

With only one challenge left, the choice depends on the score. Decision tables for both games are shown in tables 2 and 3. It can be seen that in the first to 20 game, a near-optimal policy for Player X would be to choose Option II if Player Y's score is seven or less and Option III otherwise; and that in the first to 30 game, a near-optimal policy for Player X would be to choose Option II if Player Y's score is ten or less and Option III otherwise. In both games first to N the critical score for Y is about $N/3$. If this decision rule were applied to a set in tennis, then the recommended rule would be: take every opportunity to challenge, but with only one challenge remaining, if your opponent has won two games or fewer, challenge only if you have a (relatively) high chance of success

Clearly this decision rule will alter depending on the relative ability of the players. We have used $p = 0.6$, which might be relevant for a seeded player playing a non-seeded player. This might be the case for many of the matches played on the show courts for

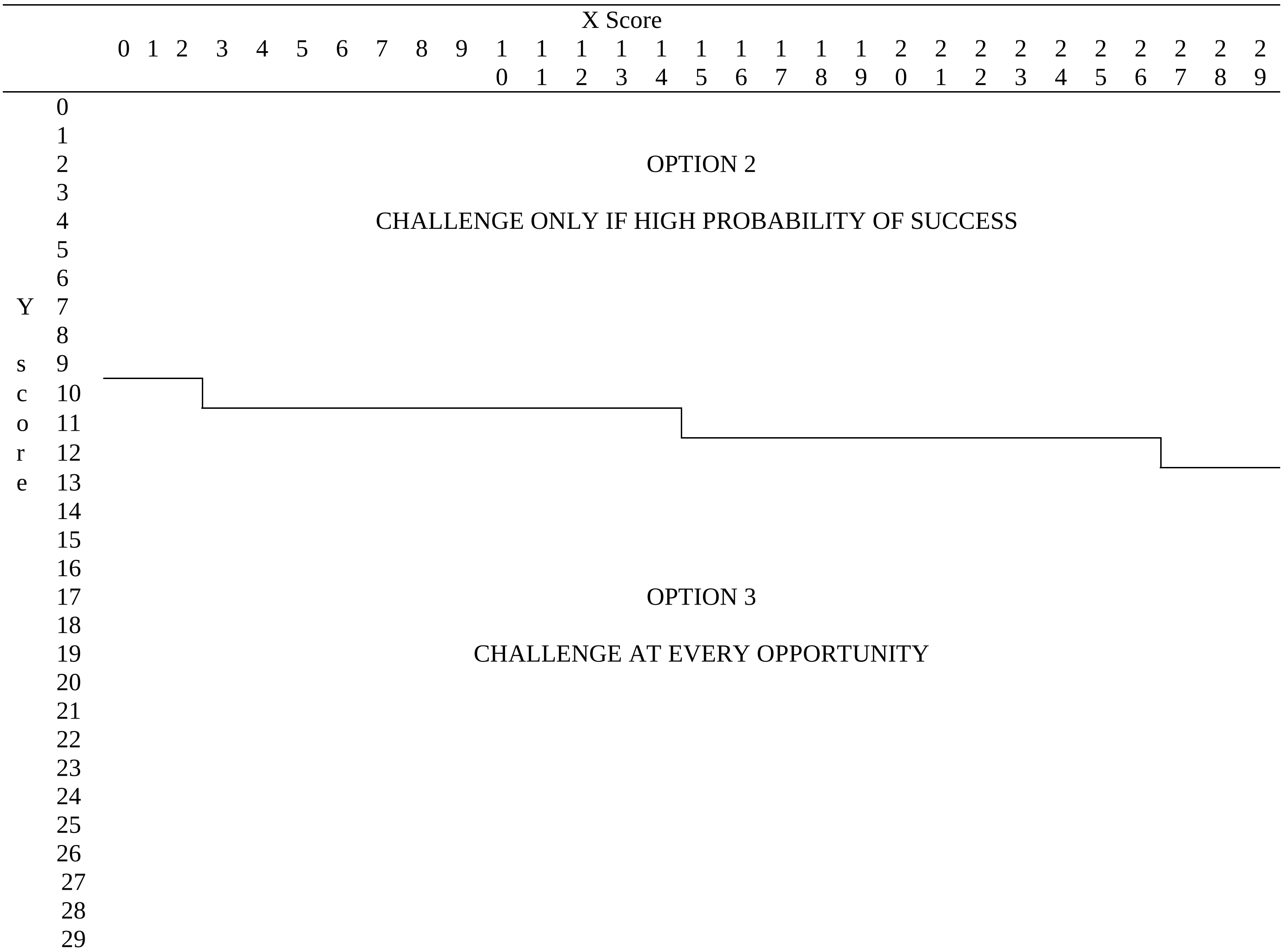


Table 3: Optimal strategy in first to 30 game, with one challenge left, p = .6

which the challenge system is used. As the player becomes weaker, his relative reward for a successful challenge becomes greater, and he should challenge more often. The seeded player's opponent, for example, (p = 0.4) should virtually always challenge. In the first to 30 game, with one challenge left nearly even players, (p = 0.5), should use option II only if their opponents score is less than about 7,and a vastly superior player (p = 0.7) only if their opponents score is less than about 16.

There are some challenges in which the successful challenger only wins a replay of the point. If an out call on a ball which the opponent could have played is successfully challenged, the point is replayed. While we have not included this possibility in the model, since the rewards for a successful challenge are not as good, clearly the player should be less aggressive in challenging such calls.

6. CONCLUSIONS

Analysing simplified rules can be helpful in generating simple rules towards optimal challenge strategy. Results suggest that in a simple 'first to' game, the optimal strategy will also be fairly simple – always challenge when you have a 'good' chance of success, and take any challenge once you get deep enough into the game that it looks as if you might not use all your challenges. However we do not expect the decision rules to be as straightforward with the nested scoring system used in tennis. In the simple game analysed here, there is no sense in saving a challenge until a more important stage of the game. Once a challenge opportunity with the maximum chance of success arises, there is nothing to be gained by saving that challenge to later in the game, as if the challenge is successful that point stays on your score until the game is over. Keeping it in case you have a similar challenge opportunity at 19 all is futile, since if you used it successfully earlier you would be 20-18. The only consideration is then how long into the match does it become unlikely that another maximum chance challenge opportunity will arise – at that stage you may take a lesser chance challenge. This is not true in the nested scoring system used in many racquet sports, in particular tennis. A set consists of first to six games, and once a game is won there is no advantage in winning it to love as opposed to winning it to 15. A challenge opportunity arising at 40-0 might not be taken up, as the game will probably be won anyway. Morris (1977) defines the importance of points in tennis. An Excel spreadsheet which solves a similar model to the above for the actual scoring system used in tennis has been developed (Clarke & Norman, 2010). Preliminary analysis of this model suggests the optimal strategy depends on the importance of the point – the more important the point in winning the set, the more likely a player should challenge. Since importance increases in later points of close games, and in later games of close sets, this implies that players should save their challenges until needed deeper into close games and sets.

References

Clarke SR and Norman JM (2010). Optimal challenges in tennis. Under review

Norman JM (1995). Dynamic programming in sport: A survey of applications. *IMA J Math Appl Business Ind* 6(December): 171–176.

Morris C. (1977). The most important points in tennis. In *Optimal strategies in sports*. S. P. Ladany and R. E. Machol (Eds.). Amsterdam, North Holland: 131-140

SOME ASPECTS OF ORDERING, RANKING AND SEEDING 1

Pollard, Graham[2,3], Pollard, Geoff[1] and Meyer, Denny[1]

[1] *Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia*

[2] *Faculty of Information Sciences and Engineering, University of Canberra, Australia*

[3] *Corresponding author: graham@foulsham.com.au*

Abstract

This is the first of two papers on ordering, ranking and seeding players or teams in sport. Such ordering may be used for player selection, seeding, handicapping, player progress evaluation and for predicting match and tournament outcomes. Pair-wise match probabilities are used to create order criteria based on match expected values, and the analyses of round robins, knock-out tournaments and ladder systems. Theoretical results are given for three and four players/teams. The total number of players/teams can affect the ratings.

Keywords: Ordering, ranking, seeding, round robins, random knock-outs, ladders, ratings.

1. INTRODUCTION

Clarke and Dyte (2000) suggest that ratings be used to seed players for tournaments, to allow entry into tournaments, to allocate prize money at the end of the year and to predict the outcome of a match. They developed a binary logistic regression model which uses the difference in ratings to predict the probability of winning a match. Blackman and Casey (1980) mention tournament placement, handicapping and player progress evaluation as other uses for tennis rankings.

As explained by Pollard and Meyer (2010), when the leading male tennis players formed the Association of Tennis Professionals (ATP) in 1973, one of their first acts was to introduce a 12 month weighted moving average world ranking system to determine which players gained entry into tournaments and to determine which players were seeded. The Women's Tennis Association (WTA), founded in 1973, introduced its computer rankings system in 1975 and a separate doubles ranking system was introduced in 1976. Prior to this time there were only informal ranking systems. Since 1989 and 1995 respectively, the ATP and WTA have used rankings that reward quantity as well as quality by selecting a player's best results from a minimum number of tournaments. Stefani (1997) gives details of the ATP tennis ranking system which is based on tournament and bonus points.

The original ranking systems were based on tournament importance as determined by prize money, and player performance was measured by the round reached. A schedule of points was agreed based on the above and a player's ranking was calculated as the average points earned for tournaments played in the previous 12 months. Musante and Yellin (1979) refined this method using the ranking of all players entered in that event to measure its importance rather than prize money while using the ranking of defeated opponents, not just the round reached, to measure the performance of any player. The concept of bonus points for defeating a higher ranked player was used by the WTA for some years, but was subsequently discontinued. Subsequently Blackman and Casey (1980) developing a ranking system similar to a golf handicap using the actual scores in all matches between the players being ranked. The difference in these rating units for any two players was shown to be a good indication of match result probabilities and could also be used to determine what handicap should be given to the weaker player to make the match more even.

Ali, Cook, and Kress (1986) rank ordered a set of players on the basis of a set of pair-wise comparisons arising from a tournament. They showed that the minimum violation ranking rule can be represented as a mixed integer generalized network program. Strauss and Arnold (1987) used maximum likelihood and moment estimates for rally winning probabilities to develop a rating system for players in a tournament, also based on a paired comparison method.

Cook, Golan, and Kress (1988) examined a series of heuristics for ranking the players in a round robin tournament. The comparison is based on a set of randomly generated n-player tournaments. The Generalized Iterated Kendall Method, produced a 9% reduction in ranking violations compared to the next best contender. This work continued with Cook and Kress (1990) presenting a model for developing a weak ranking of players in a round-robin tournament.

More recently Clarke (1994) has suggested that exponential smoothing should be used to update ratings while Klaassen and Magnus (2003), Bedford and Clarke (2000) and Barnett and Clarke (2002) have explored the use of ratings to provide models for predicting tennis match outcomes.

A group of players needs to be ordered in order to determine which players are accepted into a tournament, who is seeded first, who is second, etc. In this paper we explore various criteria for ordering players, and consider round robins, knock-outs and ladder systems for three and four players. Match result probabilities for each pairing are assumed.

2. METHODS

Ordering three players

Ordering two players is trivial. A natural way of ordering three players is to order them according to the magnitude of their row sums in Table 1. Thus, if the best player is defined as the player who has the largest row sum, then the best player is the one who has the largest value for the expected number of matches won in a round robin (RR) event, giving the definition some mathematical and statistical appeal. An example with players A and B being equally best and player C being the worst (according to this definition) is given in Table 1.

If we had been considering just the two players A and B, A would be better than B. Thus, the ordering of two players A and B within a group of three players can be different to the ordering within just the two players. Thus, using the above definition, the ordering of two players within a group of players can change when other players are added or subtracted from the group. The ordering of three players is in general not a trivial exercise.

Probability i beats j	Player A	Player B	Player C	Row sum
Player A	X	$p_{12} = 0.6$	$p_{13} = 0.6$	1.2
Player B	$p_{21}=0.4$	X	$p_{23} = 0.8$	1.2
Player C	$p_{31}=0.4$	$p_{32}=0.2$	X	0.6
Total				3.0

Table 1: An example with three players

A round robin tournament (RR) with 3 players

In a RR with 3 players, player A plays player B, player A plays player C, and player B plays player C. Either one player 'wins outright' by winning both matches, or the event is a 'draw' with each player winning one match (with 'draw' probability $D_r = p_{12}p_{23}p_{31} + p_{13}p_{32}p_{21}$). In a RR there are several criteria that might be used to order the players.

Criterion 1: (row sums/expected values) The players are ordered according to their row sum probabilities. This criterion has particular relevance when the interest is in the expected total reward for each player and the total reward for the tournament is the sum of the rewards for each match each with the same reward structure (eg. each player's earnings for a loss in any match is \$X and each player's earnings for a win in any match is equal \$Y (> \$ X). Also, the player with the greatest row sum has the greatest expected number of wins in the RR.

Criterion 2: (based on rank distributions) The players are ordered according to some aspect of their rank distributions. The relevant aspect can take on a range of forms. The rank distributions for three players are given in Table 2, which includes the numerical values relevant to Table 1.

Rank/Player	A	B	C
1 (worst)	$p_{21}p_{31}$ (0.16)	$p_{12}p_{32}$ (0.12)	$P_{13}p_{23}$ (0.48)
2	$p_{12}p_{31} + p_{13}p_{21}$ (0.48)	$p_{21}p_{32} + p_{23}p_{12}$ (0.56)	$P_{31}p_{23} + p_{32}p_{13}$ (0.44)
3 (Best)	$p_{12}p_{13}$ (0.36)	$p_{21}p_{23}$ (0.32)	$P_{31}p_{32}$ (0.08)
Total	1	1	1

Table 2: The rank distributions of three players

Criterion 3: (win outright) Here the players are ordered according to their probabilities of winning all their matches in the RR. For 3 players, this ordering is based on the sizes of $p_{12}p_{13}$, $p_{21}p_{23}$ and $p_{31}p_{32}$ for players A, B and C respectively. This criterion is particularly relevant when the outright winner (only) gets an extra reward such as the right to go on to some major tournament. *Interestingly, player A in Table 1 has a higher probability than* *player B of winning outright (0.36 Vs 0.32), even though players A and B have equal row sums.*

Criterion 4: (lose outright) Here the players are ordered according to their probabilities of losing all their matches in the RR. For 3 players, this ranking is based on the sizes of $p_{21}p_{31}$, $p_{12}p_{32}$ and $p_{13}p_{23}$ for players A, B and C respectively. This criterion is particularly relevant when the outright loser misses out on an important opportunity that the other two players receive. *Interestingly, player A in Table 1 has a higher probability than player B of losing outright (0.16 Vs 0.12), even though players A and B have equal row sums*. Note that the amount by which player B loses relative to player A in not being the outright winner (0.04) is equal to the amount by which player B gains relative to player A in not being the outright loser (0.04).

Criterion 5: (expected rank) Here the players are ordered according to the expected values of their rank distributions. These can be determined from Table 2. This ordering is relevant to the situation in which the players are paid (linearly) according to their final rank.

Criterion 6: (distribution of number of wins) Here the players are ordered according to some aspect of their number of wins distributions. Note that the aspect in question may or may not be linearly related to the number of wins. In general this criterion can be different to criterion 2, although, for the case of three players, it is in fact the same as, in Table 2, the ranks of 1, 2 and 3 in the first column are replaced by wins of 0, 1 and 2 respectively.

Criterion 7: (pair-wise comparisons): Suppose S_A is player A's probability of having more wins than player B plus his probability of having more wins than player C. S_A equals $2p_{12}p_{13} + p_{12}p_{31}p_{32} + p_{13}p_{21}p_{23}$. Player B's probability of having more wins than player A plus his probability of having more wins than player C, S_B, equals $2p_{21}p_{23} + p_{21}p_{31}p_{32} + p_{23}p_{12}p_{13}$. The corresponding value for player C, S_C is $2p_{31}p_{32} + p_{31}p_{21}p_{23} + p_{32}p_{12}p_{13}$. Under criterion 7 the players are ranked according to the sizes of S_A, S_B and S_C. Note that in Table 1 $S_A = S_B = 0.96$ and so A and B are equal players under criteria 7.. *In general, it can be shown that, for three players (but not necessarily for four players), criterion 7 gives the same ordering as criterion 1.*

A random knock-out (RKO) with three players

Here we consider the situation in which one player plays another in the first match whilst the remaining third player has a 'bye'. This match is followed by the final in which the winner of the first match plays the person who had the bye.

The probability player A wins a RKO is equal to $(2p_{12}p_{13} + p_{12}p_{23} + p_{13}p_{32})/3$, whilst the probability player B wins a RKO is equal to $(2p_{21}p_{23} + p_{21}p_{13} + p_{23}p_{31})/3$, the difference between the first expression and the second being $p_{12}p_{13} - p_{21}p_{23}$. The corresponding difference for players A and C is $p_{12}p_{13} - p_{31}p_{32}$, and for players B and C it is $p_{21}p_{23} - p_{31}p_{32}$. The difference between the probability player A wins the RKO and the probability that player B wins it equals the difference between their respective probabilities of winning a RR outright. This equality also applies to the differences between the other pairs of players. *Thus, the ordering of the three players in a RKO is always identical to the ordering under criteria 3 for a RR. (This is not necessarily true for the case of four players.) Thus, in Table 1, player A has a higher probability of winning the RKO (0.44) than does player B (0.40), even though players A and player B are 'equal players' according to criteria 1 and criteria 7 for the RR.*

An analysis of the ladder system (LS)

The LS has the following structure. With three players there is an order on the LS at the beginning of a cycle. The 'top down' cycle consists of the leader on the LS playing the second player, with the winner becoming the new leader on the LS, and the loser then playing the third player on the LS with the loser moving to the bottom of the LS. There is also a 'bottom up' cycle. The 6 possible orders (or states) at the beginning of a cycle are ABC, ACB, BAC, BCA, CAB, CBA, called states 1 to state 6 respectively. The steady state probabilities are

$$\pi_1 = (p_{23})\left(\frac{p_{12}p_{13}}{p_{12}p_{13} + p_{21}p_{23} + p_{31}p_{32}}\right),$$

$$\pi_2 = (p_{32})\left(\frac{p_{12}p_{13}}{p_{12}p_{13} + p_{21}p_{23} + p_{31}p_{32}}\right),$$

$$\pi_3 = (p_{13})\left(\frac{p_{21}p_{23}}{p_{12}p_{13} + p_{21}p_{23} + p_{31}p_{32}}\right),$$

$$\pi_4 = (p_{31})\left(\frac{p_{21}p_{23}}{p_{12}p_{13} + p_{21}p_{23} + p_{31}p_{32}}\right),$$

$$\pi_5 = (p_{12})\left(\frac{p_{31}p_{32}}{p_{12}p_{13} + p_{21}p_{23} + p_{31}p_{32}}\right)$$, and

$$\pi_6 = (p_{21})(\frac{p_{31}p_{32}}{p_{12}p_{13} + p_{21}p_{23} + p_{31}p_{32}}).$$

The probabilities for the 'bottom up' cycle are the same. Thus, the steady state probabilities that A, B, and C are on the top of the LS are respectively (π_1 + π_2), (π_3 + π_4), (π_5 + π_6), and the probabilities that they are on the bottom of the LS are (π_4 + π_6), (π_2 + π_5), (π_1 + π_3) respectively. *It would seem reasonable to assume that the sizes of these steady state probabilities give an appropriate ordering for the players in this ladder situation.*

We note a useful method for generating the steady state probabilities. The method is given here because the same procedure can be used for four players. The first state ABC has 3 implicit ordered pairs, AB, AC and BC, with 'associated probabilities' p_{12}, p_{13} and p_{23}. The product of these three probabilities is equal to $P_1 = p_{12}*p_{13}*p_{23}$. The second state ACB has the three implicit ordered pairs, AC, AB and CB with associated probabilities p_{13}, p_{12} and p_{32}. The product of these second three probabilities is equal to P_2. If the sum of the corresponding six products is denoted by S, then the steady state probability for state i is equal to P_i /S (i = 1, 2, 3, 4, 5, 6).

It follows that the steady state probabilities that player A is at the top of the LS (rank 3), and at the bottom (rank 1) are given in Table 3, which also gives the values for players B and C. In this table, D = $p_{12}p_{13} + p_{21}p_{23} + p_{31}p_{32} = p_{21}p_{31} + p_{32}p_{12} + p_{13}p_{23}$.

Each player's probability of having rank 2 can be obtained by subtraction from unity in Table 3. The expected value of player A's rank can be shown to equal $E(R_A) = 2 + (p_{12}p_{13} – p_{21}p_{31})/D = 2 + (p_{12} + p_{13} – 1)/D$, and the expected values for the other players are $E(R_B) = 2+(p_{21} + p_{23} – 1)/D$ and $E(R_C) = 2+(p_{31} + p_{32} – 1)/D$.

As the expected rankings within the LS are functions of the row sums (for three players), the expected rankings under the LS give rise to the same ordering as under criteria 1 and criteria 7. Thus, when p_{12} = 0.6, p_{13} = 0.5, p_{23} = 0.6, the row sums are 1.1, 1.0, 0.9, the values for E(Rank) are 2.1351, 2.0, 1.8649, and the above linear relationship is clear. Further, when p_{12} = 0.6, p_{13} = 0.6, p_{23} = 0.85, player A has a higher probability than player B of being at the top of the LS (0.4737 > 0.4474), although player B is better than player A under criteria 1 and criteria 7. This is because in this example $p_{12}p_{13} > p_{21}p_{23}$. *It can be seen that the players' probabilities of being at the top of the LS are in exact accordance with criteria 3.*

Probability	Top of ladder (rank 3)	Bottom of ladder (rank 1)
Player A	$p_{12}p_{13}/D$	$p_{21}p_{31}/D$
Player B	$p_{21}p_{23}/D$	$p_{32}p_{12}/D$
Player C	$p_{31}p_{32}/D$	$p_{13}p_{23}/D$

Table 3: Some details for the ladder system

The ordering of players can depend on the scoring system for each individual match

We now suppose that, in the example in Table 1, the scoring system to be used for each match between the players is the best-of-three of the earlier 'units of play' (e.g. best-of-three sets of tennis). The relevant match probabilities are now as in Table 4.

Probability i beats j	Player A	Player B	Player C	Row total
Player A	X	0.648	0.648	1.296
Player B	0.352	X	0.93925	1.29125
Player C	0.352	0.06075	X	0.41275
Total				3.0

Table 4: The example in Table 1 revisited, with each match being the 'best-of-three' units of play'

It can be seen that now the best player is player A according to criteria 1 (rather than players A and B being equal according to that criteria). *Thus, the ordering of a group of players can depend on the scoring system being used for each individual match. It follows that when data on best-of-three sets matches is used for the purpose of ordering players for a best-of-five sets competition, care should to be taken in the process of ordering the players.*

The ordering of four players

We consider the effect of adding a fourth player to the three players considered above, and see whether, by adding a fourth player, the ordering of the initial three players can be affected.

Consider the three players A, B and C in Table 1, and add a fourth player, Player D. Assume each of players A, B and C has the same probability p of beating player D, and consider eight examples.

(a) p = 1.0, (b) p = 0.75, (c) p = 0.5, (d) p = 0.25, (e) p = 0.0, (f) p_{23} is changed from 0.8 to 0.9, and p = 0.5, (g) p_{23} is changed from 0.8 to 0.8706, and p = 0.5, (h) p_{23} is changed to 0.885, and p = 0.5.

Examples (f), (g) and (h) are not chosen arbitrarily, but are chosen quite specifically to demonstrate certain facts. Case (f) is an example in which players A and B are equal under ranking criterion 3, but player B is better than player A under ranking criterion 1. In case (g), the value of p_{23} = 0.8706 has been chosen so that the probability player A wins a random knock-out equals the probability player B wins it, even though player B is better than player A under criterion 1. Further, case (h) is an example in which player A has a higher probability than B of winning under criteria 3, but has a lower probability of winning a random knock-out. The purpose of these various examples is to demonstrate that the relative ordering of (say) 2 players can depend on the scoring system used for the tournament and the criteria within that scoring system.

Thus, in case (a) players A, B and C each beat player D with probability 1, in case (c) players A, B and C each beat player D with probability 0.5, and in case (e) players A, B and C always lose to player D with probability 1. For all of cases (a) to (e), the row sum for player A equals the row sum for player B. Case (f) is the same as case (c) except that p_{23} is changed from 0.8 to 0.9, and case (g) is the same as case (c) except that p_{23} is changed to 0.8706. Case (h) is the same as case (c) except that p_{23} is replaced by 0.885.

A round robin tournament with four players

The seven criteria can be extended to four players. We note the following results

(i) Considering cases (a) to (e), provided p is not equal to 0 or 1, player A performs better than player B under criteria 7. *Thus, the addition of the fourth player, player D, has led to a (non-zero) criteria 7 difference between player A and player B, a difference which is in the same direction as the outright win or criteria 3 difference.*

(ii) For cases (a), (c) and (e) the expected value of the ranks for players A and B are equal, whereas the expected values are very slightly different in cases (b) and (d), even though players A and B have equal row sums. *Unless p = 0, 0.5 or 1.0, players A and B score slightly differently under criteria 5 (E(rank)), even though they score equally under criteria 1.*

(iii) *In case (f), the probability that player A wins outright equals the probability that player B wins outright (i.e. players A and B are equal under criteria 3 since* $p_{12}p_{13}p_{14} = p_{21}p_{23}p_{24}$*, even though player B is better than player A under criteria 1.*

(iv) In case (h) player A has a higher probability than player B of winning the RR outright (criteria 3), whilst it will be seen in the next section that he has a lower probability of winning the RKO. Thus, *the addition of a fourth player has changed the earlier result for three players… 'the ordering of three players in a RKO tournament is always identical to the ordering under criteria 3 for a RR tournament'.*

The probability that player A wins outright, the probability player B wins outright, …are given by

$$P(A) = p_{12}\,p_{13}\,p_{14}\,,$$

$$P(B) = (1 - p_{12})\,p_{23}\,p_{24}\,,$$

$$P(C) = (1 - p_{13})(1 - p_{23})\,p_{34}\,, \text{ and}$$

$$P(D) = (1 - p_{14})(1 - p_{24})(1 - p_{34})\,,$$

where these probabilities are expressed as functions of the above-diagonal elements of the p_{ij} matrix. One could use differential calculus to evaluate the effect on any of the above expressions resulting from a change in one or more of the p_{ij} values (see Pollard and Pollard (2007a and b)). Correspondingly, one could write down expressions for the player's values under criterion 7, and use the calculus similarly.

A random knock-out with four players

For the cases (a) to (d), the probability that player A wins a RKO is always greater than the probability player B wins it, even though players A and B are equal under criteria 1. In case (g), the value of p_{23} (0.870588235) has been selected so that the above two probabilities are equal.

Except for examples (f), (g) and (h), the probability players A and C are in the final equals the probability players B and C are in the final, and the probability A and D are in the final equals the probability players B and D are in the final. These equalities may have been anticipated. One can write down expressions for P(A wins the RKO), P(B wins the RKO), …The expression for player A is

$$p(A) = (\,p_{12}(\,p_{13}p_{34} + p_{14}p_{43}) + p_{13}(\,p_{12}p_{24} + p_{14}p_{42}) + p_{14}(\,p_{12}p_{23} + p_{13}p_{32}))/3$$

The ladder system with four players

There are 24 states; ABCD, ABDC, ACBD, …, DCAB and DCBA, called states 1, 2, 3, …, 23 and 24. Using recurrence methods, the steady state probabilities were found by inverting the relevant 23*23 matrix, and it was verified that those probabilities could be calculated by the method discussed above. The steady state probabilities of being on top of the LS (rank 4), second on the LS (rank 3), …, for each of the players were calculated.

Player	Rank 4	Rank 3	Rank 2	Rank 1	Sum	Expected Rank
A	0.3553	0.2763	0.2105	0.1579	1.0	2.8289
B	0.3158	0.3158	0.25	0.1184	1.0	2.8289
C	0.0789	0.1579	0.2895	0.4737	1.0	1.8421
D	0.25	0.25	0.25	0.25	1.0	2.5
Sum	1.0	1.0	1.0	1.0		

Table 5: Steady state probabilities for the ranks (case (c))

These are given in Table 5 for case (c). Player D's expected rank is 2.5 (midway between 1 and 4), and the other players' expected ranks are linearly related to their row sums (eg. 2.8289 = 2.5 + (1.7 – 1.5)*1.6447; and 1.8421 = 2.5 + (1.1 – 1.5)*1.6447.) The following results for the LS are noted.

(a) If Player A is first on the LS with a higher probability than player B, then player A wins the RR outright more often than player B.

(b) For the LS, E(Rank A) equals E(Rank B) in cases a), c) and e). However, E(Rank A) is less than E(Rank B) for case b) (as for the RR), and E(Rank A) is greater than E(Rank B) for case (d) (as for the RR), even though players A and B have equal row sums. This can be explained, for example, by the nature of the expression for P(A) below. *Thus, unless p = 0, 0.5 or 1.0, players A and B score slightly differently under criteria 5 (E(rank)), even though they score equally under criteria 1.*

The probability A is on top of the LS (rank 4) is

$$P(A) = K(p_{12}p_{13}p_{14}(p_{23}p_{24} + (1 - p_{23})p_{34} + (1 - p_{24})(1 - p_{34})),$$

where K is the reciprocal of the general expression for the sum of the various 'products' (see earlier).

An example of ordering players

Suppose we have four 'equal but not identical' players with the pij matrix given in Table 6.

p_{ij}	Player A	Player B	Player C	Player D	Sum
Player A	X	0.6	0.5	0.4	1.5
Player B	0.4	X	0.6	0.5	1.5
Player C	0.5	0.4	X	0.6	1.5
Player D	0.6	0.5	0.4	X	1.5

Table 6: An example with four 'equal' players

Now suppose p_{14} is increased from 0.4 to 0.5, and hence p_{41} is decreased to 0.5, making A clearly the best and D the worst player. Further, suppose p_{23} is changed from 0.6 to p, and hence p_{32} is changed to 1-p. We consider the ordering of the players for five scoring systems (winning the RR outright, pair-wise comparisons for the RR, winning the RKO, being top of the LS, and E(Rank) in the LS) as p increases from 0.6 up to 1. It can be shown that

(a) Based on the criteria of winning the RR outright, the ordering of the players when p is between 0.603 and 0.666 is ABCD. The ordering changes to ABDC when p is somewhere in the range 0.666 to 0.667, and it changes again to BADC as p passes through 0.75. Thus, as B's probability of beating C increases, firstly D becomes a better player than C when p is somewhere between 0.666 and 0.667, and secondly, B becomes better than A as p moves above 0.75.

(b) When p is in the range 0.675-0.684, the ordering of the players is ABDC based on winning the RR outright or being top in the LS. However, the ordering is ABCD for three other criteria (the pair-wise comparisons, the probability of winning the RKO, and the size of E(Rank) within the LS). *Thus, for a given value of p, the ordering of the players depends on the scoring system being used and the criteria for ordering within that scoring system.*

(c) In four of the five scoring systems (excluding E(Rank)) the order changes from CD to DC and then from AB to BA at different values of p, the second change always being at a higher value of p. With E(Rank) for the LS however, these two changes occur at exactly the same value of p (89/128). This 'p-range' from the 'first order change to the second order change' is smallest (ie. zero) for E(Rank) within the LS criteria, second smallest for the paired comparisons within the RR criteria, third smallest for RKO probability of winning criteria, second largest for the probability of being top in the LS criteria, and largest for the probability of winning the RR outright criteria.

Seeding for optimal knock-out outcomes

It is standard practice to seed players for a KO tournament. What is the reason for seeding? It would appear that the main reason for seeding players is to 'spread the better players' across the draw. Given four players in a KO tournament, we could quantify the reason for seeding by maximizing the probability the best two players reach the final. We demonstrate how this can be achieved for four players. Suppose the relevant matrix is as given in Table 7.

The probability that the players win a RKO are given in Table 7. The ordering of the players from best to worst is A, B, D, and C, based on the last column. One way of finding the KO draw that maximizes the probability that the best two players reach the final is as follows. Firstly, the matrix in Table 7 is re-ordered with the players in the order from best to worst. This is done in Table 8.

Probability matrix	A	B	C	D	Row Sum	P(Player wins RKO)
A	X	0.55	0.55	0.5	1.6	0.2851
B	0.45	X	0.65	0.5	1.6	0.2814
C	0.45	0.35	X	0.6	1.4	0.2135
D	0.5	0.5	0.4	X	1.4	0.22

Table 7: A random knock-out example

Calculate the maximum of $p_{13}p_{24}$ and $p_{14}p_{23}$ in this re-arranged matrix. The value of $p_{13}p_{24}$ is 0.5*0.65 = 0.325, and the value of $p_{14}p_{23}$ is 0.55*0.5 = 0.275, the maximum being $p_{13}p_{24}$. Noting the subscripts, if player A plays player D (and B plays C) in the first round, the probability that the best two players, A and B, play in the final will be maximized.

Win Probability	A	B	D	C	Sum
A	X	0.55	0.5	0.55	1.6
B	0.45	X	0.5	0.65	1.6
D	0.5	0.5	X	0.4	1.4
C	0.45	0.35	0.6	X	

Table 8: A re-arranged matrix with players in order

When is ordering/ranking of players necessary?
We have seen that if the organizers of a KO tournament wish to maximize the probability that the best two players reach the final, there needs to be a way of finding who the best two players actually are. Using the criterion that the best two players are the two who have the highest probabilities of winning the KO tournament, we have seen how this can be done. For example, with 3 players, we can consider $p_{12}*p_{13}$ for player A, $p_{21}*p_{23}$ for player B, and $p_{31}*p_{32}$ for player C. The best two players are the two with the largest values for these products. With four players we can consider differences such as the probability player A wins a RKO minus the probability player B wins it, which is given by

$$p_{34}(p_{12}p_{13}-p_{21}p_{23})+p_{43}(p_{12}p_{14}-p_{21}p_{24})$$
$$+(p_{13}p_{24}+p_{14}p_{23})(p_{12}-p_{21})$$
$$+p_{13}p_{14}(p_{42}+p_{32})-p_{23}p_{24}(p_{31}+p_{41}).$$

By considering such differences, the best two of the 4 players in a KO can be identified, and the above method used to determine the (optimal) draw. For a RR there is no obvious need to identify the best two players as the system is fair in the sense that every player plays every other player. Note however that an ordering method can be used to select which four (out of five or more) players are placed in the draw. With a KO structure however every player does not play every other player, and there is a greater need for ordering. Also, under the LS, every player plays every other player in due course in a way that is automatically determined by the LS. Thus, as for the RR, there is less need to order the players in the LS.

3. RESULTS

Several results have been noted in the methods section, and the main results have been summarized in the conclusions section. The overall result however is that the ordering of a group of players can depend on the scoring system being used for each individual match, the tournament structure and the criteria for ordering within that structure.

4. DISCUSSION

We have considered the RR, KO and LS in some detail. The KO system has a particular importance in many sports. For the KO the best player was defined as that player who has the highest probability of winning a RKO, the second best as that player with the second highest probability of winning, etc. The effect of alternative definitions might be explored. For example, given four players, the second best player might be defined as that player who, after the best player has been identified, has the next highest probability of reaching the final of a RKO. The worst player might be that player who has the highest probability of losing in the first round, or the lowest probability of reaching the final in a RKO. Alternatively the players might be ordered according to their expected earnings, or their probabilities of reaching a RKO final. These definitions may give rise to different orderings.

5. CONCLUSIONS

An ordering of players allows tournaments to decide which players to accept into the draw. For knock-out events it allows them to be seeded. The ordering of players however is not a trivial problem.
We believe that the listing of the 7 ranking criteria, and the various results for round robins (RR), random knock-outs (RKO) and ladder systems (LS) for 3 and 4 players, identifies a range of issues and conclusions that are new or not commonly appreciated. Extending these results to more players is of course not a trivial exercise, but the authors believe that this paper gives important insights into addressing this extension.
Several criteria for ordering the players within a RR were considered. One was based on the sum of a player's probabilities of beating the others. This general criterion has relevance when the interest is in the total expected earnings or the expected number of wins. A second criterion was based on a player's probability of winning outright, and this is relevant when (say) only one player moves to the next stage. A third criterion was based on a player's probability of losing outright, and this is relevant when (say) all but one of the players move to the

next stage. A fourth general criterion was based on a player's 'pair-wise comparisons' with the others.
For four players, the ordering of the players in a RR is not necessarily the same for these criteria.
The probability of winning a RKO was used as the criterion for ordering players for a KO event. For 3 players (but not necessarily 4 players) this ordering is always the same as for the second criterion above.
For the LS one criterion was the probability of being on top of the ladder. A second was the expected rank on the ladder. For three players (but not necessarily four players), this second criterion gives the same ordering as the first and fourth criteria above.
The ordering for a group of players playing best-of-three sets matches can be different to the ordering for the same group playing best-of-five set matches. Thus, when data on best-of-three sets matches is used for ordering players for a best-of-five sets event, care should to be taken in the ordering.
The overall conclusion is that the ordering of a group of players can depend on the scoring system used for each match, the tournament structure and the criteria for ordering within that structure.

References

Ali, I., Cook, W.D., & Kress, M. (1986). On the minimum violations ranking of a tournament. *Management Science, 32(6):* 660-672.

Barnett, T.J., & Clarke, S.R. (2002), Using Microsoft Excel to model a tennis match. Pp 63-68 in Cohen, G, Langtry, T eds. *Proc 6th Australasian Conference Maths and Computers in Sport*.

Bedford, A.B., & Clarke, S.R. (2000), A comparison of the ATP rating with a smoothing method for match prediction. Pp 43-51 in Cohen, G, Langtry, T eds. *Proc. 5th Australasian Conference Maths and Computers in Sport.*

Blackman, S.S., & Casey, J.W. (1980), Development of a rating system for all tennis players. *Operations Research. 28(3):* 489-502.

Clarke, S.R. (1994). An adjustive rating system for tennis and squash players. *Proc 2nd Australasian Conference Maths and Computers in Sport*, ed. Neville de Mestre, Bond University, Queensland, 43-50.

Clarke, S.R., & Dyte, D. (2000), Using official ratings to simulate major tennis tournaments, *International Transactions in Operational Research 7*, 585–594.

Cook, W.D., Golan, I., & Kress, M. (1988). Heuristics for ranking players in a round robin tournament. *Computers and Operations Research, 15(2):* 135-144.

Cook, W.D., & Kress, M. (1990). An mth generation model for weak ranking of players in a tournament. *Journal Operational Research Society, 41(12):* 1111-1119.

Klaassen, F.J.G.M., & Magnus, J.R. (2003), Forecasting the winner of a tennis match. *European Journal of Operations Research; 148*: 257-267

Musante, T.M., & Yellin, B.A. (1979), The USTA/Equitable Computerized Tennis Ranking System. *Interfaces. 9(4):* 33-37.

Pollard, G.N., & Pollard, G. H. (2007a), Importances 1:the most important sets in a match, and the most important points in a game of tennis, in *Tennis Science and Technology 3*, edited by S. Miller and J. Capel-Davies, International Tennis Federation, Roehampton, London, p281-292.

Pollard, G. N., & Pollard, G. H. (2007b), Importances 2:the most important points in a tiebreak game, and the most important games in a set, in *Tennis Science and Technology 3*, edited by S. Miller and J. Capel-Davies, International Tennis Federation, Roehampton, London, p293-300.

Pollard, G.N., & Meyer, D.H. (2010) Wiley *Encyclopedia of Operations Research and Management Science*, Ed. James J. Cochran, to appear.

Stefani, R.T. (1997), Survey of the major world sports rating systems. *Journal of Applied Statistics, 24(6):* 635-646.

Strauss, D., & Arnold, B.C. (1987), The rating of players in racquetball tournaments. *Applied Statistics, 36(2):* 163-173.

SOME ASPECTS OF ORDERING, RANKING AND SEEDING 2

Meyer, Denny[1,3] Pollard, Graham[2] and Pollard, Geoff[1]
[1] *Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia*
[2] *Faculty of Information Sciences and Engineering, University of Canberra, Australia*
[3] *Corresponding author: dmeyer@swin.edu.au*

Abstract

This is the second of two papers on ordering, ranking and seeding players and teams in sport. Such orderings may be used for player selection, seeding, handicapping, player progress evaluation and for predicting match and tournament outcomes. KAN-Soft data for eight of the top male tennis players is used in a simulation study to explore various ranking systems. These ranking systems rely on historical relative frequencies for pairwise performance in the context of two scoring systems and three tournament structures. Of special interest is the sensitivity of the rankings to individual cases.

Keywords: Ordering, ranking, seeding, round robins, random knockouts, ladders, ratings.

1. INTRODUCTION

Player orderings or rankings have administrative and motivational uses in all sports. For instance they are useful for selecting which players or teams should be allowed into elite tournaments, for seeding, handicapping, monitoring progress and predicting match outcomes (Bedford and Clarke, 2000; Klaassen and Magnus, 2003; Pollard and Meyer, 2010). Several methods have been suggested for ranking purposes. Some of these methods rely on prize money awarded in the past while others use more heuristic measures based on particular point systems (Musante and Yellin, 1979; Cook, Golan and Kress, 1988; Cook and Kress, 1990, Stefani, 1997). Evolutionary approaches for the establishment of ratings have also been considered (Bedford and Clarke, 2000; Clarke, 1994).
In this paper pairwise comparisons of match performance are used in order to establish ratings (Blackman and Casey, 1980; Ali, Cook and Kress, 1986; Strauss and Arnold, 1988). Pollard, Pollard and Meyer (2010) describe these methods in more detail and, in particular, they have explored various pairwise comparison criteria for ordering players (and teams) considering round robins, knockouts and ladder systems for three and four players. This has been done assuming that the probability that a player wins when playing against a specific opponent is known. Their results suggest that the ordering of players may depend on what tournament system is assumed and what scoring system is used for individual matches. In this paper these results are considered in the context of elite male tennis using KAN-Soft data to provide pairwise information for eight of the current top players in the world.

2. METHODS

Using the 2009 December KAN-Soft database, match results for the top eight male players were extracted and the relative frequencies for a win were determined for each possible pairing. Table 1 shows these results for all matches played between these eight players according to the KAN-Soft database. Some of these matches took place some time ago. For example the first match between Federer and Nadal took place at the French Open on 23rd May 2005. Comparing the KAN-Soft rankings for December 2009 with those of a year previously little change was observed. Only two players had dropped out of the initial eight rankings (Badopalia and Tsonga) with Sorderling and Del Potro as their replacements. The data shows that of the 77 matches he has played Federer has won 50 matches, giving him a 65% success rate overall, which has led to his dominance in the game for many years.

ID Winner	ID Loser 19	89	568	652	677	1075	5992	6465	Match Wins (2009) (Rank)
19 Federer	-	12	6	9	6	3	8	6	50 (1)
89 Roddick	1	-	2	0	2	3	4	0	12 (7)
568 Davydenko	1	1	-	3	4	4	1	3	17 (6)
652 Sorderling	0	2	6	-	2	1	1	0	12 (8)
677 Nadal	12	3	4	3	-	7	12	4	45 (2)
1075 Murray	6	5	5	0	2	-	3	5	26 (4)
5992 Djokovic	5	2	3	4	7	4	-	3	28 (3)
6465 Del Potro	2	2	0	2	3	1	0	-	10 (5)
Matches Lost	27	27	26	21	26	23	29	21	50 (1)
2008 Rank	2	6	5	17	1	4	3	9	

Table 1: Number of matches won and lost for the top eight male players as at 7/12/09 KAN-Soft data base.

As shown in Table 1, the top eight players have all played against each other at least once. However, for some pairs there are very few matches (e.g. Sorderling and Murray have met on only one occasion). This deficiency is addressed by using a modeling approach to estimate the pairwise probabilities for winning in the case of each pairing.

The model is constructed using the summary statistics given in Table 2, which shows the mean percentage of points won on serve and return for each of the eight players for all matches played amongst each other. The good performance of Federer on both service and return again justifies his position on the rankings as number one. However, Nadal is ahead of Federer on return and not far behind on serve, confirming that he is a worthy challenger. Amongst the other players Roddick is let down by his return while Davydenko and Del Potro are let down by their serves.

Player	Mean success on serve (%)	Mean success on return (%)	Rank on service	Rank on return
Federer	67.76	38.69	1	2
Roddick	65.03	31.04	2	8
Davydenko	60.95	38.53	7	3
Sorderling	62.41	32.96	5	6
Nadal	62.84	39.36	3	1
Murray	61.10	37.69	6	4
Djokovic	62.56	36.95	4	5
Del Potro	59.75	32.62	8	7

Table 2: Statistics used to predict the probability of a win on any match

Using the results in Table 2 a binary logistic regression model which allows for the prediction of match probabilities (win or loss), for all possible pairwise combinations of players, was developed. In this analysis the results were weighted according to the number of matches played by each pair of players as indicated in Table 1.

In the following equation p_{ij} represents the probability that player i will beat player j, while $Serve_i$ represents the probability of player i winning a point on serve and $Return_j$ represents the probability of player j winning a point on return. An interaction term between these two predictors failed to improve the prediction and was therefore omitted.

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = .0887Serve_i - .1526Return_j \qquad (1)$$

The odds ratio for the serve variable is 1.093 suggesting that a 1% improvement for points won on serve will increase the odds of winning by 9% on average. The odds ratio for the return variable is .859 suggesting that a 1% improvement for points won by the opponent on return will reduce the odds of a win by 14% on average. However, there is a problem with these predictions because, for any match, the predicted probabilities of a win for both players do not necessarily add to one. This deficiency in the model predictions is easily corrected by dividing by the sum of the estimated win probabilities for each pair of players. Figure 1 shows how these estimated probabilities of a win compare to the relative frequencies calculated from Table 1. In this plot the size of the points indicate the number of matches used to compute each relative frequency or “observed probability”. There is an obvious discrepancy between estimated and observed probabilities when players have played each other on only rare occasions, especially when one of the players has won all these matches. However, when two players have met many times, as indicated by the larger points, the predicted and observed probabilities appear much closer. These results suggest that the predicted probabilities are more reliable than the "observed probabilities", giving a better indication of the relative performance of each player. This is to be expected since these predicted probabilities are based on the average performance of each player on service and return for all of their Table 1 matches.

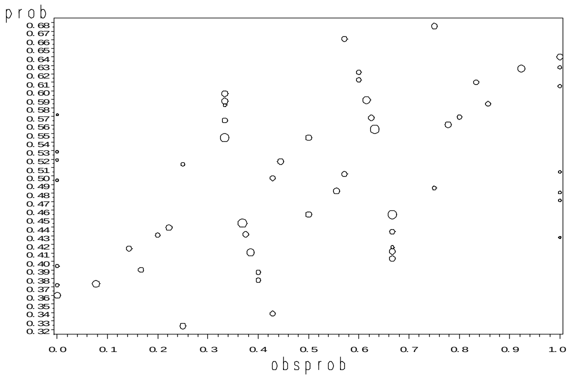


Figure 1: Predicted Probabilities (prob) of winning compared to Relative Frequencies (obsprob) from Table 1

A ranking system for these eight players can be conceptualized as a one dimensional representation of these predicted pairwise probabilities of winning. Principal component analysis is commonly used to reduce the dimension of matrices of various types, with the eigenvalues used to measure the importance of each dimension. This technique was developed by Hotelling (1933) after its original creation by Pearson (1901). A principal component analysis was therefore performed on the Table 1 matrix of predicted winning probabilities, with zeros on the diagonal, and the first eigenvalue was used as a measure of the ranking feasibility. In this paper this measure is used to compare the feasibility of a ranking system for two scoring systems and to test the importance of each of the eight players to the strength of the ranking system.

The initial analysis concerns a comparison of 3 set and 5 set scoring systems when the ranking system based on a simple row sum of winning probabilities, described by Pollard et al (2010), is used. The first eigenvalues are compared for these two scoring systems. In addition the sensitivity of the 5 set ranking system to each of the eight players is assessed by excluding each player one at a time, noting the value of the first eigenvalue and any changes in the rankings of the remaining seven players. Such a sensitivity analysis has practical applications in that it is important that injury or retirement for one of the top players should not make an existing ranking system redundant.

In the ensuing simulations performed using the estimated win probabilities, the rankings for these eight players are compared using various tournament structures. The purpose of these simulations is to not only show any differences in the expected rankings, but also to indicate the degree of variation that can be expected in the rankings obtained for the various tournament structures.

3. RESULTS

Comparison of scoring systems.

The initial ranking system developed using the predicted pairwise probabilities of winning a match is obtained using the row sum method of Pollard et al (2010). This method is used below to compare the rankings obtained for two match scoring systems, best of 5 sets and best of 3 sets.
The above predicted probabilities were based on a scoring system with the best of five sets used to determine the winner. Assuming independence between the performances of players in each set it is possible to compute the estimated probability of a win for each set (p) using the following equation for the probability of winning a five set match.

$$p^3(10-15p+6p^2) \qquad (2)$$

This allows the calculation of pairwise probabilities of winning a best of 3 sets match using the formula

$$p^2+2p^2(1-p) \qquad (3)$$

Table 3 compares the row sum of winning probabilities for each of the eight players for a best of 5 sets and a best of 3 sets scoring structure. The rankings are identical, however, there is clearly more variation in the row sum of winning probabilities for the best of 5 sets matches (2.77 – 4.25) than there is for the best of 3 sets matches (2.91 – 4.11). The first eigenvalue for the matrix of estimated pairwise winning probabilities is very similar for the two scoring systems, with a value of 55.58% for the 5 set matches and 55.79% for the 3 set matches. This means that a ranking system is only slightly more appropriate in the case of 3 set matches than it is in the case of 5 set matches, because a one dimensional model explains only slightly more of the variation in performance in the case of the best of three set matches. The shorter matches appear to better differentiate the performance of players, making it easier to create a meaningful ranking.

Player	Row Sum of Predicted Win Probabilities		Ranking	
	Best of 5 sets	Best of 3 sets	Best of 5 sets	Best of 3 sets
Federer	4.25	4.11	1	1
Roddick	3.06	3.15	6	6
Davy-denko	3.70	3.66	3	3
Sorder-ling	3.04	3.13	7	7
Nadal	4.01	3.91	2	2
Murray	3.57	3.56	5	5
Djokovic	3.58	3.57	4	4
Del Potro	2.77	2.91	8	8
Total	28	28	28	28

Table 3: Ranking based on the row sum of predicted pairwise probabilities for winning

Sensitivity of Rankings to Individual Players

Table 4 shows how the eigenvalues change when each player is removed in turn. It appears that in all cases the ranking is stronger when there are only seven players than when all eight players are included, because the first eigenvalue is slightly lower for the eight player ranking (55.58%) than for any of the seven player rankings. It appears that the strongest ranking emerges when Nadal is removed and the weakest ranking emerges when Roddick is removed. However, in only one case, when Murray is removed, is there a change in the rankings. This suggests that the row sum of winning probabilities produces a ranking system which is robust to the withdrawal of any one of the top eight players, confirming that a ranking system based on the row sum of winning probabilities is likely to be effective in practice.

Player removed	First eigen-value (%)	Ranking for remaining players using the row sum of winning probabilities							
		1	2	3	4	5	6	7	8
1.Federer	56.59	*	1	2	3	4	5	6	7
2.Nadal	56.63	1	*	2	3	4	5	6	7
3.Davydenko	56.53	1	2	*	3	4	5	6	7
4.Djokovic	56.41	1	2	3	*	4	5	6	7
5.Murray	56.43	1	2	3	7	*	5	6	4
6.Roddick	56.39	1	2	3	4	5	*	6	7
7.Sorderling	56.41	1	2	3	4	5	6	*	7
8.DelPotro	56.59	1	2	3	4	5	6	7	*
None	55.58	1	2	3	4	5	6	7	8

Table 4: Sensitivity analysis for ranking for best of 5 set matches

The results of the three tournament simulations are now considered. In these examples the estimated pairwise probabilities of winning are considered only for five set matches.

Round Robin Tournament

Results for 1000 round robin tournaments were simulated using the above predicted pairwise probabilities. The average number of matches won by each player in these tournaments appear in Table 5. The rankings based on these averages are clearly identical to those obtained in Table 3.

Also of interest is the consistency of the players. Roddick and Nadal have the lowest standard deviations for the number of matches won suggesting that these two players are slightly more consistent than the other players. However, when one looks more closely at the simulated results it is interesting to find that there is an outright winner for only 63.9% of the tournaments, with 24.5% of tournaments shared between 2 winners and the remaining 11.6% of tournaments shared between 3-6 winners. This, as well as tournament time restrictions, means that round robin tournaments are not usually a feasible option at this level of tennis. We now consider the results for a simulation of results for 10000 knockout tournaments, which are, of course, the most popular form of tournament for elite players.

	Number Matches Won		
Player	**Mean**	**Standard Deviation**	**Ranking**
Federer	4.23	1.32	1
Roddick	3.07	1.26	6
Davydenko	3.78	1.30	3
Sorderling	2.97	1.33	7
Nadal	3.96	1.28	2
Murray	3.56	1.32	5
Djokovic	3.65	1.32	4
Del Potro	2.79	1.32	8
Total	28		28

Table 5: Rankings for 1000 Round Robin Tournaments

Knockout Tournaments

Two different types of knockout tournament are considered. The first simulation assumes a random knockout in that players are randomly assigned to divisions, while in the second simulation, referred to as a "seeded" knockout, Federer and Nadal are placed in separate divisions as is common for the top seeded players, in order to improve the chances of the top seeds reaching the final.

Table 6 clearly shows that the seeded knockout tournament strategy is beneficial for the top seeds. The simulation shows Federer winning 31.6% of the seeded knockout tournaments and reaching the final in a further 22.2% of these tournaments, whereas, in the random knockout tournaments, he wins only 23.5% of these tournaments and reaches the final in only a further 13.1% of these tournaments.

	Random Knockout			Seeded Knockout		
Player	**Win**	**Final**	**Rank**	**Win**	**Final**	**Rank**
Federer	23.5	13.1	1	31.6	22.2	1
Roddick	6.2	12.5	7	4.3	5.8	8
Davydenko	10.5	16.9	5	8.4	8.8	5
Sorderling	8.0	10.5	6	5.7	8.4	6
Nadal	20.4	12.0	2	18.4	17.6	2
Murray	12.0	13.4	4	12.4	13.5	4
Djokovic	14.3	11.5	3	14.1	15.2	3
Del Potro	5.2	10.2	8	4.9	8.4	7

Table 6: Rankings for 10000 Knockout Tournaments

However, the rankings for random and seeded knockouts are similar with only Roddick and Del Potro changing places. However, a comparison of the results for knockout and round robin tournaments shows a major change in the rankings. Although Federer and Nadal are consistently in the top two positions there are plenty of differences for the other players. Djokovic and Murray move up to positions 3 and 4 in the case of knockouts, while Roddick and Sorderling and Davydenko move down.

We now consider Ladder Tournaments which assume an initial ladder and allow a number of challenges. Of particular interest in this case is the number of challenges required for the ranking to stabilize.

Ladder Tournaments

In our analyses we use the 2009 KAN-Soft rankings to suggest the initial ladder. This means that the order of play for the initial ladder challenge follows the pattern shown in Table 7. This initial series of matches is referred to as the first ladder challenge below. Figure 2 illustrates the mean average rankings for 1000 simulations for one up to 1000 such ladder challenges.

a. Federer plays Nadal deciding rank 1

b. If Federer wins then Nadal plays Djokovic. Otherwise Federer plays Djokovic. This match decides rank 2.

c.The winner of (b) plays Murray. This match decides rank 3.

d.The winner of (c) plays DelPotro. This match decides rank 4.

e.The winner of (d) plays Davydenko. This match decides rank 5.

f.The winner of (e) plays Roddick. This match decides rank 6.

g.The winner of (f) plays Sorderling. This match decides ranks 7 and 8.

Table 7: Initial Ladder Challenge based on 2009 KAN-Soft rankings

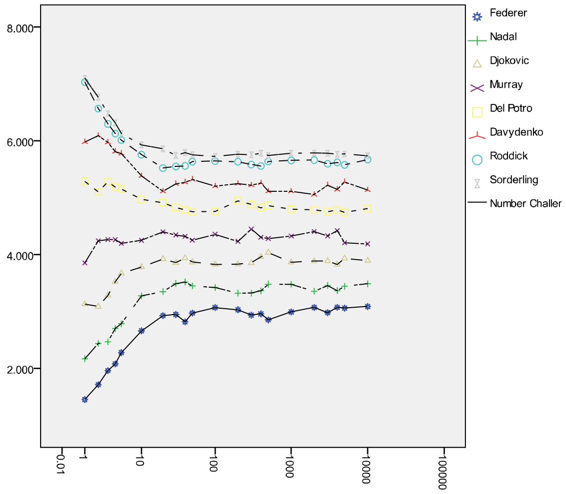


Figure 2: Average rank plotted against number of ladder challenges on a log scale

Figure 2 shows that regardless of the number of challenges the players appear to hold their positions, suggesting that it does not really matter how many ladder challenges are considered. However, for the first 10 ladder challenges the mean rankings move closer together, perhaps suggesting that rankings should be based on 10 ladder challenges.

Figure 3 shows the standard deviations for the above rankings. Clearly the standard deviations for the rankings grow as the number of ladder challenges increases, especially for the higher seeded players. However, the weaker players, such as Sorderling show more stability. Interestingly the standard deviations appear to stabilize after only 10 ladder challenges, again suggesting that this might be a realistic number of ladder challenges to use in order to establish rankings.

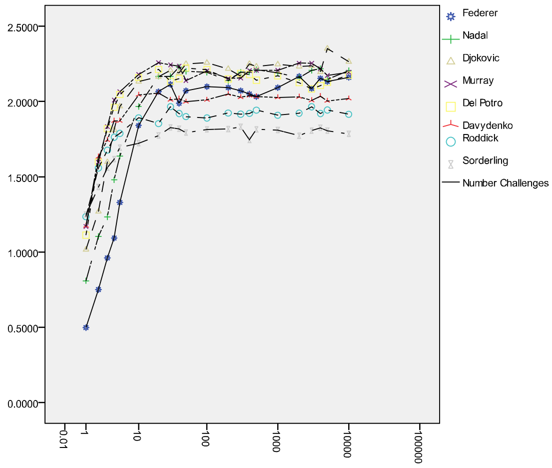


Figure 3: Rank Standard Deviation plotted against number of ladder challenges on a log scale

Finally, the sensitivity of the rankings to the initial ladder is explored in Figure 4. The initial ladder for Figure 4 has Djokovic at the top of the ladder, Federer at number 2 and Nadal at number 3, but the remaining players retain the spots allocated to them previously. The results clearly show that Federer tends to move back to the top of the rankings after a single ladder challenge and that the rankings of all the other players are unchanged. This confirms that this ranking system is not sensitive to the initial allocation of rankings and is capable of correcting any error in the initial ladder ranking. These results also suggest that it is sufficient to consider 10 ladder challenges when setting up a ranking based on the ladder tournament structure

4. DISCUSSION

This paper has developed a method for estimating pairwise probabilities for winning matches. This means that even when two players (or teams) have never met, it is still possible to estimate their relative chances of winning a match.

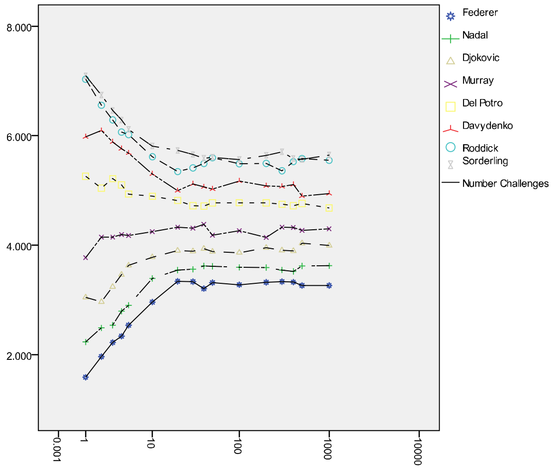


Figure 4: Average rank plotted against number of ladder challenges on a log scale when Djokovic starts at the top of the ladder

It has been suggested that a ranking system can be established by taking the row sum of these estimated pairwise probabilities of winning. A sensitivity analysis was conducted in order to assess the affect of different scoring systems and player retirement or injury on these rankings. No differences in the rankings for the best of three set matches and the best of five set matches were observed. Similarly the removal of single players appeared to have no effect on the rankings except in the case of Murray. Further sensitivity analyses could be conducted in terms of the estimated probabilities of winning and also in terms of the probabilities of winning a point on serve or return, which were used in the binary logistic regression model to estimate the pairwise probabilities of winning.

We have considered the rankings derived using simulated round robin, knockout and ladder tournaments. The purpose of these simulations was to allow for random variation in performance, thereby giving us some idea about the variability in rankings that could emerge from such systems if they were employed in practice. The only major difference between the rankings between these tournament simulations and the original ranking system, based on the row sum of predicted pairwise probabilities of winning, occurred in the case of knockout tournaments. This is strange, given the fact that the KAN-Soft data used in this analysis was collected using data from knockout tournaments. Although the positions of the top two players were not adversely affected in the case of knockout tournaments, even in the case of random knockout tournaments, it was found that the positions of the remaining six players changed when simulated knockout tournaments were used to create a ranking.

The knockout system has particular importance in many sports. In tennis in particular, almost all tournaments use the knockout structure. For this simulation the best player was defined as that player who had the highest probability of winning the random knockout tournament, the second best player was defined as the player with the second highest probability of winning, etc. We note here that the effect of alternative definitions might be explored. For example, given four players, the second best player might be considered to be that player who, after the best player has been identified, has the highest probability of reaching the final. The worst player might be that player who has the highest probability of losing in the first round, or the lowest probability of reaching the final. Alternatively the players might be ordered in terms of their probabilities of reaching, or they might be ordered strictly according to their expected earnings. In tennis for example, a typical 'winnings structure' is that the winner in the final receives $x, the loser in the final receives $x/2, and the losers in the semi-final receive $x/4. These different definitions for ordering the players may give rise to different rankings.

In the case of the round robin tournament simulations it was assumed that the best player won the most matches, the second best player won the second highest number of matches etc. For a single tournament this system will not work because tied results are so common. However, the simulation of 1000 tournaments overcame this problem. In the case of ladder tournaments one of the aims was to determine how many (complete) ladder challenges should be allowed in order to establish a final ranking. The results showed that any changes in the ranking (e.g. Djokovic placed at number 1 instead of 3) were quickly corrected by the ladder system and that rankings were not sensitive to the number of ladder challenges allowed. However, the simulations showed an increase in the standard deviations associated with the rankings from one to about ten ladder challenges, suggesting that ten ladder challenges would give a realistic picture of the ranking variability that could be expected with this system.

This analysis has considered only tennis and only the world's top eight male players. It is expected that further analyses which consider female tennis tournaments, other tournament-based sports and a doubling or tripling of the number of players or teams would provide further interesting results, particularly in the case of knockout tournament structures. Also, this paper has used a binary logistic regression to estimate pairwise probabilities of winning a match based on the success of players on service and return. Conceivably there are many other ways of estimating these probabilities, perhaps using points rather than match results or using alternative models.

5. CONCLUSIONS

The results of this study have shown that the December 2009 ranking of the top eight players by KAN-Soft is supported by the data. In particular a model has been developed which allows the estimation of pairwise probabilities of a match for the top men's tennis players in the world. Using these estimates it has been shown that a simple ranking system based on the row sum of the estimated pairwise probabilities of winning is robust to changes in scoring rules, to the loss of one of the players and to changes in tournament structures in the case of round robin and ladder system tournaments. However, there do seem to be some differences between the rankings achieved using a knockout tournament structure and the other ranking systems that have been considered (round robin tournaments, ladder tournaments and the row sum of estimated paired probabilities of winning). This is important because so many sports rely on a knockout tournament structure to ensure that tournaments are viable in terms of time and cost. However, the results have suggested that in the long run random knockout tournaments will give the same rankings as seeded knockout tournaments, where seeded players are assigned to different divisions.

Acknowledgement

The authors would like to thank KAN-soft for providing the database on which the data analyses were based. In addition they would like to thank the referees for their helpful comments.

References

Ali, I., Cook, W.D., & Kress, M. (1986). On the minimum violations ranking of a tournament. *Management Science*, *32(6):* 660-672.

Bedford, A.B., & Clarke, S.R. (2000), A comparison of the ATP rating with a smoothing method for match prediction. Pp 43-51 in Cohen, G, Langtry, T eds. *Proc. 5th Australasian Conference Maths and Computers in Sport.*

Blackman, S.S., & Casey, J.W. (1980), Development of a rating system for all tennis players. *Operations Research. 28(3):* 489-502.

Clarke, S.R. (1994). An adjustive rating system for tennis and squash players. *Proc 2nd Australasian Conference Maths and Computers in Sport*, ed. Neville de Mestre, Bond University, Queensland, 43-50.

Cook, W.D., Golan, I., & Kress, M. (1988). Heuristics for ranking players in a round robin tournament. *Computers and Operations Research, 15(2):* 135-144.

Cook, W.D., & Kress, M. (1990). An mth generation model for weak ranking of players in a tournament. *Journal Operational Research Society, 41(12):* 1111-1119.

Hotelling, H. (1933), Analysis of a complex of statistical variables into principal components. *Journal Educational. Psychology, 24,* 417-441, 498-520.

Klaassen, F.J.G.M., & Magnus, J.R. (2003), Forecasting the winner of a tennis match. *European Journal of Operations Research; 148*: 257-267

Musante, T.M., & Yellin, B.A. (1979), The USTA/Equitable Computerized Tennis Ranking System. *Interfaces. 9(4):* 33-37.

Pearson, K. (1901), On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2,* 559-572.

Pollard, G.H., Pollard, G. N., & Meyer, D.H. (2010), Some aspects of ordering, ranking and seeding A. *Proc 10th Australasian Conference Maths and Computers in Sport.*

Pollard, G.N., & Meyer, D.H. (2010) Wiley *Encyclopedia of Operations Research and Management Science*, Ed. James J. Cochran.

Stefani, R.T. (1997), Survey of the major world sports rating systems. *Journal of Applied Statistics, 24(6):* 635-646.

Strauss, D., & Arnold, B.C. (1987), The rating of players in racquetball tournaments. *Applied Statistics, 36(2):* 163-173.

A LEVEL PLAYING FIELD FOR AWD COMPETITION: CHOOSING THE PARALYMPIAN OF THE DECADE

Allison O'Riordan [a] , Chris Barnes [a,b]

[a] *Australian Institute of Sport*
[b] *Corresponding author: Chris.Barnes@ausport.gov.au*

Abstract

In elite AWD field athletic competition (World Championships, Paralympics) >20 disability classes may compete in each field event (e.g. Discus, Javelin, Shot). Logistic considerations require that classes be combined, which in turn dictates the requirement for intra-class comparison. There is a growing dissatisfaction with previous comparison methods based on a combination of recent best performances and the current world record. We show that the use of such a simple system inevitably leads to bias and a lack of fairness, due to the small number of results used; and suggest a novel, fair, system in its place.

By utilising a large number of individual performances the distribution of performances in any class can be shown to follow (asymptotically) a 2-parameter extreme-value distribution for a given event. This large number of individual performances is obtained in practice by combining results from a number of different elite competitions by modelling specifically: technical improvements over time; and differences in competition standards. This threshold-difference reduces as the number of throws in a class increases. Athletic performances are compared as percentiles of their respective distributions, modulo the uncertainty in the percentile estimate.

A class by class analysis of the throws events (discus, javelin and shot) at all Paralympic and World Championship events (a total of about 30,000 throws) showed that, for the majority of event/class combinations, a linear quartile-quartile plot was consistent with a 2-paramater extreme-value distribution. The coefficients of the linear relationship of these plots formed the basis of inter-class comparisons, and uncertainties in these estimates allowed the determination of significant difference thresholds.

The system can be used objectively to decide outcomes in mixed-class competition, and also for setting qualifying standards systematically. In a few cases, anomalous results indicated copying/transcription errors in official results or misclassification of athletes (in a number of cases these athletes subsequently underwent re-classification). A colleague (Prof. Will Hopkins, AUT, NZ) suggested a refinement of the methodology so that all parameters (class differences, technical improvements and competition differences) were determined in a single step, using a bootstrapping technique, and this will be reported on elsewhere. The uncertainty in estimating EV-distribution parameters leads to a threshold-difference, within which performances must be declared equal (a tie). This is an inescapable feature of the task of class comparison, and although usually overlooked, must be recognised in a fair system.

The methodology is quite general, and can also be used to compare relative performances across gender, events (the Meet Champion), and even time, for both AWD and able-bodied athletes ("the athlete of the decade"). Ultimately it could be used to place 'enhanced classes' (prostheses, drugs, genetics...) on an even footing with all other athletes!

DETECTION OF BODY MOVEMENTS AND MEASUREMENT OF PHYSIOLOGICAL STRESS WITH A MOBILE CHEST MODULE IN OBESITY PREVENTION

Peter Christ[a,e], Jörg Mielebacher[b], Martin Haag[c], Ulrich Rückert[d]

[a] *System and Circuit Technology, Heinz Nixdorf Institute, University of Paderborn, Germany*
[b] *Mielebacher Informatiksysteme, Bremen, Germany*
[c] *University of Heilbronn / University of Heidelberg, Germany*
[d] *Cognitronics and Sensor Systems, CITEC, University of Bielefeld, Germany*
[e] *Corresponding author: pchrist@hni.upb.de*

Abstract

The number of obese children is increasing steadily. This is mainly caused by a lack of physical exercise. Furthermore, those problems are highly correlated with the increasing prevalence of cardiovascular diseases. Health insurance companies are demanding special programs to motivate obese children to exercise frequently.

This work proposes a computer game to motivate children to physical exercise. Unlike common acceleration-based games it additionally evaluates the outcome of the cardiovascular system. Our computer game uses a portable chest module, which is mounted on a chest strap. It includes sensors for single channel ECG, 3D-acceleration, and temperature. All data is transmitted continuously to a desktop computer by using a 2.4GHz wireless connection.

Our chest module enables the user to evade randomly created obstacles in the game's scenario by bending sideways or by jumping and to project the running speed on the agent. These motion patterns are detected by smoothing and adaptively thresholding the acceleration signal. The heart rate is used to judge the game's outcome and is determined by filtering the signal and applying a heuristic scheme.

The performance of the motion pattern recognition was tested with different probands. Bending sideways is recognized with a sensitivity of 97% and a specificity of 99%. Jumping and running are detected with a sensitivity of 95% to 100%. The body movements caused a moderate to high cardiovascular strain with heart rates between 100 to 160 beats per minute. The probands described the interaction with the game as entertaining and challenging. Combining our chest module with a computer game is an attractive approach to prevent obesity among children.

Key words: Motion tracking, physiological stress, obesity prevention, cardiovascular diseases

1. INTRODUCTION

A lack of physical exercise is an increasing problem in the developed countries. The reasons for this problem can be found in mainly sitting activities in school, at work, and in personal life. Consequences are obesity, postural defects or premature cardiovascular diseases even at young age (Kurth and Rosario (2007)). In Germany 15% of the young people between the ages 3-17 have overweight (BMI > P90)[1]. Furthermore, 6.3% suffer from obesity (BMI > P97). The highest increase towards overweight can be found during elementary school age (Kurth and Rosario (2007)). In this context regular physical exercise is of importance (Opper *et al.* (2007)).

The AOK, the largest German health insurance

[1] The statistical data were taken from a study (KiGGS) of the Robert-Koch-Institute from 2003 to 2006. BMI is short for Body Mass Index. P90 and P97 stand for overweight and obesity within the percentile curve of the weight distribution.

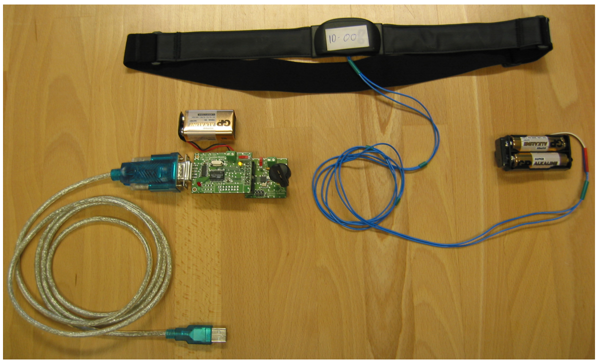


Figure 1: The module mounted on a regular chest belt with integrated electrodes. The external battery pack is optional, normally a coin cell is used. The lower part of the figure shows the receiver for transferring the data to a desktop computer.

company, cooperates with several institutions and universities to launch projects in kindergarten and school against obesity. The aim is to introduce exercise programs and create an understanding for the importance of sport.

Since 2006, starting with Nintendo Wii, several acceleration-based computer games have been brought to the market. They capture movements using acceleration sensors. High sales figures indicate the attractiveness and acceptance of this approach. On the Game Developer Conference in March 2010 Electronic Arts announced a version 2.0 of their "Sports Active" series which extends a motion based controller with the opportunity to measure the heart rate. In conjunction with a training manager the program will guide the user through a personal training plan (Steinlechner (2010)).

Besides of the use as a leisure activity there are projects utilizing miniaturized motion sensors on clinical background. "Partnership for the Heart" (Köhle, Lücke (2007)) is a project of a consortium led by the university hospital Charité in Berlin. It links different sensors wirelessly to treat patients with congestive heart failure in their home environment. If a critical situation is detected it provides the possibility to transfer the required medical data to a tele-health facility for seeking aid. The remote patient monitoring system includes an activity sensor. This module uses acceleration sensors to measure the patients motions to guide the patient through a 6-minute-walk test at home (Jehn *et al.* (2009)).

Witkowski *et. al* (2008) have developed a portable chest module for measuring body movement and physiological parameters. Figure1 shows the module mounted on a chest strap. The module includes sensors for single channel ECG (Electrocardiography), 3D-acceleration, and temperature. Data is transmitted continuously to a desktop computer using a 2.4GHz wireless connection.

This work examines how to use this module for obesity prevention. First the quality of the registered ECG and acceleration signals is evaluated to detect the heart rate and motion patterns. Furthermore, a game scenario is developed in which the module enables the user to evade randomly created obstacles using the detected motions. The scenario demands various movements such as running, jumping and bending sideways to control the agent. In this way the user is challenged to activate different muscle groups. The heart rate is used to measure the cardiovascular stress and functions as a feedback loop to adapt game parameters. The chest module and the game scenario were evaluated with probands for usability.

The system is designed to be appropriate and attractive for young people to motivate them to train regular with a training effective level.

This paper is organized as follows. Section 2 outlines the signal filtering of the ECG captured with the chest module and the determination of the heart rate. Section 3 outlines the signal filtering of the measured accelerations and the recognition of the motion patterns running, jumping and bending of the upper torso. Section 4 states the results of the performance tests on the pattern recognition and the game scenario. Section 5 discusses the results and possible improvements. Section 6 outlines options for future use of the chest module.

2. Heart Rate Detection

Registering signals on the surface of a moving body usually causes superposed artifacts of body movements. In addition one observes electromagnetic interferences (Thakor and Zhu (1991)). The pattern recognition algorithms need to react robust towards these disturbances.

Compared to clinical ECGs which are registered with self-adhesive electrodes we expect significantly more motion-related artifacts (Brüne (2008)). This is mainly caused by the slipping of the electrodes during movements. Varying contact resistance and additional myographic contributions result in noise and a fluctuating signal baseline (figure 4). Furthermore, the signal quality depends on

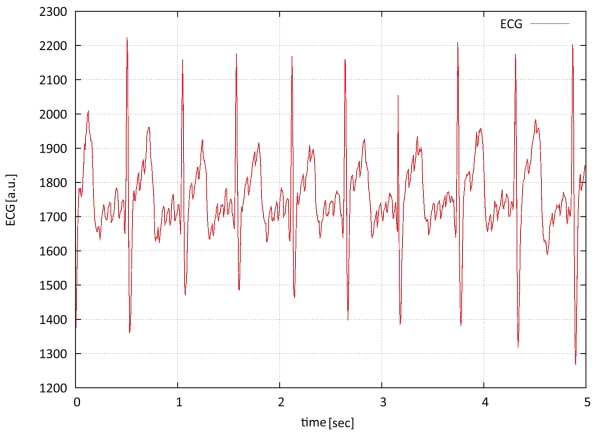


Figure 2: An undisturbed ECG registered with the chest module at rest with good conductance between electrodes and skin.

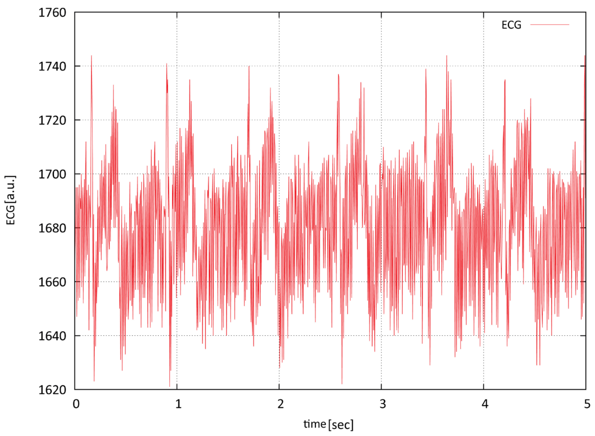


Figure 3: An ECG disturbed with electromagnetic influences (high peak at 50Hz in the power spectrum). The ECG was registered at rest with poor conductance between electrodes and skin.

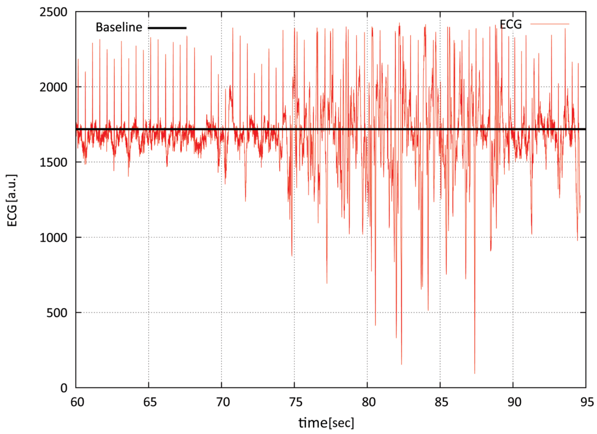


Figure 4: An ECG disturbed with motion and electrode artifacts. The ECG was registered during physical exercise with good conductance between electrodes and skin.

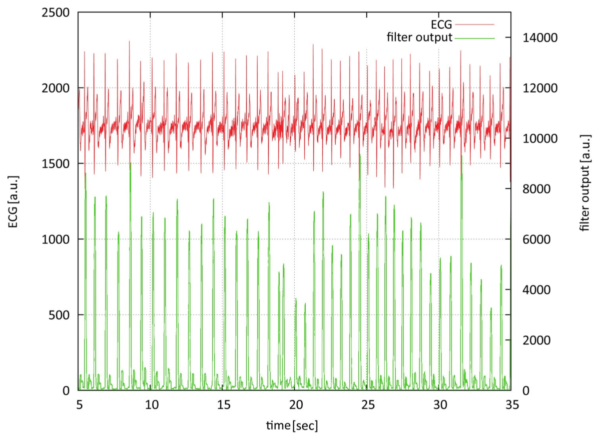


Figure 5: An undisturbed ECG (red) and the filter output (green). The ECG was registered with little movement and good conductance between electrodes and skin.

the conductance of the electrodes and the closeness to unshielded electronic devices (figure 2 and figure 3). Several algorithms for heart rate detection in clinical ECGs have been published. A popular approach for QRS[2] detection has been introduced by Pan and Tompkin (1985). We have adapted their method to the observed disturbances in the ECG-signal registered with the chest module during physical exercise.

The signals are filtered to suppress irrelevant signal components and to accentuate the pitches of the R-waves[3]. First a bandpass, implemented as a combination of an averaging and a median filter, is applied on the ECG input signal to reduce the influence of other muscles, electromagnetic disturbances and baseline variations. Second a gradient filter accentuates the R-wave and removes components with smaller gradients per time, e.g. baseline variations. The output is squared for pointing out the QRS-complex and obtaining the absolute value. Finally a long range averaging filter smoothes the signal leaving QRS-complexes as peaks in the filter output with other signal components near the baseline. Figure 5 and figure 6 show the ECG input and the filtered signal. To determine the heart rate, the filtered signal is analyzed for matching thresholding criteria. In this regard the signal filtering highly influences the performance of the detection algorithm. Strong artifacts may lead to similar characteristics in the filter's output signal and a heartbeat is false wise detected when the thresholds are exceeded. Because it is not to be expected that all disturbances can be suppressed (Brüne (2008)),

[2]The QRS-complex describes a series of waves that correspond to the contraction phase of the cardiac chambers.

[3]The R-wave has the highest pitch within a QRS-complex.

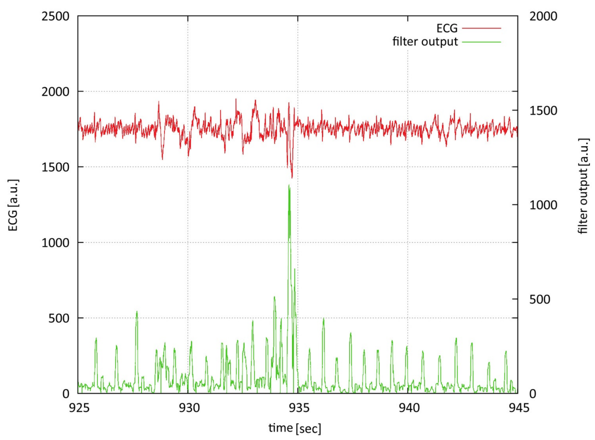


Figure 6: An ECG disturbed with strong motion and electrode artifacts (red) and the filter output (green). The ECG was registered during physical exercise and medium conductance between electrodes and skin.

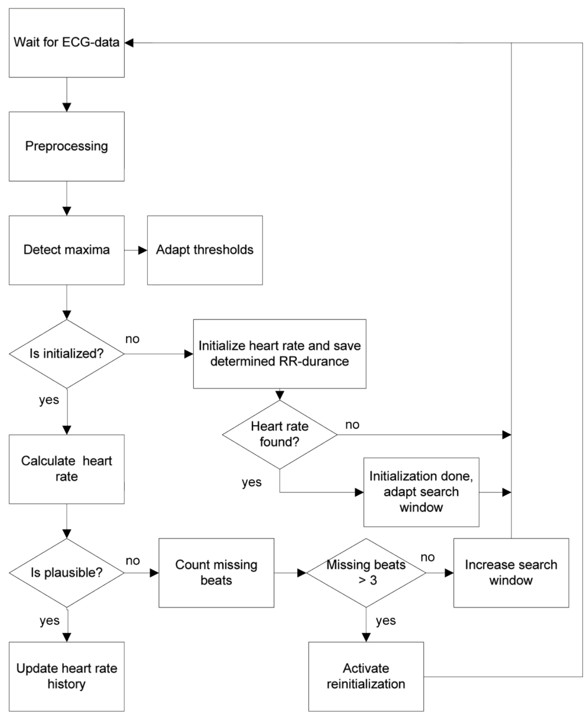


Figure 7: Flow diagram of the heart rate detection algorithm.

the detection must check the plausibility of the determined heart rate values to avoid unrealistically or rapidly changing heart rates. If distinguishing QRS-complexes from artifacts or noise is not possible, the detection must pause.

These requirements are implemented in form of a heuristic algorithm. First the detection is initialized by searching for a regular rhythm consisting of 9 QRS-complexes. In the filter output QRS-complexes are peaks with highest amplitudes. The amplitudes are depending on the physiology of the user and the conductance of the electrodes. The routine searching for these QRS-complexes updates the detection threshold to the signal quality of the last found peak. Typically at the beginning of a workout the conductance of the electrodes is poor, but during physical exercise sweat improves the quality of the captured ECG leading to sharp peaks with higher amplitudes. As a result the threshold is increased, fading out noises of smaller amplitudes.

To protect the heart rate detection against disturbances, the detection-routine is only called during a time window when a heartbeat is expected. First, a initialization phase determines the parameters of the time window by searching for peaks matching a series of heartbeats. During this phase the allowed variations from peak to peak narrow down to approach the heart rate rhythm. The search window is configured using the arithmetic average of the last three detected peaks. After the initialization phase the heart rate is updated by averaging the last 5 determined values. If there is no heartbeat found in the given time period, the search interval is increased. After continuously missing three heartbeats the initialization routine is called again, as the heart rate of the user can't be expected to be the same.

3. Motion Pattern Recognition

3.1. Detection of Running and Jumping

Typically sensors for detecting the user's running speed or jumps are placed near the foot. Companies like Suunto Oy, Garmin Ltd. and Polar Electro Oy have foot pods as accessories for their sport watches. The foot pods are placed on top of the shoes to capture the strides.

Measuring accelerations in the chest area leads to different signal shapes as strides and jumps are absorbed. In this work the placement of the acceleration sensor is limited to the chest area as we want the chest module to be easily attachable on a chest strap. Furthermore, it is important to take into account, that strides and jumps are performed differently by each user and also vary over time because the body is getting exhausted or the attention fades. This causes variations in amplitude, width and form of the signal shapes.

The filtering of the acceleration signals smoothes and accentuates characteristics of strides and jumps and suppresses signal parts of irrelevant motions.

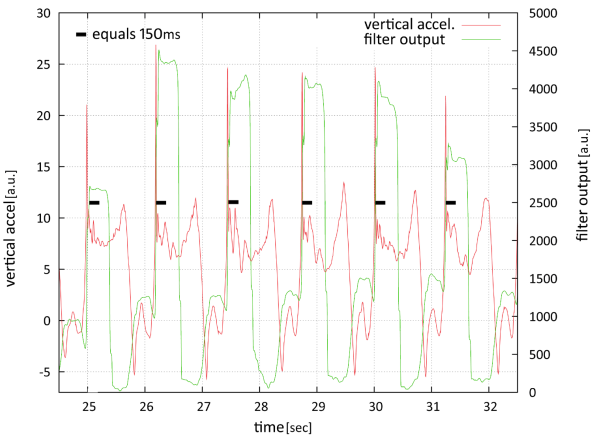


Figure 8: Measured vertical accelerations (red) and the filter output (green) of 6 jumps performed by proband A as uniform as possible. The black bar is indicating the threshold and the time period it must be exceeded for detecting a jump.

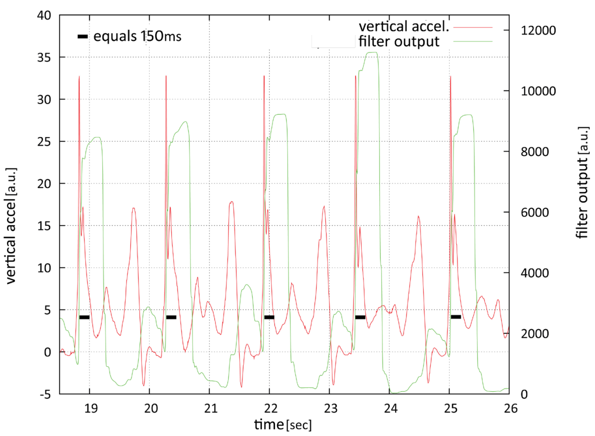


Figure 9: Measured vertical accelerations (red) and the filter output (green) of 5 jumps performed by proband B as uniform as possible. The black bar is indicating the threshold and the time period it must be exceeded for detecting a jump.

Characteristic for strides and jumps are high pitch rates and amplitudes in the registered signals. Three criteria are used to recognize the patterns running and jumping. Therefore, the filtered signal must continuously exceed a threshold for a given time period (black bars in figure 8, figure 9 and figure 10). If a stride or jump is recognized the detection is paused afterwards. We determined that in the game scenario 2 strides and 1 jump per second are distinguishable in the acceleration signals. The pause provides security for the motion pattern detection to avoid strong disturbances to be continuously recognized as a stride or jump. The settings for the criteria were derived from tests with different probands.

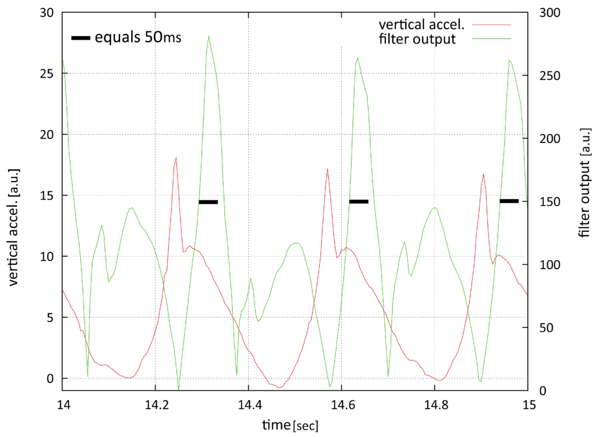


Figure 10: Measured vertical accelerations (red) and the filter output (green) of 4 strides. The black bar is indicating the threshold and the time period it must be continuously exceeded.

3.2. *Detection of the Bending Direction of the Upper Torso*

For the detection, the baseline of the acceleration signals is analyzed as it varies with the bending direction of the upper torso. The axis of the acceleration sensor which points straight down to earth has a baseline of 1 g. With changes in the orientation the gravity effects different axes causing changes on their baselines. During physical exercise accelerations from body movements overlap the underlying baseline. A median filter over a time range of 0.5 sec is used to approximate the baseline. The detection distinguishes bending forwards, backwards, sideways left and right.
In tests with probands we determined the thresholds of the attitude angle and the duration it must be exceeded. Choosing a longer time period increases the detection latency but reduces the algorithms response towards unwanted movements (specificity).

3.3. *Description of the Game Scenario*

The game scenario is meant to challenge the user on a medically sensible cardiovascular stress level corresponding to a form of endurance sport. Furthermore, the user is challenged to use different movements to train various groups of muscles. On the other side the scenario needs to be attractive to make the user enjoy the workout.
The playing field is generated at random and contains obstacles in form of ditches, holes and narrowing walls. The detected motions are used to control the game. Therefore, the user's running

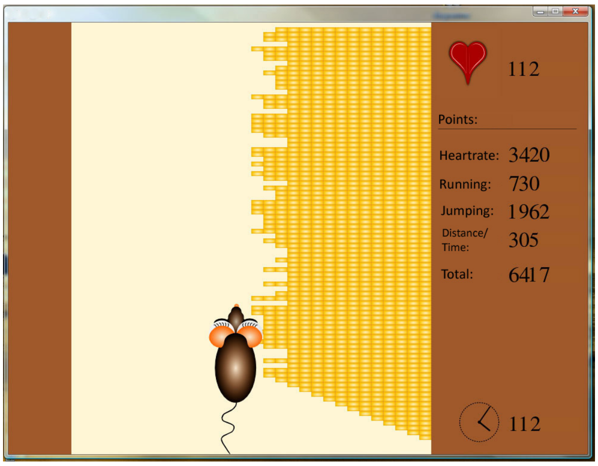


Figure 11: Playing field designed to be appropriate for young people.

class 3	QRS	no QRS	sum
QRS detected	129	0	**129**
no detection	63	192	**255**
sum	**192**	**192**	

specificity	**1,00**
sensitivity	**0,67**

class 2	QRS	no QRS	sum
QRS detected	407	0	**407**
no detection	50	457	**507**
sum	**457**	**457**	

specificity	**1,00**
sensitivity	**0,89**

class 1	QRS	no QRS	sum
QRS detected	1074	0	**1074**
no detection	6	1080	**1086**
sum	**1080**	**1080**	

specificity	**1,00**
sensitivity	**0,99**

class 3	**strong:** baseline variations, moving and electrode artifacts **more than 3 heartbeats** continuously missing
class 2	**medium:** baseline variations, moving and electrode artifacts **up to 3 heartbeats** continuously missing
class 1	few or no artifacts **no heartbeats missing**

Figure 12: Contingency table with the results on the performance test of the QRS-detection algorithm. The test classifies the recorded ECG-data into three classes. For evaluation data files containing 1729 QRS-complexes were manually analyzed.

speed is projected on the agent. Jumping and bending sideways are used to avoid obstacles. The goal is to cover as much distances as possible per time. The effort is rewarded with points covering distance per time and the complexity of the playing field. At the moment the heart rate is valued with points, too. In future it will be used to adapt the playing field and the game rules to meet a training effective heart rate zone.

4. RESULTS

The performance of the pattern detection was tested with three different probands. For the evaluation of the heart rate detection algorithm ECG records are grouped in three classes with strong (class 3), medium (class 2) and only few (class 1) disturbances (figure 12). For the evaluation records containing 1729 QRS-complexes (class 3: 192, class 2: 457, class 1: 1080) were manually analyzed. The outcome is noted in form of a contingency table and the sensitivity and specificity rate is calculated. The sensitivity rate describes the detected QRS-complexes which are correctly identified as such. The specificity rate states the negatives which are identified as such. The determined sensitivity rate in class 3 is 67% , in class 2 87% and in class 1 99%.
Missing consecutively one to three beats is compensated by the algorithm and does not lead to detection failures. Disturbances of class 3 in the ECG-signal are seldom, but can lead to deviations to the real heart rate. In conclusion the tests showed that the heart rate detection is suitable for the precision needed in this scenario. Running strides are detected with a sensitivity of 97% and a specificity of 100% (180 strides were analyzed). The algorithm's parameters are set to detect strides corresponding to running with moderate speed. It will pickup other movements as well if they match the pattern, but this has no disadvantage for our application as it shows that the user is physically active. Jumps are recognized with a sensitivity of 95% and a specificity of 100% (108 jumps were analyzed). For the evaluation the probands tried to continuously jump with the demanded intensity. Changes in the jump technique or trying out movements with similar characteristics will lead to different sensitivity or specificity rates.
The bending direction of the upper torso is detected with a sensitivity of 97% and a specificity of 99% (114 bendings were analyzed). As described, a high reliability of the recognition is achieved by taking into account signal values over a long time range of 0.5 sec. To control a game this latency is high and demands a game scenario with foresightful acting. Overall the system requests the coordination of different movements and caused a cardiovascular stress level with heart rates between 100 to 160 beats per minute.

running detection	step	no step	sum		
step detected	180	6	**186**	specificity	**0.97**
no detection	0	180	**180**	sensitivity	**1.00**
sum	**180**	**186**	**732**		

jump detection	jump	no jump	sum		
jump detected	103	0	**103**	specificity	**1.00**
no detection	5	181	**186**	sensitivity	**0.95**
sum	**108**	**181**	**578**		

leaning detection	leaning	no leaning	sum		
leaning detected	111	1	**112**	specificity	**0.99**
no detection	3	120	**123**	sensitivity	**0.97**
sum	**114**	**121**	**470**		

Figure 13: Contingency table with the results of the performance test on the motion pattern detection. For evaluation data files containing 180 strides, 108 jumps and 114 bendings of the upper torso were manually analyzed.

5. DISCUSSION

The tests showed that a correct position of the chest strap is important. It should be placed tight under the pectoral muscle to avoid slipping which causes disturbances in the ECG and acceleration signals. The user should also wet the electrodes with water in advance. This increases the conductance and reduces the noise level of the ECG signal.
As already mentioned measuring accelerations in the chest area has disadvantages for the recognition of strides and jumps. However the results show that the precision of the detection suits the requirements of the given game scenario. In future it should be researched if the accuracy can be increased by adapting the algorithm's parameters to the user and environmental conditions. Furthermore, it should be checked if this suppresses unwanted movements to greater extend and whether the latency of the bending detection can be reduced. This could be done using an initialization step in advance or by continuously learning optimal thresholds from the user's actions.
Ongoing tests should be extended with more probands to acquire statistically more sufficient data. Furthermore, tests with probands coming from non technical background are necessary to analyze the usability of the game controller.

6. CONCLUSIONS

This work analyzed the quality of the signals registered with the custom-made chest module. We examined its use as a game controller which additionally allows measuring the cardiovascular outcome. In the tests the probands felt challenged and it could be shown that the developed algorithms are able to recognize the motion patterns running, jumping and bending as well as the heart rate with a precision suitable for the game scenario. In conclusion the system could be used as part of an attractive approach to prevent obesity among children.
Different extensions or applications for the use of the system will be investigated. The game speed or the complexity level of the playing field could be adjusted to the measured cardiovascular stress level to match a personal training plan. A game scenario in form of a "Drill Instructor" could define a series of movements and monitor their correct execution. In this context not the latency of the detection but its precision is of importance. Furthermore the chest module can be used for applications in the field of ambient assisted living or tele-health, for example to transmit ECG data in case of an emergency.

References

Kurth B.-M. & Rosario A. S. (2007). The prevalence of overweight and obese children and adolescents living in Germany. Results of the German Health Interview and Examination Survey for Children and Adolescents (KiGGS). *Journal of Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 50*, 736-743.

Opper E., Worth A., Wagner M. & Bös K. (2007). The module Motorik in the German Health Interview and Examination Survey for Children and Adolescents (KiGGS). Motor Fitness and physical activity of children and young people. *Journal of Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 50*, 879-888.

Köhle, F.& Lücke, S. (2007). Partnership for the Heart. *Kardiotechnik. 4, 2007, 4*, 110.

Steinlechner P. (2010). EA Sports misst Herzfrequenz mit Active 2.0. *Golem.de, 09 April 2010*, URL: http://www.golem.de/1003/73742.html.

Jehn M., Schmidt-Trucksäss A., Schuster T., Weis M., Hanssen H., Halle M. & Koehler F (2009). Accelerometer-Based Quantification of 6-Minute Walk Test Performance in Patients With Chronic Heart Failure: Applicability in Telemedicine. *J Cardiac Failure 2009; 15 (4)*, 334-340.

Witkowski U., Wilhelm, P. & Parketny, T. (2008). Einsatz von Low-Power Netzwerken zum Monitoring leistungsdiagnostischer Daten im Teamsport. *Wireless Technologies Kongress, 23.-24. Sep 2008, Bochum, Germany*, 261-270.

Thakor N. V. & Zhu Y. S. (1991). Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. *IEEE Trans. Biomed. Eng., BME-38 (8)*, 785-794.

Pan J. & Tompkins W. J. (1985). A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng., BME-32 (3)*, 230-236.

Brüne S. (2008). Entwicklung eines robusten Algorithmus zur QRS-Komplexerkennung. *Diploma Thesis, System and Circuit Technology Group University of Paderborn*.

IDENTIFYING TAPER SHAPE AND ITS EFFECT ON PERFORMANCE

Churchill, T[a,b], Sharma, D[a] and Balachandran, B.M[a]

[a] *Faculty of Information Sciences and Engineering, The University of Canberra, ACT, Australia*

[b] *tania.churchill@canberra.edu.au*

Abstract

The aim of this study was to investigate the effect of the shape of a taper on elite cyclists' performance. Training and performance data were collected from two elite female cyclists over an extended period. Daily training load was quantified using a modified TRIMP (PTRIMP) and performance was quantified using a technique which calculates performance from power data recorded during races. The training load (PTRIMP) for the 3 days prior to each performance was analysed using a data mining approach whereby the time series is transformed into a symbolic representation. Symbolic Aggregate approXimation (SAX) was the method used for the transformation. SAX divides the distribution space of the time series into a user-defined number of equiprobable regions. Each region was assigned a symbol, and then each point in the time series was mapped to the symbol corresponding to the region in which it resides. The resulting symbolic representation thus described the shape of the training load prior to a performance. The shapes identified were grouped into 4 categories (low, high tail, low tail and high). The main effect of taper shape on performance was significant for both subjects ($F(3, 22) = 4.2$, $p < 0.05$, $MSE = 0.5$) and ($F(3,18) = 3.9$, $p < 0.05$, $MSE = 0.3$). There was considerable inter-athlete difference in optimal taper shape. The results indicate that this novel approach to identifying and examining the impact of taper shape can provide useful information to athletes and coaches in planning the most effective taper. Future work could extend this technique to identify the patterns of training load in training microcycles, and relate these patterns to fatigue and performance measures.

Keywords: taper, performance, shape, data mining

1. INTRODUCTION

The taper can be defined as a period of reduced training prior to a competition, undertaken with the aim of achieving peak performance at the desired time (Thomas, Mujika & Busso, 2009). It is of paramount importance in the preparation of athletes for competitions (Pyne, Mujika & Reilly, 2009).

The effectiveness of a taper as reported in the literature varies, however the improvement in performance is usually in the range of 0.5-6%. A realistic goal for performance improvement as the result of a taper is about 3% (Mujika & Padilla, 2003). In competitive athletes such modest improvements are important. A worthwhile improvement for top-ranked athletes is estimated to be in the range of 0.5-3.0% for events such as endurance cycling (Hopkins, Hawley & Burke, 1999).

The aim of the taper is to reduce accumulated training-induced fatigue, while retaining or further enhancing physical fitness (Bosquet, Montpetit, Arvisais & Mujika, 2007). The key elements to manipulate in determining an optimal taper include; the magnitude of reduction in training volume; training intensity; duration of the taper, and; the pattern of the taper (Pyne, Mujika & Reilly, 2009).

Uncertainty exists about the optimal design of a taper (Mujika & Padilla, 2003). In a meta-analysis study Bosquet, Montpetit, Arvisais & Mujika (2007) suggested that training volume should be reduced by 41-60% over a two-week taper, without any modification to training intensity or frequency. They found that reducing training volume elicited a performance improvement approximately twice that gained by modifying either training intensity or frequency.

A number of taper patterns have been described and investigated in the literature. Training load can be reduced in the form of a simple step – where the load is suddenly reduced and then maintained at the same low level; or it can be reduced progressively, either with a constant linear slope, or with an exponential decay (Thomas, Mujika & Busso, 2009; Mujika & Padilla, 2003). There is evidence to suggest that a progressive taper is to be preferred (Bosquet, Montpetit, Arvisais & Mujika, 2007; Banister, Carter & Zarkadas, 1999).

Little research has been done on more complicated taper patterns. One such study looked at the effect of a two-phase taper. This model study found that the last 3 days of the taper were optimised with a 20 to 30% increase in training load. In the modelled response such a two-phase approach allowed for additional fitness adaptations to be made in the final 3 days, without compromising the removal of fatigue. The magnitude of the performance gain is questionable (0.01), however, over the optimal linear taper (Thomas, Mujika & Busso, 2009).
Much of the literature provides generalised guidelines on designing an optimal taper. It must be noted, however, that individual responses to training vary. Not all athletes respond equally to the training undertaken during a taper, and tapering strategies must be individualised (Mujika, 2009). Individual profiles of training adaptation and the time course of de-training need to be considered in determining optimal taper duration (Mujika & Padilla, 2003).

The positive performance results of the two-phase taper in modelling work done by Thomas, Mujika and Busso (2009) suggests that further investigation into optimal taper shapes is warranted.

The purpose of this study was to investigate the effect of the shape of a taper on elite cyclists' performance, and to determine if individual differences in optimal taper shape exist between athletes. We aim to investigate these aspects from a novel angle – through symbolisation of taper time series information and subsequent relation of the taper shape to performance.

2. METHOD

Two female elite cyclists provided data for the study over a period of 250 and 870 days respectively. An SRM power monitor (professional model, Schoeberer Rad Messtechnik, Germany) was fitted to each subject's bike(s) over the data collection period. Power data from all rides (training and racing) was captured at 1Hz. In accordance with operator instructions, the SRM was zeroed prior to the start of each session. SRM data files were imported into a custom-built program for the calculation of training load (PTRIMP) and performance.

Quantifying Training Load

The following steps were taken to calculate PTRIMP. The power data was first smoothed by taking 3 rolling averages, of durations of 5s, 30s, and 4mins. Each of the smoothed points was then given a weight, which is calculated by determining the percentage the point represents of the athlete's Maximal Mean Power (MMP) for that duration. The percentage is then multiplied by an exponential formula (refer to equation (1)). The weight of each point, for each of the 3 smoothed datasets is then added together to determine PTRIMP (refer to equation (2)).

$$\mathrm{PTRIMP}(s) = \sum_{i=1}^{n-1} p^i \times c. \qquad (1)$$

where c = exponential curve based on MMP.

$$\mathrm{PTRIMP} = \frac{\mathrm{PTRIMP}(5s) + \mathrm{PTRIMP}(30s) + \mathrm{PTRIMP}(240s)}{1000}. \qquad (2)$$

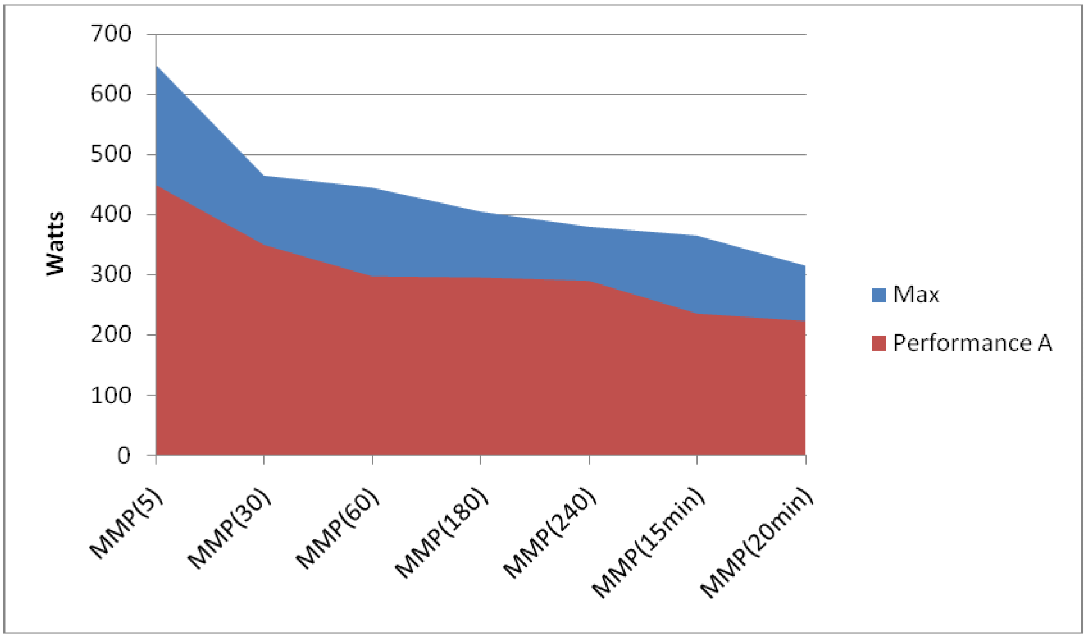


Figure 1. Comparison of the area under the MMP curve between an athlete's personal best, and a Performance A.

A record of MMPs for 5s, 30s, and 4mins achieved by the athlete was kept, and updated when a new personal best was achieved.

Quantifying Performance

Each athlete kept a training diary, which was used to identify races. Performances were included in the dataset when training data existed for the three days prior to the performance. Twenty-six performances were identified for Subject A, and 22 for Subject B.

A curve was created from the MMPs for time durations from 5s to 20min for the power file from which performance was calculated. The area under the curve was then compared to that of the athlete's maximum MMP profile at the time of the performance (refer to Figure 1). The definite integral was calculated using the trapezium rule.

The durations for which MMP were recorded were at regular intervals from 5s to 20min. These durations are chosen to represent a spectrum of energy system contributions. For durations of up to approximately 10 seconds, energy is predominantly supplied by the anaerobic alactic system. From around 10 seconds to approximately 60 seconds, energy is predominantly supplied by the anaerobic lactic system. Beyond this, the contribution of the aerobic energy system increasingly becomes the major contributor (Gore, 2000). The range in durations thus theoretically balanced out the effects of different types of races.

Symbolic Aggregate Approximation

Symbolic Aggregate approXimation (SAX) allows a time-series of arbitrary length n to be reduced to a string of arbitrary length w. The alphabet size used is also an arbitrary integer (Lin, Keogh, Lonardi & Chiu, 2003). The process involves firstly normalising the time series to have a mean of zero and a standard deviation of one. A series of breakpoints are identified that divide a Gaussian distribution up into n number of equiprobable regions. These breakpoints are used to map each data point into symbols, such that a data point lower than the smallest breakpoint will be mapped to the symbol 'a', a point greater than or equal to the smallest breakpoint but smaller than the next breakpoint will be mapped to 'b' and so on (refer to Figure 2).

A time series consisting of the three days training load (PTRIMP) in the lead up to a performance was created. The Matlab code of Lin, Keogh, Lonardi and Chiu (2002 & 2003) was used to discretise the time series of PTRIMP data. The code was customised in two areas. A change was made such that normalisation of the time series was performed based on the entire series, rather than just on the current 'window' of data. The code which performs numerosity reduction was also removed. A window size of three days and an alphabet size of three were used. An alphabet size of three was selected as a good compromise enabling the creation of a usable number of patterns whilst maintaining reasonable statistical power.

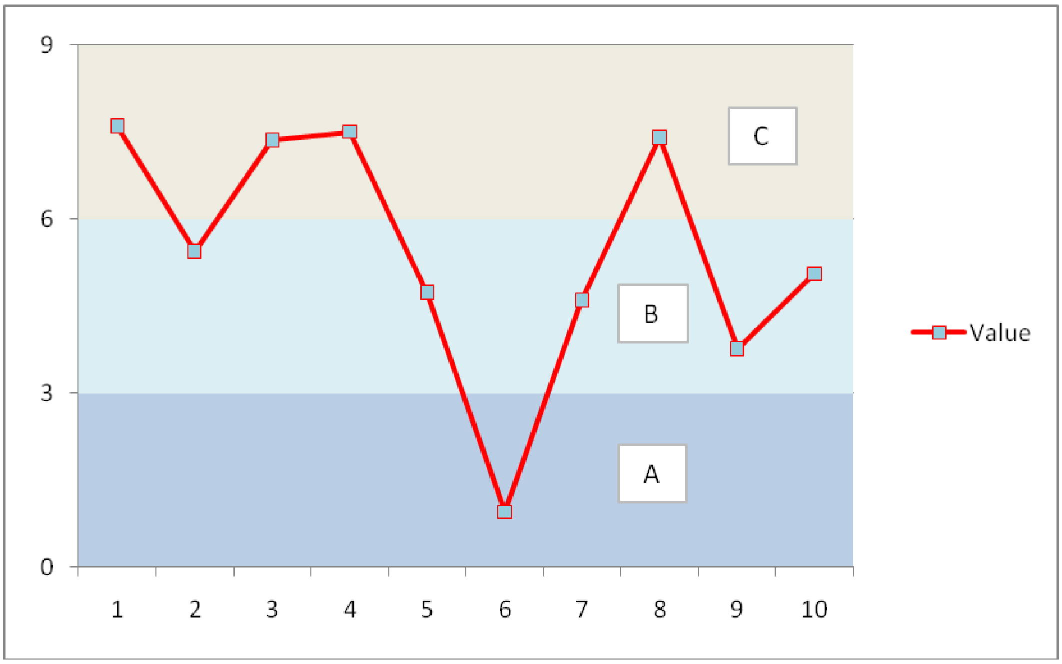


Figure 2. Predetermined breakpoints are used to break the feature space into equiprobable regions. Each data point is mapped to the symbol of the region it falls in.

The result was a sequence of three strings for each 3 day taper period, representing the shape of the taper. A result of '321', for example, represented consecutive days of high, medium and low training respectively.

Taper Shapes

The symbolised taper time series allowed different taper "shapes" to be identified. A number of categorisation schemes were tested. Analysis of the data suggested that the training load of the day before a performance had the greatest effect on the subsequent performance. It also suggested that a training load of 3 (high) on the final taper day was highly represented in tapers resulting in poor performances. Using this knowledge a grouping scheme with four categories was developed (low, high-tail, low-tail and high). A low taper contained any combination of low to medium training loads. A high tail taper contained low to medium training loads on days one and two, and a high load on day three. A low tail taper consisted of a high training load on days one and/or two, and low to medium loads on day three. A high taper included tapers with a high training load on days one and/or two, and a high load on day three.

Data Pre-Processing

SRM files occasionally include short durations of spurious power readings. Such spurious data points were identified and removed.

The Performance dataset for one subject contained an outlier. This (worst) performance was capped at -155000(au).

The Performance data was transformed by taking the natural log of Performance. A constant was added to Performance so that all values were positive prior to transformation. As performance is measured in arbitrary values (au), it was judged no information was lost in this process.

Statistical Analysis

The Shapiro-Wilk normality test was performed to verify the normality of the distribution. A two-way analysis of variance (ANOVA) confirmed that a significant interaction effect between subject and taper shape was present. Subsequently, the difference between the taper shapes was compared using a one-way analysis of variance (ANOVA). The scale proposed by Cohen (1988) was used for interpretation. The magnitude of the difference was considered either small (0.2), moderate (0.5), or large (0.8).The statistical power for the effect size was determined to indicate the probability of correctly rejecting a false null hypothesis. Statistics were calculated using the R Statistical Package Version 2.6.1 (The R Development Core) and the package Rcmdr (version 1.3-15).

3. RESULTS

Significant differences were observed in the mean of performances grouped by taper shape between subjects (for raw data refer to Table 1). The main effect of taper shape on performance was significant for both Subject A ($F(3, 22) = 4.2$, $p < 0.05$, $MSE = 0.5$) and Subject B ($F(3,18) = 3.9$, $p < 0.05$, $MSE = 0.3$). The effect of each taper group on performance showed considerable variation between subjects (refer to Figure 3).

Table 1. Mean and standard deviation (SD) of performance grouped by taper shape for each subject.

	Mean		SD		*n*	
Group	**Subject A**	**Subject B**	**Subject A**	**Subject B**	**Subject A**	**Subject B**
High	11.28	11.36	0.36	0.27	5	4
high tail	11.15	11.48	0.23	0.33	3	4
Low	11.83	11.29	0.38	0.25	8	8
low tail	11.56	11.78	0.33	0.28	10	6

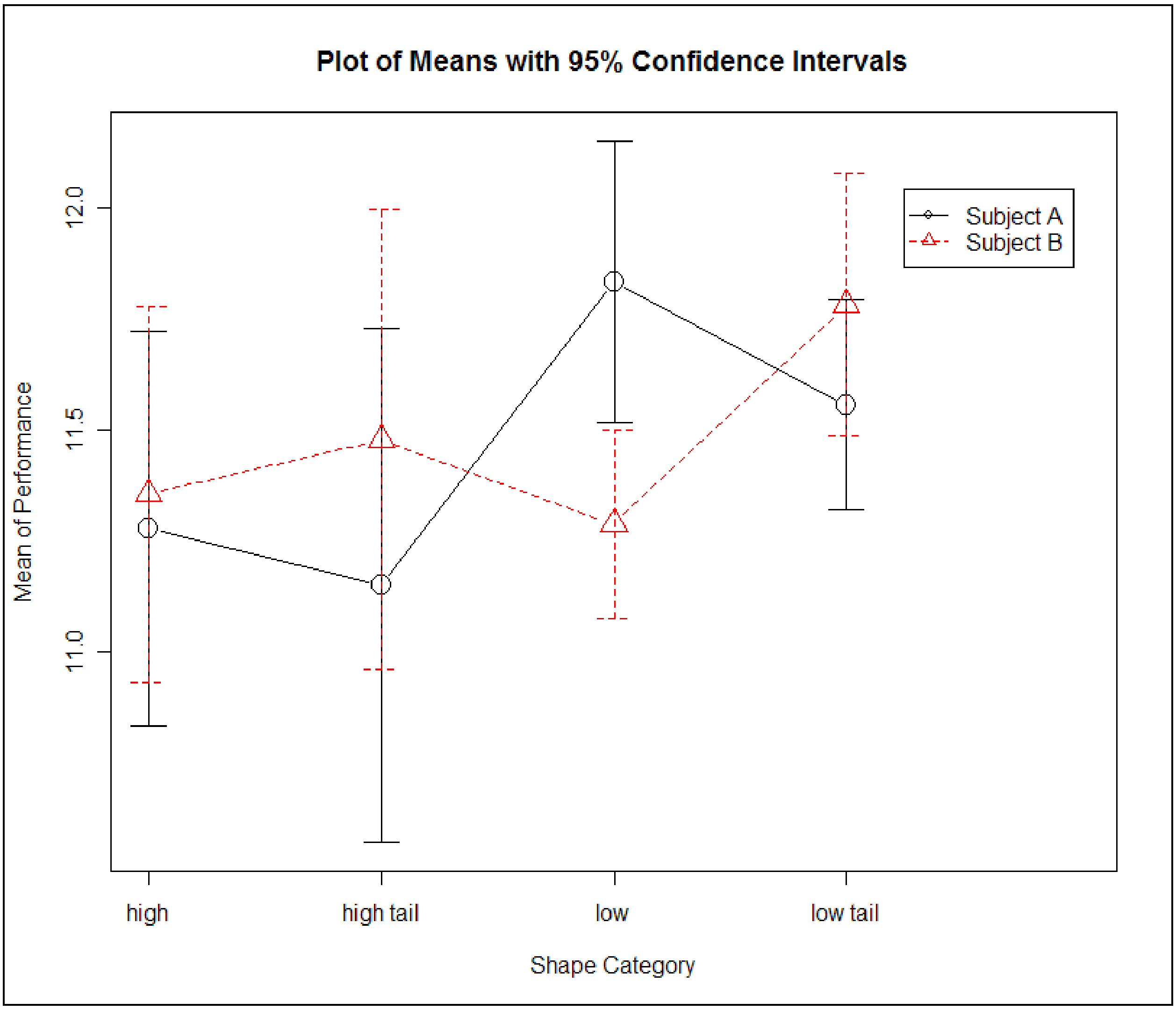


Figure 3. Plot of mean performance for each subject, grouped by category. Error bars show 95% confidence intervals. This plot shows the variation between subjects in their reaction to different taper shapes.

The most effective taper for Subject A was a “low” taper, where combinations of low to medium training days were performed. The least effective taper was a “high tail” taper, where low to medium training was performed in days one and two, and a high training load was undertaken on day three. Using Cohen’s scale (Cohen, 1988) the effect size between the “low” and the “high tail” taper was classified as large. The statistical power was 97.5% (5% error level). When expressed as a percentage, this was a 6.1% improvement in performance.

The most effective taper for Subject B was a “low tail” where a high training load was undertaken on days one and/or two, and a low training load undertaken on day three. The least effective taper for Subject B was the “low” taper. Using Cohen’s scale (Cohen, 1988) the effect size between the “low tail” and the “low” taper was classified as large. The statistical power was 92.4% (5% error level). When expressed as a percentage, this was a 4.4% improvement in performance.

4. DISCUSSION

The results show considerable variance between individuals in their response to a taper. A two-way ANOVA showed that the interaction effects between taper shape and subject was significant. Combining the results of both subjects was considered inappropriate. The subjects for this study were reasonably homogenous – both female elite cyclists. Tapering advice in the literature is frequently generalised across genders, between different sports, and between trained athletes and elite athletes. These results suggest that specific tapering advice can not be provided using a generalist model.

The “low” taper shape was associated with the highest mean performance for Subject A, and the lowest mean performance for Subject B. The optimal taper may be influenced by the intensity and volume of the training preceding the taper. Those who train harder and longer may require a longer taper to enable them to recover, while those with less training require a shorter taper to minimise loss of fitness (Kubukeli, Noakes, & Dennis, 2002). One possible explanation for the variance in taper effect observed is that Subject A tended to carry more fatigue into the final 3 days of training, due to a higher training load in the preceding training cycle. Thus, a low training load in the taper allowed for more effective dissipation of fatigue, resulting in higher performance.

The most effective taper shape for Subject B was the “low tail” shape. If Subject B went into the final 3 days with relatively low levels of fatigue, higher training loads would not have the negative effect on overall fatigue that they would for Subject A. Thomas, Mujika and Busso (2009) suggest that a moderate increase in training load in the final 3 days of a taper can allow adaptations to training to occur without compromising fatigue minimisation.

The experimental design did not consider the training load prior to the final 3 days of taper before a performance. Previous work by the researchers (Churchill, Sharma & Balachandran, 2009) suggested that training load over the 3 days before a performance had the highest correlation with performance. The current research could be extended by modelling the training load in the 4-11 days prior to the 3-day period studied. Such an extension could potentially determine whether the level of training-induced fatigue brought into the taper period affects the optimal shape of the taper, as hypothesised.

The training load quantification method (PTRIMP) aggregates the volume and intensity of training. This means it is not possible to distinguish the specific influence of training intensity during the taper. The model we have developed can determine the optimal shape of training load within a taper, but can not provide guidance on the optimal balance between training duration and intensity during the taper. There is general agreement in the literature that intensity should be maintained in the taper (e.g Mujika & Padilla, 2003; Pyne, Mujika & Reilly, 2009; Bosquet, Montpetit, Arvisais & Mujika, 2007), and this is the general practice followed by the subjects.

In this study, crisp boundaries were used in determining the breakpoints in the symbolisation process. Crisp boundaries result in some loss of data, as two similar training load values positioned either side of a boundary will be treated as different categories although the actual difference in values is small. Removing crisp boundaries in favour of fuzzy boundaries would remove this potential issue.

This research studied the effect of different taper shapes on performance in actual competitions. Few studies have used data from real world training and performance data. The relationship between performance in tests and performance in events was questioned by Hopkins, Hawley & Burke (1999), and how a change in performance in a test translates to performance change in a competition remains uncertain.

Performance tests in competition are affected by external factors, such as climatic conditions, tactics, drafting, terrain and varying competition types (time trials, hilly road races, flat road races and criteriums are all race types included in this study's dataset). Such factors are either not applicable, or can be controlled for, in a lab situation. The fact that the effect size of the different taper shape treatments is significant for both subjects indicates that the model developed is robust, however.

The use of field-derived model inputs makes this model a practical tool for the planning of a taper to maximise competition performance. The relatively minimal data collection requirements make it feasible for elite athletes to provide data without interfering with their normal training and racing.

5. CONCLUSION

This paper describes a model which uses power data from a cyclist's training and races to determine an optimal individualised taper strategy for the final 3 days before a competition. The model uses a novel application of a symbolisation technique to enable examination of the shape of a taper. The results of this study suggest that taper responses are highly individual, and can not be generalised. The use of field-derived model inputs makes this model a practical tool for the planning of a taper to maximise competition performance. Future work could extend this technique to identify the patterns of training load in training microcycles, and relate these patterns to fatigue and performance measures.

Acknowledgements

The authors acknowledge the support of the Australian Institute of Sport and the University of Canberra in providing equipment used in the study.

References

Bosquet, L., Montpetit, J., Arvisais, D., & Mujika, I. (2007). Effects of tapering on performance: a meta-analysis. *Medicine & Science in Sports & Exercise*, *39*(8), 1358.

Banister, E. W., Carter, J. B., & Zarkadas, P. C. (1999). Training theory and taper: validation in triathlon athletes. *European Journal of Applied Physiology and Occupational Physiology*, *79*(2), 182-191.

Churchill, T., Sharma, D., Balachandran, B., (2009). Correlation of novel training load and performance metrics in elite cyclists. Proceedings of the 7th International Symposium for the International Association on Computer Science in Sport (Canberra, 2009).

Cohen, J., (1988). Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, Hillsdale.

Gore, C. J. (Ed.). (2000). *Physiological Tests for Elite Athletes*. Champaign: Human Kinetics.

Hopkins, W. G., Hawley, J. A., & Burke, L. M. (1999). Design and analysis of research on sport performance enhancement. *Medicine and Science in Sports and Exercise*, *31*(3), 472-485.

Kubukeli, Z. N., Noakes, T. D., & Dennis, S. C. (2002). Training techniques to improve endurance exercise performances. *Sports Medicine (Auckland, N.Z.)*, *32*(8), 489-509.

Lin, J., Keogh, E., Patel, P. & Lonardi, S., (2002). Finding motifs in time series. In *proceedings of the 2nd Workshop on Temporal Data Mining, at the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Edmonton, Alberta, Canada.

Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. San Diego, California: ACM.

Lin, J., Keogh, E., Lonardi, S., Lankford, J. P., & Nystrom, D. M. (2004). Visually mining and monitoring massive time series. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*.

Lin, J., Keogh, E., Wei, L., & Lonardi, S. (2007). Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, *15*(2), 107–144.

Mujika, I. (2009). Tapering and Peaking for Optimal Performance. Human Kinetics.

Mujika, I., & Padilla, S. (2003). Scientific bases for precompetition tapering strategies. *Medicine & Science in Sports & Exercise*, *35*(7), 1182.

Pyne, D. B., Mujika, I., & Reilly, T. (2009). Peaking for optimal performance: Research limitations and future directions. *Journal of Sports Sciences*, *27*(3), 195-202.

Thomas, L., Mujika, I., & Busso, T. (2008). A model study of optimal training reduction during pre-event taper in elite swimmers. *Journal of Sports Sciences*, *26*(6), 643.

MODELING OF THE SNATCH WEIGHTLIFTING

M. H. Korayem [a] , S. A. Mustafa, A. H. Korayem, and A. Amanati

Robotic Research Laboratory, College of Mechanical Engineering,

Iran University of Science and Technology, Tehran, Iran

[a] *Corresponding author: hkorayem@iust.ac.ir*

Abstract

Weightlifting has been the subject of research among world scholars since a long time ago. Previous research mainly focused on biomechanical and kinesiological aspects, whereas in recent years, simulation of weightlifting motion has come to the forefront. However, there is an evident lack of data on modeling and simulation of this motion from biomechanical and robotic aspects.

In this paper, snatch technique of weightlifting has been modeled from a biomechanical perspective, which includes physical joint constraints, muscle strength and the typical snatch technique; and by applying the proposed model it is possible to model the motion, from start to finish in a given time, for an athlete of known weight and stature, and present the suitable optimum motion as output. Using motion analysis equipment and comparison of athlete's motion with that of the model, we can improve the performance of novice athletes to prepare them for medal attainment in competitions. The model output accurately replicates the biomechanical details such as the new elite athletes' technique – double knee bending in which the knee initially extends and then flexes so as to reduce the reaction torque and hence decreasing injury risks. The weightlifting robot has been modeled with a Lie algebraic formulation and the resultant problem is solved as a parameter constrained optimal control problem.

Keywords: Snatch, Weightlifting, Optimal Control, Lie Algebra, Double Knee Bending

1. INTRODUCTION

Barbell Trajectory and other dynamic characteristics of motion, like displacement and velocity of barbell during the Snatch Lift Technique have garnered widespread interest for the last few years. Several researches investigated by Isaka *et al.* (1996), Baumann *et al.* (1988), Gourgoulis *et al.* (2000), Byrd (2001), Garhammer (2001) and Schilling *et al.* (2002), evaluated optimum barbell trajectory experimentally according to the percentage of the athletes' success, aimed to answer research questions by investigating the relationships between variables using quantities data obtained in an experiment and assessing the significance of the results statistically. Theoretical approaches to answering a research question typically employ a model that gives a simplified representation of physical system under study.

The snatch technique requires the barbell to be lifted from the floor to a straight-arm overhead position in one continuous movement (Garhammer, 1989).The movement as a whole includes six phases, as shown in Figure 1: the first pull (a), the transition from the first to the second pull (b), the second pull (c), the turnover under the barbell (d), the catch phase (e) and rising from the squat position (f)(Baumann *et al.*, 1988).The first five phases are considered to be the most important phases of the lift (Baumann *et al.*, 1988). They occur in less than 1 s and, as such, involve a high power output (Isaka *et al.*, 1996). In recent years, some researches introduced optimized patterns for lifting tasks using actuating torque as a mechanical parameter.

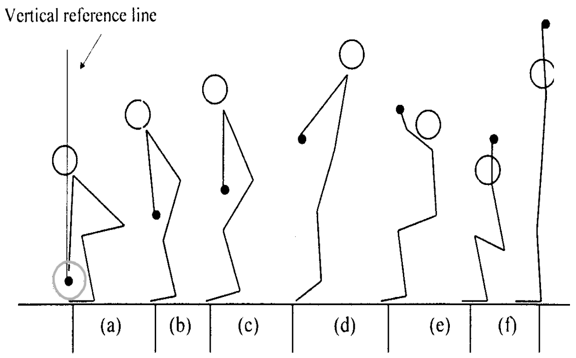


Figure 1: The phases of the snatch: (a) the first pull, (b) transition from the first to the second pull, (c) the second pull, (d) turnover under the barbell, (e) the catch phase, (f) rising from the squat position.

For example, investigating the differences in motion patterns for goal-directed lifting by considering biomechanical constraints or physiological responses and the redundancy of degrees of freedom made it possible to have an optimum motion pattern (Park *et al.*2005).But because of its complexity this method for weightlifting was not widely used.

Using double knee bending technique, which is unique to weight lifting, permits reemployment of powerful knee extensor muscles through their strongest range of motion. Capability of optimal control theory for sport activities encourage us to apply this method to optimize the whole motion of snatch lift and a mathematical model based on dynamic principles to predict the barbell trajectory which minimized the specific criterion is formulated.

2. METHODS

To effectively devise a mathematical model of the human's physical properties, the five-link model is employed to analyze lifting Tasks which have been used in several researches.

By simplifying comprehensive model for weightlifting which has been introduced by Chaffin and Anderson (1991), to a sagittal plane model, the body segments are modelled with solid links and body joints to simple revolute joints. The schematic diagram of this model which is made by five links and also five body joints, used by like Chang *et al.* (2001), Menegaldo *et al.* (2003) and Park *et al.* (2005) is shown in Figure 2. L1 to L5 represent shin, thigh, trunk, upper arm and forearm, respectively and ankle, knee, hip, shoulder and elbow are named O1 to O5 respectively. Therefore we used this model.

The redundancy of degrees of freedom makes it possible to have an optimum motion pattern.

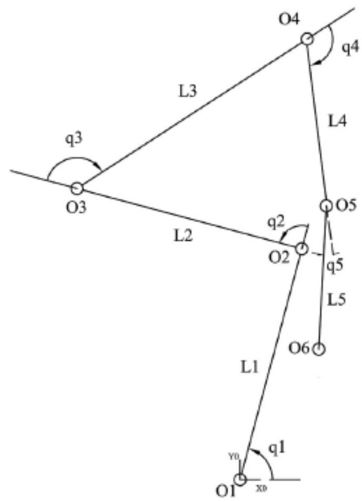


Figure 2: Biomechanical Model of a Weightlifter

The minimum-effort optimal control problem for this system describes as (Martin and Bobrow 1999):
Minimize τ where

$$J(\tau) = \frac{1}{2}\int_0^{t_f} \|\tau\|^2 \, dt \tag{1}$$

Now τ is given by motion equation as follow:

$$M(q)\ddot{q} + h(q,\dot{q}) = \tau \tag{2}$$

where $M(q)$ shows the $(n \times n)$ inertia matrix and $h(q,\dot{q})$ consists of Coriolis and gravity and friction vectors and $q \in \Re^n$, $\tau \in \Re^n$.

Bounds on the joint coordinate and initial and final conditions are primarily shown as follows:

$$\underline{q} \le q(t) \le \overline{q} \quad , \quad \underline{q}, \overline{q} \in \Re^n \tag{3}$$

$$q(0) = q_0 \quad , \quad \dot{q}(0) = 0$$
$$q(t_f) = q_f \quad , \quad \dot{q}(t_f) = 0 \tag{4}$$

where $\underline{q}$ and $\overline{q}$ are specified values and t_f is the final time.

In order to solve the optimal control problem it is assumed that joint coordinates are B-splines where their curves depend on $B_i(t)$ basic functions ,and

$P=\{p_1,p_2,\ldots,p_m\}$, vector of control inputs, where $p_i \in \Re^n \quad ,i=1,\ldots,m$. Therefore the joint trajectory turns to $q=q(t,P)$ where:

$$q(t,P)=\sum_{i=1}^{m} B_i(t)\,p_i \tag{5}$$

Considering the fact that $\sum_{i=1}^{m} B_i(t)=1$, boundary conditions of joint movements in (3) can be replaced with boundary of spline parameters p_i and therefore the parametric optimal control problem can be described as follow:

Minimize P where

$$J(P)=\frac{1}{2}\int_0^{t_f}\|\tau(t,P)\|^2\,dt \tag{6}$$

Subject to

$$\underline{q} \le pi \le q \quad , \quad i=1,\ldots,m \tag{7}$$

accompanied by the boundary conditions in (4).Joint torque can also be calculated for each P vector using amended Newton-Euler recursive dynamic approach based on Product of Exponentials which can be easily solved using computing techniques.

Our intended purpose is to propose a model for weightlifting robot with a Lie algebraic formulation which does not consider the experimental trajectory as the input of simulation and not only solve it as a constrained optimal control problem but, as well, is in accordance with the experimental results .

The force exerted by a muscle is a function of the activation level and the maximum muscle force. As indicated by Yeadon *et al*.(2006), a four parameter function consisting of two rectangular hyperbolas was used to model the torque/angular velocity relationship while the activation/angular velocity relationship was modelled using a three parameter function for high concentric velocities. The product of these two functions gave a seven parameter function which best models the torque/angular velocity, therefore we used this model and extended it for other muscles.

Control constraints term is used for the inequalities defining limitation on torques acting on the system and the redundancy of degrees of freedom in order to (i) provide a continuous motion (ii) limit the motion by human's joints and torques boundaries (iii) minimize the proposed cost function from a biomechanical point of view. Therefore an optimal control problem is formulated. We aim to generate an optimal motion by minimizing actuating torques as our performance criterion. We have used minimum-torque-change principle introduced by Uno *et al*. (1989) and penalized it with maximum muscle force by introducing cost function as:

Minimize P where

$$\begin{aligned} J(P) &= \\ &\Big[c_1 \int_0^{t_f} \|\dot{\tau}(t,P).\mathrm{cost}(\tau(t,P))\|^2\,dt + \\ &c_2.\ arclen(bar(P))\Big]_0^{t_f} + \\ &c_3.(\min(bar_y(P)) - bar_0(P))^4 - \\ &c_4.\max(bar_y(P))\Big]\ .\exp_{HF}\ .\exp_{HK} \end{aligned} \tag{8}$$

The first term of Equation (8) which describes the torque-change cost function is multiplying with another cost function which penalizes the deviation from maximum muscle torque as follow:

$$\mathrm{cost}\ (\tau(t,P)) = penalize\ (\tau(t,P)/\tau_{\max}(t,P))$$

$$penalize\ (x)=\begin{cases} x & (if\,|x|\le 1) \\ \exp(x) & otherwise \end{cases} \tag{9}$$

AS it can be seen in equation (9) this cost function penalizes the torques which is greater than the maximum muscle torque in exponential form and others will remain the same. The second term of equation (8) which shows the barbell trajectory in snatch lift helps us to optimize the barbell trajectory which minimizes the torque-change cost.

$$\begin{aligned} &arclen\ (bar(P)) = \\ &\int \sqrt{\partial bar(P)/\partial x)^2 + (\partial bar(P)/\partial y)^2\,dt} \end{aligned} \tag{10}$$

Third and fourth terms of cost function (8) represent minimum and maximum vertical height of barbell in the whole motion, respectively. Minimum vertical height of barbell, $\min(bar_y(P))$, should be greater than vertical height at the start point $bar_0(P)$ and any difference also penalizes in exponential form. Decreasing the cost function with maximum vertical height of the barbell, results in more extension of body during snatch lift.

$$\exp_{HF} = (c_5.(\max(\tau_{hip})/\max(\tau_{ankle}) - r_1))^4 + 1 \tag{11}$$

$$\exp_{HK} = (c_6.(\max(\tau_{hip})/\max(\tau_{ankle}) - r_2))^4 + 1 \tag{12}$$

Double knee bending (DKB) technique indicates in equation (11) and (12). Since applying torque on knee joint is 3 times less than on hip and is regardless to weight of athlete and barbell, expert weightlifters use this technique to reduce the reaction torque and hence decrease injury risks. r_1 and r_2 coefficients are obtained from experimental data and c_i coefficients used to equal terms of cost function. Eliminating these coefficients result in ignoring the lower order terms.

3. RESULTS

The problem was solved for a weightlifter with 60 kg mass who lifts a 143 kg barbell by snatch technique. Other dimensional parameters were calculated based on this information. Actuating torques of all joints and starting and ending angles were used as indicated in Bartonietz (1996). Barbell trajectory is considered during the start of snatch from the barbell lift-off until the start of catch phase in 1 second is shown in Figure 3.

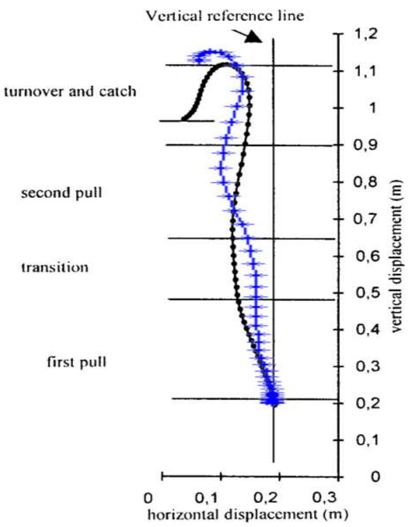


Figure 3: Optimized and experimental barbell trajectory during snatch lift

A good agreement could be seen between the optimized trajectory and experimental results found by Gourgoulis *et al*. (2000) and also this model follows the concavity and convexity compared with similar research by Nejadian *et al*. (2007, 2008).

From a biomechanical point of view, an effective snatch lift is characterized by a velocity-time relationship of the barbell in which the vertical linear velocity of the barbell increases continuously between the first and the second pull. Figure 4 shows the diagram of the vertical barbell velocity which changes with the time during snatch lift. It can be observed here that vertical velocity first increases and decreases when the barbell is moving toward the final position.

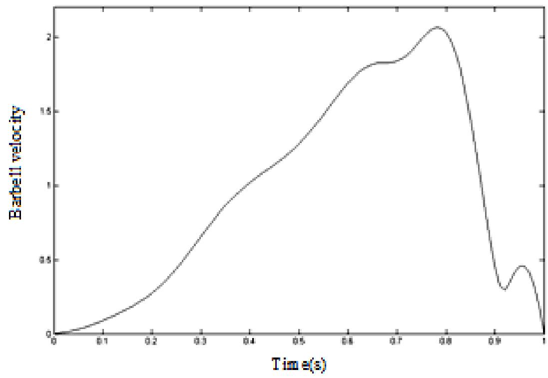


Figure 4: Optimized vertical velocity of barbell during snatch lift

Figure 5 illustrates curves for joint angular displacements including ankle, knee and trunk in the sagittal plane to determine the movement of the leg and the trunk with respect to horizontal axis (as described in experimental graphs).

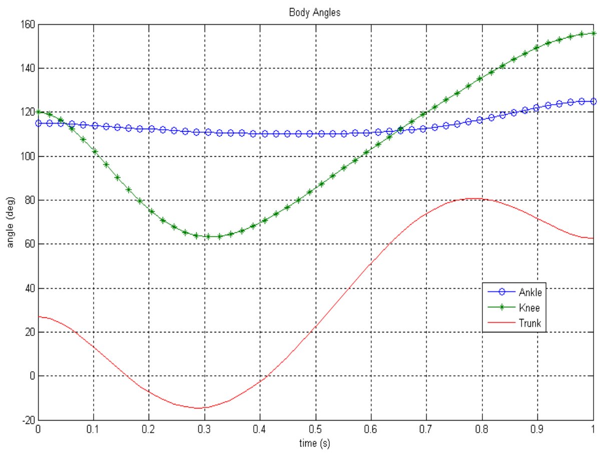


Figure 5: lower-limb angular velocities during snatch lift

The reason why the curves in above diagram are in a good agreement with experimental graphs as the barbell moves toward the final position is the similarity between simulated and real motion.

As it could be observed, during the first pull, the knee joint is extended and achieved a minimum at the end of the first pull. After achievement of the minimum knee angle, which marked the start of the second pull, the knee is explosively extended and reached its maximum extension at the end of the second pull. The ankle joint is extended continuously during the first pull and then decreased during the transition phase. During the second pull, the ankle joint was explosively extended and reached its maximum at the end of the second pull, which demonstrates essential technique like DKB in modern weightlifting.

Considering coefficients (11) and (12) in cost function (8) causes the knee joint crosses 90° twice, as is shown in Figure 5, which describes its flexion and extension. Applying waist-knee torque ratio in the cost function forces the output of simulation to provide a trajectory satisfying this constraint which was not considered in past researches.

The curves of lower-limb moments including ankle, knee and hip during snatch lift are plotted in Figure 6. As shown the great amount of moment is applied on hip joint and applying torque on knee joint is the least. The positive moment or negative moment declares that extensor muscle or flexor muscle is acting.

As the barbell approaches the end of first pull, the moment on hip joint becomes steady while the moment on knee and ankle joint decreases. The moment on the knee decreases continuously over the first pull and the knee flexor muscle acts, accordingly. By starting the second pull phase all the moments become positive again and knee and ankle moments reach their second maximums before decreasing. The joint moment-time curves are approximately similar in different athletes.

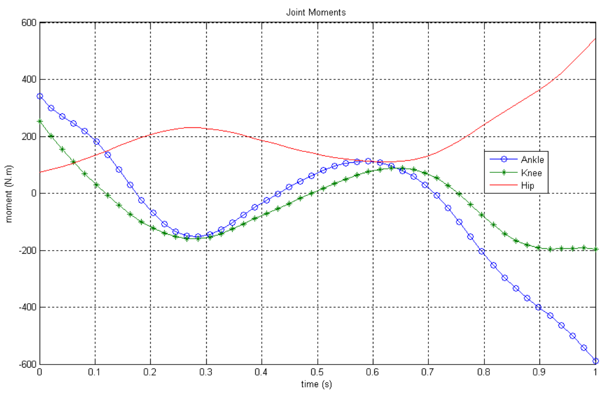


Figure 6: lower-limb moments during snatch lift

In Figure 7, the ground reaction force (GRF) during snatch lift can be observed, which decreases as the barbell accelerates upward and reaches a minimum at the end of first pull and increases continuously till the barbell reaches its highest position.

Since the foot-off phase is not considered in the biomechanical model (Fig.2), decreasing all moments and also ground reaction force to zero right after the foot-off phase, as can be observed in experimental diagrams, are missed here.

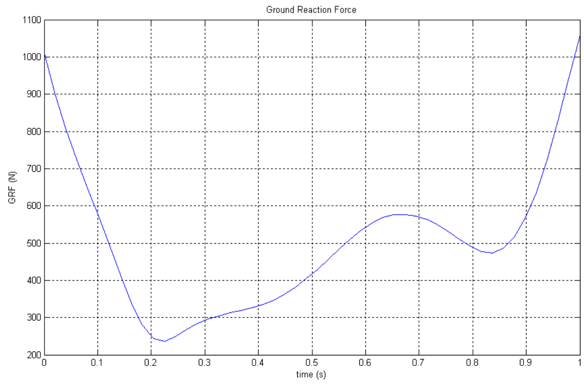


Figure 7: Ground reaction force during snatch lift

4. DISCUSSION

Our optimized model provides Barbell trajectory which can be observed as its typical form in experimental diagrams during snatch lift. Producing this optimized trajectory from system motion dynamics, penalizing actuating torque with human muscular model and considering the double knee bending technique in our model confirm that proposed model achieves a relative success to anticipate the optimal motion in comparison with other optimizing strategies. This optimization can apply for training weightlifters to act like the optimized kinematics parameters or to make their characteristics like resultant kinetic parameters.

However selecting several parameters such as time of snatch lift, energy expenditure,… as criteria which should be minimized during an optimized snatch lift results in improving the performance of weightlifter. Introducing and modifying the proper criterion which is in accordance with human motion pattern is another advantage of this study and we believe that we are successful regarding to this matter. However, lack of data on modeling and simulation of this motion from biomechanical aspects is a problem that we experienced.

5. CONCLUSIONS

Dynamic of the model help us to gain an accurate and deep understanding of motion and to control and improve it during the snatch lift. Obtained results which are in accordance with experimental results confirm the reliability of proposed model due to achieving a relative success to anticipate the optimal motion.

Applying optimized motion during snatch lift does not only help coaches to advise weightlifter about the correct velocity he should reach or the strength training he should do to compensate the weakness of particular joint, assists athletes as well to act like the optimized kinematics parameters with the purpose of reducing injury risks and maximizing desirable training effects. The importance and descending role of each joint during the snatch lift can also be identified which are good parameters to show the practical differences between an actual snatch motion of weightlifter and the ideal optimized one which weightlifter could achieve.

Considering the body movement, physical characteristics of the various sectors of the movement, muscle strength characteristics of barbell trajectory of the movement which have a certain impact on snatch weightlifting will be the focus of future research.

Acknowledgements

This paper describes a research done in Robotic Research Laboratory, College of Mechanical Engineering, Iran University of Science and Technology (IUST). Authors thank IUST for supporting the program during our research activity.

References

Bartonietz, K.E. (1996). Biomechanics of the snatch: Toward a higher training efficiency, *National Strength and Conditioning Association Journal,* 18: pp. 24-31.

Baumann, W.V., Gross, K., Quade, P., Galbierz and Schwirtz, A. (1988). the Snatch Technique of World Class Weightlifters at the 1985 World Championship. The 1985 World Championships, 4: 68-89.

Byrd, R. (2001). Barbell trajectories: three case study. *Strength and Health*, 3:40-42.

Chaffin, D.B. and Anderson, G.B.J. (1991). *Occupational Biomechanics.* John Wiley.

Garhammer, J., 2001. Barbell Trajectory, Velocity, and Power Changes: Six Attempts and Four World Records. Weightlifting USA, 19(3): 27-30.

Gourgoulis, V., Aggelousis, N., Mavromatis, G. And Garas, A.(2000).Three-Dimensional Kinematic Analysis of the Snatch of Elite Greek Weightlifters. *Journal of Sports Sciences*, 18(8):643-652.

Isaka, T., Okada, J. and Funato, K. (1996). Kinematics Analysis of the Barbell during the Snatch Movement of Elite Asian Weightlifters. *Journal of Applied Biomechanics,* 12: 508-516.

Martin, B. J. and Bobrow, J. E.(1999). Minimum-Effort Motions for Open-Chain Manipulators with Task-Dependent End-Effector Constraints, *International Journal of Robotics Research,* 18: pp. 213-224

Nejadian S. L. and Rostami, M. (2007). Mathematical Modeling of Snatch Lift Technique, *25th International Symposium on Biomechanics in Sports, Ouro* Preto – Brazil.

Nejadian S. L. and Rostami, M. and Towhidkhah F. (2008). Optimization of Barbell Trajectory during The Snatch Lift Technique By Using Optimal Control, *American Journal of Applied Sciences,* vol. 5.

Park, W., Martin, B.J., Choe, S., Chaffin, D.B. and Reed, M.P. (2005). Representing and Identifying Alternative Movement Technique for Goal-Directed Manual Tasks. *Journal of Biomechanics*, 38: 519-527.

Schilling, B., Stone, M. , O'Brayant, H.S. , Fry, A.C. , Cogllanese, R.H. and Pierce, K.C. (2002). Snatch Technique of College National Level Weightlifters. *Journal of Strength and Conditioning Research*, 16(2): 551-555

Uno, Y., Kawato, M. and Suzuki, R.(1989). Formation and Control of Optimal Trajectory in Human Multijoint Arm Movement, *Journal of Biological Cybernetics,* vol. 61.

Yeadon, M.R., King, M. A. and Wilson, C. (2006). Modelling the maximum voluntary joint torque/angular velocity relationship in human movement, *Journal of Biomechanics*, 39: 476-482.

A 'POTTED HISTORY' OF MATHEMATICS AND COMPUTERS IN SPORT: IN AUSTRALASIA (1992-2008)

John Hammond [a,b]

[a] *Department of Sport Coaching and Exercise Science, University of Lincoln, UK*
[b] *Corresponding author: john.hammond@scu.edu.au*

Abstract

This paper describes briefly the development of Mathematics and Computers in Sport in Australia and New Zealand through analysis of papers and presentations at 'MathSport' biennial conferences. In providing a context for this analysis it is proposed that Mathematics and Computers in Sport is seen as contributing to sports science both directly and indirectly through other science disciplines. Scholars and scientists associated with the 'MathSport' group began input into the body of knowledge in sports science in the early 1990s. This contribution of research and scholarship has continued up to the present. The variety of contributions in terms of different sports, theoretical frameworks and modes of analysis has increased in volume since the early days.

Keywords: performance analysis, sports science, computer science

1. INTRODUCTION

This paper describes the history of contribution to sports science by the MathSport group in Australia and New Zealand through collaboration and presentations at biennial conferences. Mathematics and Computers in Sport (MCS) is seen by the MathSport group as contributing to sports science through: mathematical and statistical modelling in sport; the use of computers in sport; the application of these to improve coaching and individual performance; and, teaching that combines mathematics, computers and sport (ANZIAM, 2008). In placing this view within science in general and sports science specifically, three models are proposed, adapted from a concept by Hughes, Hughes and Behan (2007). The models are presented in progression to illustrate how mathematics and computers contribute to the body of knowledge in sports science and to the enhancement of performance. The first of these (Figure 1) provides a view of sports science contribution to the performer, with the performer at the centre of the model in a dynamic environment which includes social, pedagogical and performance contexts.

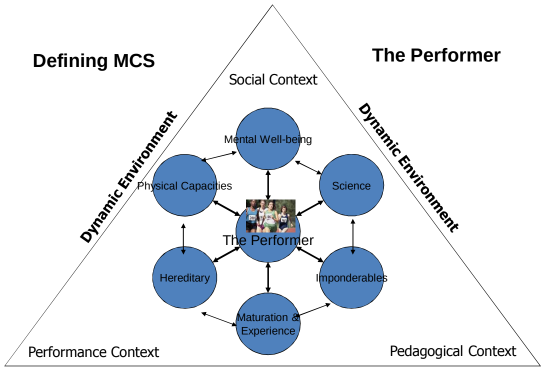


Figure 1: A model based on the performer

The second model (Figure 2) then places sports science in the same triangulated context but depicting how wider science disciplines and applied disciplines, such as performance analysis, contribute to sports science. The obvious link between sports science and performer is reduced here to a single linear relationship but this is merely to connect back to Figure 1 and does not illustrate the dynamic multi-faceted nature of the link.

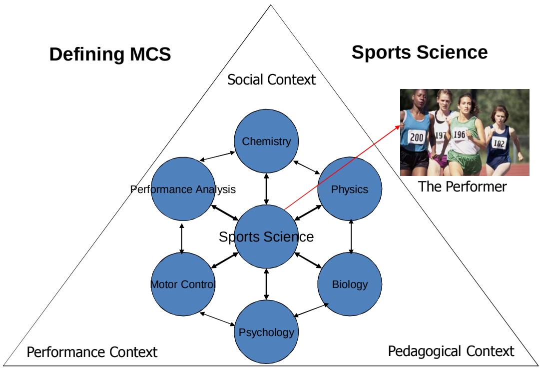


Fig. 2: A model of the contributions of science to sports science

The third model (Figure 3) depicts how mathematics and computer science then contributes to the wider science disciplines and, therefore, indirectly to sports science (Hammond & de Mestre, 2008). Again, the complexity of these relationships cannot be adequately illustrated in a two-dimensional model but the dynamic environment would continue, albeit abstractly, across the second and third models postulated here.

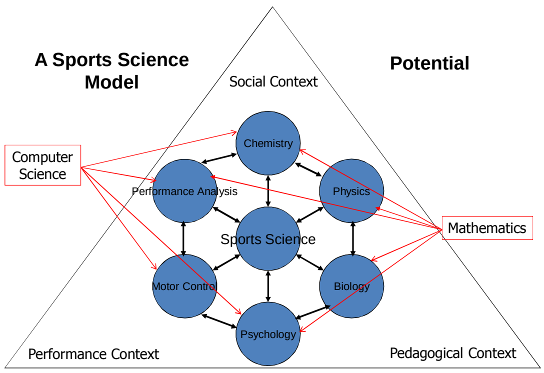


Fig. 3: The contributions of mathematics and computer science

This progressive view of intra and inter relationships in sports science set the contributions of mathematics and computer science in a wider context. Developments within these elements of sports science in Australasia over the last eighteen years, or so, are described in the next section.

2. BRIEF HISTORY OF MCS IN AUSTRALASIA

The MathSport group, set up to promote MCS, grew out of members of the Australian Mathematical Society (AustMS) who had like interests in applying their discipline to research and scholarship in sport. AustMS is the national society of the mathematics profession in Australia, whose mission is to promote and extend mathematical knowledge and its applications (AustMS, 2008). Within AustMS the Australia and New Zealand Industrial and Applied Mathematics division (ANZIAM) supports special interest groups, one of which is the MathSport group (ANZIAM, 2008). This group facilitates forums within which sports scientists, from Australian and New Zealand and the wider international community, interact. The MathSport special interest group holds biennial meetings - the Australasian Mathematics and Computers in Sport Conferences (MathSport, 2008).

The MathSport group was inspired by the publication 'Mathematics of Projectiles in Sport' and its author Neville de Mestre (1990). The group held its first conference at Bond University in 1992, i.e. 'Mathematics and Computers in Sport Conference'. The topics were varied and the program included papers on: one-day cricket (Johnston, 1992) – this subject matter that was to become synonymous with the conference; fell running (Hayes & Norman, 1992); and, combining mathematics and computer technology for improving sport, by keynote speaker Jon Patrick (1992). The 1992 conference saw a lasting relationship between the conferences and Professor Stephen Clarke who presented a paper concerning sports betting and Australian rules football (1992), topics which have been strong features of the conferences over the years. The two other keynote addresses were given by John Croucher (1992) and David Hoffman (1992) who presented papers concerning the science of winning and team rating systems, respectively.

The success of the first conference initiated requests for a follow-up conference, which came to fruition in 1994. The number of presentations increased from twelve to twenty, with topics including: seven papers about various aspects of golf (1994); dynamic control of bobsled by keynote speaker Mont Hubbard (1994); and, a computerised sports counselling program (1994). Stephen Clarke (1994) also gave a keynote address on rating systems for racquet sports. Neville de Mestre continued his involvement as conference organiser and proceedings editor in 1994, assisted by Kuldeep Kumar in 1996, with both conferences being hosted at Bond University. The 1996 conference saw the first appearance at the conference of Tony Lewis who presented a paper on the famous Duckworth-Lewis (D/L) method of calculating target scores in rain interrupted one-day cricket matches (Duckworth & Lewis, 1996). Tony Lewis has proved to be a regular attendee at MCS conferences since then, with continual updates and explanation of the statistical modelling used in the evolving Duckworth-Lewis methods. Keynote speakers were Stewart Townend, whose presentation on Mathematical Principles was not published and Hugh Morton (1996) who delivered a paper on analysis of world records. The 1996 conference also saw papers on netball (Noble, 1996) and horse-racing (Benter, Miel & Diane, 1996; Phatarfod, 1996; Kumar, Ganesalingam & Ganesh, 1996) for the first time.

Up to and Beyond the Sydney Olympics

Bond University continued to host the conference in 1998 where the decision was made to move the conference to Sydney in 2000, being the Sydney Olympic year. The 1998 conference saw the start of performance analysis papers through John Croucher's keynote address on tennis (1998). Papers in this emerging area of sports science, at the time, proved to increase in number as the conference sequence progressed. In 1998 the other two keynote speakers were Tony Lewis about the D/L model (1998) and Chris Harman (1998), who delivered a lecture on optimal baseball running. The Olympic year conference was hosted by the University of Technology, Sydney, under the auspices of Graeme Cohen, ably assisted in editing the proceedings by Tim Langtry. Of the 25 papers presented the inextricable link between mathematics, physics and biomechanics came out in the two keynote addresses (Cross, 2000; Siff, 2000). Cricket and Australian rules football featured strongly again on the program and papers on teaching mathematics emerged here (Cohen, 2000; Byun, 2000). The conference added 'Australian' to the title, becoming the 'Australian Mathematics and Computers in Sport Conference'.

The sixth conference in 2002 returned to the Gold Coast with Neville de Mestre taking up the organisational reins again, with proceedings editors again being Cohen and Langtry. Ray Stefani from California joined Australians Graeme Cohen and Stephen Gray as keynote speakers, who presented papers on rating systems (Stefani, 2002), chance in cricket (Cohen, 2002) and forecasting scores in cricket (Gray & Tuan, 2002) respectively. The 2002 conference proved another watershed in terms of influential speakers appearing for the first time, in that Graham Pollard presented the first of many papers, with his colleagues, at MCS meetings about scoring systems in tennis (Pollard & Noble, 2002).

Expanding Horizons

In 2004 the conference moved away from Australia to be hosted by Massey University in New Zealand. Hugh Morton took on the organiser role and was assisted in the editing of the proceedings by Selvanayagam Ganesalingam. The program had now seen a steady increase in the number of papers being presented to reach 30 for the first time (see Figure 4).

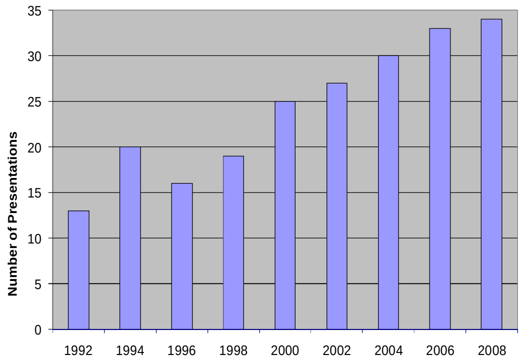


Figure 4 Number of Papers presented at MCS conferences 1992-2008

With New Zealand input now becoming more prominent, organisers changed the title of the conference replacing 'Australian' with 'Australasian'. Both keynote speakers came to the conference with high level international reputations as experts in their field of sports science. Keith Davids (NZ) presented a paper on dynamic movement systems (Davids & Button, 2004) and Mike Hughes (Wales) took the area of performance analysis perspectives to a new height at this conference (Hughes, 2004). Richard Green also increased the exposure of delegates to computer science and its application to biomechanical analysis (2004). In 2006 the eighth conference moved back to the Gold Coast in Australia but under the independent auspices of the MathSport group. Neville de Mestre and John Hammond shared both the organisation of the conference and editing of the proceedings. Roger Bartlett travelled from New Zealand to present his keynote paper on artificial intelligence and how it relates to biomechanics (2006). Stuart Morgan from the Victorian Institute of Sport, gave the second keynote address on synergies in high performance sport (2006). Following Mike Hughes's lead at the previous conference and Stuart Morgan's keynote talk, papers featuring performance analysis were more prolific at this event with Aaron Silk (Silk, Hammond & Weatherby, 2006) Didier Seyfried (2006), Martin Lames (2006) and John Hammond (Hammond &

Smith, 2006; Smith, Gilleard, Hammond & Brooks, 2006) providing a variety of applications of this increasingly influential sub-discipline. In addition to the proceedings, selected papers from the 2006 conference were published in a special edition of the Journal of Sports Science and Medicine.

At the 2008 conference Arnold Baca (President of the International Association of Computer Science in Sport) and Ian Renshaw (QUT) gave the invited keynote addresses on 'Tracking Human Motion in Sport' and 'Performance and Learning of Motor Skills' respectively. This conference saw a re-emergence of water-based sport science with papers on Bodysurfing (de Mestre, 2008) Sailing (Tonkes, 2008) and Kayaking (Janssen, Sachlikidis & Hunter, 2008) as well as scoring and prediction systems for tennis and cricket (Lewis, 2008; Stern, 2008; Bailey & Clarkea/b, 2008; Lisle, Pollard & Pollard, 2008; Pollard & Pollarda/b, 2008; Brown, Barnett, Pollard, Lisle & Pollard, 2008). There were a record number of presentations at this conference of 34 (see Figure 4).

Also of note, across the period that the MCS conferences have been scheduled, are the following outstanding contributions by authors, not previously mentioned in the chronological sequence presented above:

- Cogill's papers on the mathematics of cycling (1994 & 1998);
- Norton's papers on prediction models and advantage in Softball (1994), Tennis (1996 & 2002), Netball (2000) and Basketball (2004);
- Heazlewood's papers on predicting Olympic games performance (1996 & 1998);
- Bedford's papers on ratings-based models applied to World Cup handball (2004) and soccer (2004);
- Norman's papers on batting on 'sticky wickets' (2004) and fell walking (2006);
- Ovens's papers on mathematical modelling applied to one-day cricket (2004 and 2006).

Summary of the Presentations

The number of presentations at MCS conferences across the years 1992-2008 was illustrated in Figure 4. There has been a wide variety of sports and topics presented at MCS conferences. In breaking down the numbers into convenient categories, as can be seen from Figure 5, cricket heads the list at 39 with Tennis (32) and Australian Rules Football (29) following closely behind.

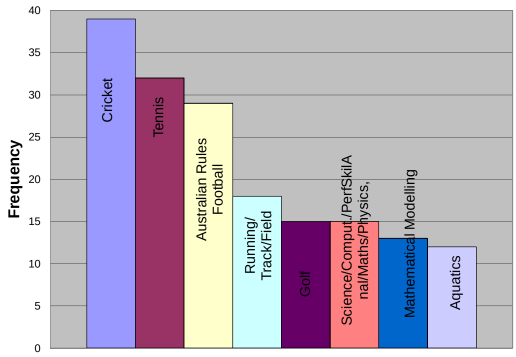


Figure 5 The Most Popular Topics at MCS conferences 1992-2008

Categorisation for the data presented in Figure 5 is mainly based on specific sports, for simplification. However, re-analysis of the data into different ways of categorisation would also see categories such as mathematical modelling, rating and scoring systems emerge as prolific as the top three sports categories. This is because papers often addressed more than one focus, such mathematical modelling of sports betting, player rating systems across or within sports, scoring systems across racquet sports and the like.

3. CONTEXTUAL SYNOPSIS

The extent and diversity of topics described here, demonstrate that mathematics and computers have an important role to play in adding to the body of knowledge of sports science in the Australasian setting. The role of these disciplines in sport can be viewed in three overlapping contexts: performance, social and pedagogical. This view is depicted in Figure 6 where the place of sport and its association with mathematics and computers is seen in a dynamic and interactive model. The levels of complexity, interconnectivity and dynamism represented in a simple two-dimensional model, such as this, cannot accurately be depicted here. However, Figure 6 does seek to provide an illustration, albeit simplified, of the mathematics, computers and sports contributions within these contexts.

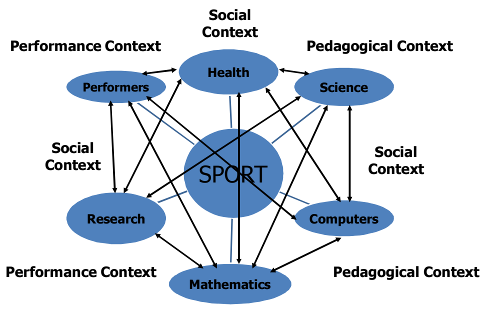


Figure 6: A Contextual Dynamic Model of the Relationship of Mathematics and Computer Science to Sports

In the performance context there is a broad scope of sports that already use or can use mathematical and computer analysis, such as performance analysis technology or in the measurement of the physical capacities that can enhance performance. Research into every aspect of performance, relies heavily on these disciplines, particularly in the recording of and the analysis of data. In the social context, more widely based issues such as the place of sport in society or the enhancement of health, can be aided indirectly, via sport and exercise sciences, and directly through the use of mathematics and computers. In the pedagogical context, analysis and research into coaching methods can lead to improvement in coach education and methods. In addition, the use of sport as medium through which to teach mathematics and computer science at all levels of education is invaluable and reciprocal. The function of computers and mathematics in teaching sports science is obvious.

4. CONCLUSION

This paper considered the place of mathematics and computers in a wider sports science context as a background to the history in Australasia of this area. Models of the performer in the sports science context, sports science in a wider scientific context and how mathematics and computer science contributes to sports science were proposed. The links between each of the models were described and it was suggested that mathematics and computers have both a direct and indirect contribution to scientific support of the sports performer. A history of the MathSport group's initiatives in promoting mathematics and computers in sport (MCS) was set out. The MCS conferences have grown in size from 12 papers in 1992 to a program of 34 papers in 2008 and the forthcoming conference in 2010 will continue the trend of 30+ papers. Among the many presenters at the conference over the years there has been a significant number of scholars of international repute, both as keynote speakers and as ordinary delegates and presenters. Whilst this is still a relatively small conference in comparison to some scientific conferences, its diverse nature in terms of sports discussed and strategy of non-parallel sessions has been productive. From this it can be reasonably suggested that the MathSport group has progressed research and scholarship in MCS both within Australasia and in the international sports science community in general.

Acknowledgements

I am indebted to colleagues in the MathSport group whose contributions have created this history.
I wish to thank the World Academic Press for their kind permission to reproduce sections of this paper from a previous publication (Hammond & de Mestre, 2008).
I wish to thank co-author of that previous paper, Neville de Mestre, for his input.

References

ANZIAM. (2008). Applying mathematics knowledge. Australia and New Zealand Industrial and Applied Mathematics. http://www.anziam.org.au/HomePage,

AustMS. (2008). The national society of the mathematics profession. Australian Mathematics Society. http://www.austms.org.au/HomePage.

Baca, A. (2008). Tracking motion in sport – trends and limitations. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 1-7.

Bailey, M. & Clarke, S. (2008)a. Predicting the number of runs per over in one-day international cricket matches. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 23-31.

Bailey, M. & Clarke, S. (2008)b. A comparison of distributions for the runs scored per over in one-day international cricket matches. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 32-36.

Bartlett, R. (2006). Artificial intelligence in sports biomechanics: new dawn or false hope. In: J. Hammond & N. de Mestre (eds.). Mathematics and Computers in Sport. Coolangatta: MathSport, 91-98.

Bedford, A. (2004). Predicting women's World Cup handball match outcomes using optimised ratings models. In: H. Morton & G. Ganesalingam (eds.) Mathematics and Computers in Sport, Palmerston North: Massey University, 66-74.

Bedford, A. & Da Costa, C. (2004). A ratings based analysis of Oceania's road to the World Cup. In: H. Morton & G. Ganesalingam (eds.) Mathematics and Computers in Sport, Palmerston North: Massey University, 75-87.

Benter, W., Miel, G. & Diane, P. (1996). Turnbough Modelling distance preference in thoroughbred racehorses. In: N. de Mestre & K. Kumar (eds.). Mathematics and Computers in Sport. Queensland: MathSport, 75-81.

Brown, A., Barnett, T., Pollard, G., Lisle, I. & Pollard, G. (2008). The characteristics of various men's doubles scoring systems. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 52-59.

Byun, K. (2000). Teaching statistics through sport. In: G. Cohen & T. Langtry (eds.). Mathematics and Computers in Sport. Sydney: University of Technology, 58-65.

Clarke, S. (1992). Computer and human tipping of AFL football - a comparison of 1991 results. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 72-78.

Clarke, S. (1994). An Adjustive Rating System for Tennis and Squash Players. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 41-48.

Coggill, W. (1994). The mathematics of bicycling. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 21-26.

Coggill, W.. (1998). The mathematics of bicycling part II. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: MathSport, 111-117.

Cohen, G. (2000). An annotated bibliography of mathematical articles on sport amenable to elementary university mathematics teaching. In: G. Cohen & T. Langtry (eds.). Mathematics and Computers in Sport. Sydney: University of Technology, 86-94.

Cohen, G. (2002). Cricketing Chances. In: G. Cohen & T. Langtry (eds.). Mathematics and Computers in Sport. Sydney: University of Technology, 1-13.

Cross, R. (2000). Mathematics and physics in ball sports. In: G. Cohen & T. Langtry (eds.). Mathematics and Computers in Sport. Sydney: University of Technology, 1-8.

Croucher, J. (1992). Winning with Science. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 1-6.

Croucher, J. (1998). Using Computers and Scientific Method to Determine Optimal Strategies in Tennis. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: MathSport, 1-7.

Davids, K. & Button, C. (2004). Variability and constraints in dynamical movement systems. In: H. Morton & G. Ganesalingam (eds.). Mathematics and Computers in Sport. Palmerston North: Massey University. 1-12.

de Mestre, N. (1990). The Mathematics of Projectiles in Sport. Queensland: Cambridge University Press.

de Mestre, N. (2008). A wave-catching model for bodysurfing. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 151-155.

Duckworth, F. & Lewis, T. (1996). A fair method for resetting the target in interrupted one-day cricket matches. In: N. de Mestre & K. Kumar (eds.). Mathematics and Computers in Sport. Queensland: MathSport, 7-12.

Gray, S. & Tuan, L. (2002). How to fix a one-day international cricket match. In: G. Cohen & T. Langtry (eds.) Mathematics and Computers in Sport. Sydney: University of Technology, 14-31.

Green, R. (2004). Augmented reality user interfaces for biomechanical data overlays. In: H. Morton & G. Ganesalingam (eds.) Mathematics and Computers in Sport, Palmerston North: Massey University, 153-164.

Heazlewood, I. & Lackey, G. (1996). The use of mathematical models to predict elite athletic performance at the Olympic games. In: N. de Mestre & K. Kumar (eds.). Mathematics and Computers in Sport. Queensland: MathSport, 74-79.

Heazlewood, I. & Lackey, G. (1998). Mathematical models that predict performance decline in elite veteran athletes 30-90 years in the sprint, distance and jump event. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: MathSport, 41-48.

Hammond, J. & de Mestre, N. (2008). Developments in mathematics and sport in Australia. In: Y. Jiang, A. Baca & H. Zhang (eds.) Computer Science in Sports, Liverpool: World Academic Press, 85-90.

Hammond, J. & Smith, T. (2006). Low compression tennis balls and skill development. In: J. Hammond and N. de Mestre (eds.). Mathematics and Computers in Sport. Coolangatta: MathSport, 273-282.

Harman, C. (1998). "Who's on first!" "What?" "What's on second!" and how 'what' got there on an optimal base running path. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: MathSport, 91-98.

Hayes, M. & Norman, J. (1992). Strategy in fell running: an analysis of the Bob Graham round. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 14-22.

Hoare, D. (1994). Sport Search - a sports counselling program for students. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 100-107.

Hoffman, D. (1992). A Team Rating System. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 7-12.

Hubbard, M. (1994). Simulating sensitive dynamic control of a bobsled. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 1-5.

Hughes, M. (2004). Notational analysis – a mathematical perspective. In: H. Morton & G. Ganesalingam (eds.). Mathematics and Computers in Sport. Palmerston North: Massey University, 13-47.

Hughes, M. D., Hughes M. T. & Behan, H. (2007). The evolution of a computerised notational analysis through the example of racket sports. International Journal of Sports Science and Engineering, 1:3-28.

Janssen, I. Sachlikidis, A. & Hunter, A. (2008). The accuracy of intra-stroke velocity using a GPS-based accelerometer in kayaking. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 156-163.

Johnston, M. (1992). An analysis of scoring policies in one day cricket. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 42-49.

Kumar, K., Ganesalingam, S. & Ganesh, S. (1996). Statistical analysis of horse racing data. In: N. de Mestre & K. Kumar (eds.). Mathematics and Computers in Sport. Queensland: MathSport, 94-99.

Lames, M. (2006). Modelling the interactions in game sports – relative phase and moving correlations. In: J. Hammond & N. de Mestre (eds.). Mathematics and Computers in Sport. Coolangatta: MathSport, 29-34.

Lewis, T. & Duckworth, F. (1998). Developments in the Duckworth Lewis method of target resetting in one day cricket matches. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: MathSport, 83-98.

Lewis, T. (2008). Use of the Duckworth/Lewis methodology to provide additional benefits for one-day cricket. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 8-13.

Lisle, I., Pollard, G. & Pollard, G. (2008). Improved scoring systems for a contest between tow tennis teams. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 37-44.

MathSport (2008). Special interest group of ANZIAM. http://www.anziam.org.au/Mathsport,

Morgan, S. (2006). Synergy in Sport – directions, convergence and opportunity in high performance sport. Presentation at the *Eighth Australasian Conference on Mathematics and Computers in Sport MathSport* (ANZIAM) Coolangatta, July 2006.

Morton, H. (1996). An analysis of world records for races run entirely in lanes. In: N. de Mestre & K. Kumar (eds.). Mathematics and Computers in Sport. Queensland: MathSport, 1-6.

Noble, D. (1996). Formation of netball teams for a series of trial matches. In: N. de Mestre & K. Kumar (eds.). Mathematics and Computers in Sport. Queensland: MathSport, 35-41.

Norman, J. & Clarke, S. (2004). Dynamic programming in cricket: batting on a sticky wicket. In: H. Morton & G. Ganesalingam (eds.) Mathematics and Computers in Sport, Palmerston North: Massey University, 226-232.

Norman, J. & Hayes, M. (2006). Naismith's rule and its variants. In: J. Hammond & N. de Mestre (eds.). Mathematics and Computers in Sport. Coolangatta: MathSport, 68-72.

Norton, P. (1994). Softball statistics. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 80-85.

Norton, P. (1996). The "leveller system" tennis tournament. In: N. de Mestre & K. Kumar (eds.). Mathematics and Computers in Sport. Queensland: MathSport, 50-55.

Norton, P. (2000). Home ground advantage in the Australian Netball League. In: G. Cohen & T. Langtry (eds.). Mathematics and Computers in Sport. Sydney: University of Technology, 168-173.

Norton, P. & Clarke, S. (2002). Serving up some grand slam tennis statistics. In: G. Cohen & T. Langtry (eds.). Mathematics and Computers in Sport. Sydney: University of Technology, 202-209.

Norton, P. (2004). Is there a difference in the predictability of men's and women's basketball matches. In: H. Morton & G. Ganesalingam (eds.) Mathematics and Computers in Sport, Palmerston North: Massey University, 233-241.

Ovens, M. (2004). If it rains, do you still have a sporting chance?. In: H. Morton & G. Ganesalingam (eds.) Mathematics and Computers in Sport, Palmerston North: Massey University, 242-252.

Ovens, M. & Bukiet, B. (2006). A mathematical modelling approach to one-day cricket batting orders. In: J. Hammond & N. de Mestre (eds.). Mathematics and Computers in Sport. Coolangatta: MathSport, 58-67.

Patrick, J. (1992). The marriage of mathematics and computer technologies for sport improvement. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 31-36.

Phatarfod, R. (1996). Betting strategies in horse races. In: N. de Mestre & K. Kumar (eds.). Mathematics and Computers in Sport. Queensland: MathSport, 100-105.

Pollard, G. & Noble, K. (2002). A solution to the unfairness of tie-break game when used in tennis doubles. In: G. Cohen and T. Langtry (eds.). Mathematics and Computers in Sport. Sydney: University of Technology, 140-145.

Pollard, G. & Pollard, G. (2008)a. Four ball best ball 1. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 208-215.

Pollard, G. & Pollard, G. (2008)b. Four ball best ball 2. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport,216-223.

Renshaw, I. (2008). Performance and learning of motor skills: a constraints-led perspective for studying human movement systems. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 83-87.

Scott, J. & de Mestre, N. (1994). Direct golf putting dynamics and strategies. In: N. de Mestre (ed.). Mathematics and Computers in Sport. Queensland: Bond University, 273-282.

Seyfried, D. (2006). The use of mathematics and computer science in the assignment of optimal individual stroke rate parameters in elite swimmers. In: J. Hammond & N. de Mestre (eds.). Mathematics and Computers in Sport. Coolangatta: MathSport, 12-18.

Siff, M. (2000). Biomechanics as an erogenic aid in strength training. In: G. Cohen & T. Langtry (eds.). Mathematics and Computers in Sport. Sydney: University of Technology, 9-27.

Silk, A., Hammond, J. & Weatherby, R. (2006). Resting toucher: a time and motion analysis of elite lawn bowls. In: J. Hammond & N. de Mestre (eds.). Mathematics and Computers in Sport. Coolangatta: MathSport, 19-28.

Smith, C., Gilleard, W., Hammond, J. & Brooks, L. (2006). The Application of an exploratory factor analysis to investigate the inter-relationships amongst joint movements during performance of a football skill. In: J. Hammond & N. de Mestre (eds.). Mathematics and Computers in Sport. Coolangatta: MathSport, 283-292.

Stefani, R. (2002). The fundamental nature of differential male/female world and Olympic winning performances, sports and ratings systems. In: G. Cohen & T. Langtry (eds.). Mathematics and Computers in Sport. Sydney: University of Technology, 32-48.

Stern, S. (2008). Ranking international limited-overs cricket teams using a weighted, heteroskedastic logistic regression with beta distributed outcomes. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 14-22.

Tonkes, E. (2008). The influence of crew weight on sailing performancein taipan catamarans. In: J. Hammond (ed.). Mathematics and Computers in Sport. Coolangatta: MathSport, 164-167.

LONG-DISTANCE RELATIONSHIPS: POSITIONAL CLASSIFICATION OF AUSTRALIAN FOOTBALL LEAGUE PLAYERS

Jonathan Sargent [a] and Anthony Bedford [a]

[a] *School of Mathematical & Geospatial Sciences*
RMIT University
[b] *Corresponding author: jonathan.sargent@student.rmit.edu.au*

Abstract

This research demonstrates how Australian Football League (AFL) players can be accurately and efficiently classified into four recognised playing positions (Defence; Midfield; Forward; Ruck) after each match, using only a handful of collected game-related performance variables. By maximising the Mahalanobis distance between a linear combination of thirteen performance variables and their respective centroids, 7,744 individual player cases in the 2009 AFL season are assigned to one of the four positions, without any prior knowledge of that player's movement within the match. Once the discriminant functions have been developed, Bayesian probabilities are then calculated to highlight each player's level of activity across the four positions in each match. This information is crucial when developing a set of position-dependent rules with which to measure AFL player performance. The research then progresses to intra-position analysis where each player is further classified based on the Squared Euclidian distance between position-specific elements derived from that player's performance covariance matrix (PCM). A case study details how forwards can be segmented into *discrete* and *continuous* playing roles based on the distances between covariance couplets. This information is of high importance for coaching staff and pundits alike as post-match deductions can be made, not only about a player's influence on the match, but also a player's influence within each position. An appealing aspect of the research is that only a few simple game-related statistics are required to gauge a player's positional performance, without having to resort to audio-visual tools and complex mapping.

Keywords: Classification, Mahalanobis distance, Euclidian distance, Covariance matrix

1. INTRODUCTION

Australian Rules football, or AFL, is a game played between two teams with twenty-two players each; a regular season consisting of sixteen teams playing twenty-two matches. Eighteen players per team are on the field at any one time – six forwards, six defenders, four midfielders and two "ruckmen", responsible for giving their team advantage at "ball-ups" (similar to a jump ball in basketball), with four reserves on an interchange. The ultimate goal of the game is to kick an oval ball between two taller sets of posts at either end of an elliptic ground; a goal is worth six points. The team with the most points at the end of the match is declared the winner. The dynamics of the game are similar to world football (soccer), the main exception being that AFL players are permitted to you use their hands to punch the ball (handball) to the advantage of another player.

The continuous nature of AFL makes accurate statistical analysis of team and in particular player performance exceedingly difficult. Oliver (2004) makes mention of the relative ease of analysing discrete games such as baseball due to the game's slow pace, incremental progress towards score advantage (runs) and limited interaction between players. He goes on to describe (American) football, from an analytical perspective as definitely more elegant and complex than either of basketball or baseball. AFL, it can be argued is more complex than all three of these sports. Of the world's most recognised football codes (world football or soccer, rugby and gridiron), AFL boasts the most number of players on the field at any one time (36 with a further four on the bench per side for rotations).

Players are selected to play in particular defensive, attacking or midfield positions on the ground, however the frenetic nature of the game demands that a player must operate in different zones at different match stages. One flaw in the measurement of AFL player performance is the assumption that a recognised positional player will play the entire match in his expected position. Quantitative analysts may assume prior to a match that an established forward will play entirely at the attacking end of the ground, however he may play a temporary defensive role which demands an additional set of rules by which to measure his performance in the defensive position. Without match vision, how is it possible to realise, let alone measure this change in position?

This paper demonstrates that by maximising the Mahalanobis distance between thirteen recorded AFL performance variables and their positional centroids, it is possible to quickly and efficiently allocate each player to one of four positions [Defence (D); Forward (F); Midfield (M); Ruck (R)]. Bayesian probabilities are then applied to establish a player's "time spent" in each of the four positions in each of his matches. Analysis is carried out on every player's m x X_i performance covariance matrices (PCM) from the 2009 season where X_i is performance variable i = 1,..,13 for match m. Intra-position characteristics are then examined using Squared Euclidian distance to further demonstrate the differing roles assumed in the four groups. A case study is provided using the covariance matrices of players allocated to the forward group.

Classification analysis is not uncommon in team sports. Mahalanobis distances are used effectively by Chatterjee and Yilmaz (1999) to observe differences in the performance characteristics of MVP players in the NBA. Sampaio et al (2006) employ discriminant analysis to maximize the average dissimilarities in game statistics between guards, forwards and centres in the National Basketball Association (NBA). Also with the aid of discriminant analysis, Fratzke (1976) was able to determine basketball player ability and position using varying biographic data, while Marelic et al (2004) observed "block" and "spike" in volleyball were the most important predictors of team success. Pyne et al (2006) concluded fitness assessments at the AFL draft involving statistical analysis on physical qualities such as height, mass and agility were useful in determining future player position.

The information drawn from this research becomes important for coaching staff and pundits alike as immediate post-hoc deductions can be made, not only about a player's influence on the match, but also a player's influence within each position. Player rating systems become more accurate as performance variable weights can be adjusted to reflect the relative influence of covariates in the position in which a player is classified (Sargent and Bedford, 2007). Another appealing aspect of the research is that only a handful of simple game-related statistics are required to gauge a player's positional movements; no prior knowledge or vision of an AFL match is required. This is particularly advantageous when analysing previous seasons' match data.

2. METHODS

i. Classification by performance

Hughes and Bartlett (2002) identify performance as any combination of quantifiable variables within a sporting match that, when aggregated constitute team play. Moreover, they discuss the concept of notational analysis, or the performance of a team or its individual members based on "open skills" (kicks, goals etc). The combination of game-related skills is an important determinant in classifying team success (Ibanez et al, 2009) as well as in the measurement of individual performance in team sports (Koop, 2002). This paper introduces an important concept for improving quantitative estimates of team player performance by assigning each player, post-match to one of k game-related positions, where k depends on the sport in focus. The classification method outlined in this paper made it possible to assign each AFL player to the group vector $Pos = k = [D,F,M,R]$ by linearly combining thirteen recognised AFL performance variables, X_i: Kick (KCK); Mark (MRK); Handball (HBL); Handball Receive (HBR); Inside 50 (I50); Rebound 50 (R50); Goal (GLS); Tackle (TKL); Clearance (CLE); Loose Ball Get (LBG); Hard Ball Get (HBG); Spoil (SPL); Hit-out (HIT). Each variable was entered into a stepwise model to arrive at k – 1 discriminant functions:

$$d_{up} = b_o + \sum_{i=1}^{13} b_i X_{imp} \qquad (1)$$

where d_{up} is the u^{th} discriminant function for player p, X_{imp} is the value of player p's performance variable i after match m, b_o is a constant and b_i are discriminant coefficients selected in the first discriminant function to maximise the Mahalanobis distance between the four positional centroids in k. The second discriminant function is selected so as to be orthogonal to the first, and the third discriminant function orthogonal to the second (Johnson & Wichern, 2007). Each player was assigned to the position to which his Mahalanobis distance from the positional centroids was the smallest (Sampaio et al, 2006). The Mahalanobis distance is a measure of distance between two points in the space defined by two or more correlated variables, and is in some sense a multidimensional *z*-score (James, 1985) measured by:

$$D_m(X) = \sqrt{(X-\mu)' S^{-1} (X-\mu)} \qquad (2)$$

where $X = (X_1,\ldots,X_{13})$, $\mu = (\mu_1,\ldots,\mu_{13})$ and S is the common covariance matrix (see (5)).
The classification judgements were also supported on the values of the overall structure coefficients b_i; higher values were better contributors to the classification process (Sampaio et al, 2006). Table 1 displays the overall model classification coefficients by position, with (*) indicating the strongest discriminatory positional predictors.

X_i	Position *k*			
	D	**F**	**M**	**R**
KCK	0.431	0.460	0.366	0.320
HBL	0.414	0.386	0.449	0.402
GLS	0.453	1.206 *	0.532	0.683
TKL	0.513	0.532	0.620 *	0.333
HIT	0.036	0.048	-0.045	1.386 *
R50	0.516	0.140	0.228	0.128
I50	0.099	0.253	0.230	-0.028
CLE	-0.600	-0.609	-0.337	-0.464
HBG	0.056	0.085	0.227	0.015
LBG	0.148	0.136	0.233	0.020
HBR	-0.169	-0.248	-0.083	-0.216
SPL	0.935 *	0.552	0.389	0.498
Constant	-6.756	-6.356	-7.030	-15.74

Table 1: Classification function coefficients by position

Mark (MRK) was not a significant predictor in the classification process and was removed from the model. Figure 1 displays how the first (d_1) and second (d_2) discriminant functions from our model have classified a random sample of 200 player matches from Round 22 into the four positions.
Higher values for d_1 were associated with classification of the ruck position (it is safe to conclude the ruck position was the most accurately classified based on their unique role), while higher d_2 values classified the midfield positions. Defenders were best classified by negative values of d_2 while forwards fell at the intercept.

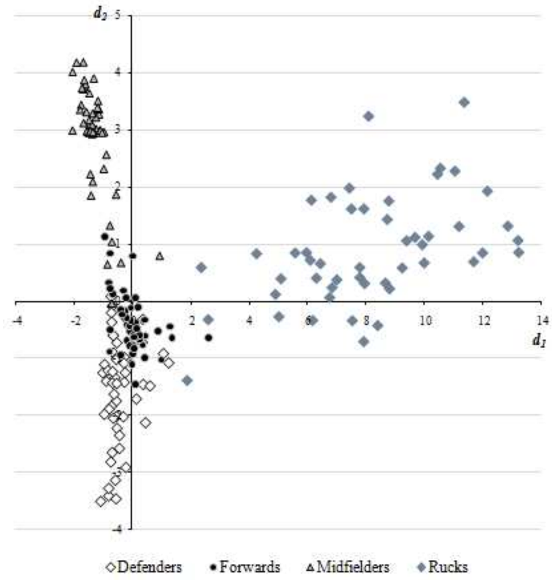


Figure 1: Canonical discriminant function graph of *n*=200 players

ii. Posterior Probabilities

Where much classification research is content to draw conclusions from the predictive path to the classified groups, a measure of classification assurance is an interesting and important extension (James, 1985). For the purposes of AFL player performance measurement, a player may be correctly assigned to a position, but how realistic is it to assume he played the entire match there? To mathematically ascertain this knowledge, posterior probabilities were calculated for each player's defensive, forward, midfield and ruck roles in each match. Once the discriminant functions (1) have been calculated these probabilities are produced by:

$$P(Pos_k \mid x) = \frac{\exp[\max(d_u) - d_k] / P(Pos_k)}{\sum_{j=1}^{4} \exp[\max(d_u) - d_j] / P(Pos_j)} \qquad (3)$$

where: $\sum_{k=1}^{4} P(Pos_k \mid x) = 1$

and: $P(Pos_k)$ is the prior (pre-match) probability of playing in position k.
The research found significant correlations between the posterior probabilities of classification to positions D, F and M and the number of disposals (KCK + HBL) the player achieved in the defensive,

forward and midfield zones respectively. In Table 4, Jonathan Brown of the Brisbane Lions is abundantly classified as a forward for the first 5 rounds of the 2009 season, but in a smaller capacity for Round 4 where he played a greater midfield role than the other matches [$P(M \mid x) = 38.12\%$]. The majority of his disposals are in the forward zone for higher $P(F \mid x)$ but for Round 4, his Forward to Midfield disposal ratio (For:Mid) decreases below 1.0 in line with his decrease in $P(F \mid x)$ and increase in $P(M \mid x)$. Conversely, Brown's most prominent forward performance [$P(F \mid x) = 99.31\%$] in Round 3 returned his highest For:Mid ratio for the 5 rounds. This data supports our discriminant model selection.

iii. Intra-Position Analysis

With the discriminant model accurately classifying each player into position by his accumulation of performance variables, the research benefitted from further investigation into the performance characteristic differences exhibited by players within the established positions. With this approach it was worthwhile considering the variability in each player's performance variables (Chatterjee and Yilmaz, 1999). Considering the pair of performance variables $\{X_i, X_j\}$ for player i to match m, the covariance $Cov(X_i, X_j)$ is a measure of the linear coupling between X_i and X_j (James, 1985). If entries in the column vector:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix} \quad (4)$$

are random variables, each with finite variance, then the covariance matrix Σ is the matrix whose (i, j) entry is the covariance:

$$\sum_{ij} = \mathrm{cov}(X_i, X_j) = E[(X_i - \mu_{ikm})(X_j - \mu_{jkm})] \quad (5)$$

where $\mu_{ikm} = E(X_i)$ in position k in match m, and $\mu_{jkm} = E(X_j)$ in position k in match m are the expected values of the i^{th} and j^{th} entry in (4). Incorporating the match mean vector in (5) for variable X_i rather than the league mean at Round n standardises the distances from the performance variable mean vectors for matches that may exhibit unusually high or low variable means, for example, wet weather having a negative impact on total disposals.

Chatterjee and Yilmaz (1999) favour the use of covariance matrices, rather than correlation matrices, in performance measurement, because they express variability in the performance variables' commonly used scales. With (5) covariance matrices were established at a league, position [from (1)] and player level (see Table 5) allowing analysis of matrix elements, for example the covariance between Kicks and Goals [KCKGLS]. Figure 2 illustrates how it is possible to compare sets of covariate couplets, for example, [KCKGLS, HBRGLS] hence, allowing analysis of four covariates in a two-dimensional space (Gordon, 1981). Moreover, by assessing the Squared Euclidean distances between these covariate couplets, the positions classified by (1) could be further segmented to enhance the knowledge of intra-position performance characteristics. The Squared Euclidean distance formula is defined as:

$$d_{ij} = \sum_{p=1}^{n} (x_{ip} - x_{jp})^2 \quad (6)$$

where: x_{ip} and x_{jp} denote the values taken by the p^{th} player on covariate couplet i and j respectively.

3. RESULTS

Having used (1) to classify all players after Round 22, a dissimilarity matrix was defined using (6) to determine robust classifiers within the forward position (Table 2). The largest distance (*) was between MRKGLS and HBRGLS, implying forwards could be classified into two further groups: *discrete play* forwards who predominantly kick goals after taking a mark (MRKGLS), and *continuous play* forwards who set up or kick goals through handball receives (HBRGLS).

	HBR-GLS	MRK-GLS	LBG-GLS	HBG-GLS	TKL-GLS
HBR - GLS	0.00				
MRK-GLS	335.74 *	0.00			
LBG-GLS	108.53	257.77	0.00		
HBG-GLS	72.67	229.84	48.06	0.00	
TKL-GLS	84.07	283.08	65.02	41.13	0.00

Table 2: Dissimilarity matrix of Squared Euclidean distance between forwards' $cov(X_i,X_j)$.

Figure 2 displays how all forwards are positioned after the final round in 2009 based on MRKGLS and HBRGLS. By maximising the Mahalanobis distance between the centroids for these covariate couplets using (2), it was possible to segment the forwards into sub-positions, F_1 and F_2, where F_1 contains the *discrete play* forwards and F_2 contains the *continuous play* forwards. A small cluster of players around [HBRGLS=0] and [MRKGLS=0], implying little variability through the season, proved difficult to classify. Recognised forwards however, show the highest variability and largest distance from the couplet centroids. Jonathan Brown could be considered the best *discrete play* forward based on his largest distance from the MRKGLS centroid. Akermanis could be considered the best *continuous play* forward resulting from his distance from the HBRGLS centroid. However, Brown and Akermanis are contrasting forwards, given the large distance in Table 3, measured by (6). This hypothesis is supported by Brown being considered a highly rated *key* forward and Akermanis, a highly rated *small* or *roving* forward by the AFL.

	16 Br Brown
1 12 St Riewoldt	2.933
2 16 Br Brown	0.000
3 12 Ri Richardson	4.897
4 25 Ca Fevola	7.708
5 23 St Koschitzke	10.814
6 16 Po Tredrea	14.671
:	:
44 21 Wb Akermanis	48.317

Table 3: Dissimilarity matrix of Squared Euclidean distance: Brown and Akermanis [MRKGLS, HBRGLS]

The right mixture of these two types of forwards in a team is important from a coaching perspective. Successful sides generally have two *key* forwards and at least two *small* forwards. This modelling can assist in the selection process.

4. CONCLUSIONS

The classification model described in this paper transforms a typical AFL match statistics sheet into a meaningful and accurate representation of the central and transitory responsibilities that different players assume within a match. Positional classification is important for player rating models, while intra-position distance analysis allows coaching staff to assess the optimal structure for their defensive, attacking and midfield zones. This classification model is accurate whilst only requiring a small post-match set of game-related statistics for each player.

Acknowledgements

We wish to thank Prowess Sports for the use of their data.

References

Chatterjee, S. & Yilmaz, M.R. (1999). The NBA as an evolving multivariate system. The American Statistician, 53, 257-262.

Fratzke, M.R. (1976). Discriminant analysis of basketball skill tests and biographic data. Paper presented at the American Alliance for Health, Physical Education and Recreation Central District Meeting (March 19-20).

Gordon, A.D. (1981). Classification, London: Chapman and Hall.

Hughes, M.D. & Bartlett, R.M. (2002). The use of performance indicators in performance analysis. Journal of Sports Sciences, 20, 739-754.

Ibanez, S.J., Garcia, J., Fue, S., Lorenzo, A. & Sampaio, J. (2009). Effects of consecutive basketball games on the game-related statistics that discriminate winning and losing teams. Journal of Sports Science and Medicine, 8, 458-462.

James, M. (1985). Classification Algorithms, New York: John Wiley & Sons.

Johnson, R.A. & Wichern, D.W. (2007). Applied Multivariate Statistical Analysis (6e), New York: Prentice-Hall.

Koop, G. (2002). Comparing the performance of baseball players: a multiple output approach. Journal of the American Statistical Association, 97, 710-720.

Marelic, N., Resetar, T. & Jankovic, V. (2004). Discriminant analysis of the sets won and the sets lost by one team in A1 Italian Volleyball League – A case study. Kinesiology, 36, 75-80.

Oliver, D. (2004). Basketball on Paper, Washington, D.C.: Brassey's Inc.

Pyne, D., Gardner, A., Sheehan, K. & Hopkins, W. (2005). Fitness testing and career progression in AFL football. Journal of Science and Medicine in Sport, 8, 321-332.

Sampaio, J., Janeira, M., Ibanez, S.J. & Lorenzo, A. (2006). Discriminant analysis of game-related statistics between basketball guards, forwards and centres in three professional leagues. European Journal of Sport Science, 6, 173-178.

Sargent, J. & Bedford, A. (2007). Calculating Australian Football League player ratings using a T4253H Elo model. Proceedings of the first IMA International Conference on Mathematics in Sport. Manchester, UK. 186-191.

Player	Round	Classified Position	$P(D \mid x)$	$P(F \mid x)$	$P(M \mid x)$	$P(R \mid x)$	Disposals			
							Defence	Forward	Midfield	For:Mid
16 Br Brown	1	F	0.86%	80.95%	17.67%	0.51%	0	8	7	1.14
16 Br Brown	2	F	0.19%	94.18%	5.46%	0.17%	2	10	6	1.67
16 Br Brown	3	F	0.02%	99.31%	0.62%	0.05%	0	8	4	2.00
16 Br Brown	4	F	4.62%	56.45%	38.12%	0.82%	1	4	8	0.50
16 Br Brown	5	F	1.07%	81.69%	16.76%	0.48%	1	5	4	1.25

Table 4: Classification posterior probabilities for Jonathan Brown

	KCK	**HBL**	**MRK**	**HBR**	**GLS**	**TKL**	**HIT**	**I50**	**R50**	**CLE**	**HBG**	**LBG**	**SPL**
KCK	28.62												
HBL	2.97	5.13											
MRK	19.22	0.92	16.26										
HBR	-2.54	1.54	-3.40	3.73									
GLS	9.38	-0.07	7.13	-1.94	5.89								
TKL	-2.67	-0.07	-2.33	1.28	-1.69	1.23							
HIT	1.15	0.58	0.15	0.39	-0.06	0.11	0.65						
I50	4.24	1.72	2.46	0.31	0.48	-0.14	0.58	2.06					
R50	-1.05	-0.21	-0.91	0.34	-0.59	0.24	-0.01	-0.13	0.17				
CLE	1.41	0.79	0.24	0.33	0.43	0.08	0.00	0.38	0.04	0.83			
HBG	1.76	0.28	0.79	-0.52	0.85	-0.19	0.23	0.36	-0.09	0.22	1.10		
LBG	5.04	1.93	2.27	0.00	0.72	0.17	0.38	0.69	-0.22	0.55	0.48	3.03	
SPL	0.47	0.46	0.21	0.16	0.19	0.03	-0.06	0.03	0.03	0.32	-0.10	0.23	0.29

Table 5: Performance covariance matrix (PCM) for Jonathan Brown

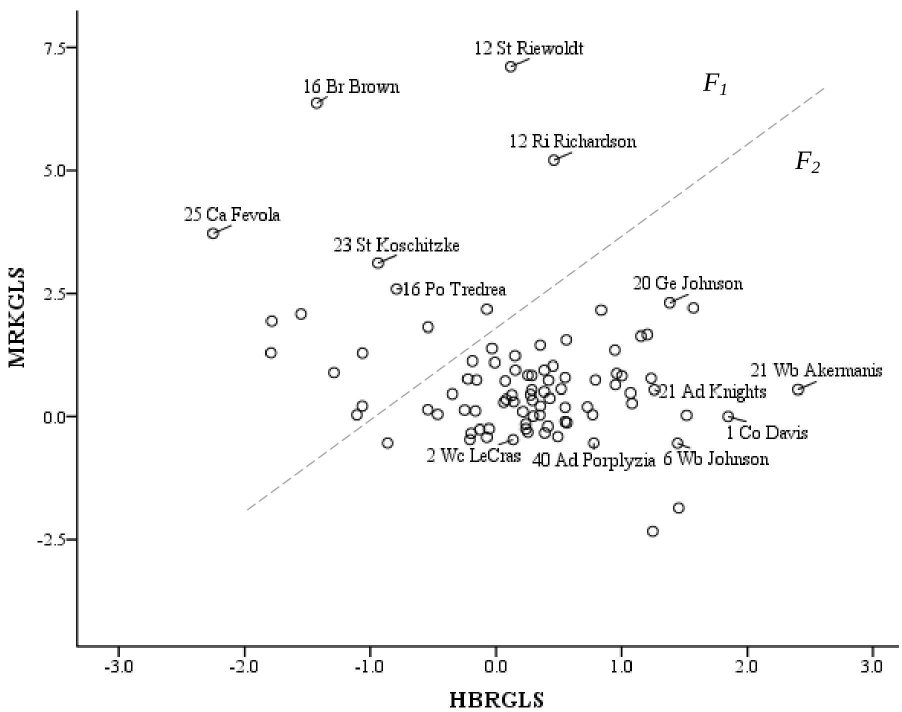


Figure 2: Reclassification of forwards using Squared Euclidean distance

REVISITING THE DRAFT: A REVISED PROBABILITY BASED APPROACH FOR ALLOCATING DRAFT SELECTIONS IN AUSTRALIAN RULES FOOTBALL

Adrian J. Schembri[a, d] and Anthony Bedford[a]

[a] *School of Mathematical and Geospatial Sciences, RMIT University*
[d] *Corresponding author: adrian.schembri@rmit.edu.au*

Abstract

The existing method of allocating draft selections in the AFL is based on inverse ladder position at season's end. A criticism levelled at this system is that it provides teams who are unlikely to figure in the finals with an incentive to lose, with greater rewards being provided to teams who win fewer matches. In 2006, we proposed a probability-based system that allocates a score when a team wins an 'unimportant' match. Calculation of unimportance was based on the likelihood of a team making the final eight following each round of the season. A limitation of this model was that higher picks were awarded to teams who won a string of matches late in the season and just missed out of playing in the finals. Furthermore, teams who won very few matches and finished in the bottom ladder positions were awarded few draft allocation points, and thus were unlikely to receive high draft selections. In this paper, further refinement of the original model is undertaken to address these shortcomings. Draft point allocations are moderated based on the quality of opposition being played and the number of matches the team has won during the season to date. Using this revised approach, we simulated 100,000 seasons based on actual season data for the 2009 AFL home-and-away season. We investigate the distribution of draft picks awarded to each ladder position, and highlight teams who are more or less likely to be awarded top draft picks under the revised system.

Keywords: AFL, Draft, Importance, Incentives, Probability

1. INTRODUCTION

Several issues remain within the AFL code that generate controversy year after year. These include an uneven fixture, frequent rule changes, and a draft system that provides teams who are unlikely to figure in the finals with an incentive to lose. In 2006, we proposed a probability-based system that rewards teams for winning 'unimportant' matches (akin to Carl Morris' definition of importance; Morris, 1977). Calculation of unimportance was based on the likelihood of a team making the final eight following each round of the season (Bedford & Schembri, 2006). Since this model was first published, the system utilised by the AFL to allocate draft selections has remained unchanged, with first round draft picks allocated on the basis of reverse ladder position at the end of the season. Subsequent rounds of the draft replicate the first round. Up until 2006, teams who won fewer than five matches received a priority pick that was awarded before pick 1, and thus a team who finished last with fewer than five wins would essentially receive the first two picks in the draft. To address such as high incentive being available for teams to not win five matches, the AFL revised the priority pick system to only grant teams a priority pick if they won fewer than five matches for two consecutive seasons (www.afl.com.au).

Whilst it was hoped that implementation of this new priority pick system would eradicate much of the controversy that has surrounded the AFL draft, disillusionment remains high in AFL channels given the events of the 2007, 2008, and 2009 seasons. In the final round of the 2007 season, Melbourne played Carlton in a match that triggered significant debate within the AFL community. A win to Melbourne would see the club receive fourth pick in the draft rather than second pick, and the club would also lose the right to a potential priority pick in 2008. The stakes were much higher for Carlton, with a win costing them a priority pick and first pick in the draft. The eventual result was a win to Melbourne by 33 points, a result which was met with despair rather than elation by many Melbourne club members and supporters. In the final Round of 2007, Richmond entered a match against St Kilda knowing that a win would cost the club Pick 2. Months after the game, Richmond coach Terry Wallace conceded that he did little to coach the Richmond side to win, stating that “It was a no-win situation for everyone in the coach’s box. We decided the best way to operate was just to let the players go out. I didn’t do anything. I just let the boys play” (Stevens & Ralph, 2009). By 2009, the Melbourne Football Club could win only one of its final seven matches in order to receive a priority pick and the first pick in the national draft. At the end of the season, Melbourne forward Russell Robertson commented “I know this whole tanking vibe has disgruntled a few players at the Melbourne Football Club. You can’t blame the coach. It is more just the way the AFL is at the moment with the [priority pick] systems that are in place. I’m not saying we tanked, I’m just saying players were played out of position…I don’t think it was in our best interests to win” (Stubbs, 2009).

The AFL code is not alone with respect to controversy surrounding player draft selections. The American National Basketball Association (NBA), National Hockey League (NHL), Major League Baseball, and the National Football League (NFL) have encountered criticism regarding the draft systems utilised in their respective sports. Refer to Bedford and Schembri (2006) for a review of these draft systems. Although reverse ladder position and lottery based systems dominate player draft systems worldwide, neither system is considered ideal (Bedford & Schembri, 2006), yet despite the need for a revised system, a paucity of research existed in this area in 2006, and very few studies have been published in more recent years. In one study, Gold (2010) suggested that losing can assist teams in the NHL with an opportunity to acquire more tickets for the draft lottery, thus encouraging supporters to cheer for opposition teams. This author employed a mathematical elimination technique to enforce a competitive atmosphere on teams who performed poorly, whilst those teams eliminated earliest had more opportunity to earn higher draft picks.

Borland, Chicu, and Macdonald (2009) presented on an exploratory study into trends in the AFL prior to and following the introduction of the current draft system. Results indicated that there was no evidence that clubs had engaged in ‘tanking’ (a colloquial term used to describe a club that is losing intentionally). Several reasons were highlighted for this finding, including the few benefits that exist for clubs who engage in tanking, and that few opportunities exist for tanking to occur in AFL football. Despite these findings, controversy and general disenchantment remains in the AFL community, with journalists, players, coaches, and general supporters of the game expressing their concerns in recent years. The introduction of two new teams into the AFL in 2011 and 2012 will result in draft concessions being available to the new teams, and therefore even fewer incentive to tank exists for AFL clubs. Many have argued that this presents the AFL with a period of respite, and an opportunity to revise the current draft system, whilst others have suggested that the tanking issue will escalate following introduction of the new clubs (Leech, 2008).

In this paper, we build upon the draft model presented in Bedford and Schembri (2006). Several weaknesses of the original model are addressed, with a series of moderating variables being introduced to ensure that poorer teams receive adequate reward for winning unimportant matches. We will begin our work by defining the criteria used to develop our original model, and the modifications made for the revised system. Simulation data is then presented in consideration of the operational aspects of the revised model

2. METHODS

To begin, we present the mathematical and probabilistic elements that were fundamental to the original DScore model published in 2006.

A large amount of detail is given in our earlier work (Bedford and Schembri (2006)) on the finer details of the model. Here we summarise it briefly. At its heart, the model simply aimed to reward teams with a reduced probability of making the final eight with a higher draft score (DScore) if they win. Firstly, the projected wins (ParWins) to make the final eight is calculated, and from this the binomial distribution is used to determine each teams chance of making the final eight. The unimportance of a match is then calculated simply by

$$U_i(r) = 1 - b(ParWins_1; 22 - (r+1), p) \qquad (1)$$

where $b(x,n,p)$ is the discrete binomial distribution function. From this the DScore is given by the sum of the draft points reward for team i , so

$$DScore_i(r) = \sum_{i=7}^{r} U_i(r)\left(1 - P(F8 \mid r)\right) \qquad (2)$$

noting that points are only awarded if a team wins and $r > 6$. F8 is the number of wins needed to make the final 8, and r the round.

Although this model provided several positive steps in the development of the DScore draft system, several weaknesses were evident. Firstly, the top pick in the draft was frequently awarded to a team who, whilst unlikely to figure in the finals, won a string of matches at the end of the season. Whilst these teams typically finished between ninth and twelfth on the ladder, they were awarded the top draft pick due to the high number of unimportant matches that they had won. Another limitation of the model was that teams who performed poorly and won very few matches for the entire season were only able to obtain a maximum of one DScore point for every game that they did win. This limited their opportunity to climb the DScore ladder and receive a top five draft pick at the end of the season.

Moderator Variables in the 2010 DScore Model

In order to address these shortcomings, several moderating variables were introduced into the model to reduce the magnitude of reward for teams who were climbing the ladder late in the season, and also to amplify DScores for those teams who won very few matches throughout the season, particularly late in the season when there was little incentive to win. The following moderators were introduced into the revised model:

Quality of opposition measure.

To obtain a measure of the quality of opposition being played and in order to reward the poorly performing team accordingly, a moderator was introduced based on the ratio of the opposition team's pre-match percentage and the team in question's percentage prior to the match being played. As an example, Melbourne (16th) played Port Adelaide (9th) in Round 15 of the 2009 season (refer to Table 1).

Melbourne's percentage going into the game was 71.5% whilst Port Adelaide's was 92.1%. For Melbourne, dividing Port Adelaide's percentage by their own percentage results in a DScore scaling factor of 1.288, whilst for Port Adelaide, dividing Melbourne's percentage by their own percentage results in a DScore scaling factor of 0.776. Introduction of this variable enabled the DScore model to take the quality of the opponent into consideration for each round.

Position	Team	W	D	L	%
9	Port Adelaide	7	0	7	92.1
16	Melbourne	2	0	12	71.5

Table 1. Ladder position of Melbourne and Port Adelaide prior to Round 15, 2009.

Difference in wins when compared with 8th position.

To further unearth the team in question's likelihood of making the finals at the time of each match, and in essence, measuring the unimportance of each match, the ratio of the number of wins obtained by the team in eighth place and the number of wins obtained by the team in question was considered. The following equation was utilised to compute this:

$$\Omega = \frac{8^{th}\,Position\,Wins - Team\,i\,wins}{8^{th}\,Position\,Wins} \qquad (3)$$

For example, going into the Round 15 encounter with Port Adelaide, Melbourne had won two matches up to that point in the season. Going into the round, Essendon was in eighth place with seven wins. For Melbourne, the ratio of Essendon's wins to their own resulted in a scaling factor of 0.714 (that is, [7-2]/7 = .714). Relative to teams who had won more matches up to that point in the season, this scaling factor would further amplify Melbourne's Draft Point Reward (DPR) for the round if they were to beat Port Adelaide.

Moderating factor based on ladder position.

To ensure that teams in the bottom four positions of the ladder (13th, 14th, 15th and 16th) receive maximum reward for winning late in the season, and to reduce the reward received by teams in 12th position and above, a third scaling factor was introduced. Simply put, those teams who are placed in the bottom four positions of the ladder at the time they won a match receive their allocated DScore multiplied by one (that is, it remained unchanged), however those teams placed in 12th position and above have their allocated DScore for that round halved. This scaling factor was introduced to prevent teams who are of sufficient quality from receiving a top draft pick. For example, during the 2009 season, West Coast won 4 of its first 17 matches, and 4 of its last 5 matches. For this latter string of wins, the first two matches were won in 13th and 14th position on the ladder, however the final two were won from 12th position. By implementing the moderating factor based on ladder position, West Coast received their full DScore compliment for the first two wins, yet their DScore was multiplied by 0.5 for the wins where they were in 12th position, that is, when they were not one of the bottom four teams on the ladder.

Other moderating factors that were trialled.

Several alternative moderating factors were also trialled and not implemented into the final model. To ensure that the top four draft picks were more likely to be awarded to the bottom four teams, a trial simulation was undertaken whereby only teams who were in the bottom four positions on the ladder at the time they won a match would receive a DPR for winning the match. Simulation results revealed that this approach reduced the number of teams in contention for the top four draft choices, and also resulted in most teams not receiving a DPR for the entire season, given that only a handful of teams are actually in the bottom four between rounds eight and 22.

Introduction of a cumulative DScore was also considered whereby teams would be awarded all possible DScore points that had accumulated since their last win. For example, a team in 16th position who has not won a match for five rounds upon entering Round 20 would be rewarded with their DPR for Rounds 15, 16, 17, 18, 19 and 20. Simulation results indicated that this variable increased the volatility of the entire draft system, since lowly teams who had performed consistently (e.g., winning one in every four matches) were overhauled by a team who did not win many matches during the year, but won a match late in the season (e.g., did not win between rounds eight and 20, but won a match in Round 21).

The allocation of DPR based on a countback system at the end of the season was also contemplated by the authors however this had the potential to reduce the competitive nature of the DScore system throughout the season, given that teams would not be aware of their current standing at the end of each round.

3. RESULTS

We begin our review of the revised DScore model by examining the 2009 season in detail. Where possible, the revised model is compared with the current AFL draft system and the 2006 DScore model.

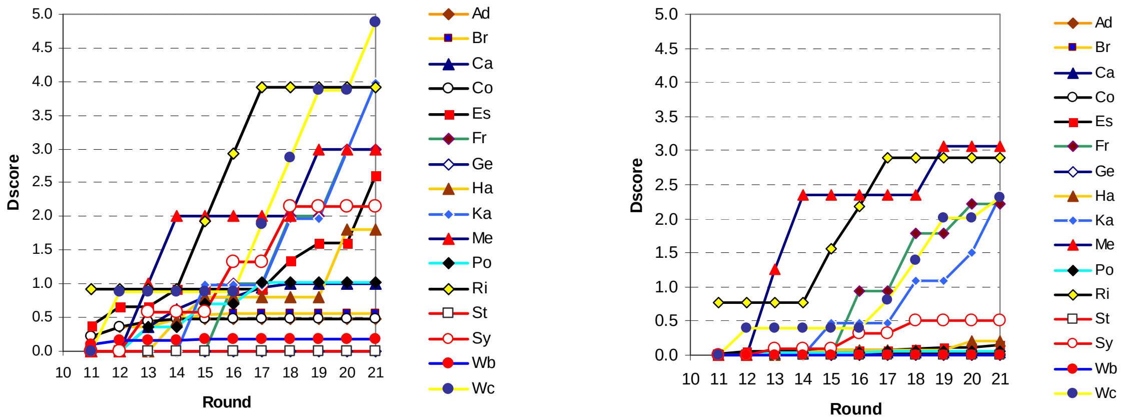
The 2009 AFL Season

In 2009, the AFL season was highly competitive, with the makeup of the finals uncertain until the final round. With five rounds remaining in the season, eleven teams were still vying for a finals position. In addition, only six premiership points (or one and a half wins) separated 12th and 16th on the ladder, and thus the race for the wooden spoon (awarded to the team that finishes last) was still wide open. At seasons end, Melbourne finished last on the ladder with only four wins, and was granted a priority pick given that they had only won three matches in 2008. Table 2 presents a summary of the 2009 AFL ladder at the end of the home-and-away season, with results of the AFL Draft system, the 2006 DScore model, and the revised 2010 DScore model.

Under the 2006 DScore model, West Coast were awarded the top draft pick following a string of wins late in the season. A series of back-to-back wins was also achieved by North Melbourne, and therefore they received pick 2 under the 2006 model and pick 3 under the revised 2010 model. The bottom two teams won four and five matches respectively in 2009, with the majority of these wins being achieved in the second half of the season. Based on the DScore model being fundamentally based on poorly performing teams having an incentive to win, the newly introduced moderating factors had a desired effect in the 2009 season, with the bottom two teams being awarded picks 1 and 2 under the revised draft model. Under the revised system, North Melbourne and West Coast were rewarded substantially for their performances late in year, receiving picks 3 and 4.

The 2006 DScore model plot demonstrates the effects of teams winning a string of consecutive matches late in the season, with West Coast and North Melbourne rising to the top of the DScore ladder late in the season. This was not the case under the revised model, as only moderate increases in DScore rankings were evident late in the season.

This finding demonstrates that the revised model has the potential to take the quality of the team into consideration, by moderating the DScore value being awarded to teams who finish between 9th and 12th position on the ladder.



2006 DScore Model

2010 DScore Model

Figure 1. Progressive DScores for the 2009 AFL season for the 2006 and 2010 DScore Models.

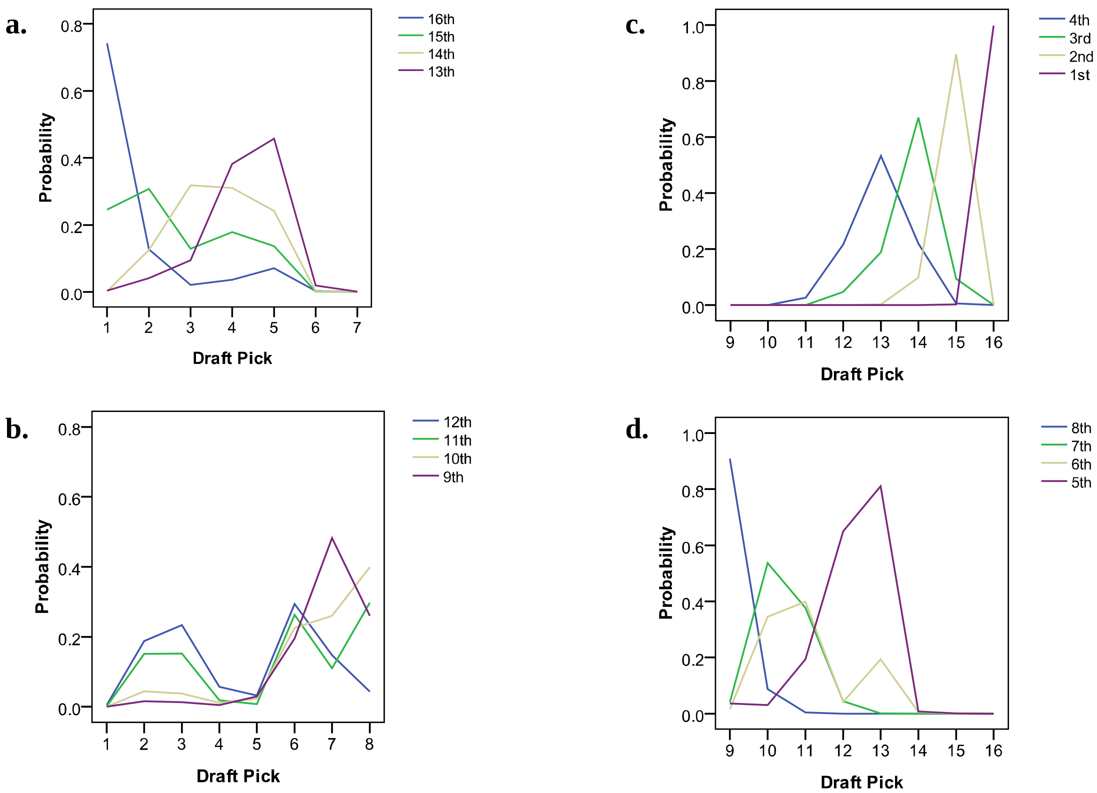


Figure 2. Probability distribution of draft choices for each ladder position at the conclusion of the 2009 AFL home-and-away season.

Team	W-D-L	%	AFL	2010 Model DScore	2010 Model Pick	2006 Model DScore	2006 Model Pick
St	20-0-2	155.7	16	0.0002	16	0.0000	16
Ge	18-0-4	127.4	15	0.0002	15	0.0000	15
Wb	15-0-7	122.6	14	0.0069	14	0.1775	14
Co	15-0-7	122.3	13	0.0196	13	0.4903	13
Ad	14-0-8	117.6	12	0.0233	12	0.4903	12
Bri	13-1-8	106.7	11	0.0291	11	0.5545	11
Ca	13-0-9	110.5	10	0.0404	10	0.9961	10
Es	10-1-11	97.8	9	0.1553	9	2.6049	9
Ha	9-0-13	92.5	8	0.2064	7	1.8065	7
Po	9-0-13	88.7	7	0.0618	8	1.0196	8
Wc	8-0-14	93.3	6	2.3197	4	4.8790	1
Sy	8-0-14	93.1	5	0.5136	6	2.1548	6
Ka	7-1-14	83.4	4	2.3219	3	3.9768	2
Fr	6-0-16	77.3	3	2.2111	5	3.0001	5
Ri	5-1-16	74.3	2	2.9037	2	3.9249	3
Me	4-0-18	74.7	1	3.0613	1	3.0002	4

Table 2. Draft Pick Comparison.
(W=win; D=draw; L=loss, %=score for team/score against team)

Simulation results based on the 2009 AFL season
As described in the methodology, a 100,000 iteration simulation was conducted on the 2009 AFL season. The purpose of the simulation was to examine the distribution of scores obtained under the revised DScore model, and the likelihood of teams finishing in different ladder positions receiving high or low draft choices. Figure 2 displays the probability distribution of obtaining each draft pick for every ladder position at the conclusion of the 2009 AFL home-and-away season.
As shown, the team that finished last on the ladder had a probability of .77 of obtaining the first draft pick, and a probability of .15 of obtaining the second draft choice. Ladder positions 13th to 15th had a high probability of obtaining a top five draft choice, however the team who finished 13th was most likely to receive pick 4 or 5. Part b of Figure 2 demonstrates that it was most probable for ladder positions 9th to 12th to receive picks 6 to 8, however teams who finished 11th or 12th regularly received picks 2 or 3. Parts c and d of Figure 2 indicate that teams who finish between 6th and 8th position receive an assortment of picks 9 to 11, whilst those teams who finish in ladder positions 1 to 5 have a very high probability of obtaining a draft pick that corresponds to their reverse ladder position (e.g., first position on the ladder receives pick 16). Figure 3 displays the cumulative probability of obtaining a top five draft pick for teams finishing in the bottom eight positions on the ladder, that is, those teams who did not qualify for the finals.

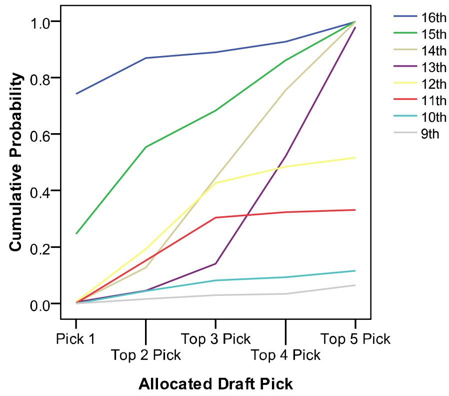


Figure 3. Cumulative probability of obtaining a top five draft pick for ladder positions 9 to 16.

Teams who finished between 13th and 16th had a very high probability (between .97 and .99) of obtaining a top 5 draft pick, however the order of the top 5 picks varied considerably. To illustrate, ladder positions 13 and 14 were unlikely to receive a top 2 pick (below .20), whilst positions 15 and 16 had a respective probability of .55 and .85 of receiving one of these picks. Of note, it was very unlikely that teams who finished in positions 9th through to 12th would receive a top two draft pick, however a top five pick was possible for these teams. In effect, the distribution of cumulative probabilities in this figure demonstrates that all teams in the bottom five positions on the ladder have great incentive to win matches late in the year under the revised DScore model, with the greatest incentive being for those teams in the bottom two positions on the ladder.

Distribution of actual DScores and difference scores for the 2009 AFL season
In addition to examining the distribution of ladder position and corresponding draft choice allocations, a review of the distribution of actual DScores is necessary. Examination of the difference in actual DScores enables an assessment of the number of teams who are in contention to receive draft selections. Table 3 presents mean and standard deviation scores for the DPR allocated to each team over the course of the 2009 AFL home-and-away season for the 100,000 iteration simulation.

DScore allocations decreased as ladder position increased, with the team finishing on top of the ladder having a mean DScore of .0002 and a standard deviation of 0.0000. All teams that finished in the bottom four positions on the ladder had a

AFL Pos.	Min.	Max.	Mean	SD
16	.4139	4.6186	2.8850	.5115
15	.1527	4.5644	2.5831	.6578
14	.5616	3.8516	2.1037	.3228
13	.2152	3.5411	2.0054	.3577
12	.0590	3.5476	1.4367	.9053
11	.0306	3.5405	.9770	.9538
10	.0302	3.5196	.4888	.6423
9	.0283	2.6995	.3699	.4530
8	.0284	2.3196	.1986	.1682
7	.0100	.3392	.0404	.0318
6	.0002	.1493	.0352	.0241
5	.0002	.2078	.0300	.0278
4	.0002	.0244	.0159	.0070
3	.0002	.0239	.0075	.0060
2	.0002	.0198	.0004	.0007
1	.0002	.0002	.0002	.0000

Table 3. Draft point allocation descriptive statistics for each ladder position following the 100,000 iteration simulation.

mean DScore between two and three points, and therefore each have the potential to receive the top draft choice. As would be expected based on the newly introduced scaling factors, a marked decrease in mean DScores was evident for teams who finished in 11th or 12th position, with a further decrease for those teams who finished in 10th position or above. It is likely that teams finishing in 11th and 12th position won a proportion of their matches whilst in the bottom four positions on the ladder, thus their DPR was somewhat higher than teams who finished higher on the ladder. However, winning a proportion of matches when outside the bottom four resulted in these teams having substantially lower DScores than teams in the bottom four ladder positions. This finding is further evidenced by the higher standard deviation scores for teams finishing in 11th or 12th, with results indicating standard deviations of almost one DScore point.

To further analyse the distribution in scores and the number of teams in contention to win the top draft pick, the difference between the top DScore (Rank 1) and DScore ranks 2 to 8 were computed, refer to Figure 4.

On average, difference scores between ranks 1 and 2 were approximately 0.6 DScore points, which is the equivalent to one win. This finding indicates that the second highest DScore rank could have potentially won the top draft pick in the majority of cases had the team won one more match. The difference in DScores between Rank 1 and ranks 2 and 3 were approximately one DScore point, which equates to between one and two wins. Teams between ranks 1 and 5 were clearly in contention for the top draft choice based on difference scores, however a substantial increase in difference scores was evident for ranks 6, 7, and 8, with difference scores exceeding 2.5 DScore points (equivalent to between three and four wins).

As such, teams who finish in the bottom five positions on the ladder are typically in contention for the top draft choice, however on average, this is not likely for teams with a DScore ranking from rank 6 and above.

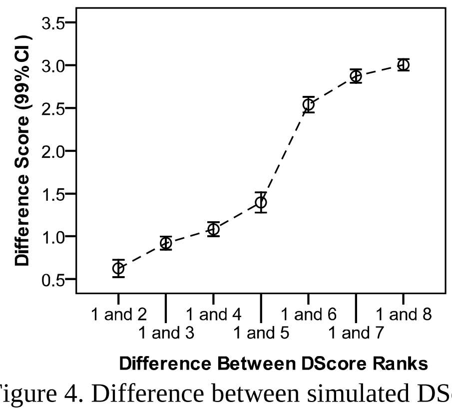


Figure 4. Difference between simulated DScore ranks for the 2009 AFL home-and-away season.

Implications of the DScore Model on Consistently Poor Performers

One criticism that has been levelled at the DScore model relates to teams who are very poor and thus, are not capable of winning many matches throughout a season. These teams do not have the capacity to obtain DScore points and subsequently, do not receive high draft choices. It could be argued that these are the teams that require high draft choices the most given that they lack sufficient player quality to win matches. Figure 5 displays an example of this occurring in an actual season, that being the 2007 AFL home-and-away season. In this season, Carlton (finished 15th on the ladder) won three matches and had one draw in the first eight rounds, but did not win a match for the remainder of the season.

In this season, Richmond (finished 16th on the ladder) received the top draft pick, followed by Melbourne who finished 14th. There was a close battle for picks 3, 4, and 5 between teams who finished between 9th and 13th on the ladder. Of note, Carlton did not receive any DScore points throughout the season given that they did not win a match from Round 11 onwards, which is the round where the DScore model commences. Whilst it might be argued that Carlton lacked sufficient team strength to win matches, they did win three matches and draw another in the first eight rounds. In addition, the team had little incentive to win based on the current draft system since any more wins in the final 14 rounds would have made the club ineligible to receive a priority pick at the end of the season.

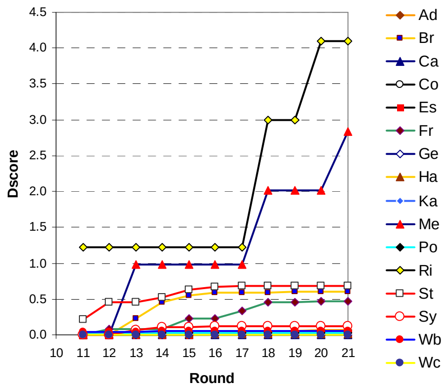


Figure 5. Progressive DScores for the 2007 AFL season for the 2010 DScore Model.

This season exemplifies that poorly performing teams need to win at least one to two matches to obtain a high draft choice, given that in this season, Richmond won only two matches after the DScore model commenced, yet had sufficient DPR without their win in round 21 to be awarded the top draft choice.

4. DISCUSSION

Results of the current paper have indicated that the revised DScore model builds on the model presented in 2006 in several ways. Firstly, the quality of the opposition being played is taken into consideration prior to each match. The inclusion of this moderating variable enables the DScore model to reward a team with additional DPR if a superior team is defeated (e.g., last defeating a team who is currently in the top eight) as opposed to a team who is of similar quality (e.g., last defeating second last on the ladder). Taking into account current ladder position also prevents teams who win a series of consecutive matches late in the season from receiving the top draft choices, when clearly there are other teams in the league that are in greater need of such high draft picks. Despite this, teams who win a string of matches are not penalised, but rather the simulation results indicate that these teams are awarded a pick in the latter half of the top 5 (e.g., pick 3, 4, or 5), thus there is still a strong incentive to win consecutive matches and finish higher on the ladder for teams capable of doing so.

Examination of the distribution in scores indicates that under the revised model, teams who finish in the bottom four positions of the ladder, yet win between three and five matches, typically receive very high draft choices. Whilst it could be argued that little difference exists between the results of the revised DScore system and the current reverse order model employed by the AFL, a fundamental principle differentiates the two approaches to allocating draft selections. This difference lies in the incentive of each team to win. The current AFL model promotes teams who are out of finals contention to lose matches, and recently retired players and coaches have attested to this. By contrast, the revised DScore model provides teams in the bottom half of the ladder with a strong incentive to win football matches right up to the final match of the season.

Simulation results have demonstrated the stability of the revised DScore model, with additional trials undertaken by the authors over the past 10 AFL seasons demonstrating the robustness of this draft system. Analysis of difference scores indicates that, in most cases, the bottom five teams are all in contention for the top draft pick, and the possibility of a bottom five team receiving a top five pick is very high (exceeding 0.90). In effect, the introduction of several moderator variables into the revised DScore model has enabled lowly teams with a greater opportunity to receive high draft picks.

Additional work may also pertain to adopting the DScore model to other sporting codes, including NBA basketball, the NHL, and the Australian National Rugby League.

5. CONCLUSIONS

In conclusion, the current paper has presented on a revised DScore model that provides a unique system for the allocation of draft choices in AFL football. This model provides poorly performing teams with an incentive to win matches throughout a season, which is particularly critical once a team is no longer in contention to play in the finals. By incorporating Morris' (1977) definition of unimportance, this model vastly reduces the frequency of 'dead rubbers', that is, matches which have no consequence on the season. Most importantly, this model provides coaches and players with a reason to be motivated to win, and club supporters with an opportunity to continue encouraging their team to win.

REFERENCES

Bedford, A., & Schembri, A. J. (2006). A probability based approach for the allocation of player draft selection in Australian Rules Football. *Journal of Sports Science and Medicine*, *5*, 509-516.

Borland, J., Chicu, M., & Macdonald, R. (2009). *Do teams always lose to win? Performance incentives and the player draft in the Australia Football League*. Unpublished manuscript.

Gold, A. M. (2010). NHL draft order based on mathematical elimination. *Journal of Quantitative Analysis in Sports*, *6*.

Leech, F. (June 20, 2008). Tanking will only worsen with two new AFL clubs. Sourced from http://www.crickey.com.au/2008/06/20/tanking-will-only-worsen-with-two-new-afl-clubs/

Morris, C. (1977). The most important points in tennis. In Ladany, S.P and Machol, R.E. (Eds.), *Optimal strategies in sport* (pp. 131-140), North-Holland Publishing Company: Amsterdam.

Stevens, M., & Ralph, J. (July 21, 2009). Terry Wallace coached Richmond to snag Trent Cotchin. Sourced from http://www.heraldsun.com.au/news/terry-wallace-coached-richmond-to-snag-trent-cotchin/story-0-1225752922510.

Stubbs, B. (September 26, 2009). Robbo talks tanking, low pay. Sourced from http://www.themercury.com.au/article/2009/09/26/99815_tasmania-news.html.

Taylor, B.A and Trogdon, J.G. (2002). Losing to win: Tournament incentives in the National Basketball Association. *Journal of Labor Economics* **20**, 23-41.

ON THE VALUE OF AFL PLAYER DRAFT PICKS

Darren O'Shaughnessy [a]

[a] *Ranking Software: darren@rankingsoftware.com*

Abstract

AFL (Australian Football League) clubs are allocated player selections ("picks") in the National Draft in reverse order of their final position in the preceding season. Clubs which perform below a certain threshold in a single season are allocated an additional pick, while clubs which meet that threshold in two successive seasons receive a more valuable pick. These somewhat arbitrary thresholds lead to a discontinuous performance / reward relationship, where it is clearly in a club's best interest to lose certain matches. The natural suspicion and speculation around "tanking" detracts from the integrity of the game, in the eyes of the AFL. However, a recent paper by Borland, Chicu & Macdonald (2009) concludes that there is little evidence of systematic "losing to win" in that league.

A natural and flexible valuation scheme for draft picks is proposed and tested, using extreme value statistics pioneered by Gumbel (1954) in what could be regarded as a variation on Galton (1902)'s Difference Problem. It removes the arbitrary discontinuities while continuing to support competitive equalisation via higher picks for genuinely struggling clubs. This draft pick method does not enforce a constant order to be followed in every round. As a corollary, the scheme suggests a method for clubs to value their picks when developing trading strategies. It could also furnish the AFL with an alternative means of compensating clubs for the loss of key players to start-up teams, and penalising clubs for transgressions.

While this scheme has direct applicability to the AFL, it is easily portable to other sports' player drafts, such as the NFL, MLB and NBA.

Keywords: Australian Rules Football, AFL, Player Draft, Extreme Value Theory, Galton's Difference Problem

1. INTRODUCTION

The Australian Football League (AFL)'s annual National Draft is the only way for existing clubs to add players to their squads[1], and is therefore crucially important to their prospects. In common with many other sports, as part of its competition equalisation policy the AFL allocates draft picks in order of reversed final position on the ladder (AFL Development, 2010). More controversially (see e.g. Sheahan (2009)), clubs are allocated "priority" picks if they are considered to be direly uncompetitive. A team which fails to win more than four matches is given an extra pick between the first and second rounds of the draft, while a team falling below this threshold in two consecutive seasons has its priority pick upgraded to above the first round. In this way Melbourne Football Club received both of the first two draft picks in the 2009 draft after finishing last with exactly four wins in the second of its dire seasons. Certainly the reward for Melbourne losing just its last game was immense: access to the two best players in the country, rather than one.

With the addition of new clubs over the next two years, the AFL has proposed a formula to compensate existing clubs for the loss of star players. Wilson (2010) suggests that for the very best players, the AFL may provide two first-round picks instead of one, with the club able to choose which year it exercises the extra picks, but only after its existing first-round pick (the position of which will vary from year to year). For a club in this situation, there is a great deal riding on the AFL's decision, and the quantum of compensation is rather large – they cannot have 1.5 first-round picks, for instance.

In this paper I develop a valuation system for draft picks and advise how the arbitrary thresholds in the system might be abolished without losing the ability to help truly uncompetitive teams.

AFL Draft Research

Borland, Chicu & Macdonald (2009) examined the teams faced by these perverse incentives for deliberately losing (also known as "tanking") and concluded that there is no significant change in behaviour. A dreadful season can lead to loss of sponsors and members, and fewer lucrative TV slots when the fixture is drawn up. Rielly (2009) reported on commissioned research by Mitchell et al (2009) that found good correlation between draft order and subsequent player performance for the first round only, with very weak correlation after pick number 16. Bedford & Schembri (2006) proposed a system where clubs not in contention for finals are rewarded for winning formally unimportant matches with an improved draft position.

Other Leagues' Draft Research

Professional US leagues such as the NFL, NBA and MLB have similar annual drafts. Burger & Walters (2009) point out that there is high risk and a lot of money at stake: only 8% of players picked in the first ten rounds of the MLB draft become established Major League players. Barzilai (2007) thoroughly analyses the empirical value of NBA Draft pick players. Berri & Simmons (2009) give an example of teams not using their choices wisely, with only a weak correlation between NFL amateur draft order and performance. Massey & Thaler (2005) state that NFL clubs overvalue the right to choose, and pay too much for the first pick in the draft.

The NFL Draft appears to receive the most attention, likely due to a famous "Draft Value Chart" developed around 1990 (Trotter, 2007) under Dallas Cowboys head coach Jimmy Johnson, anecdotally (Crowe, 2009) with help from mathematicians although the exact derivation is unknown. The chart gives a rule-of-thumb value that clubs should place on their draft picks when they are considering trading, so for instance the 1st pick is worth 3,000 points, 2nd pick 2,600, 16th pick 1,000, etc., down to the 224th pick worth 2 points. Recently there have been a number of comprehensive analyses assessing and adjusting this chart (e.g. Stuart (2008), Patterson (2009), Vance (2009)), based on performance ratings of the players picked at those positions, but none presenting an underlying theoretical model.

Extreme Value Theory

Francis Galton (1902) asked the question: if a competition has a £100 pool and there will be prizes for first and second, "How should the £100 be most suitably divided between the two?" His answer of roughly £75 : £25 is based on the expected value of the "excess merit" of someone in those positions, compared to the third-place competitor. Subsequent extensions to *n* prize-winners in large pools of competitors grew into a branch of Extreme Value Statistics, pioneered by Gumbel (1935) and Fisher & Tippett (1928).

I draw a parallel to the "competition" between potential draftees. The potential talent pool consists of young men with diverse aptitude for football. In a demographic sense, it is reasonable that an aptitude score for the population cohort of 18-year-old Australian men should be approximately normally

[1] Player trading is only permitted within the framework of the Draft, and usually in exchange for draft picks

distributed[2], like many other broad-based attributes. The players drafted would then form one extreme tail of that distribution, assuming that clubs can make an efficient assessment of that aptitude.

I propose a system in which the k^{th} draft pick has a value proportional to the k^{th} order statistic of a suitably large normal population. This paper shows the necessary calibrations and consequences.

2. METHODS

Galton's method of allocating prizes considered firstly a population of n=10. His simple assumption was that the most probable values of merit Θ for the ten competitors correspond to equidistant values of the cumulative distribution function (CDF), namely 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95. By looking up numerical probability integral tables, he discovered that the ratio of first's advantage over third compared to second's advantage over third was about 72.8:27.2. As he increased n, the ratio approached a limit of about 75.4:24.6. Therefore his proposal was that the most appropriate prize for first is about 75% of the pool.

Estimates for Extremal Values

ABS (2009) shows that at September 2009 there are approximately 772,070 males between the ages of 15-19 in Australia. The eligible demographic passing through the annual AFL Draft window is approximately one-fifth of that, indicating an appropriate n = 155,000. While men can nominate for multiple drafts, in theory they should be drafted when first eligible as their inherent aptitude is considered to be constant.

The modal value of the k^{th} extremal of a normal distribution is (Gumbel, 1954, equation (3.32)):

$$u_k = \Phi^{-1}\left(k/(n+1)\right) \tag{1}$$

where Φ is the CDF of the normal distribution with PDF ϕ. The mean is slightly higher (ibid.):

$$\mu_k = u_k + (\log k - S_k + \gamma)/n\phi(u_k) \tag{2}$$

where $\gamma \approx 0.577216$ is the Euler-Mascheroni constant and

$$S_k = \sum_{j=1}^{k-1} \frac{1}{j} \tag{3}$$

is the k^{th} harmonic number. Blom (1958) generalised the different equidistant formulas of Galton and Gumbel into an approximant for the mean:

$$m_k = \Phi^{-1}\left(\frac{k-\alpha}{n-2\alpha+1}\right) \tag{4}$$

and proposed $\alpha = 3/8$ as a rule-of-thumb constant between Galton's $\alpha = 1/2$ and Gumbel's modal $\alpha = 0$, although Harter (1961) pointed out that α actually varies with n.

With such a large n, it is worth considering whether the asymptotic ($n \to \infty$) form[3] is appropriate. Ideally, there should not be a dependency on the ABS's latest demographic trends each year in order to calibrate the draft. Fisher & Tippett (1928) point out that the tendency toward asymptotic form is exceedingly slow in the normal case (David & Nagaraja, 2003), while Dronkers (1958) proposes that it should only be used when the extremal index $k \ll \sqrt{n}$.

Cramér (1946) equation (28.6.16) gives the asymptotic formula for the mean of the k^{th} extremal:

$$\sqrt{2\log n} - \frac{\log\log n + \log 4\pi + 2(S_k - \gamma)}{2\sqrt{2\log n}} \tag{5}$$

The choice of formula to measure the value of each draft pick makes a significant difference to the first few picks, but little difference to the rest. In the table below, the difference between pick one and pick two is compared to the difference between pick two and pick ten.

Method of Valuation	$\frac{m_1 - m_2}{m_2 - m_{10}}$
Galton (α = 0.5)	0.55
Mode (α = 0)	0.41
Mean (n = 155,000)	0.63
Asymptotic Mean ($n \to \infty$)	0.70

Table 1: Relative Value of First Pick by Method

Under the assumptions outlined, the average aptitude or merit of the best young players in the country should follow the "Mean" valuation method. Consider however the assumption that clubs have perfect skills in assessing that hidden variable. If clubs are not efficient assessors, the impact of the

[2] Galton makes the same proposal for merit

[3] The characteristic distribution of the extremal is the Gumbel Distribution with CDF = $\exp(-e^{-\frac{x-\mu}{\sigma}})$

error will fall most heavily on the clubs with the early picks. In particular, the club with the first pick can only obtain full value by choosing the best player in the pool. The club with the second pick has a non-zero chance of doing better than its allocation, if the first club makes an error of choice, but could also make an error and choose a player worse than the second-best. Perhaps this effect is evident in the findings discussed in the introduction, where the first pick is empirically overvalued.

I therefore propose to use the modal (or most likely) value in the valuation method. This keeps the dependency on n, but the valuation ratios do not vary materially from year to year.

The Worthless Pick

Galton decided that in the simplest version of his question, there should only be two prizes. Every competitor from third onwards was treated the same. Having decided on the shape of the valuation method, I also need to set the zero. Every potential player below a certain level of aptitude is the same, as far as the clubs are concerned. This is essentially an empirical judgement – when do the clubs decide their next pick is worthless?

In the AFL National Draft, clubs may take between four and eight players. In 2009, both Melbourne and Fremantle elected not to use their early fifth-round picks (#66 and #68 respectively), effectively declaring them valueless. The last pick used was #95, compared to #83 in 2008 and #75 in 2007. Geelong traded away two unwanted players (Steven King and Charlie Gardiner) in 2007 for pick #90, which they did not use. For the purposes of constructing the model, I will draw the line after six rounds, i.e., pick #97. The exact zero point does not have much of an effect on the valuation scheme, because the difference between subsequent picks near the zero is quite small.

At the other end of the scale, I will conform to the NFL convention and arbitrarily value the first pick at 3,000 “Draft Points”. Therefore the linear transformation to pick values v_k ($0 < k < 97$) is

$$v_k = 3000 \cdot \frac{m_k - m_{97}}{m_1 - m_{97}} \tag{6}$$

Allocating Draft Picks to Clubs

Based on their season performance, clubs are allocated a certain number of Draft Points. The simplest version of the model replicates the current draft, with club c (numbered from 1^{st} on the ladder to 16^{th}) receiving Draft Points $P_{c,1}$ according to:

$$P_{c,1} = \sum_{k=17-c,33-c,49-c,\dots}^{97-c} v_k \tag{7}$$

The second index indicates the number of Draft Points club c has prior to pick i. To determine the draft order after the season, the following algorithm is run for each pick i:

1. Find the club t with the most remaining Draft Points, i.e., $t : P_{t,i} = \max_c\{P_{c,i}\}$
2. Club t owns pick i and has v_i Draft Points removed from its total: $P_{t,i+1}=P_{t,i} - v_i$

In this simple model, each club receives a pick in reverse ladder order for every round.

Note that Draft Points are positive real numbers.

Measuring Need

Draft Points could also be allocated in a completely different way, for instance through a formula which rates a team for its ladder position, number of wins, and/or percentage (points for / points against). Often there are a number of clubs in the middle of the ladder with similar win-loss records. In 2009, Sydney won just one fewer match than Hawthorn and had a better percentage, yet received picks 6, 22, 38, … compared to 9, 25, 41, … because they finished three rungs lower on the ladder. On the Draft Point scale, Sydney were allocated 4,435 points to Hawthorn’s 3,938 – 12.6% more – despite the difference in quality between the two being virtually undetectable. A formula which rated the middling teams closer together would see a more balanced allocation of draft picks.

The philosophy of the draft is to adequately support struggling clubs, so that they can become average clubs. In the past, the reward for finishing last in consecutive seasons has tended to overcompensate the dire clubs and allowed them to compete at the top of the ladder within 6-8 years (Mitchell et al, 2009). It should not unduly punish the premier – the current allocation of the last pick in each round would remain the standard.

A possible formula to achieve these ends is as follows:

- The eight finalists are allocated Draft Points as per (7)
- Non-finalists are given an initial “Need Rating” (8) based on their number of wins and percentage. Points for-versus-against percentage is considered a safe indicator, as teams do not deliberately set out to be thrashed. It is demoralising for the players and supporters, and a percentage below 70% points to dire need
- The Need Rating is topped up with a fraction of the club’s previous season Need Rating, from 7.5% (9^{th}) to 60% (16^{th}). Clubs which made the finals in the previous season have no carry-over rating

- The Need Rating is linearly transformed into Draft Points (9) using constants dependent on (6)

Need Rating $$NR_c = 94 - \frac{PC}{2} - PP \qquad (8)$$

where PC is the club's percentage and PP is their premiership points (four for a win and two for a draw). 94 is chosen so that a team with the competition average 44 points and 100% is not considered in need. If NR_c is calculated at less than zero, it will be taken to be zero.

$$P_{c,1} = 4003 + 6710 \cdot \frac{NR_c}{\sum_j NR_j} \qquad (9)$$

3. RESULTS

Table 2 displays the complete set of Draft Points for a 16-club, six-round draft.

Pick	Points	Pick	Points	Pick	Points	Pick	Points
1	3,000	25	977	49	504	73	213
2	2,593	26	950	50	489	74	203
3	2,348	27	924	51	475	75	193
4	2,171	28	899	52	461	76	183
5	2,032	29	874	53	447	77	173
6	1,918	30	851	54	434	78	164
7	1,820	31	828	55	420	79	154
8	1,734	32	806	56	407	80	145
9	1,658	33	784	57	394	81	135
10	1,590	34	763	58	382	82	126
11	1,528	35	743	59	369	83	117
12	1,471	36	723	60	357	84	108
13	1,418	37	704	61	345	85	99
14	1,369	38	685	62	333	86	91
15	1,323	39	667	63	321	87	82
16	1,280	40	649	64	310	88	73
17	1,239	41	631	65	298	89	65
18	1,201	42	614	66	287	90	56
19	1,164	43	597	67	276	91	48
20	1,129	44	581	68	265	92	40
21	1,096	45	565	69	254	93	32
22	1,065	46	549	70	244	94	24
23	1,034	47	534	71	233	95	16
24	1,005	48	518	72	223	96	8

Table 2: Value of the k^{th} Draft Pick

The only publicly available comparison for this theoretical model is the NFL Draft Value Chart:

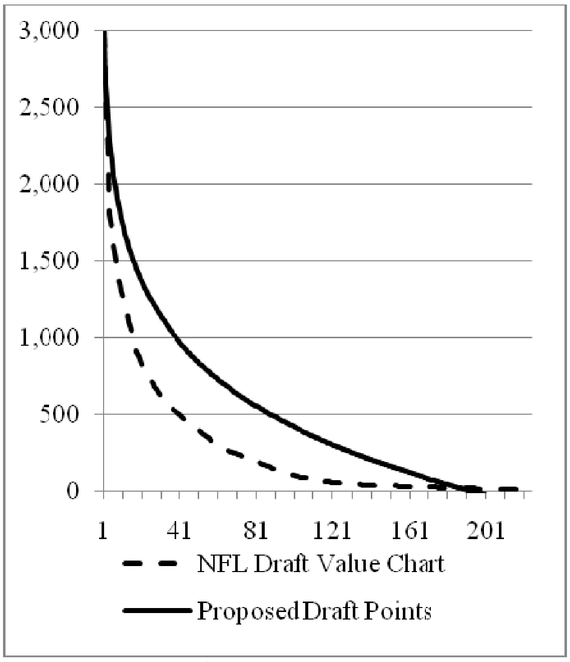


Figure 1: The NFL Draft Value Chart (Crowe, 2009) compared to the proposed valuation scheme.

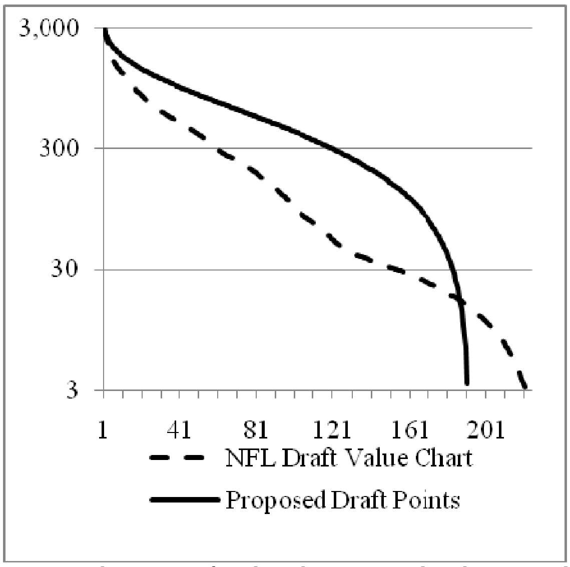


Figure 2: The NFL Draft Value Chart compared to the proposed valuation scheme on a logarithmic scale.

The extreme-value model clearly does not fit the published NFL charts, even after taking the USA's larger population into account. The mid-range choices on the chart are substantially undervalued in comparison. It appears from the log-scale Figure 2 that the NFL chart may have been drawn from a simple logarithm then smoothed from about pick 130 to asymptotically approach zero. Potentially there is merit in this smoothing, as late picks have some residual value due to the rare good player who

is still uncovered at that late stage, but most will be close to the minimum league standard.
The divergence between the two curves is similar to that seen by Stuart (2008), who used empirical career data to rate the actual picks from 1970 to 1999.

2009 Season Example
Table 3 compares the number of Draft Points clubs would receive under various scenarios. The second column is a regular ladder without any priority picks, the third column is how the points were allocated after Melbourne received the priority pick, and the fourth column shows what clubs would have received under the formula of the previous section.

Club	Regular	2009 Draft	Proposed
Geelong	3066	2961	3066
St Kilda	3176	3066	3176
W Bulldogs	3289	3176	3289
Collingwood	3407	3289	3407
Adelaide	3530	3407	3530
Brisbane	3659	3530	3659
Carlton	3795	3659	3795
Essendon	3938	3795	3938
Hawthorn	4091	3938	4324
P Adelaide	4256	4091	4446
West Coast	4435	4256	4687
Sydney	4633	4435	4422
North Melb	4858	4633	4603
Fremantle	5123	4858	5132
Richmond	5459	5123	4978
Melbourne	5961	8459	6225

Table 3: Draft Points comparison

The extraordinary boost received by Melbourne for not winning its last game of 2009 is evident here: an extra 2,498 Draft Points. There are several differences in the proposed scheme, with Fremantle and West Coast (14^{th} and 15^{th} in 2008) carrying some Need Rating over to 2009. Melbourne would have received 6475 Draft Points in 2008, before winning an extra game with a substantially superior percentage in 2009.

Figure 3 shows how the picks are allocated in a traditional draft (left) compared to one derived from the points of the last column of Table 3:

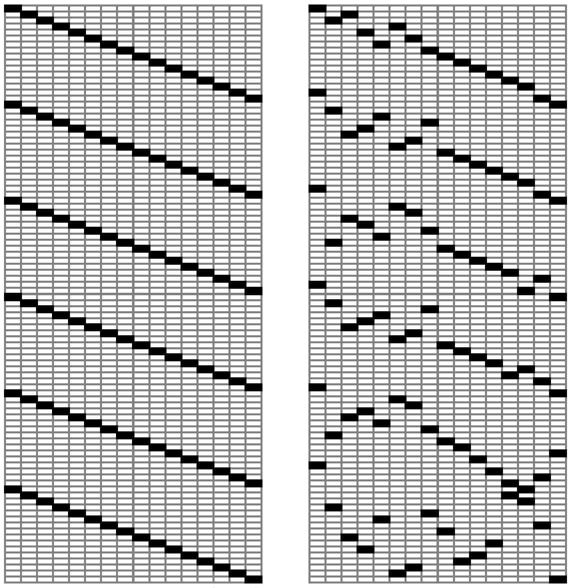


Figure 3: A regular draft (left) compared with one run according to the proposed valuation scheme. Columns are the clubs in reverse ladder order; each row is a draft pick from top to bottom.

Melbourne receives picks #1, #15, #31, #47, #64 and #77 in the proposed scheme. Its second pick pre-empts (*c*=2) St Kilda's first pick, which becomes #16. As compensation, St Kilda receives its third pick ahead of Western Bulldogs (*c*=3). Note also that although Fremantle received pick #2, the subtraction of that pick's value means its next pick is not until #22 (6^{th} in the "round"). By the time the draft gets to the last round, the order is unrecognisable.

4. DISCUSSION

Applications of the Draft Point valuation scheme are numerous. They provide trade utility, not being grossly quantised like players or full picks. Additionally, clubs that lose a star player to a new franchise could be appropriately reimbursed with Draft Points by making the existing discrete compensation formula continuous.
Clubs which transgress against salary cap regulations or other AFL rules could be penalised in Draft Points, not necessarily completely excluded from the draft.
The AFL National Draft is followed by a Rookie Draft and Pre-season Draft. While these have not been mentioned in the methodology, they should be brought into the same system. Mitchell et al (2009) assert that players selected early in the Rookie Draft can have an impact similar to a second-round National Draft pick.
We may also judge past and future trades and player selections against the measuring stick of Draft Points. As an example, during 2009 Trade Week complex negotiations between Hawthorn, Essendon

and Port Adelaide involving star players Shaun Burgoyne and Mark Williams had reached an impasse because the teams could not agree on the number of the draft pick. Geelong entered the discussions and provided an acceptable draft pick in exchange for a number of lower selections. Geelong's contribution can be accounted for thus:

Transaction	Draft Points	
Sell pick #33	-784	
Sell pick #97	-0 (not used)	
Receive pick #40		+649
Receive pick #42		+614
Receive pick #56		+407
Net Gain		**+886**

Table 4: Geelong's Pick Trading in 2009

Geelong made an extraordinary 886 point gain on the trade, the equivalent of an extra #29 pick. One might think they had a mathematician in the negotiations!

5. CONCLUSIONS

This paper has presented a mathematical basis for valuing selections in a sports league draft. Potentially there are other applications where allocations of choice are made, for instance in game theory. Calibration of the model for other sporting leagues should be relatively easy and robust.

Mitchell et al (2009) have examined AFL performance data relative to draft position and identified two disjoint trends, for high and low picks. It would be interesting to re-examine their data for fit against this model.

A possible extension would be to include a stochastic model of the clubs' ability to choose the next most talented player in the pool. Another candidate for adjusting the model would consider that many draftees never play, despite having an aptitude very close to AFL standard, so their actual value to the club is lower than the extreme-value model suggests. These may smooth the characteristic curve in a way similar to the NFL Draft Value Chart.

It is hoped that mathematicians can play our part in removing the taint of tanking from the AFL, if only to give journalists something more edifying to write about.

References

AFL Development (2010). About NAB AFL Draft. *afl.com.au*. Available at http://www.afl.com.au/about/tabid/13514/default.aspx.

ABS (2009). Australian Demographic Statistics, September Quarter 2009. *abs.gov.au*.

Barzilai, A. (2007). Assessing the relative value of draft position in the NBA draft. *82games.com*. Available at http://www.82games.com/barzilai1.htm.

Bedford, A., & Schembri, A. J. (2006). A probability based approach for the allocation of player draft selections in Australian Rules football. *Journal of Sports Science and Medicine, 5*, 509-516.

Berri, D. J., & Simmons, R. (2009). Catching a draft: on the process of selecting quarterbacks in the National Football League amateur draft. *Journal of Productivity Analysis,* published online 18 September 2009.

Blom, G. (1958). Statistical estimates and transformed beta-variables. Published by John Wiley & Sons.

Borland, J., Chicu, M. & Macdonald, R. D. (2009). Do teams always lose to win? Performance incentives and the player draft in the Australian Football League. *Journal of Sports Economics, 10*, 451-484.

Burger, J. D. & Walters, S. J. K. (2009). Uncertain Prospects: Rates of return in the baseball draft. *Journal of Sports Economics, 10,* 485-501.

Cramér, H. (1946). Mathematical methods of statistics. Published by Princeton University Press.

Crowe, S. (2009). Where the New England Patriots stand on the NFL Draft Trade Value Chart. *New England Patriots Examiner (blog), April 23, 2009,* available at http://www.examiner.com/examiner/x-1324-New-England-Patriots-Examiner~y2009m4d23-Where-the-New-England-Patriots-stand-on-the-NFL-Draft-Trade-Value-Chart

David, H. A. & Nagaraja, H. N. (2003). Order statistics. Published by Wiley-Interscience (Wiley Series in Probability and Statistics).

Dronkers, J. J. (1958). Approximate formulae for the statistical distributions of extreme values. *Biometrika, 45,* 447-470.

Fisher, R. A. & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proceedings of the Cambridge Philosophical Society, 24,* 180-190.

Galton, F. (1902). The most suitable proportion between the values of first and second prizes. *Biometrika, 1,* 385-399.

Gumbel, E. J. (1935). Les valeurs extrêmes des distributions statistiques. *Ann. Inst. H. Poincaré, 5,* 115-158.

Gumbel, E. J. (1954). Statistical theory of extreme values and some practical applications. *National Bureau of Standards Appl. Math. Series, 33,* 1-51.

Gumbel, E. J. (1958). Statistics of extremes. Published by Columbia University Press.

Haldane, J. B. S. & Jayakar, S. D. (1963). The distribution of extremal and nearly extremal values in samples from a normal distribution. *Biometrika, 50,* 89-94.

Harter, H. L. (1961). Expected values of normal order statistics. *Biometrika, 48,* 151-165.

Lenten, L. J. A. & Winchester, N. (2009). Optimal bonus points in the Australian Football League. *University of Otago Economics Discussion Papers, No. 0903*.

Massey, C. & Thaler, R. H. (2005). Overconfidence vs market efficiency in the National Football League.

NBER Working Paper, No. W11270. Available at http://www.nber.org/papers/w11270

Mitchell, H., Stavros, C. & Stewart, M. F. (2009). AFL Recruitment Prospectus. *RMIT University (only a flyer has been made public)*.

Patterson, T. (2009). Sabermetrics 101: How to value prospects and draft picks. *ywacademy.com*. Available at http://www.ywacademy.com/2009/11/sabermetrics-101-how-to-value-prospects.html.

Rielly, S. (2009). Draft pattern gives recruiters cold. *The Weekend Australian newspaper, May 09 2009*, available at http://www.theaustralian.com.au/news/draft-pattern-gives-recruiters-cold/story-e6frg7mx-1225710683235

Sheahan, M. (2009). AFL must rejig its priority-pick system. *Herald Sun newspaper, July 14 2009*, available at http://www.heraldsun.com.au/news/afl-must-rejig-its-priority-pick-system/story-0-1225750046260

Stuart, C. (2008). The draft value chart: right or wrong? *pro-football-reference.com*. Available at http://www.pro-football-reference.com/blog/?p=527.

Trotter, J. (2007). Value chart has NFL talking points. *San Diego Union Tribune, April 18, 2007*, available at http://legacy.signonsandiego.com/sports/nfl/20070418-9999-1s18nfltrade.html

Vance, J. (2009). The draft trade chart – version 2.0. *bloggingtheboys.com*. Available at http://www.bloggingtheboys.com/2009/4/12/823024/the-draft-trade-chart-version-20.

Wilson, C. (2010). League to suggest champions worth two draft picks. *The Age newspaper, April 16, 2010*, available at http://www.theage.com.au/afl/afl-news/champions-worth-two-draft-picks-20100415-shkl.html

FITTING PROBABILITY DISTRIBUTIONS TO REAL-TIME AFL DATA FOR MATCH PREDICTION

Richard Ryall [a,b] , Anthony Bedford [a]

[a] School of Mathematical and Geospatial Sciences, RMIT University, Australia
[b] *Corresponding author: Richard.ryall@student.rmit.edu.au*

Abstract

In this research a Generalized Logistic Model (GLM) is used to model outcomes of Australian Rules football matches in real-time. Incorporating difference in team quality and score difference the outcome of the model is the probability of victory at each of the quarter time breaks. The parameters of the GLM are a function of opponent quality and are optimized via simulation for each quarter. Archival AFL data was obtained from seasons 2000 to 2009 which consisted of year, round, quarter, (nominal) home team, away team, home team score and away team score. Seasons 2000 to 2004 are used as a training set for the forward prediction of seasons 2005 to 2009. Comparisons are made throughout against a simple Brownian motion model. Both models are then evaluated on predicted and actual probabilities of winning.

Keywords: AFL, real-time, optimization, generalized logistic model, curve fitting, prediction

1. INTRODUCTION

Sports commentators in game sports constantly talk about the likelihood of either team winning at any point in time rarely with any empirical evidence to support their suggestions. Comments such as “Boston Celtics rarely loses the match if they are leading at three quarter time” are very common. It has been shown by Cooper, DeNeve and Mosteller (1992) that either team leading after three quarters of the game in Basketball (NBA), Football (NFL) and Hockey (NHL) won approximately 90% of the time. This is of course without making any adjustments for quality of the two competing teams. However, Australian Rules football (AFL) is known for its high level of uncertainty intra-match with the team leading at three quarter time winning approximately 85% of matches. The colloquial saying “the match is not over until the final siren is blown” has never been more appropriate. It is this uncertainty that draws spectators to matches and entices academics to try and explain it.

The outcome of an AFL match is dependent on several important factors which may include but not limited too: home ground advantage, injuries, current form, the opposition, weather and match time (day, night and twilight). Some factors may have more effect than others while others may have no effect at all. However, once the match is underway these so called important factors are likely to have less and less influence as the match progresses. Other factors start to have more influence such as current score and time remaining in the match.

Falter and Pérignon (2000) determined the in-game probability of winning a soccer match using binary-probit models. Matches were split into five sub-periods of 15 minutes with the model being re-estimated after each period as more information becomes available. Objective (pre-game) variables included home advantage and the standing (ranking) of both teams, intra-match variables included goal differential dummy variables. The coefficients of the objective variables were not assumed to be constant (Bayesian updating process). As the game progressed home advantage had less influence on the outcome of the match. The team’s ranking was most critical in the first and the last period. Goal difference is increasingly crucial towards the end of the match. Notably, not much is known about the interaction between the objective variables and current score.

Bailey and Clarke (2008) predict run total of the batting team in One Day International (ODI) cricket matches while the game is in progress. Using multiple linear regression to obtain the pre-match Margin of Victory (MOV) in conjunction with the Duckworth-Lewis method to determine resources remaining, an updated MOV could be obtained at the conclusion of each over. This model was then implemented to investigative the efficiency "in-play" betting markets in ODI with marginal evidence that punters may over/under bet particular teams relative to their chances of winning.

Klaassen and Magnus (2003) develop *TENNISPROB* a computer algorithm which amongst other things instantaneously calculates the in-game probability of either playing winning. Provided points are independent and identically distributed (i.i.d) the match probability can be determined by the probabilities either player wins a point on their serve, the type of tournament, current score and the current server. A previous study by Klaassen and Magnus (2001) found that although points in tennis are not i.i.d. the deviations are so small that the i.i.d. assumption is justified.

More relevant studies of intra-match prediction in game sports include Stern (1994) and Glasson (2006). Stern (1994) considered using a simple Brownian Motion Model (BMM) to provide an in-game probability estimate of team's chance of victory in baseball and basketball. The model incorporates a pre-game point estimate (μ) to adjust for home advantage, score difference (l) and time elapsed (t). For the BMM model to be valid the score difference $X(t)$ is assumed to be normally distributed with mean μt and variance $\sigma^2 t$; and $X(s)-X(t)$ for some $s>t$ is independent of $X(t)$. Stern found reasonably evidence to suggest the data for basketball were reasonably consistent with BMM in the first three quarters with the home team outscoring the visiting team by approximately 1.5 points and a standard deviation of approximately 7.5 points. However the fourth quarter was remarkably different with only a slight advantage to the home team. A possible explanation given by Stern is that if a team has a comfortable lead they might ease up or use less skilful players.

In basketball, observing the probability of the home team winning (y-axis) against score difference (x-axis) at each of the quarter time breaks reveals an inverted S-shape pattern [Stern (1994), p1132]. The shape of the two bumps in the S-shape pattern is more pronounced as the match progress which reflects the increase significance of being ahead on the scoreboard late in the game.

Glasson (2006) builds on this and applies the model defined in Stern (1994) to Australian Rules football replacing the pre-game point estimate (μ) with the bookmaker's line to adjust for home ground advantage *and* the quality of the two teams. Again the data seemed relatively consistent with the BMM, with the errors for each quarter between the bookmaker's line and the non-cumulative score difference for each quarter approximately equal to zero. That is,

$$\varepsilon_t = X(t) - X(t-1) - \frac{\mu}{4} \approx 0 \qquad [1]$$

However, the error term defined in [1] is biased because it measures the average error of the bookmaker's line and score difference, and fails to take into account their interdependence. For example, consider two scenarios (1) a heavy favourite is *ahead* on the scoreboard at time t and exceeding expectations at time $t+1$, that is, $\left[X(t+1)-X(t)>\mu/4\right]$ and (2) a heavy favourite is *behind* on the scoreboard at time t and falling short of these expectations at time $t+1$, that is, $\left[X(t+1)-X(t)<\mu/4\right]$ averaging these two scenarios out leads to $\left[X(t+1)-X(t)\approx\mu/4\right]$. Table 2 shows the in-game scoring behaviour of Australian Rules football teams and provides evidence that $X(s)-X(t)$ for some $s>t$ is ***not*** independent of $X(t)$.

Quarter	Favourite	Next Quarter $\overline{X}(t)$
1	Ahead	+4.69
	Behind	+3.21
2	Ahead	+6.32
	Behind	+1.12
3	Ahead	+6.12
	Behind	+1.38

Table 1 Mean difference between the favourite and corresponding underdog team score $\overline{X}(t)$ in the 2nd, 3rd and 4th quarters as a function whether the favourite was ahead or behind at the end of the previous quarter, 2000 to 2009.

Although increasingly more research is being directed towards real-time predictions in game sports, no such literature to date takes into account the interdependence between opponent quality, current score and time remaining in the match. Therefore the purpose of this paper is to develop more statistically robust non-linear (S-shaped) functions for real-time prediction in AFL that provide a superior fit and account for the interdependence between score difference and quality of opponent at each of the quarter time breaks.

This manuscript is divided into four sections; (2) describes the methodology of curve fitting and the reasoning of applying the Generalized Logistic Function to our data; (3) the methodology of finding the optimal parameters of the non-linear function to maximize the predicted probability of winning; (4) details the results of the model including predicted accuracy and describes the change in distribution of the non-linear function subject to time elapsed, score difference and the difference in quality between the two teams; (5) conclusions of the research including future directions.

2. CURVE FITTING

Akin to regression analysis, curve fitting is the procedure of fitting a probability distribution which gives the best fit to a series of data points. Typical probability distributions used in curve fitting include Beta, Exponential, Gamma, Generalized Logistic, Gompertz, Linear, Lognormal and Weibull. Kuper and Sterken (2006) apply the inverted S-shaped Gompertz function to model the development of world records in running. They find that the point of inflection, that is, the period of greatest gain [Nevill and Whyte (2005)] in running was in the 1940's and 1950's. Due to the asymptotic behaviour of the Gompertz function, implied limits of world records could be deduced.

To aid in the selection of a probability distribution to predict the outcome of an AFL match in real time, the actual probability of winning was plotted against the score difference at each of the quarter time breaks for varying differences of opponent quality. The results suggest a non-linear (S-shaped) function as suitable with a lower asymptote of zero and an upper asymptote of one.

Therefore, the four-parameter Generalized Logistic function was utilized which is given by

$$\Pr_{i,SD}(t) = \frac{1}{\left[1 + Qe^{-B_i(SD - M_i)}\right]^{1/v_i}} \qquad [1]$$

where $\Pr_{i,SD}(t)$ denotes the probability of team *t* winning at quarter *i*, for given score difference *SD*. Four parameters for each quarter *i* are estimated. B controls the rate of growth, M shifts the time of maximum growth, Q depends on the value $\Pr_{i,0}(t)$ and v affects which asymptote maximum growth occurs.

Figure 1 illustrates the effect each of the parameters (excluding *Q*) has on $\Pr_{i,SD}(\text{home})$ keeping all the other parameters constant.

3. METHODS

This papers analysis is based on seasons 2000 to 2009. AFL data was gathered from ProEdge a statistical package developed by ProWess Sports. Data consisted of year, round, quarter, (nominal) home team, away team and home team margin. A pre-game point estimate (or *LINE*) was calculated for each match for seasons 2000 to 2009 using the ratings model developed by Ryall and Bedford (2010). For example, in round 9 2010 Essendon (home) played Richmond (away) the *LINE* was +28 points in favour of Essendon. That is, prior to the start of the match Essendon are expected to win by 28 points.

Four parameters B, M, Q and v of the Generalized Logistic function need to be optimized for each quarter *i*. However the model given in [1] does not incorporate opponent quality herein referred to as the *LINE*. Therefore each of these parameters is replaced by a simple linear equation. A linear model was selected for simplicity and also seemed suitable after plotting the data. Since Q depends solely on the value $\Pr_{i,0}(t)$, and $\Pr_{i,0}(t) = 0.5$ when $LINE = 0$ (i.e. probability of winning equals 0.5 when scores are level and quality of both teams is the same), therefore $M = 0$ when this occurs and thus Q becomes a function of v given by:

$$Q_i = \sqrt[1/v]{2} - 1 \qquad [2]$$

Therefore we now have five variables to be optimized for each quarter *i*.

$$\begin{aligned} B_i &= B1_i + B2_i\left[ABS(LINE)\right] \\ M_i &= M2_i\left[ABS(LINE)\right] \\ v_i &= v1_i + v2_i\left[ABS(LINE)\right] \end{aligned} \quad [3]$$

Since every match has a nominated home team and a nominated away team the sum of these two probabilities must equal one for quarter *i* and given score difference *SD*. That is, for every match

$$\Pr_{i,SD}(\text{home}) + \Pr_{i,SD}(\text{away}) = 1 \quad [4]$$

Therefore,

$$\Pr_{i,SD}(\mathrm{t}) = \begin{cases} \dfrac{1}{\left[1+Qe^{-B_i(SD-M_i)}\right]^{1/v_i}}, & \text{if } t = \text{home} \\ 1 - \dfrac{1}{\left[1+Qe^{-B_i(-SD-M_i)}\right]^{1/v_i}}, & \text{if } t = \text{away} \end{cases} \quad [5]$$

As appose to Stern (1994), the probability of winning at time *i* given in [5] allows for the interdependence of team quality (*LINE*) and score difference (*SD*), since the contribution of team quality is dependent on score difference and vice versa.

There are several different loss functions that can be utilized for evaluating prediction models (Witten and Frank, 2005). In this research we will concentrate on quadratic loss or more specifically the *Brier Score* (Brier, 1950) which is given by:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2 \quad [6]$$

Where p_i is the forecast probability and o_i is the outcome variable (win=1, draw=0.5, loss=0). Seasons 2000 to 2004 were used as a training set in the forward prediction of seasons 2005 to 2009. Simulations were carried out utilizing the Monte Carlo algorithm using RiskOptimizer an add-in for Excel.

Several constraints are placed on the parameters given in [3]. Firstly the parameters B_1, B_2, v_1 and v_2 should all be nonnegative (or zero) and M_2 should be negative (or zero) because as *LINE* increases so too should $\Pr_{i,SD}(\text{home})$. Upper bounds are also placed on the parameters to reduce the total possible number of combinations and speed up convergence. These bounds were determined from figure 1.

4. RESULTS

Figure 2 displays the empirical probability of winning as a function of score difference (*SD*) at each of the quarter time break for varying levels of quality of opponent. Akin to Stern (1994) as the match progresses score difference (*SD*) has more influence whilst quality of opponent has less.

Various measures can be used to evaluate the performance of prediction models in game sports. Some commonly used measures in the literature include Average Absolute margin of Error (AAE), number of predicted winners and Return on Investment (Bailey and Clarke, 2004). Since the number of predicted winners will tend towards one as the match progress an alternative measure is needed to evaluate the performance of the GLM.
Akin to Stephani and Clarke (1992) this research will compare the predicted probabilities and actual probabilities winning and their corresponding proportions for the GLM compared to the BMM. Firstly the predicted probability of the in-game favourite winning is banded into five subgroups. The number of games and the actual probability of winning for each subgroup of predicted probabilities are shown in Table 2. For example, in the first quarter the BMM had 22.9% of all matches as a 50-59% favourite, teams that fell in this category had a 56.1% chance of victory on average.

The importance of having a high level of predicted probabilities without compromising the actual probability of winning is critical from a betting perspective (Ryall and Bedford, 2010). Therein they show that long term profits are attainable when there is a significant advantage in the bettor's favour where advantage is defined as:

$$A = (P \times MO) - 1 \quad [7]$$

where *A* = advantage, *P* = predicted probability of winning, and *MO* = market odds.

Although the number of winners predicted by the two different models is approximately equal the distribution of predicted probabilities for the GLM is heavily skewed towards one (win). Interestingly the GLM increasingly outperforms the BMM as the match progresses.

Similarly in some cases of the BMM the actual probability of winning falls above the predicted range particularly late in the game. This can be attributed to assumption that future scoring behaviour is independent of current score.

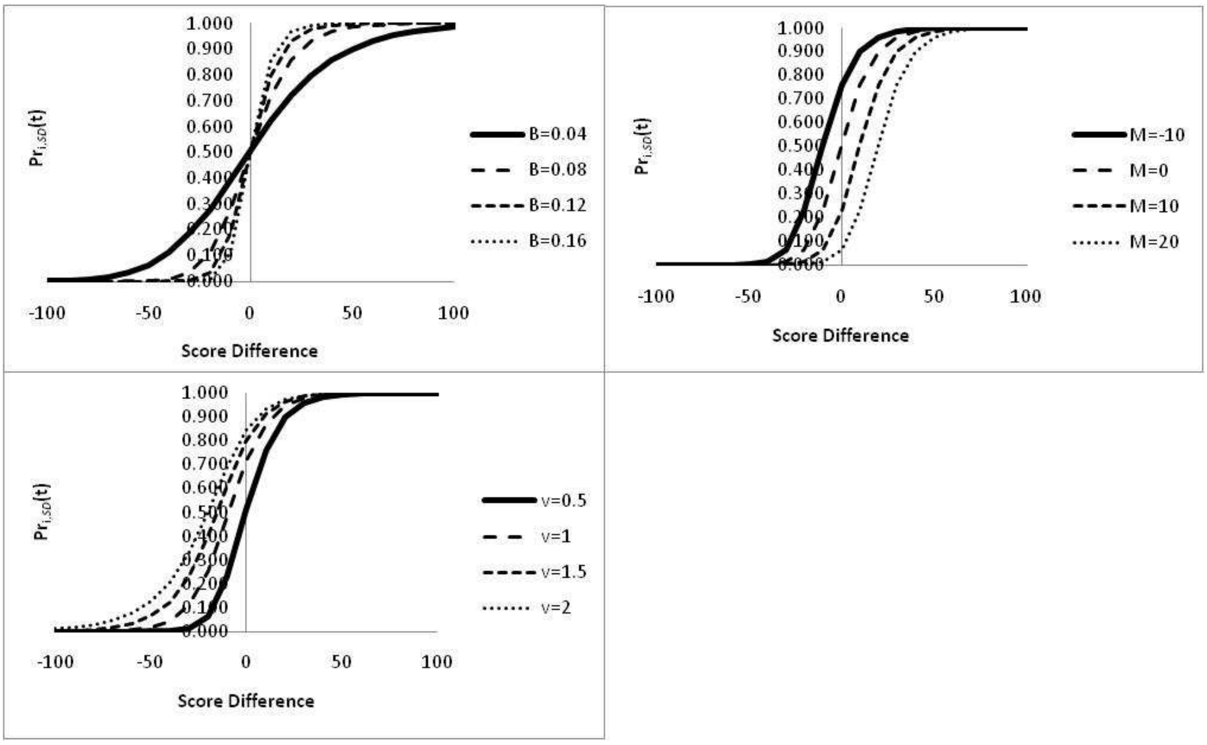


Figure 1 The Generalized Logistic Function for varying parameter values keeping other parameters constant

	Quarter 1		Quarter 2		Quarter 3	
Predicted Probability	BMM	GLM	BMM	GLM	BMM	GLM
0.50 – 0.59	0.561(0.256)	0.468 (0.213)	0.602 (0.192)	0.574 (0.153)	0.500 (0.117)	0.588 (0.076)
0.60 – 0.69	0.615 (0.225)	0.634 (0.206)	0.697 (0.171)	0.699 (0.127)	0.684 (0.107)	0.603 (0.071)
0.70 – 0.79	0.756 (0.230)	0.713 (0.235)	0.815 (0.170)	0.740 (0.169)	0.802 (0.142)	0.720 (0.092)
0.80 – 0.89	0.859 (0.167)	0.823 (0.228)	0.889 (0.212)	0.843 (0.222)	0.945 (0.163)	0.822 (0.152)
0.90 – 1.00	0.959 (0.110)	0.958 (0.107)	0.968 (0.243)	0.951 (0.318)	0.980 (0.461)	0.970 (0.598)

Note. BMM = Brownian Motion Model, GLM = Generalized Logistic Model.

Table 2 Predicted and actual probabilities of winning and corresponding proportions (in parenthesis), seasons 2005-2009

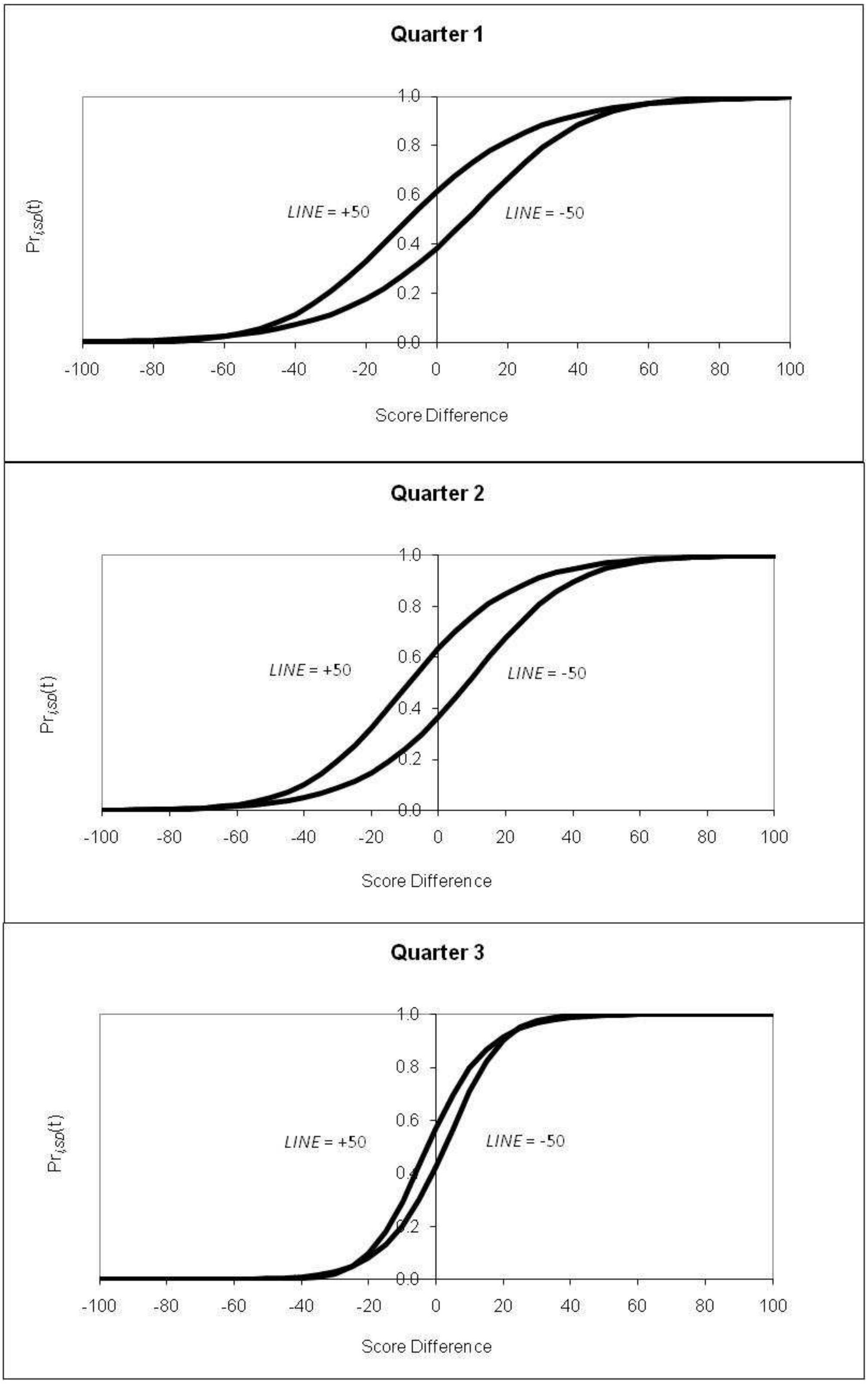


Figure 2 Smooth curves showing the probability of winning an AFL match at quarter time, half time and three quarter time for given score difference *SD*

5. CONCLUSION

This research shows that current score and opponent quality in Australian Rules football are interdependent variables which depend on time remaining in the match.

This research opens the door for future research into the modelling of high scoring sports in real-time. The authors suggest any future work in this area should incorporate time as a continuous variable and thus account for the interdependence between difference in opponent quality, score difference *and* time elapsed.

References

Bailey, M. and Clarke, S.R. (2004). Deriving a profit from Australian Rules football: a statistical approach. In proceedings of the Seventh Australian conference on Mathematics and Computers in Sport, H. Morton and S. Ganesalingam, (Eds.), Massey University: Palmerston North. 48-56.

Bailey, M. and Clarke, S.R. (2008). Predicting the match outcome in one day international cricket matches while the game is in progress. In proceedings of the Ninth Australian conference on Mathematics and Computers in Sport, J. Hammond, (Eds.), Tweed Heads: New South Wales. 160-169.

Barnett, R. and Clarke, S.R. (2005). Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*, 16, 113-120.

Brier G.W. (1950). Verification of weather forecasts expressed in terms of probability, *Monthly Weather Review* 78, 1–3.

Cooper, H, DeNeve, K.M. and Mosteller, F. (1992). Predicting professional sports game outcomes from intermediate game scores. *Chance: New directions for statistics and computing*, 3-4(5) 18-22.

Falter, J.M., Pérignon, C. (2000). Demand for football and intramatch winning probability: an essay on the glorious uncertainty of sports. *Applied Economics*, 32(13), 1757 – 1765.

Glasson, S. (2006). A Brownian motion model for the progress of Australian Rules Football scores. in proceedings of the 7th *Australasian Conference on Mathematics and Computers in Sport*, 216-225.

Klaassen, F.J.G.M. and Magnus, J.R. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96(454), 500-509.

Klaassen, F.J.G.M. and Magnus, J.R. (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research*, 148, 257-267.

Kuper, G.H. and Sterken, E. (2006). Modelling the development of world records in running. Working paper, accessed on 25/03/10 http://ccso.eldoc.ub.rug.nl/root/2006/200604/

Nevill, A.M. and Whyte, G. (2005). Are there limits to running world records? Medicine and Science in Sports and Exercise, 37(10), 1785-1788.

Ryall, R. and Bedford, A. (2010). An optimized ratings based model for forecasting Australian Rules football. International Journal of Forecasting. (in press)

Stephani, R.T. and Clarke, S.R. (1992). Predictions and home advantage for Australian Rules Football. *The Journal of Applied Statistics*, 19(2), 251-261.

Stern, H. S. (1994). A brownian motion model for the progress of sports scores. *Journal of the American Statistical Association*, 89, 1128-1134.

Witten I.H., and Frank, E. (2005). Data mining: Practical machine learning tools and techniques, Elsevier, San Francisco, CA.

A CONCEPTUAL FRAMEWORK FOR GLOBAL POSITIONING SYSTEMS WITHIN THE AUSTRALIAN FOOTBALL LEAGUE: A DISCUSSION OF TWO MODELS

Kelly Foreman [a b d]

[a] *School of Computer and Information Science, University of South Australia*
[b] *Reconfigurable Computing Lab*
[d] *Corresponding author: Kelly.Foreman@unisa.edu.au*

Abstract

Elite sporting clubs are using advanced technology such as Global Positioning Systems (GPS) to assist with issues like fitness and strength conditioning, tactical and strategic analysis of the game and its players as well as rehabilitation (Catapult 2009). These issues have arisen as the game of Australian Rules Football has become more physically demanding and faster; leading to increased likelihood of injuries to elite athletes (Le Grand 2008). This is done by measuring and analysing data from sensors, including multi-axis accelerometers and heart rate monitors (Catapult Innovations Pty Ltd 2008).

In this paper a conceptual framework is proposed as an outcome from a literature review and prototype study of Australian Rules Football Clubs in the AFL. The conceptual framework comprises of four models; 1) fitness and strength conditioning model, 2) tactical and strategic analysis model, 3) rehabilitation model and 4) injury management model. This paper explores the first two models in the conceptual framework showing clubs how GPS can be applied to fitness and strength conditioning as well as tactical and strategic analysis.

The fitness and strength conditioning model is comprised of six components encompassing coaching staff setting benchmarks for players, the completion of specific fitness tasks before and after the game as well as determining if players are fatigued. The model assists coaching staff with player's fitness and strength ensuring they are ready after the game to commence normal training and are not suffering any after effects of the game such as fatigue.

The tactical and strategic analysis model captures the performance of individuals and teams live during the game assisting coaches in modifying tactical or strategic practices or determining if players are in the correct position on the field. It also assists coaching staff to make decisions about moving players to other positions based on the benchmarks they are or are not meeting.

The outcome of this research has been to develop a model for each of the three Catapult Innovations (2008) research areas; fitness and strength conditioning, tactical and strategic analysis and rehabilitation as well as a model for injury management, which is another important application of GPS within the AFL.

Keywords: AFL, Global Positioning Systems, Tactical and Strategic Analysis, Fitness and Strength Conditioning

1. INTRODUCTION

The aim of this research is to develop a conceptual model for the application of GPS within the areas of fitness and strength conditioning and tactical and strategic analysis in the Australian Football League. This will be developed with information from a literature survey and then fine-tuned with results from a prototype study. At present there are no conceptual models that outlines the application of GPS within the two research areas fitness and strength conditioning and tactical and strategic analysis.

2. BACKGROUND

Fitness and Strength Conditioning

Player performance can measured and monitored through GPS with systems measuring the velocity of a player and their movements; the distance players run; the accelerations a player makes throughout the game; the player load through both running and non-running activities and the heart rate of the player (Catapult 2009). This allows coaches to replicate and simulate game demands allowing them to make more informed decisions based on a specific set of circumstances whilst aiding players in decision making (Dawson et al. 2004). Furthermore the physical intensity during training sessions can be measured by comparing data from games to the data received from the devices worn by players at training sessions. Coaching staff can then view data retrieved from the units in both situations and narrow down which areas players need to improve in order to replicate game demands, such as 100 meter sprinting, intensity, the force of tackles laid and player work rate. A study conducted by AFL researchers in 2008 tracking 1500 players using GPS devices during matches found that "players now were not running at top speed (above 25 km/h) as much as they were and they were running a slightly less overall distance and they were running at a higher average rate" (Gleeson 2009, para. 5). This is supported by information such as the average player "travel[s] approximately 15kms during a game but half of that is generally jogging or walking, so they run 8-10kms and run intensely for 3-4kms" (Sydney Swans 2006, para. 10). This allows coaching staff to determine how far a player can travel at a particular speed during a match before coaching staff need to rest them and rotate them on the field for a player who can perform more optimally.

"At the most basic level, the strength and conditioning professional is concerned with maximising physical performance and must therefore conduct programs that are designed to increase muscular strength, muscular endurance, and flexibility" (National Strength and Conditioning Association 2008, p.4). When examining fitness and strength conditioning GPS devices allow coaching staff to measure players' physical effort and compare it to the effort of the players on match day. It also allows coaches to design a training program for individual players as well as the whole team as player's physical limits are known. Furthermore the exertion of players is transparent which allows coaching staff to know exactly what is happening with an individual player at a given time (2009, pers. comm. 25 August).

It can be determined from the literature that the use of GPS for fitness and strength conditioning provides the potential for coaching staff to replicate game demands in a controlled environment and enables clubs to minimise the risk of injury to their players during training drills. This simulates players' match experiences in training drills whilst minimising the chances of injury during peak playing performance. This is done by slowing the training session down or resting players when GPS software informs coaching staff they are exceeding specific benchmarks and thresholds. "That in turn means that in matches our players can run faster and for longer periods and are not as fatigued when they get the ball, which means they are more likely to execute skills properly" (Pavlich 2008, para.21).

Monitoring player fatigue is a critical component of fitness and strength conditioning. Wilmore, Costill and Kenney (2008, p. 113) define fatigue as "the decrements in muscular performance with continued effort". When applying this definition to elite Australian sport, fatigue is the decrease in muscular performance with continued effort over a period of time (Wilmore et al 2008) which can be caused by a number of factors including low blood glucose, depletion of muscle fuels, overheating and dehydration. Slobounov (2008, p. 77) adds that fatigue "can refer to both physical and mental exhaustion due to prolonged stimulation or exercise". GPS data can aid in determining when a player is fatigued. Depending upon the predetermined benchmarks, the GPS software informs coaching staff when a player reaches their fatigue benchmarks allowing them to decide whether the player should be removed from the field. Although there is literature about the impact of fatigue on players and the increased risk of injury there is no mention on prevention or management either with the use of GPS devices.

Symptoms of fatigue include: performance decrements; decreased effort during exercise; decreased work rate and lactate threshold; sore or aching muscles; muscle weakness; slowed reflexes and responses; impaired decision making and judgement; impaired hand to eye coordination, poor concentration and a reduced ability to pay attention to the situation at hand (Garrett & Kirkendall 2000 ; Victorian Department of Human Services 2007). Therefore when a player shows evidence of these symptoms and continues their role on the field, it could have an adverse affect not only on the player, with the potential for injuries, but for the team and the result of the match. Collingwood Football Club fitness coach David Buttifant believes “the potential to monitor players' fatigue levels could mean shorter injury lists” (Quayle 2007, p. 5) whilst GPSports founder and managing director Adrian Faccioni supports this by arguing “the technology would help reduce injuries caused by fatigue” (Quayle 2007, p. 5). Reducing the risk of an injury occurring can take place by Ways to reduce the risk of injury are to removing and/or rotating players on and off the field as coaching staff are alerted when players is fatigued.

A study conducted on the affect of fatigue on the decision making of water polo players when shooting goals found “that incremental increases in fatigue differentially influenced decision-making, technical skill performance, and accuracy and power of a goal-shot” (Royal, Farrow et al. 2003, p. 83). Similarly in AFL football fatigue affects the accuracy and power of a goal shot although different parts of the body are used to score goals eg. the arm for water polo and leg for football. The effects of fatigue within both sports are still the same. When this concept is directly applied to AFL this means that each time a player’s level of fatigue increases their ability to make appropriate decisions, the way they perform skills and actions, their body motions and accuracy are all heavily affected and impaired. Therefore it is suggested that when players become fatigued, which is flagged by the use of GPS software, they are immediately removed from the field. If players remain on the field it can result in poor decision-making resulting in opposition goals, injuries and performance degradation. Which has an impact on the overall team not just individual player(s) experiencing fatigue.

Fatigue encompasses more than the decrease in physical performance. Marcora, Staiano and Manning (2009, p.857) define mental fatigue as “a psychobiological state caused by prolonged periods of demanding cognitive activity and characterized by subjective feelings of ‘tiredness’ and ‘lack of energy’”. Further, through higher perception of effort rather than cardio respiratory and muscular mechanisms, mental fatigue limits exercise tolerance in humans (Marcora et al. 2009). Hence, players who become mentally fatigued perceive they are still operating at optimum performance, but this is merely a perception of the effort they are putting in rather than the actions their bodies are making in order to achieve this performance. This reflects their perception of the effort they are exerting rather than their actual performance levels (Marcora et al. 2009). AFL Chief Operations Manager Adrian Anderson has determined from GPS data "Team-based success closely related to team work rate - harder-working teams finished higher on the premiership ladder” (Barratt 2009, para 3). This statement highlights how critical it is for the GPS data to be monitored during games ensuring that when a player’s work rate declines they are replaced by a player who can perform at a higher work rate. Hence, it is critical that players are removed from the field at the onset of fatigue when detected and replaced with a fit player who can perform at optimum pace.

As it can be determined from the literature there is not presently a model or guide for the ‘best practice’ implementation of GPS for fitness and strength conditioning. More specifically one that encompasses all aspects of the area such as the ability to set benchmarks for players based on statistics or information received from previous training sessions, the ability to determine if a player is fatigued and the measurement of players performing specific tasks.

Tactical and Strategic Analysis

Tactical and strategic analysis using GPS data allows coaches to examine team structure and player movement (Catapult 2009) which may include clusters of players, known as the forward, midfield or backline. This type of analysis may be combined with synchronisation of video footage (Catapult 2009). Further, it allows coaching staff to manage individual player interchanges for greater effectiveness.

As AFL has become more physically demanding and faster (Le Grand 2008), the role of the interchange bench has also evolved. In the 2004 AFL season, for example, teams averaged 30 player rotations, while they are now (2008) averaging 80 rotations per game, which is an increase of over 250% on previous years (Wellman 2008). Just in the past years alone (2008) player rotations have increased by more than 30% (Wellman 2008). The large increase since 2004 (Wisbey et al. 2008), is attributed to a combination of multiple factors which have been noted from GPS data, including monitoring of player heart rates; the decision to rest players when they reach specific benchmarks, the ability for coaches to set individual benchmarks and thresholds for players; player recovery time; player fatigue with respect to distance travelled; and the G-force measured from tackles (Wisbey et al. 2008). Currently there is no literature in regards to individual or types of benchmarks set by AFL coaching staff for elite athletes, this is at the discretion of fitness staff based on their individual goals for players and the team.

When using GPS data during training, the coaching team set benchmarks for individual players based on their fitness; aerobic strength; heart rate; previous or current injuries and the distance they can cover at certain speeds before experiencing fatigue (Edgecomb & Norton 2006). During the match, when a player exceeds their benchmark, the coaching team is informed, providing the opportunity to rest the player. There is some evidence suggesting the decision is automatic, for example Le Grand (2007, p.43) states, “whenever a player starts to ‘red-line’ on any given measure, he is brought to the bench and replaced by a team-mate with fresh legs, enabling the game to continue at maximum speed and intensity”. Fremantle Football Club’s strength and conditioning coach Ben Tarbox adds “from a physiological perspective, you can determine how hard they can run for a certain period of time and how long they need to recover” (Le Grand 2007, p.43). However, in practice, if a player is not making undue errors, or if other players are more fatigued, the player may remain on the field. Ultimately the decision remains with the coaching staff. However, GPS is likely to provide more accurate information, upon which they can make these decisions,

Dawson et al. (2004, p. 292) states that “sports scientists and coaches spend a great deal of time planning training drills and programs that are designed to stimulate game demands by replicating the physiological skill and decision-making requirements of actual competition”. The ability to recreate game demands during training sessions provides the potential for coaches and their staff to gain competitive advantage from information received by GPS units and incorporating the information into training drills designed to simulate game demands and focus on core components of player’s fitness and skills.

As it can be determined from the literature there is not presently a model or guide for the ‘best practice’ implementation of GPS for tactical and strategic analysis, more specifically a model that encompasses all aspects of the area such as the capture of team performance, monitoring of player structure and movement and a post-game GPS data review.

3. METHODOLOGY

A prototype survey comprised of sixty-eight qualitative and quantitative questions divided into six categories; Players and teams, data, GPS units, injuries, training and restrictions was given to a group of ten coaching staff who are currently or have previously coached clubs in the South Australia National Football League (SANFL), Junior South Australian State team coaches and coaches of Under 18 teams. The group of ten participants completed the survey and made suggestions regarding the readability, structure and usability of the survey. This resulted in changes to the way seven questions had been worded to easier comprehension.

Two data analysis techniques were used in this research 1) statistical analysis to analyse the quantitative answers and 2) the use of Leximancer software to analyse the qualitative answers.
The survey was then posted to the senior coach of each of the 16 AFL clubs to complete.

The use of statistical analysis allows the calculation of the percentage of ‘yes’ and ‘no’ answers received for each question and how the two percentages correlate. Leximancer assists and confirms concepts and relationships the researcher has already discovered and established. The software “takes a substantial body of text and rapidly consolidates it into meaningful ‘Themes', ‘Concepts' and their associated relationships” (Leximancer 2009, para. 1).

4. DISCUSSION

Response Rates

Twelve responses were received from AFL clubs out of a possible sixteen. Three of these are deemed unusable as two respondents selected both yes and no responses for some of the quantitative questions whilst the other returned a partly completed survey. This resulted in a valid response rate of 56% therefore statistical analysis cannot take place as there is insufficient data to perform statistical analysis.

Selected clubs chose to follow up and provide further detailed information for the survey questions via phone and email. This assisted in ensuring that survey answers were correctly interpreted validated and the results obtained by the research were correct.

Fitness and Strength Conditioning

The fitness and strength conditioning model was developed based on the knowledge that coaches are setting benchmarks based on information and training data from the pre-season, as well as past seasons for players who have been at the club longer than a year then using this information after a game to determine if a player is fatigued. If a player is experiencing fatigue then it is recommended their training program is redesigned until they have recovered. This will ensure their risk of receiving an injury is not heightened and the player won't go into the next game without completely recovering. It demonstrates the processes of using GPS data from training sessions and previous seasons to measure fitness and performance before and after a game. It also shows a link between the application of GPS and managing player fatigue post game; this can be seen in Figure 1.

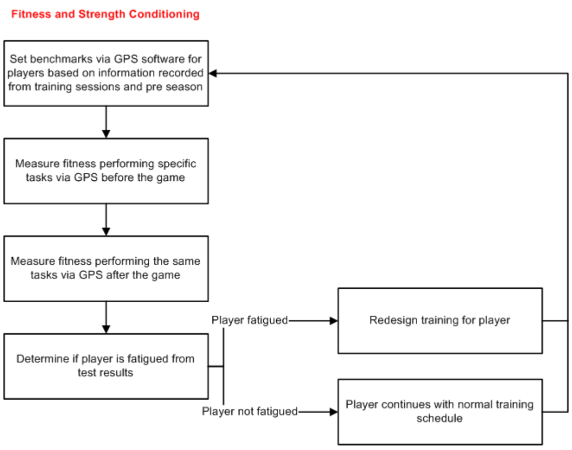


Figure 1 - Fitness and Strength Conditioning Model

The first component of the model involves coaching staff setting benchmarks in GPS software for players based on results from pre-season training and for players who have been at the club for more than one season, the previous years games and training sessions; for example; how far a player can run before experiencing fatigue, bands of running speed and efforts made during a contests and heart rate. Individual benchmarks are set for each player based on ability, age and their position/role on the field and within the team. Benchmarks alert coaching staff when upper or lower thresholds are not being met or are being exceeded, they also informs coaching staff when players are fatigued through a set of alarms and customisable rules within the software.

This links to the second component of the model, which measures a player's fitness/ability to perform a designated task while wearing the GPS device immediately before a game. This assists coaching staff in determining if a player is fatigued post game. Examples of tasks a player might be asked to perform include: a short distance sprint, a jog for a designated period of time and/or distance or agility tasks to measure heart rate.

The third component of the model is conducted immediately following the game and involves the player performing the same tasks conducted during the second component.

The fourth component involves coaching staff determining if the player(s) wearing GPS device(s) are fatigued, taking into account they have just played a game. Individual clubs have their own methods of calculating the results and taking the effect of the game into consideration. The results from the calculation determine if the player(s) who wore GPS unit during the game are fatigued.

The next component depends on whether the player is determined as being fatigued or not fatigued. If the player is fatigued they will move to component five of the model which is involves fitness coaches redesigning or modifying the players training program and workload load for the following training sessions until it has been determined the player(s) are no longer suffering from fatigue. This ensures that the player will not heighten the risk of injuring themselves while suffering from fatigue experienced during the game. It will also ensure they are fit to play the following week and resume their normal training program.

If it has been determined that the player(s) are not fatigued, then component six of the model is applied and they will continue with their normal training schedule to prepare for the following week's match.

Tactical and Strategic Analysis

The tactical and strategic analysis model was developed based on the knowledge that after the game is complete the GPS data is synchronised with the match video and an after match review is performed. Players training programs may then be modified and tactical practices changed depending on the results of the match review. Also during the game coaching staff monitor a player's time on the field, their structure and movement and make rotations when certain benchmarks are not met or have been exceeded or a player is experiencing fatigue. This can be seen in Figure 2.

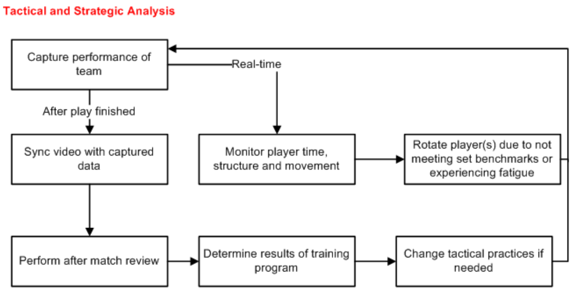


Figure 2 - Tactical and Strategic Analysis Model

The first component of the model involves capturing the team's performance through GPS technology and video. Once this step is complete the model forks into two different paths one which takes place after match play has been completed (component 2) and real-time during the game (component 6)
After match play has finished video that was captured from the game is synchronise with players GPS data making it easier for coaching staff to be able to see what led to specific data produced by the player in any situation. Overall it allows coaching staff to watch both the video and GPS data side by side or through an overlay view. This makes up component two of the model.

The third component of the model involves performing an after match review which involves the coaching staff firstly reviewing the video and GPS data as discussed previously by themselves and then taking 'snippets' of the video to show individual or groups of players in firstly their own individual player reviews and then their playing group reviews ie. The ruckmen, forwards, backline or midfielders.

Determining the results of the training program designed by coaching and fitness staff makes up the fourth component of the model. This examines how players performed during the game and if the current training plan is meeting the requirements the coaching staff have set for players and their performance.

If tactical practices are required to change this takes place as part of component five in the model after the results of the training program as discussed in component four. Once this process is complete the model begins again from process one with the inclusion of the new tactics.

The second fork in the model takes place real-time during the game (component six). Player's time on the field, structure and movement is observed to check they haven't been on the field too long ensuring they are constantly playing an effective role on the field and within the team. Structure is monitored to confirm that zones that are run or specific tactical procedures run/used by players are being executed with full precision and each player in the structure is at their required location for the execution of the play. Movement is also monitored to ensure when a play such as a zone is being run coaching staff can watch the exact execution of the play and the flow of movement between players and their team mate or how individual players are moving on the field.

The seventh and final component of the model is the rotation of player(s) due to their inability to meet set benchmarks by the coaching staff or the onset of fatigue. If this occurs the player(s) is either rotated to another position on the field depending on their position/role within the team or will be sent to the bench and replaced by another play from their team. One this phase of the model is complete the process begins again at stage one.

6. RESTRICTIONS

The ability to implement both the fitness and strength conditioning and tactical and strategic analysis models would be a lot simpler for coaching staff if they were not confined to the ten unit restriction on game day. This restriction only allows ten units to be worn during a game by players on each team. If 22 units could be worn by players on game day, it would allow the models displayed in Figures 1 and 2 to be further developed, leading to more informed decision making on behalf of the coaching staff.

This theory is supported by Gill (2009, para.9) stating that "with the [GPS] technology, in theory you can have 22 players with a GPS [device] on and you can actually see all of them on a computer screen". This allows coaching staff to see the structure and movement of any one player at anytime (when the club is permitted to wear 22 units).

Current restrictions are hampering exactly how this information can be used and the quality and quantity of information coaching staff receive with a limit of ten players on the field being able to wear the units in the AFL.

7. CONCLUSIONS

This research provides two models (fitness and strength conditioning model and tactical and strategic analysis model from a framework to practically apply GPS to both training and games for AFL clubs.

Data gathered using GPS devices influences the fitness and strength of players through the ability for coaching staff to determine the maximum physical effort a player can output during a training session or game before coming fatigued. It also allows the opportunity for players to be rested when these benchmarks are met allowing coaching staff to constantly keep ‘fresh’ players on the field. The GPS results from the game allow coaching staff to modify training sessions for fatigued players, which provides extra recovery time and prevents over-training for players who have had a heavy running game.

Tactical and strategic analysis is significantly aided by the use of GPS. It gives coaching staff the ability to see exactly where any one player is at any one time live during the game, as well as being able to see the formation of tactical plays such as the rolling zone to determine how it is being executed and make any changes if needed.

Acknowledgements

I wish to thank the twelve AFL clubs who participated in the study as part of my honours thesis project.

References

Catapult 2008, The Application of GPS in AFL, Micro Technology CRC, Melbourne, viewed 31 July 2008, <http://www.microtechnologycrc.com/afl.html>

Catapult 2009, Catapult, Melbourne, viewed 2 January 2009, <http://www.catapultinnovations.com/minimaxx.html>

Dawson B, Hopkinson R, Appleby B, Stewart G & Roberts C 2004, 'Comparision of training activities and game demands in the Australian Football League', Journal of Science and Medicine in Sport, vol. 7, no. 3, pp. 292-301.

Edgecomb S & Norton K 2006, 'Comparison of global positioning and computer-based tracking systems for measuring player movement distance during Australian Football', Journal of Science and Medicine in Sport, vol. 9, no. 1-2, pp. 25-32.

Garrett W & Kirkendall D 2000, Exercise and Sport Science, 1st edn, Lippincott Williams and Wilkins, Philadelphia.

Gill K 2009, Crows Investing in Technology, Adelaide, viewed 2 May 2009, <http://www.afl.com.au/news/newsarticle/tabid/4417/newsid/75570/default.aspx>.

Gleeson M 2009, AFL researches changes, Melbourne, viewed 15 February 2009, <www.realfooty.com.au/news/news/afl-researches-changes/2009/02/12/1234028205772.html >.

Le Grand C 2007, 'AFL's hi-tech future within reach', Australian, 7 April 2007, p. 43.

Le Grand C 2008, Malthouse sets pace of change, The Australia, viewed 13 October 2008, <http://www.theaustralian.news.com.au/story/0,25197,23431882-2722,00.html>

Leximancer 2009, Leximancer, Leximancer, Brisbane, viewed 1 April 2009, <https://www.leximancer.com/products/>.

Marcora S, Staiano W & Manning V 2009, 'Mental fatigue impairs physical performance in humans', Journal of Applied Physiology, vol. 106, no. 3, pp. 857-864.

National Strength and Conditioning Organisation 2008, Essentials of strength training and conditioning, 3rd edn, Human Kinetics, Champaign.

Pavlich M 2008, Five days on the training track with Pav, Perth, viewed 31 March 2009, <http://www.fremantlefc.com.au/tabid/7009/default.aspx?newsid=70533&print=true>.

Quayle E 2007, 'Satellites to keep track of players', Sunday Age, 27 May 2007, p. 5.

Respondent 3 2009, AFL Study, Kelly Foreman, Adelaide, p. 1.

Royal K, Farrow D, Mujika L, Halson S & Pyne D 2003, The Effects on Fatigue on Decision-Making and Technical Skill Performance in Water Polo Players, Elsevier, National Convention Centre, Canberra, pp. 83-85.

Slobounov S 2008, Injuries in Athletics: Causes and Consequences, 1st edn, Springer US, New York.

Sydney Swans 2006, GPS helping the Swans, Sydney, viewed 31 March 2009, <http://www.sydneyswans.com.au/Season2007/News/NewsArticle/tabid/7106/Default.aspx?newsId=25048>.

Victorian Department of Human Services 2007, Faituge Explained, Better Health Channel, Victoria, viewed 5 May 2009, <http://www.betterhealth.vic.gov.au/BHCV2/bhcArticles.nsf/pages/Fatigue_explained?OpenDocument>.

Wellman S 2008, Speed in defence, Melbourne, viewed 31 March 2009, <http://www.afl.com.au/news/newsarticle/tabid/208/newsid/65747/default.aspx>.

Wilmore J, Costill D & Kenney L 2008, Physiology of sport and exercise, 4th edn, Human Kinetics, Array Champaign, Illonois.

Wisbey B, Rattray B & Pyne D 2008, Quantifying Changes In AFL Player Game Demands Using GPS Tracking - 2008 AFL Season, FitSense, Melbourne.

AN INTEGRATED MONITORING AND ANALYSIS SYSTEM FOR PERFORMANCE DATA OF INDOOR SPORT ACTIVITIES

Per Wilhelm[a,d], Patrick Thomas[a], Emad Monier[a], Robert Timmermann[b], Michael Dellnitz[b], Felix Werner[c], Ulrich Rückert[c]

[a]*System and Circuit Technology, Heinz Nixdorf Institute, University of Paderborn*
[b]*Applied Mathematics, Department of Mathematics, University of Paderborn*
[c]*Cognitronics and Sensor Technology, Center of Excellence Cognitive Interaction Technology, Bielefeld University*
[d]*Corresponding author: per.wilhelm@hni.upb.de*

Abstract

The wish of many sports scientists and trainers is accessing performance diagnoses data of athletes during training or competition. This data concerns the external conditions (e.g. speed and distance) as well as the internal (physical) strain of the players. For collecting this performance data, we have developed an analysis system, consisting of a high resolution video-system together with a wireless sensor network.
In order to record the physiological data (heart rate) of the athlete, a custom-built sensor module has been developed and integrated into a sports shirt. The integrated sensors collect the physiological data. Following the data collection some signal processing is optionally performed and the data is transmitted via a wireless communication technology to a central computer.
We use an adopted Suunto Oy Foot Pod to measure online the current speed of an athlete and compute its overall distance through integration.
In physical and tactical analysis of indoor sport games path information of the players is of great importance. In order to acquire players' path information, a training session or game is captured by a video-system consisting of two cameras which are mounted in the ceiling of a sports hall. The video data is post-processed in order to identify positions of the players and to track all players on the field.
The recorded data of the mobile devices can be processed and visualised online. For example during the sports event, the heart rate can be monitored and the trainer can decide on substituting a player based on his heart rate profile. Another application of our system is the substantial evaluation of the covered distance of basketball players per quarter. The results of this study will be presented in this paper.

Key words: Performance analysis, video tracking, wireless sensor network

1. INTRODUCTION

For individual sports, mainly endurance disciplines, products are already available for data recording and analysis in various forms. However, a gap exists in the area of performance diagnoses in different types of sports, especially team sports, in which complex movement patterns are common and where contacts between sportsmen occur.
Video-based analysis is a common tool for analysing sport games in technical and tactical aspects. In recent years, video analysis also became an instrument for measuring performance parameters such as the overall covered distance per athlete. Individual performance analysis of players and team strategy investigations require information about the athletes' positions during the games. Hence, many European soccer clubs have equipped their sport grounds with multi camera-systems for player tracking. The advantage of video-based tracking systems is that they are entirely passive; the athlete does not have to wear any kind of sensor or marker.
The most popular tracking system is Amisco Pro distributed by MasterCoach Int. GmbH (AMISCO, 2010). The game is captured with up to eight cam-

eras for a rough online and detailed offline analysis. One or two days after a game, the coach receives a complete analysis including performance data of the players. The disadvantage of Amisco Pro is that it is based on infrared cameras. Thus, the tracking results need to be visualised in a virtual environment. Moreover, many cameras including their synchronisation effort makes the system unaffordable for clubs with less financial resources.

In this paper we describe the Sports Performance Analyzer (SPA). SPA aims on providing a new platform for analysing team sports. With only two ceiling-mounted cameras it is ideally suited for indoor sport activities such as handball, volleyball, basketball, and (ice-) hockey which are the most popular sports in Germany besides soccer (DOSB, 2009).

One part of SPA is a video tracking system that reliably computes the positions of the players during a game. Based on this data we can calculate external information, e.g. overall covered distance, speed, and acceleration of single players. This information is used for further higher level analysis such as team strategy as well as performance and fitness of the players which can help the coaches to improve their training methods.

In addition to the external information, SPA also considers the internal (physical) strain of the players indicated for example by their heart rate (HR). For monitoring the HR we cannot avoid equipping the sportsmen with a sensor module. We have developed a custom-built sensor module which is integrated into a sports shirt in order to minimise the impact on the player. Instead of transmitting the processed features (e.g. HR) to a watch and store the data on this device, we transmit the data wirelessly to a central computer. One additional attribute of SPA is the online measurement of speed and distance: We adopt the commercially available Foot Pod sensors[1] to measure the current speed of an athlete which lets us compute his overall covered distance through integration. All recorded data of the mobile devices is processed and visualised online (in real time).

As shown in figure 1, the SPA system has three main modules: data acquisition (video-system (2.1.2) and wireless sensor network (2.1.1)), tracking (2.1.3) and analysis/visualisation. The acquisition module is responsible for recording video streams from two cameras and the data of the wireless sensor nodes. The recording of video and wireless sensor data streams is synchronised to make further analysis of the data easier. Because the amount of wireless sensor data is small compared to the video data, it can be processed and visualised online. For example during a sports event, the heart rate can be monitored and the coach can decide to substitute a player based on his heart rate profile.

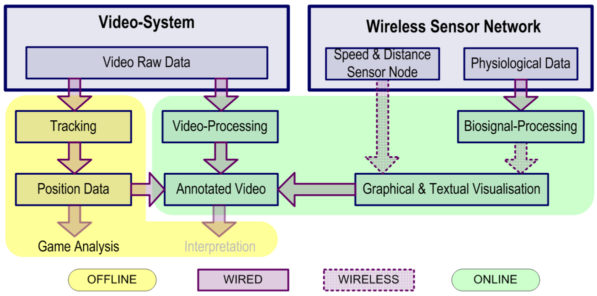


Figure 1: System structure of the Sports Performance Analyzer (SPA).

The video tracking module works offline to extract position data of the players. It utilises the two video streams to produce the positions of the players in real world coordinates (meters) which can be postprocessed to gain further information. The analysis/visualisation module processes three inputs: video, wireless sensor, and position data. It produces different visualisations such as graphs and (interactive) videos with annotated information.

In section 2.2 we present a method for the identification of breaks during a basketball game: For five given player trajectories, we introduce different methods for velocity computation to define a game's state, consisting of position and velocity information for every time step. We use this data to train a Gaussian mixture model that can be used to classify the states of an unclassified basketball game as either game or break.

One application of our system is the substantial evaluation of the covered distance of basketball players per quarter. We have recorded and analysed 14 German major league basketball games of the team Paderborn Baskets with regard to the covered distance of every player of the team. Schmidt (2003) presents in a previous study a value of approximately 23km for the covered distance per team and game. Because this analysis was done manually, the statistical data base is only one game. The results of our study will be presented in 3.3.

[1]Suunto or Garmin Foot Pod based on an acceleration sensor of Dynastream Innovations.

2. METHODS

2.1. MONITORING AND ANALYSIS SYSTEM

A schematical representation of our system is shown in figure 2. The system consists of two data acquisition modules, namely the *video-system* and the *wireless sensor network*. Both modules are integrated in one software solution (SPA) and they can work independently or together.

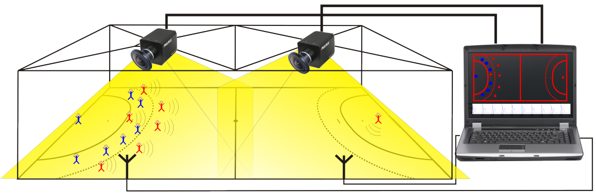


Figure 2: System for capturing match action and recording physiological data.

2.1.1. WIRELESS SENSOR NETWORK

One group in our sports department conducts research on skin temperature and skin conductance with the aim to better understand the interaction between physical and mental stress (Baumeister, 2008). To fulfil their request to acquire more relevant data of the athlete than the heart rate (still the most important physiological parameter for sport scientists), we have developed an advanced brest belt module. Our solution is extremely mobile (lighter than 50g) and can collect skin temperature and conductance, heart rate, and additional information from an onboard 3-axis acceleration sensor. The module itself consists of a motherboard with

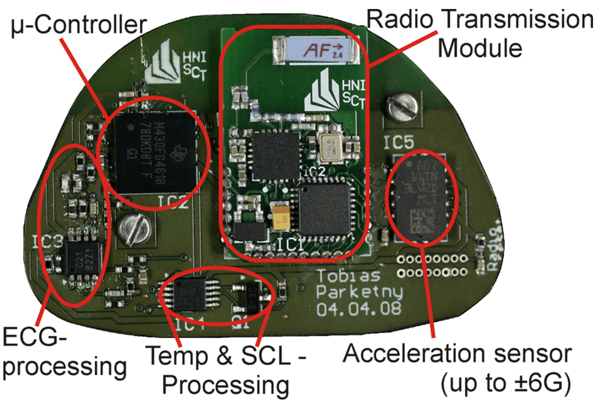


Figure 3: Breast belt module with integrated sensors, evaluation and communication unit as well as power supply.

an additional radio transmission module (daughterboard). For online processing we have equipped the motherboard with a 16bit low-power RISC microcontroller, containing 12bit A/D converters and three operational amplifiers. The microcontroller is powerful enough for processing algorithms, e.g. heart rate detection. Together with the communication stick (daughterboard), the battery lifetime for the complete module is more than 24 hours in operation (general cell coin - 220mAh).

Technical Data Radio Transmission Module:

- Topology: multipoint-to-point (star)
- Frequency band: 2.4GHz
- Range: 30m
- Max. number of sensor nodes: 30
- Less than 10% packet loss
- Power consumption:
  - 35mW in operation (TX mode)
  - 21μW in sleep (power down mode)
  - 70.5μW average[2]

For rough online results of the covered distance (±10%), we adopt the commercially available Foot Pod products. For a detailed later (offline) analysis, motion capturing (video tracking) methods are used.

2.1.2. VIDEO-SYSTEM

The afore mentioned indoor team sports are played on field sizes up to 40m × 20m. Assuming a minimum hall height of 7m, a field of vision of more than 150 degrees is required. No commercially available lens is able to map this range of vision without distortion. Even using one single fisheye lens will lead to too much information loss close to the back lines. To solve this problem, we have installed two video cameras. They are placed at the hall ceiling, one over the middle of each half of the field, recording the game from a bird's eye view. A fisheye lens is used in order to capture the required view. The selected megapixel cameras are equipped with a Bayer CCD sensor and a Gigabit-Ethernet interface. Each camera is capable of delivering up to 30 frames per second (fps) which causes a data rate of more than 30MB/s. With an up-to-date desktop processor a live preview can only be done with a reduced resolution and/or reduced frame rate. For a preview in high definition with full frame rate a hardware support by graphic accelerators or FPGAs is necessary.

[2] Provided that each packet requires 300μs to be transmitted (32Bytes@1 Mbit/s) and the packet rate is 5Hz.

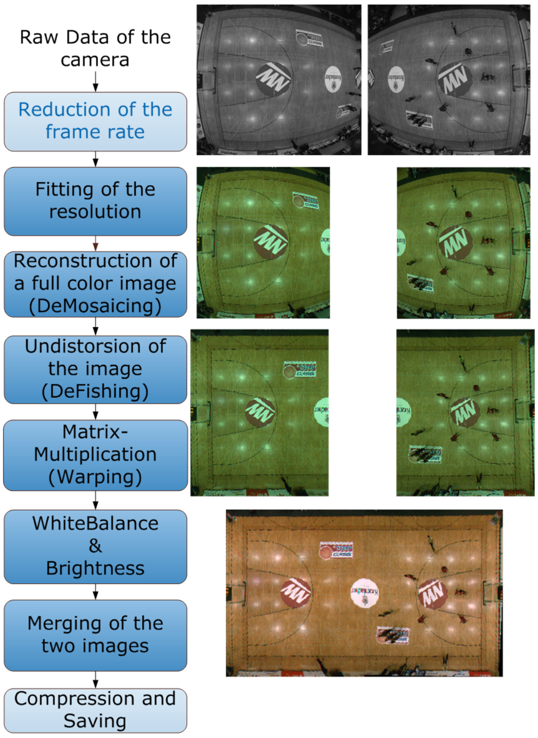


Figure 4: Image pre-processing steps.

As input data for the image pre-processing serve the full raw images of the camera. First, a region of interest (ROI) is selected and a colour reconstruction (DeMosaicing) is done for this part of the image. Then geometrical transformations are executed for producing undistorted images (DeFishing by Look-Up-Table, Warping by Matrix-Multiplication). Finally, a white balancing is performed, the two images are merged together for an online preview, and the video can be saved in a file (avi-container with MPEG4 codec). For later (offline) tracking the raw data is saved with a lossless compression algorithm (lagarith codec).

2.1.3. VIDEO TRACKING

The tracking algorithm used in our system is Template Matching (Lewis, 1995). This method is used to find the parts of an image which match with a reference image (template). A template of the upper part of the body (head and shoulders) is used to search for the player in the next frame. The player's shape changes slightly between two consecutive frames so the template is adapted. The template image is compared to all parts of the searched image and a measure of similarity is computed in each comparison step. The position with the highest value of similarity is the possible position of the template in the searched image.

A strategy based on partitioning of the search space is used to handle the tracking of multiple players. The tracking is done under human supervision to correct the errors that cannot be handled automatically. A detailed description of our tracking algorithm can be found in Monier (2009). The tracking itself is performed on the distorted raw images because no benefit can be achieved by pre-processing the images. As mentioned above, the images are recorded using a fisheye lens. For the purposes of creating distortion-free images and converting the tracked image-positions to real-world positions, a number of transformations have to be applied. The series of steps is presented in figure 5, steps 1-4:

The first step is the undistortion of the points in the fisheye image (**DeFishing**). Because of small variations in the camera position and viewing-angle, a software calibration (**Warping**) has to follow. The corrected world coordinates are mapped from the players **head to** his **foot** position. The final position describes the foot position of the player on the field. All further steps of transformation (5-8) are needed during tracking if a player changes between the two field sides, because the tracking will be continued with the other camera. On the opposite side of the field, the transformations have to be inverted to map the real-world coordinates to the distorted image. As a final step, we smooth the calculated foot coordinates in the world coordinate space using a moderate zero-phase low pass filter[3].

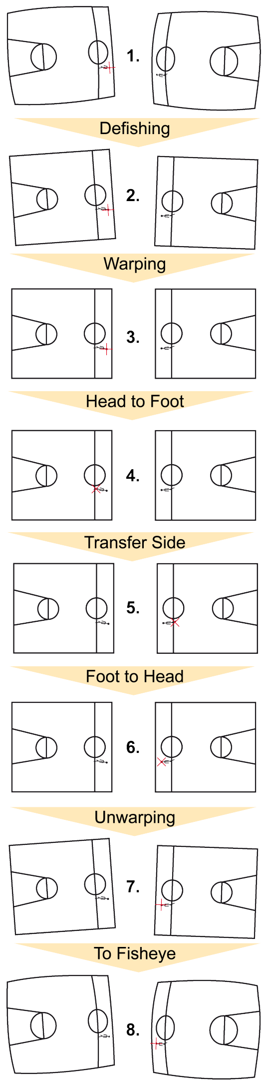


Figure 5: Coordinate Transformations.

[3]Digital FIR-filter, order = 16, $\omega_n = 0.1$.

2.2. NET-TIME COMPUTATIONS

In this section we present a method to divide each quarter of a basketball game into *action* and *break* parts. These results can be used for post-game analysis, for example to calculate the distances covered by the players or their average velocities. A similar application can be found in (Perse, 2009).
Although the parameters in the presented method are fine tuned for basketball games, this method could be adapted to other team sport games.

2.2.1. DATA

We have tracking data from a number of basketball games available. This data consists of the positions of all active players over each quarter. It is used to construct a state $(position, velocity)$ of the game for every timestep t_i, $i = 1, \dots, N$.
For the position information we use the five player positions $(x_{i,1}, y_{i,1}) \dots (x_{i,5}, y_{i,5})$ of the five active players in each timestep t_i and calculate the average position of the players via

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \frac{1}{5} \sum_{p=1}^{5} \begin{pmatrix} x_{i,p} \\ y_{i,p} \end{pmatrix} . \tag{1}$$

We propose a total of five different methods to compute the velocity data. Two of these methods generate a scalar value whereas the other three methods yield a two dimensional vector as a result. We will compare these methods in section 3.2. For all methods we need a time interval of $t_{s,i} = t_i - t_{i-1}$ to scale velocities to m/s. In our case the length of all time intervals is constant and equal to the frame rate of about 1/30th of a second, thus $t_{s,i} = t_s = const.$

I. The first method is to calculate the vector from one average trajectory point (as introduced in (1)) to the next:

$$v_i = \frac{1}{5\,t_s} \sum_{p=1}^{5} \left(\begin{pmatrix} x_{i,p} \\ y_{i,p} \end{pmatrix} - \begin{pmatrix} x_{i-1,p} \\ y_{i-1,p} \end{pmatrix} \right) .$$

II. For the second method, we simply compute the norm of velocity I.

$$v_i = \frac{1}{5\,t_s} \left\| \sum_{p=1}^{5} \left(\begin{pmatrix} x_{i,p} \\ y_{i,p} \end{pmatrix} - \begin{pmatrix} x_{i-1,p} \\ y_{i-1,p} \end{pmatrix} \right) \right\|_2 ,$$

thus reducing the dimension of the velocity information to one.

III. The third method is a variation of II.:

$$v_i = \frac{1}{5\,t_s} \sum_{p=1}^{5} \left\| \begin{pmatrix} x_{i,p} \\ y_{i,p} \end{pmatrix} - \begin{pmatrix} x_{i-1,p} \\ y_{i-1,p} \end{pmatrix} \right\|_2 .$$

This value will always be greater or equal than method II. It represents the varying velocities of the players better, because the velocities of two players moving in opposite directions do not cancel each other out.

The last two methods represent the velocity information in polar coordinates. Both utilise the angle φ between the x-axis and the velocity vector (see I.), which is defined[4] as

$$\varphi = \operatorname{atan2}\left(\sum_{p=1}^{5} \left(x_{i,p} - x_{i-1,p} \right), \sum_{p=1}^{5} \left(y_{i,p} - y_{i-1,p} \right) \right) .$$

IV. The fourth method is a combination of the angle φ and method II.:

$$\begin{pmatrix} v_{i,1} \\ v_{i,2} \end{pmatrix} = \left(\frac{1}{5\,t_s} \left\| \sum_{p=1}^{5} \left(\begin{pmatrix} x_{i,p} \\ y_{i,p} \end{pmatrix} - \begin{pmatrix} x_{i-1,p} \\ y_{i-1,p} \end{pmatrix} \right) \right\|_2 , \varphi \right)^T .$$

V. Finally, the fifth version is a combination of φ and method III.:

$$\begin{pmatrix} v_{i,1} \\ v_{i,2} \end{pmatrix} = \left(\frac{1}{5\,t_s} \sum_{p=1}^{5} \left\| \begin{pmatrix} x_{i,p} \\ y_{i,p} \end{pmatrix} - \begin{pmatrix} x_{i-1,p} \\ y_{i-1,p} \end{pmatrix} \right\|_2 , \varphi \right)^T .$$

For each timestep we get a three or four component vector X_i that characterises the current state of the game:

$$X_i = (x_i, y_i, v_i) \in \mathbb{R}^3 \text{ or } \quad X_i = (x_i, y_i, v_{i,1}, v_{i,2}) \in \mathbb{R}^4 .$$

Additionally to the data described above, we divided 23 quarters manually into action and break sections. This knowledge can be used to train an appropriate model for the labelling of new quarters.

2.2.2. GAUSSIAN MIXTURE MODEL

We use two Gaussian mixture models (see (Rasmussen, 2006)), one to characterise the action states and one for the break states. These models are trained using the data from section 2.2.1 and

[4] atan2 is a variation of the inverse tangent function and places angles correctly in all four quadrants.

used for the classification of unclassified data. In a Gaussian mixture model a sum over k Gaussian distributions is used to approximate the state of a system

$$P(X_i|m_l) = \sum_{j=1}^{k} \alpha_j P(X_i|\mu_j, \Sigma_j)\,,$$
$$\text{and } \sum_{j=1}^{k} \alpha_j = 1\,,$$

with $m_l \in \{action, break\}$. We separate our data into two sets: those states when the game's time was running, and those states when the clock was stopped. With these sets, two Gaussian mixture models are trained using the EM-Algorithm (see section 3.2 for details). The training returns two sets of parameters, $(\alpha_j^{(a)}, \mu_j^{(a)}, \Sigma_j^{(a)})$, $j = 1, \ldots, k$, for the points belonging to the action set and $(\alpha_j^{(b)}, \mu_j^{(b)}, \Sigma_j^{(b)})$, $j = 1, \ldots, k$, for the points belonging to the break set.

Our goal is to divide a new quarter into two sets action and break. This can now be done by classifying each state using the models explained above. We use Bayes's rule to compute the probability of a given state X_i of the game to belong to either action or break:

$$P(m_l|X_i) = \frac{P(X_i|m_l)P(m_l)}{P(X_i)}\,. \tag{2}$$

Because the length of one quarter is fixed to 10 minutes, $P(action)$ equals 600s divided by the total length of the quarter in seconds, and $P(break) = 1 - P(action)$. We say, a state belongs to the action phase, if $P(action|X_i) \geq P(break|X_i)$ and it belongs to the break phase, if $P(action|X_i) < P(break|X_i)$. Thus it is not neccessary to know $P(X_i)$ in equation (2) to classify a state.

Results of this method will be presented in 3.2.

3. RESULTS

In this section we are going to present evaluation results of the different modules of our SPA system. Our developed heart rate sensor node is capable of transmitting every single heart beat so that a beat-to-beat analysis becomes possible. Moreover, all wireless sensor data can be visualised online in our SPA software. Unfortunately, the use of these breast belt modules is not allowed in official basketball games, so that we do not have any physiological data to augment our tracking data.

3.1. VIDEO TRACKING RESULTS

Regarding the video-system, the processing rate without correction (f_{auto}) for tracking five players (NoP) is 10fps. The average number of corrections (cR) is 0.004 corrections per frame and player. The average correction time (cT) for one error is 3.3 seconds. Finally, the frame rate for tracking including correction (f_{corr}) is 6fps (Monier, 2009).

$$f_{corr} = \frac{1}{1/f_{auto} + cT \cdot cR \cdot NoP}$$

Compared to the source frame rate of the video (30fps), the processing time is five times longer than the gross playing time. Considering the accuracy of the tracking system, we ran several test cycles which resulted in an accuracy of above 94% (Paier, 2009).

One main application of our system is the evaluation of the covered distance to generate an individual profile for each player in basketball games. The primary output of the video tracking is the filtered position data which is used to calculate the covered distance of the players in the game. For the gross covered distance, we consider all players of the host team for the complete game including breaks[5]. To extract the net covered distance from the position data, we have tested automated methods. Before we present our results in section 3.3 we are going to validate the methods introduced in section 2.2.

3.2. NET-TIME COMPUTATION RESULTS

The numerical computations are carried out using Matlab and its Statistics Toolbox. The `gmdistribution.fit` function, which is part of this toolbox, estimates the parameters for the two Gaussian mixture models using the expectation maximization (EM) algorithm.

To judge the effectiveness of our algorithms, we compare the automatic labelling using the Gaussian mixture model with the manual labelling. We compute the number of correctly classified trajectory points (those points, where automatic and manual labelling yield the same result) and divide it by the total number of points. This correctly classified ratio measures the performance of our algorithm.

[5]Except for official team timeouts.

	velocity computation method				
k	**I**	**II**	**III**	**IV**	**V**
3	0.7888	0.7724	0.8279	0.7725	0.8409
4	0.8076	0.8088	0.8587	0.8092	0.8603

Table 1: Ratio of correctly classified points using the five different velocity computation methods for $k = 3$ and $k = 4$.

Several parameters have to be adjusted in order to maximise the performance of the classification algorithm. The first parameter is the number of Gaussian distributions for the mixture model. We tested $k = 3, 4, 5$ distributions, whereas for $k = 5$ the EM-Algorithm did not converge. The second parameter is the velocity computation method. We carried out all computations using the five proposed methods. The results for $k = 3, 4$ and for the five velocity computation methods are shown in table 1.

Since the best results could be achieved for $k = 4$ and velocity computation method 5, we use these values throughout the rest of the computations.

As a result of the process described in section 2.2.2 we receive a classification of a quarter into action and break parts, that still contains unrealistically many switches between the two. To overcome this problem, we perform a two step post-processing. First, we filter the resulting data using a zero phase digital lowpass filter. Secondly, we remove break sequences, that are less than 5s long. Our analysis of the manually labeled quarters shows us, that only 3 of 407 or 0.74% of all breaks are less than 5 seconds long. Hense, to cut off below 5s seems reasonable.

Using the above described post-processing steps, the correctly classified value improves to an average of 90.42% correctly classified points.

In table 2 our final testing results are shown. We computed the sum of the net distances the five players cover in each quarter using the manual and the automatically computed action and break division. On average the automatically computed value differs from the manually extracted values by 7.75%.

Game	Qtr	d_{man} [km]	d_{auto} [km]	Dev. [%]
1	1	5.702	5.664	0.67
	2	6.112	5.471	11.73
	3	5.513	5.308	3.86
	4	5.388	5.406	0.33
2	1	5.693	5.428	4.87
	2	6.205	5.675	9.34
	3	5.361	5.098	5.16
	4	5.659	5.327	6.24
	5	2.992	2.600	15.09
3	1	6.063	5.416	11.93
	2	5.900	5.455	8.15
	3	5.751	5.496	4.64
	4	6.095	6.609	7.78
4	1	6.219	5.533	12.40
5	1	5.556	5.182	7.22
	2	5.238	4.991	4.95
	3	5.522	4.986	10.76
	4	5.291	4.988	6.09
6	1	6.367	5.425	17.36
7	1	5.910	5.435	8.73
	2	5.400	5.299	1.90
	3	5.995	5.344	12.18
	4	6.290	5.889	6.80
avg.		**5.662**	**5.306**	**7.75**

Table 2: Comparison of the covered net distances of the basketball players. We have manually divided a total of 23 quarters of 7 games into action and break parts. We used this data to calculate the cumulative net distances d_{man} covered by the five players in each quarter (see column three). In column four we present the net distances d_{auto}, that were calculated using the algorithms presented in section 2.2. In column five the deviation of d_{auto} from d_{man} is shown. The last row shows the average of all distances and deviations.

3.3. COVERED DISTANCES

We have analysed a total of 56 quarters from 14 randomly chosen games over two and a half years. Table 3 and figure 6 present the results in textual and graphical form respectively.

	Q1	Q2	Q3	Q4	SUM
GROSS	6996.3	7386.6	7243.0	7901.0	29524.0
NET	5867.0	5545.0	5648.6	5580.6	22641.2

Table 3: Mean gross and net covered distance of a basketball team per quarter.

We can confirm the results of Schmidt (2003), who calculated 23185.6m in average for one team per game.

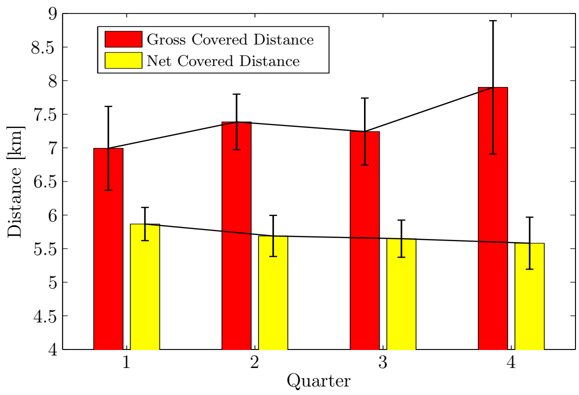


Figure 6: Mean gross and net covered distance of a basketball team per quarter and the standard deviation.

4. DISCUSSION

An application of the tracking system is the substantial evaluation of the covered distance of basketball players per quarter. Additional information, e.g. the covered distance of the players classified by their positions and into different speed ranges can be computed with little effort. Moreover, considering the players' positions, a position specific performance profile can be generated.
As an enhancement of the sensor network we apply receiver diversity technology for better energy-efficiency and communication reliability, respectively. Our system also provides large flexibility for further design improvements, e.g. the implementation of the 3-axis acceleration sensor presented in Christ (2010).
A small drift of the template out of the tracked part presents a problem in the video tracking system. To overcome this problem and to reduce the number of corrections we are going to enhance the tracking algorithm by making use of colour information in future versions. As an alternative to the existing tracking method, we actually test different tracking algorithms (e.g. Particle Filter tracking).
The net-time computation algorithm presented in 2.2 already works quite reliable, but its accuracy could be further improved by training the model with more data. Instead of using manually generated data, we are going to use statistical *play-by-play* data, provided by the Beko BBL.

5. CONCLUSIONS

In this paper we have presented our Sports Performace Analyzer (SPA) system. SPA consists of a video tracking system for indoor sport activities and a sensor network that measures physiological parameters of the players. It makes it possible to visualise the actual performance of the players during training or competition. This knowledge can be used by sport experts to optimise training patterns and game strategies.

Acknowledgements

We wish to thank Paderborn Baskets 91 e.V., Beko Basketball Bundesliga (Beko BBL), Swiss Timing Ltd., the Paderborn Center for Scientific Computations, and many students for their help and support.

References

AMISCO 2010. *http://213.30.139.108/sport-universal/uk/sources/PDF/sup.pdf*.

Baumeister J., Reinecke K., Schnittker R., Weiß M. (2008). Influence of cognitive stimuli during physical load on skin conductance response. *European Journal of Applied Physiology (in preparation)*.

Christ, P., Rückert, U., & Mielebacher, J. (2010). Detection of body movement and measurement of physiological stress with a mobile sensor node in obesity prevention. In proceedings of *The 10th Australasian Conference on Mathematics and Computers in Sport*, Bedford, A. and Ovens, M., eds., Darwin, Australia, 5 - 7 Jul 2010, 67-74.

DOSB - Deutscher Olympischer Sportbund. (2009). Bestandserhebung 2009, *www.dosb.de/de/service/statistiken*.

Lewis, J. P. (1995). Fast normalized cross-correlation. *Vision Interface. Canadian Image Processing and Pattern Recognition Society*.

Monier, E., Wilhelm, P., & Rückert, U. (2009). Template Matching Based Tracking of Players in Indoor Team Sports. *Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2009)*, Como, Italy, 30 - 2 Sep 2009.

Paier, D., Schnittker, R., Reinecke, K., Wilhelm, P., Preis, R., Weiß, M., Baumeister, J. (2009). Physiologische Spielbeobachtung - Testgüte des Videotrackings im Sports Performance Analyzer (SPA). *Deutsche Zeitschrift für Sportmedizin 60(7-8)*.

Perše M., Kristan M., Kovačič S., Vučkovič G., Perš J. (2009). A trajectory-based analysis of coordinated team activity in a basketball game. *Computer Vision and Image Understanding, Volume 113, Issue 5*, 612-621.

Rasmussen, C., Williams, C. (2006). Gaussian Processes for Machine Learning. MIT Press. Cambridge, Massachusetts.

Schmidt, G., & von Benckendorf, J. (2003). Zur Lauf- und Sprungbelastung im Basketball. *Leistungssport, 33 (2003)*, 42-48.

Wilhelm, P., Monier, E., Xu, F., & Witkowski, U. (2008). Analysis of Indoor Team Sports Using Video Tracking and Wireless Sensor Network. *Current trends in Performance Analysis: World Congress of Performance Analysis of Sport VIII*, 345-348.

TOWARDS AUTOMATED FOOTBALL ANALYSIS: ALGORITHMS AND DATA STRUCTURES

Joachim Gudmundsson[a], Thomas Wolle[a,b]

[a]*NICTA Sydney*
Locked Bag 9013, Alexandria NSW 1435, Australia
[b]*Corresponding author: thomas.wolle@nicta.com.au*

Abstract

Analysing a football match is without doubt an important task for coaches, clubs and players; and with current technologies more and more match data is collected. For instance, many companies offer the ability to track the position of each player and the ball with high accuracy and high resolution. Analysing this position data can be very useful. Nowadays, some companies offer products that include simple analyses, such as statistics and basic queries. It is, however, a non-trivial task to perform a more advanced analysis. In our research, we assume that we are given only the position data of all players and the ball with high accuracy and high resolution. In this paper we present two tools.

Our first tool automatically extract (from the position data) a list of certain events that happened during the football match. These events include kick-offs, corner kicks, passes etc. In experiments we could observe that our method is very fast and reaches a high level of correctness. We also learned that errors in the event detection are hard to avoid completely, when looking at only the position data.

Our second tool aims at analysing a single player's trajectory (the sequence of all positions during a game). More precisely, we look for movements of a player that are repeated often (so called subtrajectory clusters). For example a left wing attacker runs from the centre-line along the left side of the field towards the opponent's goal. And this attacker might repeat this type of movement very often during a game (or perhaps multiple games). Our goal is to detect this kind of frequent movements automatically. Experiments showed that this method is computationally expensive. Nevertheless, it reliably identifies subtrajectory clusters, which then could be used for further analysis.

Key words: team sport, position data, trajectory analysis, event detection, clustering

1. INTRODUCTION

Recent years have witnessed massive improvements in tracking technologies that allow for recording the positions of moving entities with high spatial and temporal resolution and also with high accuracy. As a consequence such systems are used more and more often in many domains, including sports, urban planning and development, defence, location based services and animal research. In the world of football (also known as soccer), there are already many companies that provide the capabilities to track the movement of all football players during a match, and they also provide tools to analyse this data to a certain degree.

In our project we focus on applications in team sports and in particular on football. We look at these applications mainly from a computer science point of view. Assuming we are given the trajectories, i.e. the data describing the movement, of all players and the ball, we aim at a more sophisticated analysis that takes the interaction and relationships of different trajectories into account, recognises formations and perhaps even more general patterns and trends. Over the past few years we have developed several algorithms and tools for analysing trajectories. Two of them are presented in this paper.

Our first tool, introduced in Section 3, is algorithmically rather simple. Nonetheless, we consider it as an important tool that is necessary for a more advanced automated analysis. From the trajectories of all players and the ball, it automatically extract a list of basic events that happened during the football match; detected events include kick-offs, corner kicks, passes etc. Experimental results show that our method is very fast and reaches a high level of correctness for many types of events. However, some types of events seem to be hard to detect automatically when looking only at the position data.

Our second tool, presented in Section 4, aims at analysing a single player's trajectory. However, it can be easily extended to take multiple trajectories as input. In the input trajectory we look for subtrajectory clusters, which are movements of a player that are repeated often. This topic is motivated by questions such as "How is the ball transported from the defense region to the attack region?" We present a prototype where a user can specify certain parameters of the cluster, and then the prototype will reliably detect the clusters according to the chosen parameters. The current version of the prototype might not be suited as an interactive tool, because the time to answer a query can be rather long.

We believe our algorithms and implementations belong to the still immature research area that aims at automated football analysis. Despite the young age of this area, it becomes more and more popular and important as can be seen by the increasing amount of related work. For example, Kang, Hwang and Li (2006) propose a method to quantitatively evaluate the performance of football players. Their approach is based on four different measures that include different regions for each player and the kicks the players perform. Another approach that is also based on regions is presented by Fujimura and Sugihara (2005). Their regions are based on the generalised Voronoi diagrams. Grunz, Memmert and Perl (2009) address the analysis of actions in a football match. Their actions are coarser than our basic events, and the used techniques for the detection are very different.

2. PRELIMINARIES

The position of a moving object can be described by the spatial coordinates x, y and possibly also z at a certain time t. Together, these values form the *sample* (t, x, y, z). A sequence of such samples, ordered with respect to time, is called a *trajectory*. A trajectory describes the movement of an object.

The data that we used in our experiments is data from a real football match. It was anonymised and kindly provided to us by ProZone. It includes the trajectories of all the players. These trajectories have a spatial resolution of a decimeter and a temporal resolution of at least ten samples per second. However, we do not know the accuracy of the data. For the analysis we aim to do we also need the trajectory of the ball. Unfortunately, it is not included. It does, however, include a list of annotations of the match, which is a list of events such as "touch" and "pass" – very similar to the events we want to detect. These annotations were created manually by people watching the match (and/or a video thereof). From these annotations, we re-construct the ball's trajectory by using the time and spatial coordinates of each event to create samples of the ball's trajectory. The resolution of the annotations are only one meter and one second, so the obtained ball trajectory is only a rough estimate of the ball's movement. We believe that the results in Section 3 can be improved considerably if the ball trajectory would be given with higher accuracy.

Both our algorithms and prototypes have been implemented in Java. Our experiments were performed on an off-the-shelf PC with an Intel dual core processor running at 2.33 GHz and 2GB of main memory.

3. BASIC EVENT DETECTION

3.1. METHODS

For computing basic events of a football match, we need both the ball's and the players' trajectories. Here, we briefly describe how our algorithm for detecting events works.

Our event detection works on different levels of events. The bottom level is the physical event level. These are events that can be detected without any knowledge of the football rules (we do, however, require knowledge of the dimensions and lines on the football pitch). Bottom level events are for instance "ball-out", "ball-in" and "touch" events. The first two occur when the ball moves out of (or: back into) the pitch. "Ball-in" events can be refined into "throw-in", "corner kick" etc., depending on

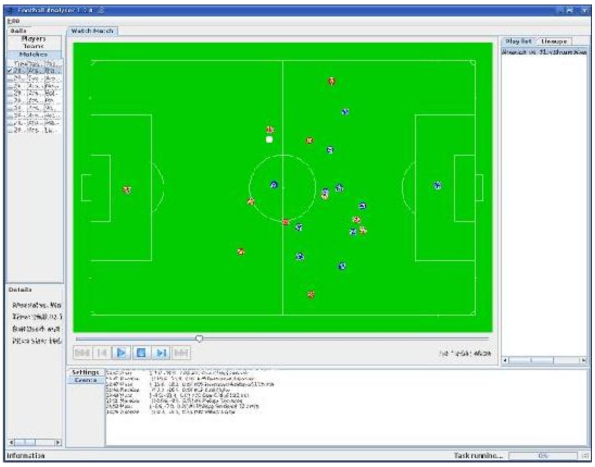


Figure 1: Screenshot of the programme to animate a match. The detected events are displayed in the area below the pitch.

where the “ball-in” event takes place. A “touch” event is detected when the ball changes its speed and/or direction. For these events we only need the ball’s trajectory.

Now, if we also take the player’s trajectories into account, we can refine the events. For instance, the player that is closest to the ball when a “touch” event happens is assumed to have touched the ball (although there are exceptions). Or the player that is closest to the ball, when a “throw-in” happens is assumed to have performed the throw-in. As a result we obtained a list of more advanced events.

These more advanced events can be refined in a second round. For instance, two consecutive “touch” events that were performed by different players from the same team, constitute a “pass” event. In a similar way we can detect “possession”, “pass interception”, “shot on goal”, etc.

In yet another refinement round, we arrive at the highest event level, which includes events such as “free-kick”, “substitution”, “offside”, “foul”, “red-card”, etc.

3.2. EXPERIMENTAL RESULTS

Figure 1 shows the main screen of the event-detecting prototype. It includes a football pitch, where the match can be viewed as an animation. The area below this pitch is the area where the events are displayed synchronously linked to the animation. An example snippet of the list of events is shown in Figure 2.

```
34:51 |Intercept       |( 37.6, -23.9,  0.0)| Player04 | Player08
34:51 |Touch           |( 38.9, -24.6,  0.0)| Player08
34:53 |Ball Out        |( 59.0, -26.0,  0.0)|
35:13 |Corner Cross    |( 57.2, -29.1,  0.0)| Player23
35:13 |Touch           |( 57.2, -29.1,  0.0)| Player23
35:13 |Shot            |( 57.2, -29.1,  0.0)| Player23
35:15 |Touch           |( 54.3, -5.3,  0.0)| Player09
35:15 |Goalkeeper Catch |( 54.3, -5.3,  0.0)| Player09
35:15 |Pass            |( 54.3, -5.3,  0.0)| Player09 | 14.1 m/s
35:16 |Receive         |( 36.3, -2.0,  0.0)| Player17
```

Figure 2: An extract of the list of detected events, showing for each event: time, type, coordinates and involved players etc.

Event type	*falsePositive*	*falseNegative*	F_1-score
touch	7	3	0.979
pass	2	2	0.955
intercept	1	1	0.960
ball-out	1	4	0.935
throw in	0	2	0.963
corner kick	0	1	0.968
goal kick	1	2	0.909
kick off	0	1	0.960
shot	1	3	0.875
goalkeeper catch	0	2	0.941
goal	0	1	0.960
foul	3	6	0.823
offside	5	3	0.556
free kick	0	3	0.953

Table 1: Summary of some statistical results for some of the detected events. The higher the F_1-score the better.

Computing all the events for an entire match takes only a couple of seconds. Hence the running time does not seem to be an obstacle in practice, and therefore we focus more on the study of the correctness of the detected events.

As mentioned above, together with the data that we used for our experiments, we were given a list of annotations. To estimate the correctness of our event detection, we compare our list of events to these annotations, where we consider the annotations as 100% correct (even though they might not be).

The measure that we use to report the level of correctness is the F_1-score, given as:

$$\frac{2 \cdot truePositive}{2 \cdot truePositive + falseNegative + falsePositive}$$

The values of *truePositive*, *falseNegative* and *falsePositive* were simply counted when comparing the list of detected events with the annotations.

Table 1 shows the values of *falsePositive*, *falseNegative* and the resulting F_1-score for some of the detected events. Note that some events are omitted from this table as they did not occur in the match, such as “penalty kick” or “red card”. Also, some events are not considered because they are not possible to detect without additional information, for example “yellow card” events.

From the numbers in Table 1, we can conclude that automated basic event detection is possible. Some events, especially the bottom level events can be detected with very high accuracy. Higher level events such as “foul” and “offside” are not reliably detected.

There are several factors that impact the correctness of our results. We could observe that the annotations themselves contain errors, which is also true for some trajectories. But more importantly, the ball’s trajectory had to be re-constructed from the annotations. This might have huge positive and/or negative effects on the detection of certain events. All in all, with the current data it is very hard, perhaps even impossible, to estimate the accuracy of our event detection methods. As a consequence, we did not put more efforts into fine tuning our methods, as this would possibly increase the accuracy, but only for the given data with the lack of a real ball’s trajectory.

4. SUBTRAJECTORY CLUSTERING

4.1. METHODS

The techniques proposed by Buchin, Buchin, Gudmundsson, Löffler and Luo (2008) are the basis of our tools. In the following, we briefly review those techniques and then, in Section 4.2, we will consider the experimental results.

The similarity between two trajectories can be defined in different ways, for example using the Longest Common Subsequence model (Vlachos, Gunopulos and Kollios, 2002), a combination of parallel distance, perpendicular distance and angle distance (Lee, Han and Whang, 2007) and the average Euclidean distances between paths (Nanni and Pedreschi, 2006). We will use the Fréchet distance, which is a distance measure for continuous shapes such as curves and surfaces, and is defined using reparameterisations of the shapes. Because it takes the continuity of the shapes into account it is generally regarded as being a more appropriate distance measure than the Hausdorff distance for curves (Alt, Knauer and Wenk, 2004).

The Fréchet distance can be intuitively explained in the following way: Imagine a person walks their

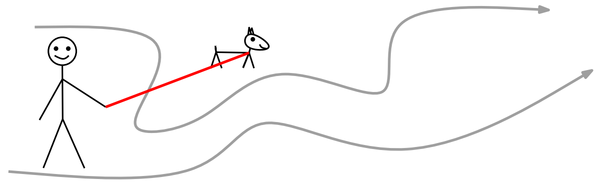


Figure 3: Illustrating the leash length between a person and their dog walking along their trajectories.

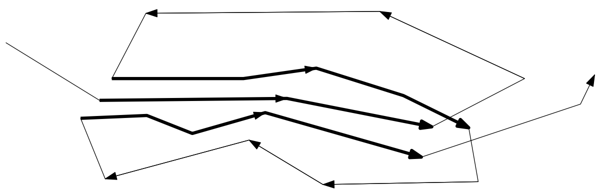


Figure 4: A subtrajectory cluster (indicated fat) of a trajectory.

dog on a leash (see Figure 3). The person will follow a certain trajectory or path T_p, while the dog follows a different path T_d. The Fréchet distance between T_p and T_d is the smallest length of a leash that allows the person and the dog to walk on their paths, where the person and the dog can change their speed or even pause, but they are not allowed to backtrack.

The Fréchet distance comes in two flavours: continuous and discrete. Intuitively, in the continuous version, the person and their dog walk on their paths in a continuous movement, while in the discrete version, they “jump” from one vertex to the next vertex of the path. Note that the continuous version can be approximated by the discrete version when using paths with many vertices, i.e. having data with high resolution. We chose the *discrete Fréchet distance*, because the corresponding algorithms are easier to implement (Buchin, Buchin, Gudmundsson, Löffler and Luo, 2008).

During a match, a player might move along certain paths multiple times. Hence, when given the trajectory T of that player, certain subtrajectories of T might form a subtrajectory cluster (see Figure 4). We follow Buchin, Buchin, Gudmundsson, Löffler and Luo (2008) who define a subtrajectory cluster depending on three parameters: m, ℓ and d. We say that a *subtrajectory cluster* consists of m non-overlapping subtrajectories $T_1, ..., T_m$ of T. At least one subtrajectory has length ℓ, and the distance between the subtrajectories is at most d.

Computing subtrajectory clusters exactly turns out to be a hard problem (Buchin, Buchin, Gudmundsson, Löffler and Luo, 2008). Deciding whether a trajectory T contains a subtrajectory cluster with specified parameters m, ℓ and d is NP-hard. The problem to maximise the number of subtrajectories or to maximise the length of the subtrajectories (while the other parameters are fixed), is also NP-hard, even when computing an approximation where m or ℓ are approximated within certain factors and d is approximated within < 2. (Intuitively, for an NP-hard problem, there is no known efficient algorithm to solve it.) That is why we look at approximation algorithms where d is approximated within a factor of ≥ 2.

The main result by Buchin, Buchin, Gudmundsson, Löffler and Luo (2008) that we will be using is the following: Given a trajectory T, there is an algorithm to compute, under the discrete Fréchet distance, a subtrajectory cluster of maximum length, where the distance d is approximated by a factor of 2 (i.e. we allow the subtrajectories to have a distance twice as large as specified by the parameter d). This algorithm runs in $O(n^2 + nm\ell)$ time and uses $O(n\ell)$ space, where n denotes the number of vertices of T, ℓ denotes the maximum number of vertices of a subtrajectory in the subtrajectory cluster, and m denotes the number of subtrajectories in this cluster.

The implemented prototype can be configured to only report subtrajectory clusters according to the following parameters: *distance* (this is the maximum allowed Fréchet distance between subtrajectories), *minimum cluster size, duration* and *length* (these are thresholds to avoid reporting very small and meaningless clusters).

4.2. EXPERIMENTAL RESULTS

By running experiments on the subtrajectory clustering, we would like to find out how well the clustering works, i.e. how much useful information is found and also how fast the clustering works, i.e. do the run times allow for an interactive tool, where an analyst specifies the parameters of a cluster and the tool will "quickly" find the corresponding clusters.

4.2.1. Usefulness

One way to evaluate the usefulness of the clustering is by examining the results visually, i.e. drawing the clusters together with the trajectory into an image

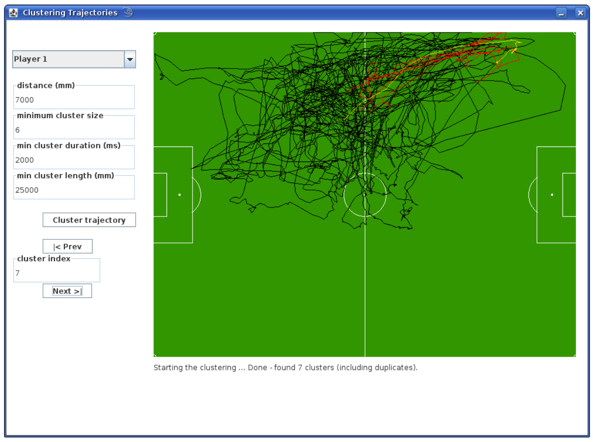


Figure 5: Screenshots showing the trajectory and a subtrajectory cluster of a left-wing player moving forward.

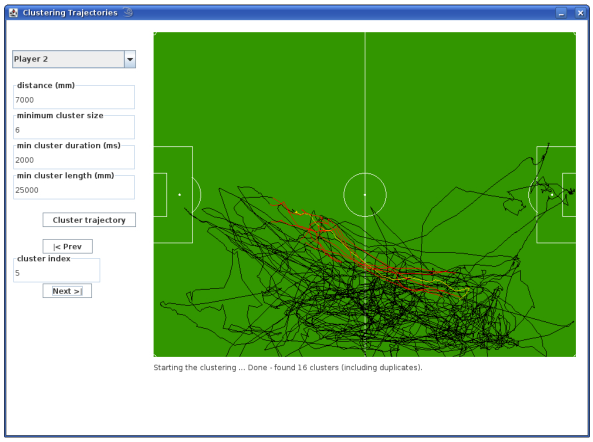


Figure 6: Screenshots showing the trajectory and a subtrajectory cluster of a right-wing player moving backward.

of a football pitch. In Figures 5-8, we provide example screenshots for the ball's trajectory and the trajectory of two players. Each screenshot shows the entire trajectory of one half of the match. Also, in each screenshot, one cluster is highlighted in red and yellow. (The yellow subtrajectory is the *representative subtrajectory* for this cluster (Buchin, Buchin, Gudmundsson, Löffler and Luo, 2008).) For instance, in Figure 5, the trajectory of a left wing player (Player1) and the cluster indicates that several (at least six) of his attacking runs in the first half of the game started near the middle of the left half of the centre line, describing a long arc going forward and further to the left of the field and then making a sharp turn back towards the middle of the field.

From the screenshots we can conclude that the subtrajectory clustering reliably finds the clusters ac-

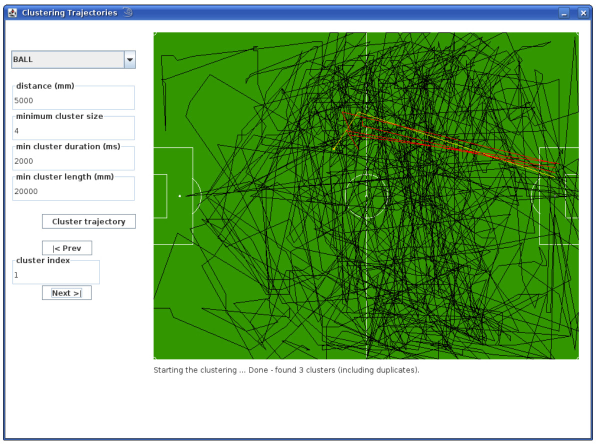


Figure 7: Screenshots showing the ball's trajectory and a subtrajectory cluster of goal kicks.

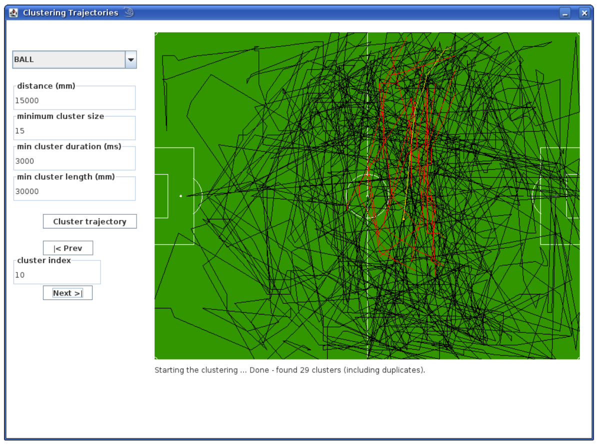


Figure 8: Screenshots showing the ball's trajectory and a subtrajectory cluster of ball movements from top to bottom.

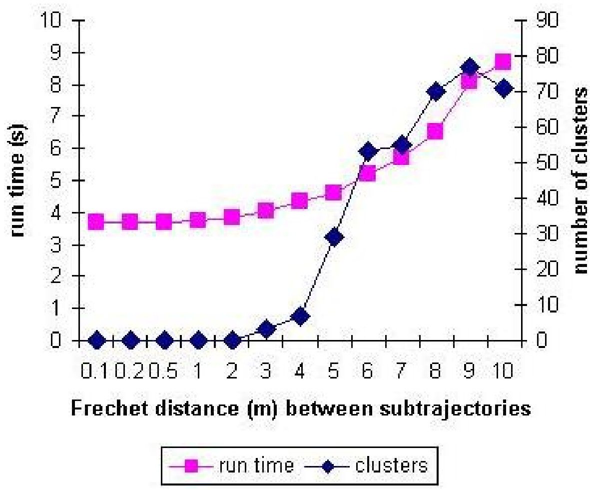


Figure 9: Chart of the running time and the number of clusters depending on the chosen Fréchet distance.

cording to the set parameters. Therefore, from a computer scientist's point of view, this confirms that the subtrajectory clustering algorithms work well. However, longer and bigger (in terms of the number of subtrajectories) clusters would probably be of more interest to domain experts.

One way to give clusters more meaning is to consider the trajectories of multiple matches, and then to look for clusters taking all these trajectories into account. In doing so, we are likely to find longer and larger clusters which can be used to identify interesting emerging patterns. However, we are also likely to face much longer running times (see Section 4.2.2). Another way to increase the meaning of the clustering is to not only look at purely geometry-based clustering, as is done in this study. Instead, one could consider a combination of geometry-based clustering and event-based clustering.

4.2.2. *Running times*

From the way the algorithm works and from the way it has been implemented we would expect certain parameters to have no, or only marginal, impact on the running time. These parameters are the *minimum cluster size, duration* and *length*, which could be confirmed in preliminary experiments. That is why we only report on the running time behaviour depending on the distance d and the number n of vertices in the trajectory, see Figure 9 and 10.

Interestingly, the chosen Fréchet distance d has an impact on the running time (see Figure 9), even though this is not reflected by the results from the theoretical worst-case analysis (see Section 4.1). This dependency can be explained by the way the algorithm works. It iterates over the vertices of the trajectory and in each iteration it keeps track of all other vertices that have a distance of at most d. For larger values of d, this strategy inevitably results in larger running times, as more and more vertices need to be considered. A bigger impact on the running time is caused by the size of the trajectory (see Figure 10). To achieve different sizes, we simplified the trajectory to various degrees, by removing vertices that are less crucial for the overall shape of the trajectory. The theoretical analysis gives us a quadratic dependency, which we could also observe in our experiments.

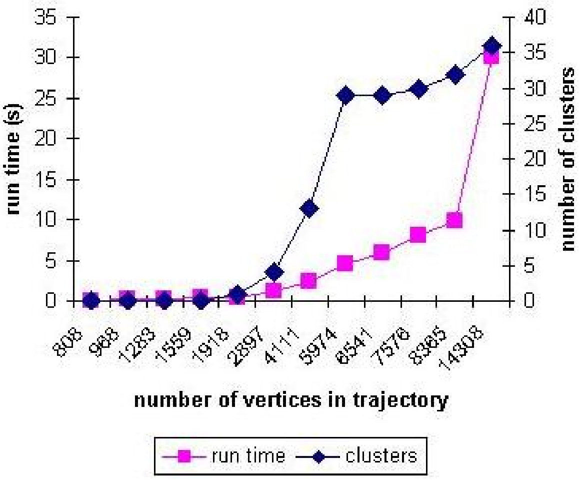


Figure 10: Chart of the running time and the number of clusters depending on the size of the trajectory.

We also see a perhaps suprising dependency of the number of clusters that are found. Clearly, a small Fréchet distance makes it harder for clusters to exist. A large Fréchet distance, on the other hand, can give rise to long clusters. As clusters may not overlap, such long clusters might hence prevent subsequent clusters to exist. Furthermore, the fact that, for small trajectory sizes, many clusters will not be found, is caused by our choice to use the discrete Fréchet distance. In our version, all subtrajectories of a cluster must start and end with vertices that are close to each other. Because of the simplification, we cut out vertices making it less likely to identify clusters for small size of the trajectory.

From the graphs in Figure 9 and 10, we can conclude that finding clusters could potentially be done in an interactive tool for certain parameters and small input sizes. However, the current implementation does not seem to be suitable for interactive use for more interesting settings, e.g. where we have large input or where we allow a large Fréchet distance.

One should note, however, that the current implementation was not optimised for speed. It should be mainly seen as a prototype for a feasibility study that reflects at least certain asymptotic running time behaviour. Also, choosing different or more sophisticated algorithms might result in very different running times, and has the potential to offer drastic speedups.

5. CONCLUSIONS

Our presented algorithms and prototypes are first steps towards completely automated football analysis. Of course, it is not clear that a completely automated analysis will ever be reached, but the more tools we can provide to coaches, the better their team might perform.

We have seen that basic event detection works well and efficiently, and that the clustering also works well, but can be expensive and might have only limited use for coaches. We will continue to work on similar problems and to make our current tools more accurate. Also increasing the speed of the clustering algorithm is a topic for further study.

The area of football analysis is a mix of many different domains. If we want to fully benefit from this mix and contribute to this area, we need a stronger and more effective dialog between experts in the different domains.

Acknowledgements

We wish to thank our students Susan Howlett, John Jiang, Sidong Liu, Cahya Ong, David Rizutto and Merlin Zhang who worked with us on this project during the past few years.

References

Alt, H., Knauer, C., & Wenk, C. (2004) Comparison of distance measures for planar curves. *Algorithmica*, 38(2):45–58.

Buchin, K., Buchin, M., Gudmundsson, J., Löffler, M., & Luo, J. (2008). Detecting commuting patterns by clustering subtrajectories. *Proceedings of the 19th International Symposium on Algorithms and Computation (ISAAC 2008)*, pages 644–655, Berlin, Heidelberg. Springer-Verlag. To appear in International Journal on Computational Geometry and Applications.

Fujimura, A., & Sugihara, K. (2005). Geometric analysis and quantitative evaluation of sport teamwork. *Systems and Computers in Japan*, 36(6):49–58.

Grunz, A., Memmert, D., & Perl, J. (2009). Analysis and simulation of actions in games by means of special self-organizing maps. *International Journal of Computer Science in Sport*, 8(1):22–36.

Kang, C.-H., Hwang J.-R., & Li, K.-J. (2006). Trajectory analysis for soccer players. *ICDMW '06: Proceedings of the Sixth IEEE International Conference on Data Mining - Workshops*, pages 377–381, Washington, DC, USA. IEEE Computer Society.

Lee, J.-G., Han, J., & Whang, K.Y. (2007). Trajectory clustering: a partition-and-group framework. *Proc. ACM SIGMOD international Conference on Management of Data*, pages 593–604.

Nanni, M., & Pedreschi, D. (2006). Time-focused density-based clustering of trajectories of moving objects. *Journal of Intelligent Information Systems*, 27(3):267–289.

Vlachos, M., Gunopulos, D., & Kollios, G. (2002). Discovering similar multidimensional trajectories. *Proc. 18th International Conference on Data Engineering (ICDE)*, pages 673–684.

MEASURING EFFICIENCY OF FOOTBALL TEAMS BY COMMON SET OF WEIGHTS IN DEA

Reza Kiani Mavi [a], Neda Kiani Mavi [b]

[a] *Department of industrial management, Islamic Azad University (IAU), Qazvin Branch, Qazvin, Iran.*
Postal address: Islamic Azad University (IAU), Nokhbegan Street, Qazvin, Iran. Tel: (+98) 281 3665275-7
[b] *Department of sports and physical education, Islamic Azad University (IAU), Qazvin Branch, Qazvin, Iran.*
[c] *Corresponding author: Rezakianimavi@yahoo.com*

Abstract

Managers continually seek improved methods to measure the performance of their organizations because they are committed to improve efficiency and effectiveness in their operating units. Data Envelopment Analysis (DEA) has proven itself to be both theoretically sound framework for performance measurement and an acceptable method by those being measured. Since football is without doubt one of the most important kinds of sports in the world, therefore by taking data for the season 1999/2000 from Haas et al. (2004) we study the efficiency of football teams in the German Bundesliga by DEA. In this paper, we applied a practical common weight compromise solution methodology for measuring efficiency of football teams and then selecting the best one. We find that discriminating power of this method is more than conventional DEA and efficiency score of teams are more reliable. The non-parametric Spearman test of relationship (r_s) and the Kendall's Tau test (τ) of correlation verify the results of DEA and compromise solution.

Keywords: Common weights, DEA, Ranking, Football team

1. INTRODUCTION

Football has become the largest leisure activity and has great social and economic importance. The growth of these leisure sectors in the 90s has been directly linked to the process of media liberalization and, more specifically, the growth of pay-per-view television. There is also consensus that these rates of growth cannot continue indefinitely, and that the market growth has currently stalled (Bosca et al., 2009)

During the last quarter of the 20th century, football became a more favorable sport then many other branches (such as volleyball, handball, basketball, athletics and boxing) and local games (baseball, golf, cricket, rugby), all over the world from Europe to South America and from Africa to Asia. In the struggle of many branches of sports, the advantages of football are its easiness, low cost, and no need for expensive tools. Football, the unique branch of sport in the countries is becoming an important element of the culture and community. The referee's decision, technical director, team's tactics, and player's faults are the causes of argument. Gossip spreads everywhere, such as to TV, newspapers, on to streets, to business, schools and homes. So after the success of national teams, great celebrations are arranged on the city streets. The traffic stops and the streets become full of flags, peoples and cars. The amount of joy is as great as the victory. In this view, football has an important role of including the people in the national idea. Thus, by the association of people, football binds many people to build a community. As a raising trend, football has become a commercial source. With the supporters of teams, sponsor income and publication, the truth income comes to a budget of hundreds of millions of dollars. For example, the income of Manchester United was 200 million euro from the English Premiere League in 2000, from German Bundesliga Bayern München's income was 150 million euro.

Performance evaluation in football might be arisen in a few critics:

i) The efficiency of the manager and technical director (coach): This kind of evaluation takes into consideration the input and output that depends on the experience of managers and directors.

ii) The efficiency of the team in a match: This evaluation deal with inputs and outputs, inputs may be the players ownership of the ball, corner kicks, kicks, penalty kicks, and outputs may be the number of goals and the result of the match.
iii-) The efficiency of the team or teams in the season: In this point of view the season is appreciated as a process and cumulative efficiency is calculated with the data at the end of each week. Alternative approaches might be considered depending on the aim (ALP, 2006)
Haas et al. (2004) propose that many people all around the world and especially in Europe and South America consider (European) football not only as a game, but nearly as a kind of religion. Hence, football is without doubt one of the most important kinds of sports in the world. Besides passion, fanatic support and joy, when the supported team wins, or pain, when it loses, football in Europe is a major industry and significant business.
Many teams have been listed on stock markets in the course of the nineties to finance investments, merchandising has more and more become a major source of revenues for the teams and the value of broadcasting rights has increased dramatically. Single teams exhibit considerable revenue figures.
Football is also the source of endless discussions on strategies, referee decisions and performance among experts, hence, among every single fan. The dispute of millions of would-be team managers or trainers not only runs along the simple line on who will win the championship, but also on relative performance: Do teams with seemingly inferior initial positions, with smaller budgets and less talented players ‘play over their heads’? How should a team, of which most experts expected to win the championship, be judged, when it ends up as second or third after the last round? These questions are without doubt of economic origin, because they dwell upon one of the central economic concepts, namely efficiency.
Thanasis (2009) suggests that the analysis of the football clubs’ operating efficiency measurement is static and includes both the estimation of teams’ (TE) technical efficiency level as well as the identification of its sources. The analysis is characterized as static due to the fact that it refers to a specific period of time. The technical efficiency term is connected directly to the use of inputs by a Decision Making Unit (DMU) during a period of production and estimates any possible waste of resources for a given technology.
The extravagance of resources (especially of money) is particularly evident in the football clubs’ case, where the cost minimization choice is not always the first priority of a team’s director. Accordingly, the efficiency notion is usually discussed by the football fans, especially when they are convinced that their team should have performed better (accumulate more points) according to the budget spent.
Considering that the economic targets achievement is not always the first priority of a Football Club’s manager, the technical efficiency estimation can be deservedly characterized as adequate only via the analysis of the technological relations that describe a team’s operation. Additionally, the fact that there are no market prices for a team’s products, which are goals scored and the inverse of goals conceded, enhances the choice of the analysis’ specific form.
This paper is devoted to analyze efficiency-related questions on the team level for the ‘Deutsche Bundesliga’ (German Federal Football League), which is one of the most important professional football leagues in the world. We are interested in the season overall efficiency of teams, since it seems to be most important from an economic point of view and least investigated.

2. MEASURING THE TECHNICAL EFFICIENCY OF FOOTBALL TEAMS:

2.1 Efficiency: Meaning and Measurement

Economically, efficiency refers to the relationship between scarce factor inputs and outputs of goods and services. This relationship can be seen and evaluated in term of either physical output or cost. If we plan to identify and determine the best possible (optimal) combination of inputs to produce a given level of output in physical term, then we are talking about technological or technical efficiency. With regard to technical inefficiency, it is caused by the failure to achieve the best possible output levels and / or usage of an excessive amount of inputs.
On the other hand, if we want to determine the optimal combination of inputs that will minimize the cost of producing a given level of output, then we are talking about economic efficiency or cost efficiency. This kind of efficiency requires the availability of input prices like the price of labor and capital. According to Drake and Hall (2003), in the absence of accurate data on input prices, performance analysis should be focused on technical efficiency (Ismail, 2004).
All that the DEA efficiency measure tells us is whether or not a DMU can improve its performance relative to the set of DMUs to which it is being compared. To this end, DEA computes a scalar measure of efficiency and determines efficient levels of inputs and outputs for the organization under evaluation. To begin, it is very essential to understand the concept of efficiency.

2.2 Data Envelopment Analysis

Football is a competitive sport with two teams of 11 players. The winning team must score more goals than the other team during the game. While there are various styles of competition, the national leagues involve each of the teams playing each other during the season. Each team plays every other team twice; once in its own 'home' ground, and once 'away' in the opponent's ground. Victories are rewarded with three points, draws receive one point, and defeats do not receive any points—therefore it is a non-zero sum game. The team with the most points wins the league and any ties at the end of the season are broken in various ways in each league.

There are incentives for occupying the highest possible position in the league at the end of the season. Of course, if being first were the only criteria then teams would lose their incentive to win when they realize that first place had become unachievable. Therefore, the best-positioned teams are rewarded with the opportunity to play in European competitions in the following season, and the lowest teams are relegated to the league below (Bosca et al., 2009).

There are two main approaches to measure efficiency: the econometric or parametric approach and the nonparametric or DEA - Data Envelopment analysis approach. The former is mainly based on econometric techniques and measures the difference between the benchmark and the inefficient entities by the residuals, while the latter is based on linear programming techniques. Some of the advantages of SFA over DEA are: firstly, the fact that it accounts for noise and secondly, the fact that it can be used in order conventional tests of hypotheses to be conducted (Thanasis, 2009).

Barros and Barrio (2008) state that unlike the econometric stochastic frontier approach, DEA allows the use of multiple inputs and outputs, but does not impose any functional form on the data; nor does it make distributional assumptions for the error term. Both methods assume that the production function of the fully efficient decision unit is known. In reality this is not the case and the efficient iso-quant has to be estimated from the sample. Under these conditions, the frontier is relative to the sample considered.

DEA is a new methodology that calculates the performance scores of various decision making units with the operation research technique. Performance evaluation has a multi-variable and complex structure. DEA is a technique that is an entirely objective way of performance evaluation (ALP, 2006).

Barros and Barrio (2008) measured the Efficiency of the English football Premier League with a random frontier model. Hoefler and Payne (2006) applied stochastic production frontier model for analyzing NBA association clubs. Barros and Leach (2007) used technical efficiency effects model in soccer clubs in the English Premier League. Barros and Leach (2006a) used DEA-CCR and BCC model in soccer clubs in the English Premier League and stochastic frontier model in soccer clubs in the English Premier League (Barros and Leach (2006b). Haas (2003a) applied DEA-CCR and DEA-BCC model for 12 US soccer clubs and Haas (2003b) used DEA-CCR and DEA-BCC model for 20 Premier League clubs. Barros and Santos (2003) applied DEA-Malmquist index on 18 training activities of sports federations.

In this paper, we will assume that clubs have access to the same level of technology, but differ in their levels of efficiency, and that this may explain differences in productivity. It seems reasonable to assume that the technology used in football (tactics, plans, physical training, etc.) is homogeneous and basically well-known among industry professionals. For this reason, non-parametric optimization techniques, specifically DEA models, are used.

2.2.1 Production possibility set

The relative comparison in DEA is examined within a production possibility set (PPS). The production possibility set (PPS) is defined as the set of all inputs and outputs of a system in which inputs can produce outputs, as:

$$T = \left\{ (x, y) \middle| x \in R_+^m \ \ can \quad produce \ \ y \in R_+^s \right\}$$

In fact, based on information about the existing data on the performance of the units and some preliminary assumptions, a PPS is built and DEA forms an empirical efficient surface (frontier). If a DMU lies on the surface, it is efficient; otherwise it is inefficient. DEA also provides efficiency scores and reference units for inefficient DMUs. Reference units are units on the efficient surface which can be regarded as target units for inefficient units. They are obtained by projecting an inefficient DMU radially or nonradially to the efficient surface (Jahanshahloo et al., 2009).

2.2.2 CCR Model

The production possibility set of CCR model which is built on the assumption of Constant returns to scale (CRS) can be constructed from the observed DMUs as semi–positive vectors $(x, y) \neq (0,0)$ as follows.

$$T_c = \left\{ (x, y) \middle| \begin{array}{l} x \geq \sum_{j=1}^{n} \lambda_j x_j, y \leq \sum_{j=1}^{n} \lambda_j y_j \\ \lambda \geq 0, \quad j = 1,2,...,n \end{array} \right\} \quad (2.1)$$

The CCR model, as introduced by Charnes et al. (1978), measures efficiency relative to the canonical convex monotone hull of the observations. This efficiency estimate can be computed using straightforward linear programming.
The input oriented CCR model estimates maximum radial input contraction of the evaluated DMU such that the projection of it is within the PPS. The corresponding linear programming problem, called envelopment model for efficiency estimation of DMU_p (Input oriented Envelopment model or IE model) is as follows:

$$
\begin{aligned}
&CCR_{IE}) \quad Min \quad \theta_p \\
&subject \quad to \\
&\sum_{j=1}^{n} \lambda_j x_{ij} - \theta_p x_{ip} \le 0, \quad i = 1,2,\ldots,m \quad (2.2)\\
&\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{rp}, \quad r = 1,2,\ldots,s\\
&\lambda_j \ge 0, \quad j = 1,2,\ldots,n\\
&\theta_p \; free
\end{aligned}
$$

The dual problem, called multiplier model (Input oriented Multiplier model or IM model), will also be used afterwards:

$$
\begin{aligned}
&CCR_{IM}) \quad Max \quad \sum_{r=1}^{s} u_r y_{rp} \\
&subject \quad to: \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1,2,\ldots,n\\
&\sum_{i=1}^{m} v_i x_{ip} = 1 \quad (2.3)\\
&u_r \ge 0, \quad r = 1,2,\ldots,s\\
&v_i \ge 0, \quad i = 1,2,\ldots,m
\end{aligned}
$$

Where the variables u_r and v_i are dual variables, in linear programming terms, to the constraints relating to the r^{th} output and i^{th} input in model CCR_{IE}, respectively.

The optimal value u_r^* of u_r can be seen as the imputed value per unit of output r. Similarly, the optimal value v_i^* of v_i can be seen as the imputed value per unit of input i. Likewise, the output oriented CCR model estimates maximum radial output expansion of the evaluated DMU such that the projection of it is within the PPS (Zohrehbandian, 2005).

Definition 1 (Efficiency):
DMUp ; $p \in \{1,2,\ldots,n\}$ is CCR efficient, if and only if $\theta^* = 1$ in (2.2) and the sum of slack variables $(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+)$ is equal to zero.

Definition 2 (Pareto efficiency):
DMUp ; $p \in \{1,2,\ldots,n\}$ is Pareto Efficient if and only if $\theta^* = 1$ in (2.2) and all constraints (except the nonnegative constraints) are binding at all optimal solutions (Jahanshahloo et al., 2009).

2.3 Multi-criteria decision-making (MCDM) problem

The MCDM problem may be represented as:

$$
\begin{aligned}
&Max \; f(x) = \{f_1(x), f_2(x), \ldots, f_k(x)\} \\
&subject\, to: \quad (2.4)\\
&x \in X = (x \mid x \in R^n, g_c(x) \le 0, c = 1,2,\ldots,t)
\end{aligned}
$$

where x is an n-dimensional vector of decision variables, f_1, f_2, . . ., f_k are k distinct criteria functions of decision vector x, and g_1, g_2, . . ., g_c are inequality constraints, and X is the feasible set of constrained decisions. (Malczewski and Jackson, 2000).

2.3.1 Compromise Solution Approach

Compromise solution approach is an example of methods which do not need any inter objective or other subjective preference information. In order to solve the problem (2.4) (identifying the efficient solutions), suppose f_j^*, j = 1, ... , k, are the optimum value of the following models:

$$
\begin{aligned}
&Max \; \{f_j\,(x)\,; \quad j = 1, \ldots, k\} \quad (2.5)\\
&subject \quad to: \quad x \in X
\end{aligned}
$$

Then an efficient solution is defined as the one which minimizes the deviations from these optimum values (ideal solution) as follows:

$$
\begin{aligned}
&\text{Min} \quad D_p = \left\{ \sum_{j=1}^{k} \left[\frac{f_j^*(x) - f_j(x)}{f_j^*(x)} \right]^p \right\}^{\frac{1}{p}}, \quad p \ge 1 \\
&subject \quad to: \quad x \in X \quad (2.6)
\end{aligned}
$$

where the value of p is based upon the utility function of the DM. in the above Model, for the smallest value of P=1, every deviation is being weighted equally. As p increases, more weights are given to the larger deviations. Ultimately, the largest deviation completely dominates when P=∞. There are three values of p, p=1, 2, and ∞, which have special mathematical properties and are worthy of consideration (Hwang and Masud, 1979).

2.4. COMMON WEIGHTS ANALYSIS (CWA)

Liu and Peng (2008) state that conventional data envelopment analysis (DEA) assists decision makers in distinguishing between efficient and inefficient decision making units (DMUs) in a homogeneous group. Its characteristic is to focus on each individual DMU to select the weights attached to the inputs and outputs, and to locate the envelopment surface. A set of weights for the inputs and outputs is determined by the DEA program to show each DMU in its most favourable light as long as the efficiency scores of all DMUs calculated from the same set of weights do not exceed 1.

As a considerable number of DMUs are usually categorized as efficient, a procedure for ranking the efficient units is sometimes necessary. This is especially important when each DMU represents one alternative for consideration. Even for inefficient units, there are cases in which they need to be ranked. For example, in allocating government subsidies, the ranking of the units is often required. However, an inefficient unit with a smaller efficiency score does not necessarily mean poorer performance than one with a larger efficiency score because only the units under the same frontier facet are comparable.

Using different sets of weights to classify the DMUs as efficient or inefficient is acceptable to the practitioners. However, if different sets of weights are used for ranking, most practitioners may not agree because every DMU believes that other DMUs will take this advantage to defeat it. Therefore, the major purpose for generating common weights in DEA is to provide a common base for ranking the DMUs, both the efficient and inefficient ones. Note that a common set of weights means one frontier hyperplane. All DMUs lie beneath that hyperplane. The idea of common weights in DEA was first introduced by Roll et al. (1991). There are many ways to generate common weights. The efficiency score calculated from the standard DEA model is the target for each DMU to achieve. The DMUs select a common set of weights which yields the shortest distance between the vector of efficiency scores calculated from this set of weights and the target. Based on the common set of weights, all DMUs are compared on one scale (Kao and Hung, 2005). Kiani Mavi et al. (2010) proposed a CSW compromise solution approach for technology selection.

3 PRACTICAL COMMON WEIGHT COMPROMISE SOLUTION APPROACH FOR TECHNOLOGY SELECTION

Jahanshahloo et al. (2005) presented the following multiple objective fractional programming problem for finding CSW.

$$\begin{aligned}
&Max\ \left\{\frac{\sum_{r=1}^{s}u_r\,y_{r1}+u_0}{\sum_{i=1}^{m}v_i\,x_{i1}},\ldots,\frac{\sum_{r=1}^{s}u_r\,y_{rn}+u_0}{\sum_{i=1}^{m}v_{in}}\right\}\\
&subject\quad to:\\
&\qquad\frac{\sum_{r=1}^{s}u_r\,y_{rj}+u_0}{\sum_{i=1}^{m}v_i\,x_{ij}}\le 1,\quad j=1,2,\ldots,n \qquad (3.1)\\
&\qquad u_0\ \ free\\
&\qquad u_r, v_i\ge\varepsilon,\quad r=1,2,\ldots,s\quad ,i=1,2,\ldots,m
\end{aligned}$$

Haas et al. (2004) propose that coaches– in particular their tactical as well as motivational skills– are very often considered to be of utmost importance for a team's performance. Evidences indicate that coaches make significant contributions to a team's success in the field.

The proposed model in Jahanshahloo et al. (2005) is nonlinear. But we used one input in the example so that the model is converted into linear one.

For convenience and simplicity, one input is considered in efficiency measurement. For this reason we took coaches' wages as the single input of the model. Consider the following MOLP problem with respect to multiple outputs and a single exact input for efficiency measurement.

$$\begin{aligned}
&Max\ \left\{\frac{\sum_{r=1}^{s}u_r\,y_{r1}}{x_1},\ldots,\frac{\sum_{r=1}^{s}u_r\,y_{rn}}{x_n}\right\}\\
&subject\quad to:\\
&\qquad\frac{\sum_{r=1}^{s}u_r\,y_{rj}}{x_j}\le 1,\qquad j=1,2,\ldots,n \qquad (3.2)\\
&\qquad u_r\ge\varepsilon,\qquad r=1,2,\ldots,s
\end{aligned}$$

where $\frac{\sum_{r=1}^{s}u_r\,y_{rj}}{x_j}$, the weighted ratio of outputs to input, may be assumed as performance of unit j.

Now, consider the following model (3.3). θ_j^* is the optimum performance of unit j.

$$\begin{aligned}
&Max\quad\frac{\sum_{r=1}^{s}u_r\,y_{rj}}{x_j}=\theta_j^*\\
&subject\quad to:\\
&\qquad\frac{\sum_{r=1}^{s}u_r\,y_{rj}}{x_j}\le 1,\qquad j=1,2,\ldots,n \qquad (3.3)\\
&\qquad u_r\ge\varepsilon,\qquad r=1,2,\ldots,s
\end{aligned}$$

We can solve the MCDM formulation (3.2) by using compromise solution approach as follow:

$$Min \left\{ \sum_{j=1}^{n} \left[\theta_j^* - \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j} \right]^p \right\}^{\frac{1}{p}}, \quad 1 \le p \le \infty$$

$$\begin{aligned} & subject \quad to: \\ & \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j} \le 1, \qquad j = 1,2,\ldots, n \qquad (3.4) \\ & u_r \ge \varepsilon, \qquad r = 1,2,\ldots, s \end{aligned}$$

Because, all of $\frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j}, \theta_j^*$ are of efficiency type, then there is no need to normalization. Let us consider two cases in which $p=1,\quad p=\infty$.

If $p=1$ then we have:

$$Min \quad \sum_{j=1}^{n} \left[\theta_j^* - \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j} \right]$$

$$\begin{aligned} & subject \quad to: \\ & \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j} \le 1, \qquad j = 1,2,\ldots, n \qquad (3.5) \\ & u_r \ge \varepsilon, \qquad r = 1,2,\ldots, s \end{aligned}$$

By solving model (3.5), u_r^* r =1,…,s, are calculated which are common weights thus we can calculate efficiency of all units by using these weights.

If $p=\infty$ then we have:

$$Min \left\{ \underset{1 \le j \le n}{Max} \left[\theta_j^* - \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j} \right] \right\}$$

$$\begin{aligned} & subject \quad to: \\ & \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j} \le 1, \qquad j = 1,2,\ldots, n \qquad (3.6) \\ & u_r \ge \varepsilon, \qquad r = 1,2,\ldots, s \end{aligned}$$

Model (3.6) is a special linear problem but by defining a variable Z, simply, converted to a linear problem as follows:

$$\begin{aligned} & Min \ Z \\ & subject \quad to: \\ & Z \ge \theta_j^* - \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j}, \qquad j = 1,2,\ldots, n \\ & \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j} \le 1, \qquad j = 1,2,\ldots, n \qquad (3.7) \\ & u_r \ge \varepsilon, \qquad r = 1,2,\ldots, s \\ & Z \ge 0 \end{aligned}$$

Furthermore, for complete ranking of DMUs in models (3.6),(3.7), we can define set A as follow:

$$A = \left\{ j \middle| DMU_j \ is \ efficient \ by \ (3.6),(3.7) \right\}$$

And then, model (3.7) is converted as follows: (Jahanshahloo et al., 2005)

$$\begin{aligned} & Min \ Z \\ & subject \quad to: \\ & Z \ge \theta_j^* - \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j}, \qquad j \in A \\ & \frac{\sum_{r=1}^{s} u_r y_{rj}}{x_j} \le 1, \qquad j \notin A \qquad (3.8) \\ & u_r \ge \varepsilon, \qquad r = 1,2,\ldots, s \\ & Z \ge 0 \end{aligned}$$

The proposed MCDM approach has two advantages in comparison with DEA-based approaches proposed in the literature for the similar problem. First, the proposed approach evaluates all alternatives by common weights, therefore overcomes the unrealistic weighting scheme problem that is common to DEA, because DEA assumes that each DMU selects the best factor weights. Second, it identifies the best team with less computation compared with DEA-based approaches.

Result: By the optimal solution of (3.7), at least one $\theta_j;\ j=1,2,\ldots,n$ equals to 1.

Proof: Suppose that (u^*, v^*) is the optimal solution of (3.7) and is CSW. There is one $j;\ j=1,2,\ldots,k$ for which the second inequality in (3.7) is binding. Because, if it is not the case, there is a sufficiently small value $\delta > 0$ where

$$(\overline{u}, \overline{v}) = \left(u^* + [\delta, 0, \ldots, 0]_{1\times s}^T, v^* - [\delta, 0, \ldots, 0]_{1\times m}^T \right)$$

satisfy the set of restrictions in (3.7). Because (u^*, v^*) is the optimal solution, then, this is a contradiction. Therefore, there is one $j;\ j=1,2,\ldots,k$ for which we have:

$$\sum_{r=1}^{s} u_r y_{rj} - x_j = 0$$

This means that relative efficiency of at least one DMU equals to 1 and CP can lead to finding relative efficiencies by CSW. ■

4. NUMERICAL EXAMPLE

Data of this paper are based on Haas et al. (2004). We have four outputs and one input. In this paper, we assumed that coaches' wage is the most important input factor for team's success then we considered it as the input. Table 1 shows data.

For analyzing the efficiency of teams compromise solution approach with P=1 and P=∞ is applied. CCR efficiency score of teams are presented in column 2 of table 2. Columns 3,4 show the efficiency score of them by CSW. Applying proposed method on the data results in the following CSW.

Table 1: Raw data for the season 1999/2000

Team	Coach's monthly wage in 1000 DM	(O)points	Spectators in 1000s	Stadium utilization in %	Total revenues in Mio. DM
Bayern München	300	73	894	83.5	220
Bayer Leverkusen	180	73	382	89.7	85
Hamburger SV	125	59	703	76.6	61
1860 München	160	53	555	51.8	42
1. FC Kaiserslautern	200	50	684	96.9	75
Hertha BSC	100	50	809	62.8	42
Vfl Wolfsburg	80	49	292	83.5	40
Vfb Stuttgart	100	48	500	65.3	52
Werder Bremen	30	47	507	84.5	63
SpVgg Unterhaching	30	44	163	76.6	14
Borussia Dortmund	100	40	1099	93.7	150
SC Freiburg	50	40	420	98.8	31
FC Schalke	70	39	689	65.4	64
Eintracht Frankfurt	80	39	605	58.3	40
Hansa Rostock	35	38	275	66	32
SSV Ulm	22	35	371	97	26
Arminia Bielefeld	50	30	335	74.4	32
MSV Duisburg	42	22	257	50.1	28

Source: Kicker Sportmagazin, Olympia Verlag

Table 2: Efficiency score of teams.

Team	CCR-Efficiency	Efficiency with P=1	Efficiency with P=∞
Bayern München	0.3492	0.1760	0.1762
Bayer Leverkusen	0.2577	0.1258	0.1256
Hamburger SV	0.3328	0.3322	0.3325
1860 München	0.2085	0.2048	0.2051
1. FC Kaiserslautern	0.2024	0.2022	0.2022
Hertha BSC	0.4787	0.4775	0.4783
Vfl Wolfsburg	0.3857	0.2166	0.2159
Vfb Stuttgart	0.3044	0.2955	0.2957
Werder Bremen	1.0000	1.0000	1.0000
SpVgg Unterhaching	0.9219	0.3238	0.3216
Borussia Dortmund	0.7143	0.6488	0.6498
SC Freiburg	0.5033	0.4977	0.4968
FC Schalke	0.5824	0.5812	0.5820
Eintracht Frankfurt	0.4475	0.4466	0.4471
Hansa Rostock	0.6842	0.4656	0.4647
SSV Ulm	1.0000	0.9997	0.9974
Arminia Bielefeld	0.4022	0.3968	0.3962
MSV Duisburg	0.3637	0.3622	0.3619
Average	0.5077	0.4307	0.4305

Table 3: Nonparametric Correlation test between CCR ranking and CP ranking (N=18)

Test	CP with P=1	CP with P=∞
Kendall's tau_b	.721**	.721**
Spearman's rho	.840**	.840**

** Correlation is significant at the 0.01 level (2-tailed).

$P = 1$: $u_1^* = 0.0001$, $u_2^* = 0.0589$, $u_3^* = 0.0014$, $u_4^* = 0.0001$

$P = \infty$: $u_1^* = 0.0001$, $u_2^* = 0.0591$, $u_3^* = 0.0001$, $u_4^* = 0.0001$

In Column 2 of table 2, CCR efficiency scores are shown. Columns 3,4 show the compromise programming scores that have been obtained by CSW. Because we applied two norms 1 and infinity then CSW and resulting scores are different from each other that is obvious in columns 3,4. Same ranking order of teams in compromise programming with norms 1 and infinity is resulted from their little difference. If the differences were significant, then different rankings were expected.

4.1 VALIDATION OF RESULTS AND DISCUSSION

To verify the results of DEA and compromise solution, the non-parametric Spearman test of relationship(r_s) and the Kendall's Tau test (τ) of correlation is employed (Azadeh et al., 2009). Because we want to test correlation between ranks obtained from two different methods then rank correlation coefficient is calculated. If data are not normally distributed or have ordered categories, choose Kendall's tau-b or Spearman, which measure the association between rank orders.

Table 3 reports the non-parametric Spearman test of relationship (r_s) between CCR rankings and CP rankings with P=1 and P=∞ which result in the rejection of H_0 at 0.01 levels.

Also, the Kendall's Tau test (τ) of correlation verifies this finding at the same level of significance. There is a direct relationship between CCR and CP results. The Spearman test statistics is 0.840 for P=1 and 0.840 for P=∞ because ranking of two approaches are same. This result shows a strong direct relationship between CCR and CP ranks. Because the number of efficient DMUs on a common weight basis is reduced, discriminating power of our approach is higher than CCR.

5. CONCLUSION

This paper developed new methodology based on MCDM and DEA for generating Common set of weights (CSW) to assess all the DMUs on the same scale. Because average of ranks in the proposed methods is less than conventional CCR model, we can say that the discriminating power of proposed methods is more than that of CCR model. Also, this finding is proved by nonparametric correlation tests.

In general, the rankings of these methods indicate that the results are reasonable. In addition to this, they are more informative. They not only differentiate the efficient units, but also detect some abnormal efficiency scores calculated from the CCR model.

Some teams have acquired higher efficiency scores with CCR model but when all of teams are considered based on common weights, then, their efficiency score has decreased. For example, SpVgg Unterhaching obtained efficiency score of 0.9219 by CCR model but its efficiency is 0.3238 and 0.3216 by common weights. This implies that, some of efficiencies obtained by conventional DEA model are abnormal. This case is true for some other teams as Bayern München, Bayer Leverkusen, Vfl Wolfsburg, Borussia Dortmund and Hansa Rostock. In all cases the amount of decrease is not important that can be attributed to the CSW but in other ones like SpVgg Unterhaching, this decrease is considerable that shows abnormal scores obtained by CCR model.

To verify the results of DEA and proposed methods, the non-parametric Spearman test of relationship (r_s) and the Kendall's Tau test (τ) of correlation is employed.

Some recommendations for doing future studies are provided.

- Useful extensions of the proposed methodologies can be developed, which enable the decision-maker to consider imprecise output data denoted by fuzzy numbers, interval data, ordinal preference information & their mixture.
- Researchers can extend the proposed methods to other models in DEA such as BCC-CCR, CCR-BCC, Additive model and etc.

References

ALP, I. (2006) Performance Evaluation of Goalkeepers of the World Cup. G.U. Journal of Science, 19(2), 119-12

Azadeh, A.,Ghaderi, G.F. and Omrani, H. (2009) A deterministic approach for performance assessment and optimization of power distribution units in Iran. Energy Policy, 37, 274–280

Barros, C.P and Barrio, P.G. (2008) Efficiency measurement of the English football Premier League with a random frontier model. Economic Modelling, 25, 994–1002

Barros, C.P., Leach, S., 2006a. Performance evaluation of the English Premier League with data envelopment analysis. Applied Economics 38 (12), 1449–1458.

Barros, C.P., Leach, S., 2006b. Analysing the performance of the English F.A. Premier League with an econometric frontier model. Journal of Sport Economics 7 (4), 391–407.

Barros, C.P., Leach, S., 2007. Technical efficiency in the English Football Association Premier League. Applied Economics Letters 14 (10), 731–741.

Barros, C.P., Santos, A., 2003. Productivity in sports organisational training activities: a DEA study. European Journal of Sport Management Quarterly 1, 46–65.

Bosa, J.E., Liern, V., Martinez, A. and Sala, R. (2009) Increasing offensive or defensive efficiency? An analysis of Italian and Spanish football. Omega, 37, 63 – 78

Charnes A., Cooper W.W. and Rhodes E. (1978), Measuring the efficiency of decision making units, European Journal of Operational Research, No.2, pp.429-444

Drake L., and Hall M.J.B. (2003) Efficiency in Japanese banking: An empirical analysis, Journal of Banking and Finance, No.27, pp.891-917.

Haas, D., Kocher M.G. and Sutter, M. (2004) Measuring Efficiency of German Football Teams by Data Envelopment Analysis. Central European Journal of Operations Research, 12, 251-268

Haas, D.J., 2003a. Technical efficiency in the Major League Soccer. Journal of Sport Economics 4 (3), 203–215.

Haas, D.J., 2003B. Productive efficiency of English football teams — a data envelopment approach. Managerial and Decision Economics 24, 403–410.

Hoefler, R.A., Payne, J.E., 2006. Efficiency in the National Basketball Association: a stochastic frontier approach with panel data. Managerial and Decision Economics 27 (4), 279–285.

Hwang C.L. and Masud A.S. (1979), Multiple objective decision making: Methods and applications, Springer- verlag, Berlin Heidelberg.

Ismail M. (2004), A DEA analysis of bank performance in Malaysia, 4th International Symposium of DEA, Aston Business School, Aston University, UK

Jahanshahloo G.R., Memariani A., Lotfi F.H., Rezai H.Z. (2005), A note on some of DEA models and finding efficiency and complete ranking using common set of weights, Applied Mathematics and Computation, No.166, pp.265-281.

Jahanshahloo G.R., Shirzadi A. and Mirdehghan S.M. (2009) Finding strong defining hyperplanes of PPS using multiplier form, European Journal of Operational Research, No.194, pp.933–938

Kao C., Hung H.T. (2005), Data envelopment analysis with common weights: the compromise solution approach, Journal of the Operational Research Society, No.56, pp.1196-1203.

Kiani Mavi, R., Makui, A., Fazli, S. and Alinezhad, A. (2010) Practical common weights compromise solution approach for technology selection. International Journal of Procurement Management, 3(2), 214-230

Liu F.H.F. and Peng H.H. (2008) Ranking of units on the DEA frontier with common weights, Computers and Operations Research, No.35, pp.1624-1635

Liu F.H.F., Peng H.H. and Chang H.W. (2006) Ranking DEA Efficient Units with the Most Compromising Common Weights, The Sixth International Symposium on Operations Research and Its Applications (ISORA’06) Xinjiang, China

Malczewski J. and Jackson M. (2000) Multicriteria spatial allocation of educational resources: an overview, Socio-Economic Planning Sciences, No.34, pp.219-235

Roll T., Cook W.D., Golany B. (1991), Controlling factor weights in data envelopment analysis, IIE Transactions, Vol.23, No.1, pp.2-9.

Thanasis, B.P. (2009) Analyzing the operating efficiency of Greek football clubs. MSc. Thesis, Department of Economics, University of Macedonia, THESSALONIKI, GREECE

Zohrehbandian M. (2005) Identification of efficient frontier and its use in data envelopment analysis, Ph.D thesis, Islamic Azad University, Science and Research branch.

AN IMPROVED WORLD CUP TOURNAMENT DESIGN

Darren O'Shaughnessy [a]

[a] *Ranking Software: darren@rankingsoftware.com*

Abstract

An alternative design for World Cups is presented, applicable to many sports. The proposed structure performs better than the incumbent design under several important metrics, including likelihood of the best team winning, number of matches between teams of similar ability ("competitive balance"), number of matches between high-quality teams, and potential value of television broadcast rights. The design is evaluated against other metrics, including those proposed by Scarf, Yusof & Bilbao (2009), and assessed for its practicality in the existing tournament timeframes. Ultimately, the duration of the tournament is the most important constraint.

The incumbent format of round-robin groups followed by a "knockout" is reversed. Instead I propose a preliminary classification phase, elite round-robin and repechage, followed by regular semi-finals and final. The 2010 FIFA World Cup took 16 days – half its length and 75% of its matches – to reduce the initial field of 32 teams to 16, but only four days for each subsequent halving. It is reconstructed as an exemplar.

Keywords: World Cup, Tournament Design, Repechage, Football, Cricket, Tennis, Rugby

References

Scarf, P., Yusof, M. M. & Bilbao, M. (2009). A numerical study of designs for sporting contests. *European Journal of Operational Research, 198,* 190-198.

RE-SCHEDULING TO MINIMISE THE POTENTIAL FOR MATCH FIXING IN THE FIFA WORLD CUP

Schembri[a,b], A. J., Golkhandan[a], S. R., and Bedford[a], A.

[a] *School of Mathematical and Geospatial Sciences, RMIT University*

[b] *Corresponding author: adrian.schembri@rmit.edu.au*

ABSTRACT

Football is arguably the most popular sport worldwide, yet its history been marred by incidents of match fixing at both international and club level. In this paper, we review the potential for match fixing during the group stage of the FIFA World Cup. Specifically, the potential for match fixing was assessed among those teams who had qualified for the Round of 16 after only two group stage matches, and therefore had little incentive to win their final group stage match. The 1998, 2002, and 2006 World Cups are reviewed in detail, in order to identify the different scenarios where there was the potential for match fixing to occur. Both FIFA rankings and Elo ratings were used to determine each team's relative ranking prior and during the group phase. Given the presence of four teams in each group, simulations were carried out for six potential schedules of games, with the aim of identifying the optimal draw in regards to minimising the potential for match fixing situations. Results indicated that the potential for match fixing is dependent on the schedule as well as the standard deviation of team rankings within each group. Specifically, groups that have a high standard deviation in Elo ratings are more likely to be susceptible to match fixing situations in the final match of the group stage. This is particularly the case if the top two ranked teams play each other in the first group stage match. The potential for match fixing can be minimised when the top two ranked teams play each other in the final group stage match.

Keywords: Match fixing, FIFA, Football, World cup, Elo ratings

1. INTRODUCTION

Football (referred to as soccer in Australia and the United States) is arguably the most popular world sport, with professional competitions established in a wealth of countries worldwide. The pinnacle of the football calendar is undoubtedly the FIFA World Cup, a tournament that is held once every four years. All registered football nations are eligible to compete for a place in the tournament, with 204 nations competing for a place in the next instalment of the FIFA World Cup, which will be held in South Africa in June/July, 2010 (FIFA, 2010a; Torgler, 2006). The World Cup is challenged only by the Olympics in terms of global attention, financial investment, and worldwide viewing as a sporting tournament, with more than 700 million viewers tuning in to the last World Cup held in Germany in 2006 (FIFA, 2010b).

Despite the popularity and financial prosperity of the sport, football has been marred by several controversies both on and off the field throughout its history (Caruso, 2009). Perhaps the most controversial is the tendency for the sport to be embroiled in match fixing scandals, with both club and international football retaining a long history of match fixing incidents (Caruso, 2009; Preston

& Szymanski, 2003). By definition, match fixing occurs when a game is played in the context of a partially or completely pre-determined result. One of the most well known match fixing incidents in football occurred at club level, in the Italian Serie A, where world powerhouses Juventus, AC Milan, Fiorentina, and Lazio were implicated in a match fixing scandal that involved selection of favourable referees. Juventus were ultimately stripped of their 2004-2005 and 2005-2006 Serie A titles, were relegated to Serie B, and were banned from participating in European club competitions in 2006-2007. Penalties were also handed down to AC Milan, Fiorentina, and Lazio, although these were not of the severity allocated to Juventus (British Broadcasting Corporation, 2006a, 2006b). The World Cup has not been exempt from match fixing accusations, with a book published by Declan Hill alleging that during the 2006 World Cup, one group match (Ghana vs. Italy), a round of 16 match (Ghana vs. Brazil), and a quarter final (Italy vs. Ukraine) were fixed by Asian betting syndicates, with final scores being known prior to the matches being played (The Age, 2008).

In one of the most infamous match fixing scandals in World Cup history, West Germany played Austria in the final group stage match of the 1982 World Cup (Caruso, 2009). In this match, a 1 or 2 goal win to West Germany would have enabled both West Germany and Austria to qualify, however a win to West Germany by 3 or more goals would have eliminated Austria and enabled Algeria to qualify. In addition, a draw or a win to Austria would have eliminated West Germany, with Algeria again qualifying as a result. In the end, West Germany scored once in the 10th minute, with neither team scoring for the remainder of the match. The lack of attacking movements from either team after West Germany scored raised significant questions about whether the match was fixed to ensure that both West Germany and Austria progressed beyond the group stage, simultaneously eliminating Algeria (Caruso, 2009). Although most match fixing scandals have been associated with financial incentives, this example demonstrates that other incentives are often present to intentionally draw or lose a match. This is particularly salient in tournament situations where a team has already qualified for the next round and therefore has little incentive to win its remaining pool matches (other than the psychological advantages associated with maintaining momentum and a winning culture).

The structure of the World Cup tournament is particularly vulnerable to this type of match fixing. In the World Cup, the 32 qualifying teams are split into eight groups of four (Groups A to H). Teams are allocated to groups by first clustering all 32 countries into four pots. For the allocation of teams for the upcoming 2010 World Cup, the first pot consisted of the eight seeded teams (that is, the host nation and seven teams who FIFA deemed to have the highest ranking), the second pot consisted of teams from the European continental zone, the third pot consisted of teams from South America and Africa, and the fourth pot consisted of teams from Asia, Oceania, North and Central America, and the Caribbean. One team from each pot was allocated to each group, resulting in a mixture of teams from different confederations and variations of highly ranked and lower ranked nations within each group. During the group stage, each team plays each other team within their group once, resulting in each team playing in three group stage matches, with six matches being played across the entire group. The teams who finish first and second in the group qualify for the Round of 16, whilst the teams who finish third and fourth are eliminated.

The schedule of the group stage in the World Cup is critical in terms of the potential for match fixing. Depending on the fixture, teams often qualify for the Round of 16 if they win their first two matches, which results in a dead rubber (that is, a match where the outcome does not impact on the team's standing in the tournament) in the final group stage match for at least one team. In this scenario, a team may be presented with one of several incentives to draw or lose a match. For example, if a team has already qualified for the Round of 16, they may intentionally draw or even lose their third match to ensure that they finish second in the group rather than first, or to ensure that the team they are playing in the final match qualifies for the Round of 16. A common question that arises is why a team would rather finish second in their group or why they would bother assisting another team with qualification for the Round of 16. Simply put, in the Round of 16, the first place team in a group plays the second placed team in another group. As such, if a team recognises that the team who finished second is more dangerous to play in the Round of 16, they may prefer to finish second in order to play the team who finished first in that group.

To illustrate, consider the state of Group B prior to the final group match in the 2002 World Cup.

Table 1. Group B at the 2002 World Cup prior to the final round of group stage matches.

Team	W	D	L	GF	GA	P
Spain	2	0	0	6	2	6
South Africa	1	1	0	3	2	4
Paraguay	0	1	1	3	5	1
Slovenia	0	0	2	1	4	0

For this group, Spain were on top and had already qualified for the Round of 16, and played second placed South Africa in their final group match. Meanwhile, Paraguay (third) played bottom placed Slovenia, who were already out of contention for the Round of 16. Based on their relative rankings and form going into the match, Paraguay were highly fancied to beat Slovenia, whilst it was thought that Spain would beat South Africa. Although South Africa were highly fancied to be defeated by Spain, a draw would have prevented Paraguay from qualifying for the Group of 16 (FIFA, 2010c). Thus, Spain had an incentive to at least draw with South Africa, given that it would eliminate the higher ranked team from the tournament without impacting on Spain's position in the group. Instances such as this are not uncommon, with scenarios that had similar ramifications occurring in four groups at the 2006 World Cup, and three and four groups for the 2002 and 1998 World Cups respectively.

In this paper, we examine the potential for match fixing during the group stage of the FIFA World Cup. FIFA rankings and Elo ratings are used to determine each team's relative ranking prior to and during the group stage of the 1998, 2002, and 2006 World Cups, with these rating models used to determine the likelihood of match fixing depending on variations in the fixture. The potential for match fixing was evaluated across the six possible schedules of matches, with the probability of teams playing in a dead rubber being determined.

2. METHODS

The schedules of group stage matches for the last three World Cups (1998, 2002, and 2006) were downloaded from the Official FIFA website (2010a). Each team's Elo rating (calculated following their final friendly before each World Cup) was considered as the rating for their first group stage match. Given that Elo ratings are updated after each match, the following formula was applied to all teams in order to calculate their Elo ratings prior to the second and third round of group stage matches:

$$R_n = R_o + K\,(W - W_e) \qquad (1)$$

In the above formula, R_n is the new rating, R_o is the pre-match rating, K is equal to 60 for World Cup finals (see www.eloratings.com for details), W is dependent on match result (i.e. 1 for a win, 0.5 for a draw, and 0 for a loss), and W_e refers to the expected result:

$$W_e = 1 / (10^{(-dr/400)} + 1) \qquad (2)$$

dr = difference between the two teams' Elo ratings.

In order to provide a more reflective measure of team strength, 100 points is typically added to the Elo rating of the home team. Therefore, for each World Cup, the Elo rating of the host(s) was equal to their actual rating plus 100. Moreover, as part of the analyses, it was essential to identify the best two teams in each group. Given that teams, on occasion, had equal Elo ratings before the start of the World Cup (e.g. Iran and Mexico prior to the 2006 World Cup), FIFA rankings (which are published several weeks before the start of the World Cup, and are not updated during the competition) were also recorded as an alternative measure of team strength.

Based on each group containing four teams, six potential schedules were created for the group stage matches, refer to Table 2.

In schedules 1 and 3, the best two teams, in terms of Elo ratings, play each other in their last match. In schedules 2 and 5, the best team plays against the second and third teams in the first two matches. Finally, in schedules 4 and 6, the match between the first and third teams is played in the third and final group game.

Table 2. The six possible schedules during the group phase of the FIFA World Cup.

Schedule	Group Ranking			
	1[a]	2	3	4
1	3[b], 4, 2	4, 3, 1	1, 2, 4	2, 1, 3
2	3, 2, 4	4, 1, 3	1, 4, 2	2, 3, 1
3	4, 3, 2	3, 4, 1	2, 1, 4	1, 2, 3
4	4, 2, 3	3, 1, 4	2, 4, 1	1, 3, 2
5	2, 3, 4	1, 4, 3	4, 1, 2	3, 1, 2
6	2, 4, 3	1, 3, 4	4, 1, 2	3, 1, 2

[a] Indicates the ranking of the team within the group

[b] Indicates that the highest ranked team played the third ranked team in their first match

In order to investigate the likelihood of match fixing, Elo ratings were used to simulate the first two rounds of group matches for each World Cup. @RISK software was used to simulate group matches for 10000 iterations, with points tables being computed following each iteration.

3. RESULTS

A preliminary simulation of group matches, using randomly generated Elo ratings for the four teams in a group, resulted in 13 possible patterns of points at the end of the second group stage match:

(a) 6-6-0-0; (b) 6-4-1-0; (c) 6-2-1-1; (d) 6-3-1-1;
(e) 6-3-3-0; (f) 4-3-2-1; (g) 4-3-3-1; (h) 4-2-2-1;
(i) 4-4-1-1; (j) 4-4-2-0; (k) 4-4-3-0; (l) 3-3-3-3;
(m) 2-2-2-2.

Out of the above possibilities, conditions (a), (b), and (c) were considered potential situations for match fixing:

(a) The first two teams play each other in their last match, and either team can lose in order to finish second in the group rather than first (refer to Table 3 for an example).

Table 3. Group H of the 1998 World Cup prior to the final round of group stage matches

Team	W	D	L	GF	GA	P
Argentina	2	0	0	6	0	6
Croatia	2	0	0	4	1	6
Japan	0	0	2	0	2	0
Jamaica	0	0	2	1	8	0

(b) The highest ranked team, which plays against the third ranked team in their last match, can reduce the chances of the second ranked team qualifying by losing against the third ranked team (refer to Table 4 for an example).

Table 4. Group C of the 2002 World Cup prior to the final round of group stage matches

Team	W	D	L	GF	GA	P
Brazil	2	0	0	6	1	6
Costa Rica	1	1	0	3	1	4
Turkey	0	1	1	2	3	1
China PR	0	0	2	0	6	0

(c) The team on six points has already qualified as the top team in the group. As such, a loss against the team in second position can ensure that teams in either third or fourth position are unable to qualify (refer to Table 5 for an example).

Table 5. Group A of the 1998 World Cup prior to the final round of group stage matches

Team	W	D	L	GF	GA	P
Brazil	2	0	0	5	1	6
Norway	0	2	0	3	3	2
Scotland	0	1	1	2	3	1
Morocco	0	1	1	2	5	1

As a preliminary analysis, the prevalence of each points pattern over the course of the 1998, 2002, and 2006 World Cups was examined, refer to Figure 1.
As shown, the most frequently occurring points pattern was the 6-4-1-0 pattern, which occurred on

over 20% of occasions. Of note, this is a points pattern that has a potential for match fixing. The 6-6-0-0 points system was equal second in terms of most frequent points pattern, whilst the 6-2-1-1 system was equal fifth, occurring on less than 10% of occasions. Cumulatively, the potential for match

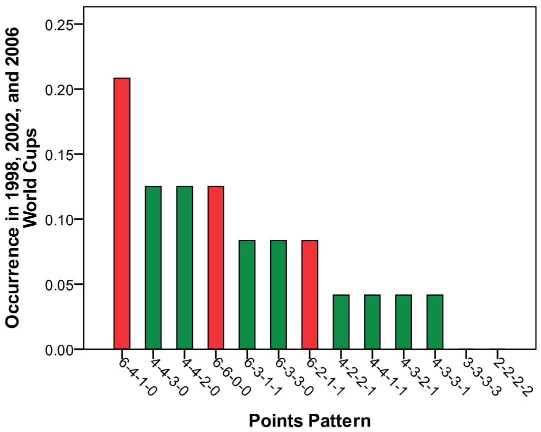


Figure 1. Proportion of occurrence of all possible points patterns over the 1998, 2002, and 2006 World Cups.

fixing was evident in 42% of all groups over the past three World Cups.

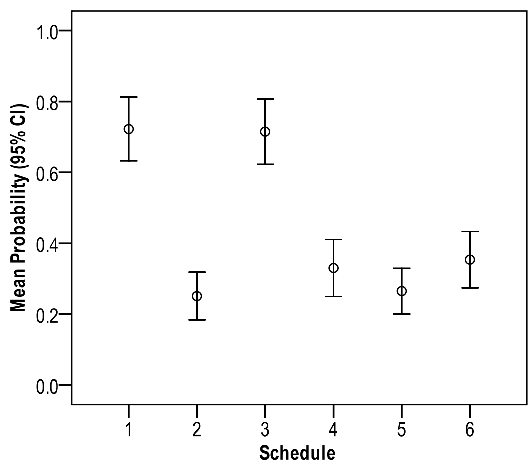
In order to investigate the probability of conditions (a), (b), and (c) occurring at the end of the second round of group stage matches, 10000 simulations using actual Elo ratings at the time of a particular World Cup were carried out for each of the six potential schedules in a group. For the purpose of these simulations, each team's Elo rating in their first match was considered as the average of the normal distribution of that team's ratings. Moreover, for each World Cup, the standard deviation of all 32 teams' Elo ratings was applied to the distribution of each team's Elo ratings. The outcome measure, which was the probability for a potential match fixing condition to occur in each specific schedule, was computed by adding the probabilities of point patterns (a), (b), and (c) in the simulation outputs.

Before reporting the simulation results, it should be noted that in the preliminary simulation that used random Elo ratings for group teams, the cumulative percentage of match fixing conditions ranged from 25.2 to 26.7% across all schedules. However, in almost all simulations using the six different schedules, using teams' real Elo ratings, the cumulative percentage for match fixing conditions was greater than 27% across all different patterns of points.

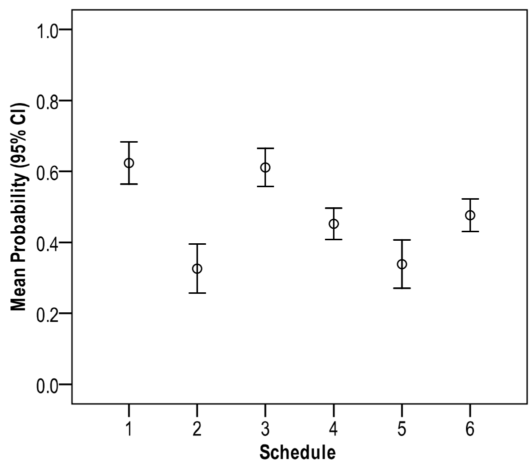
Figure 2 displays the mean probability of the potential for match fixing across the six possible schedules of the group stage of the 1998, 2002, and 2006 World Cups.

A general pattern was found in all three World Cups. Schedules 1 and 3 always resulted in the highest probability of match fixing, ranging from 55 to 88% in the eight groups of the 1998 World Cup, 53 to 72% in 2002, and 50 to 72% in 2006. In contrast, schedules 2 and 5 led to the lowest likelihood of potential match fixing, ranging from 12 to 38% across all groups in the 1998 World Cup,

1998 World Cup



2002 World Cup



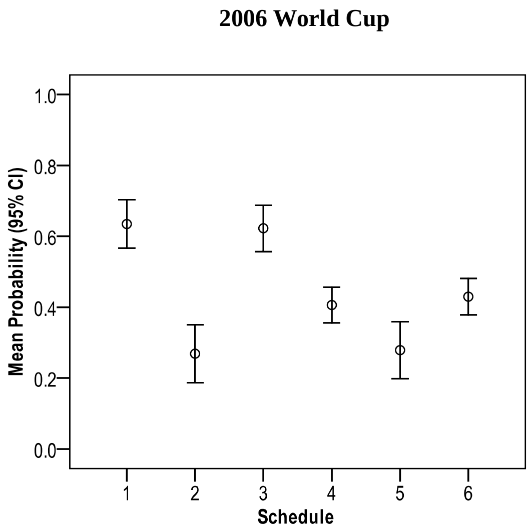


Figure 2. Mean probability of the potential for match fixing across the six possible group schedules for the 1998, 2002, and 2006 FIFA World Cups.

22 to 44% in 2002, and 12 to 41% in 2006. The outcome measure corresponding to schedules 4 and 6 were moderate in relation to the likelihood of potential match fixing situations. Across the eight groups of the 1998 World Cup, between 19% and 45% had the potential for match fixing, in 2002 these probabilities were between 39% and 57%, and between 33 and 51% in 2006. With respect to all groups within the three World Cups, the potential for match fixing was highest during 1998, followed by the 2006 and 2002 World Cups respectively.

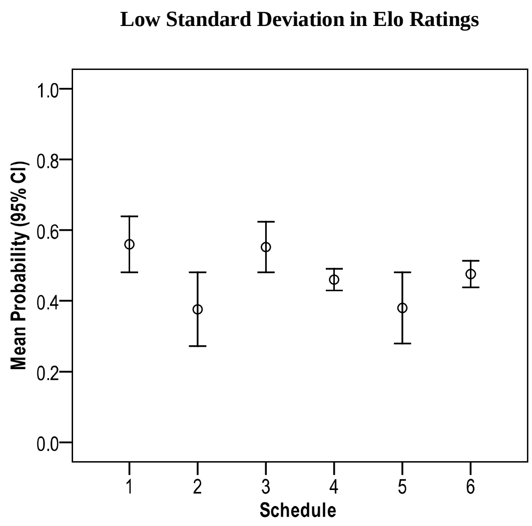
Although particular schedules consistently resulted in the highest (e.g., schedules 1 and 3) and lowest (e.g., schedules 2 and 5) potential for match fixing, notable variations were evident in regards to the likelihood of possible match fixing situations arising across different groups. While the difference between the highest and the lowest probability of potential match fixing in simulation outputs (a, b, and c) was less than 9% for some groups, it exceeded 60% in other groups. Correlational analysis revealed that there is a strong relationship (r = .89) between the standard deviation of teams' Elo ratings in a particular group and the difference between the highest and the lowest likelihood of match fixing for that group.

Based on this finding, the potential for match fixing situations was re-analysed by clustering the eight groups in each World Cup into high and low standard deviation subsets. To be categorised into the low standard deviation group, a standard deviation of less than 115 Elo rating points was required. By contrast, a standard deviation of greater than 140 Elo rating points was required to be categorised in the high standard deviation group.

As shown in Figure 3, when teams within a group were relatively even (that is, their Elo ratings maintained a low standard deviation), the likelihood of match fixing was relatively low, irrespective of the schedule for the group. By contrast, when the Elo rankings of teams were uneven (and thus the standard deviation within the group was high), the likelihood of match fixing was higher. This was particularly the case for schedules 1 and 3, whereby the top two ranked teams played each other first. Schedules 2 and 5 yielded the lowest potential for match fixing, and this corresponded to the two highest rank teams playing each in the final group game.

4. DISCUSSION

Results of this paper have revealed that potential match fixing situations were evident in three of the 13 possible group standings following each team playing two matches in the group stage of the FIFA World Cup. Whilst it may be argued that match fixing can occur irrespective of the state of the group, in the four circumstances highlighted in this



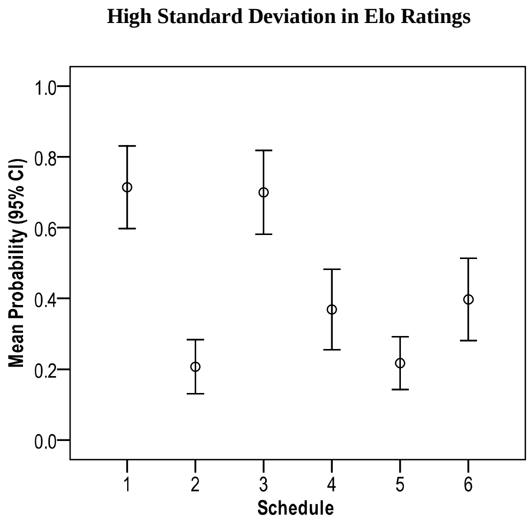


Figure 3. Mean probability of the potential for match fixing across the six possible group schedules for groups with a low and high standard deviation in Elo ratings.

paper, engaging in match fixing would be of no negative consequence to the losing team since qualification for the Round of 16 has already been achieved. In effect, the team's standing within the tournament would not be detrimentally impacted by drawing or losing their third group-stage match.

Simulation results have indicated that the schedule of matches has the potential to moderate the likelihood of match fixing opportunities in the World Cup. Specifically, whether the top ranked team played the second ranked team in the opening match, or the second or third group-stage match considerably impacted on the potential for match fixing. It is evident that the greatest risk of potential match fixing occurs when the top two ranked teams play each other in their first group stage match. In this scenario, if rank 1 defeats rank 2 in the opening match, and rank 1 then goes on to win their second match, rank 1 has the opportunity to influence whether rank 2 qualifies for the Round of 16 in their third group stage match. This is because rank 1 has already qualified for the Round of 16 prior to the match, and therefore can lose to rank 3 or rank 4 in order to prevent rank 2 qualifying.

The lowest risk of potential match fixing was evident when the two highest ranked teams played each other last. This is because the two highest ranked teams are most likely to win (or at least draw) their first two matches against weaker opposition, and then play against each other to decide who finishes on top of the group. Thus, playing each other last prevents either team from fixing a match against weaker opposition in order to prevent a higher ranked team from qualifying for the Round of 16.

A limitation of the two highest rank teams playing each other in the final group stage game should be noted. If the two highest ranked teams win their first two matches (which is probable given that the matches are against the third and fourth ranked teams in the group), then the two teams that will qualify for the Round of 16 will be known prior to any of the third and final group stage matches being played. In effect, these matches will only be played to determine who will finish first in the group, with the match between third and fourth being a dead-rubber. Whilst designing the schedule in a way that reduces the potential for match fixing has its incentives, the spectacle of the tournament may be diminished somewhat if a substantial number of teams are eliminated from the tournament prior to the third round of group stage matches. Whilst this occurs in a small number of groups in the competitions current format, engaging in dead rubbers in a large number of groups would not be beneficial for the tournament.

Based on this conundrum, the question that arises is when should a schedule be designed on the basis of match fixing potential? The answer may lie in the level of volatility within the group, and the difference in quality among the four teams. In effect, teams who have greater disparity in Elo rankings also had the greatest potential for match fixing. However, groups that contained teams who were relatively even (often referred to as groups of death) were less likely to have potential match fixing situations arise. This is typically because points awarded for wins and draws are likely to be more evenly distributed among the four teams in the group, with fewer teams qualifying for the Round of 16 prior to the final group stage match. As such, it appears most salient to adjust the schedule for groups that have high disparity in rankings across the four teams as opposed to adjusting the schedule for groups that are more evenly balanced.

5. CONCLUSION

In summary, this paper has presented on the potential for match fixing to occur during the group stage of the FIFA World Cup. Results have indicated that the likelihood of potential match fixing situations is dependent on the schedule and on the differential quality of teams within the group. Specifically, groups that have a high standard deviation in rankings are more likely to be susceptible to match fixing situations in the final match of the group stage, particularly if the top two ranked teams play each other in the first round of group stage matches.

6. REFERENCES

Caruso, R. (2009). The basic economics of match fixing in sport tournaments. *Economic Analysis and Policy, 39*, 355-377.

British Broadcasting Corporation (14 July, 2006a). Italian trio relegated to Serie B. Sourced from http://news.bbc.co.uk/sport2/hi/football/europe/5164194.stm

British Broadcasting Corporation (26 July, 2006b). Punishments cut for Italian clubs. Sourced from http://news.bbc.co.uk/sport2/hi/football/europe/5215178.stm

FIFA (2010a). Sourced from http://www.fifa.com/

FIFA (2010b). Sourced from http://www.fifa.com/aboutfifa/marketing/factsfigures/tvdata.html

FIFA (2010c). Sourced from http://www.fifa.com/worldcup/archive/edition=4395/results/index.html

Preston, I. and S. Szymanski (2003). Cheating in Contests, *Oxford Review of Economic Policy, 19*, 612-624.

The Age (September 1, 2008). World Cup matches fixed, says author. Sourced from http://www.theage.com.au/news/sport/world-cup-matches-fixed-says-author/2008/08/31/1220121049179.html

Torgler, B. (2006). Historical Excellence in Soccer World Cup Tournaments: Empirical Evidence with Data from 1930 to 2002, *Rivista di Diritto ed Economia dello Sport*, *2*, 101-117.

HOME CONTINENT ADVANTAGE ON THE AMERICAN AND EUROPEAN PROFESSIONAL GOLF TOURS

Michael Bailey[a], Stephen Clarke[b], Donald Forbes

[a] Department of Epidemiology & Preventive Medicine, Monash University
[b] Swinburne University
Corresponding author: michael.bailey@med.monash.edu.au

Abstract

Using all regular tournament rounds of golf played on both the American and European golf tours between 2000 and 2009, 860 golfers were identified as having played on both tours. By assigning the continent in which each player played the most rounds as being the 'home' continent, a mixed linear model was used to determine the effects of a 'home tour' advantage. Over 255,000 rounds of golf were examined. Overall scoring on the US tour was marginally lower than in Europe (71.58 vs 71.74 p=0.001). There was a significant difference between rounds (p<0.0001), with third round scores lowest (71.43) followed by second (71.63), first (71.78) and then final round (71.79). Golfers that played more often on the US tour were consistently over half a stroke better than golfers that played more regularly on the European tour (71.39 vs 71.93 p<0.0001). There was a significant interaction between continent and home tour (p<0.0001) with US tour golfers averaging 71.21 on the US tour and 71.58 on the European tour, while European tour golfers averaged 71.91 on the European tour and 71.95 on the US tour. Average scores appeared to have increased by about a stroke over the past 10 years, while experience and familiarity are worth about half a stroke per thousand rounds for experience and about half a stroke per sixty rounds that is played at each specific tournament. Significant differences exist between both the European and US golf tours and the players that play on these tours. There is also statistical evidence to support the theory that experience and familiarity are predictive of golf score.

Keywords: Sports, Golf, Home Advantage

1. INTRODUCTION

The role of home ground advantage (HA) has been shown to play an integral role in any analysis of sporting events. Whilst many different approaches have been used to quantify HA, the underlying reason why HA exists has been reduced to three basic principles; travel, familiarization and crowd support. When considering the sport of golf, professional golfers travel long distances between tournaments, have greater familiarity with courses that they play on more regularly and the unique proximity of the crowd and players that occurs in golf would suggest that home crowd should be a contributing factor.

This manuscript will seek to investigate the effects of a 'home tour' advantage that may exist in professional golf. By comparing scoring for players that have played on both the US and European golf tours and assigning the 'home tour' as being the continent played on the most, it is possible to quantify a 'home tour' effect and determine if it is consistent across both continents. Whilst data are not available to tease out the travel and crowd components of HA, by considering player performance relative to the number of times a player has played in each tournament it is also possible to derive a surrogate marker for familiarity component of HA.

2. BACKGROUND

Professional golf tournaments played on both the US and European tours are traditionally 72 hole events, played over 4 days, commencing on the Thursday of each week and finishing on the Sunday. Unless inclement weather conditions intervene, each daily round will consist of 18 holes of golf. Whilst tournament numbers vary, there are usually about 150 players that commence each tournament with numbers 'cut' after the first 2 rounds. Whilst 'making the cut' criteria may differ from tournament to tournament, generally speaking, only the leading half of players will go on to play the final 2 rounds.

The concept of HA has long been recognised as a phenomenon in sport and has been subject of much research. Schwartz and Barsky (1977) provided one of the first detailed studies examining HA in baseball, gridiron, ice hockey and college basketball. Subsequent studies have shown HA to exist to some degree in many other professional sports. Pollard (1986), Barnett & Hilditch (1993) and Clarke & Norman (1995) confirmed the existence of HA in professional soccer. Courneya & Carron (1992) provide a helpful taxonomy, listing a range of studies covering soccer, hockey, baseball, basketball and gridiron. HA is not limited to sporting events in the northern hemisphere. Lee (1999) has confirmed HA in Australian Rugby League, whilst Stefani & Clarke (1992) explored HA in Australian Rules football. This result was further confirmed by Bailey & Clarke (2004) and Clarke (2005). Neave and Wolfson, (2003) further identified a significant difference in salivary testosterone levels with soccer players playing at home having higher levels than those playing away.

All of the above studies relate to team sports, and little work has been done in investigating the extent to which home advantage is present in individual sports. In an overview of research on home advantage, Nevill & Holder (1999) list 50 references, of which only two (their own work on tennis and golf) were concerned with individuals rather than teams. Using regression methods Holder & Nevill (1997) found little evidence of home advantage in either the major tennis or golf tournaments held in 1993. Using log-log regression on the same data, Neville, Holder et al (1997) came to the same conclusions. However both these studies had very limited data, comprising only four golf tournaments in a single year. In the only other study to explore HA in golf, Wright, Christie et al. (1991) compared the performances of British (home) and foreign (away) representatives in the British Open Golf Championships from 1946 to 1980. In contrast to the majority of studies into HA, they found there was a home country disadvantage for British golfers that they attributed to pressure from home crowds. However this result must be treated with some scepticism, as they did not account for the different abilities of British and overseas players.

3. METHODS

A database was compiled of all 18 hole rounds of professional golf played in the past 10 years (2000-2009) on both the US (177,790 rounds) and European (170,912 rounds) golfing tours. This database was then reduced to a list of 860 golfers that had officially played golf on both tours, leaving a database of 255,453 rounds of golf. Information was collected on date, tournament, player, round and score. By sorting the data in chronological order, the number of rounds played by each player was generated to provide a marker of player experience. Similarly, the number of rounds played by each player at each tournament was also generated to provide a surrogate marker for the familiarity component of HA. Players were assigned the continent in which they had played the most rounds as being the 'Home Tour'. To further explore the effects upon the more experienced golfers, a subset analysis was performed on the 128 players that had played a minimum of 50 rounds of golf on both tours.

Statistical Analysis

Statistical analysis was performed using SAS Version 9.1 (SAS Institute Inc., Cary, NC, USA). Round scores were found to be well approximated by a normal distribution, which facilitated the use of mixed linear modelling using the PROC MIXED procedure in SAS. The MIXED procedure uses a restricted maximum likelihood algorithm that enables specific modelling of the within-player covariance structure and further enables effects to be treated as FIXED or RANDOM. A multivariate model was fitted including fixed effects for year, round, tour, home tour, tour*home tour, experience and tournament experience (familiarity) with individual players and tournaments treated as random effects.. Results have been reported as means ± standard errors with error bars on graphs representing standard errors. A two sided p-value of 0.05 was considered to be statistically significant.

4. RESULTS

The database comprised a total of 255,453 rounds played by 860 golfers. On the US tour, there were 141,868 rounds played by 452 golfers, whilst on the European tour there were 113,680 rounds played by 408 golfers. Whilst the number of golfers playing per year remained consistently around 400 on both tours, the number of rounds played per year showed a slight increase. (US 2001: 12,677, US 2009: 14,284. Euro 2001: 10,939, Euro 2009: 12,682). US golfers played a median of 564 rounds of golf in the US and a median of 16 rounds of golf in Europe, while European golfer played a median of 482 rounds in Europe and a median of 32 rounds in the US. As each tournament is held annually, the maximum number of rounds played per tournament was 40 and the median number of rounds played per tournament was 5 on both tours.

Overall golf scores were normally distributed with a mean of 71.4 and a standard deviation of 3.3. (median & mode = 71). (Figure 1)

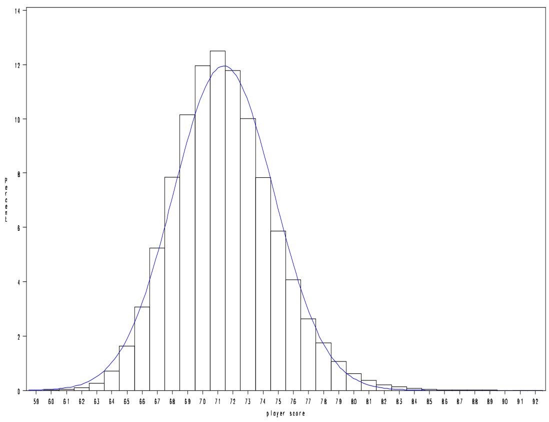


Figure 1: Histogram of golf scores on US & European tours 2000-2009 (*n*=255453)

Scoring on US vs European Tours

When considering all 860 players that had played on both tours, overall scoring on the US tour was marginally lower than scores in Europe (71.58±0.07 vs 71.74±0.07 *p*=0.001). When looking specifically at the 128 players that had played a minimum of 50 rounds on both tours, the overall scoring was lower although the difference between tours was roughly the same with scores on the US tour 0.18 strokes lower than scores in Europe. (70.93±0.09 vs 71.11±0.09 *p*=0.005). (Figure 2) As prize money on the European tour is lower, it is realistic to expect that the US tour would attract the better players which would explain the slightly lower scoring on the US tour.

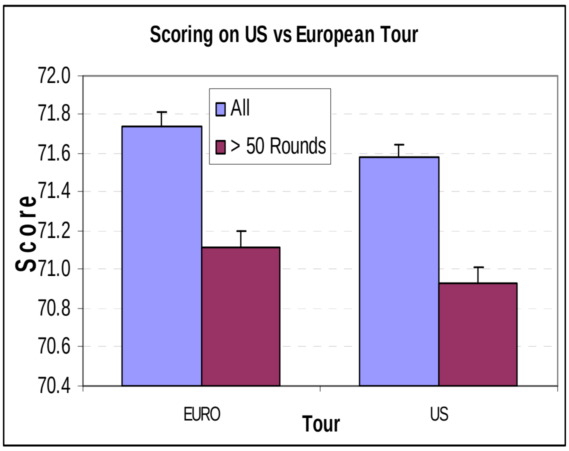


Figure 2: Scores per round in Europe and the US for all 860 golfers that played on both tours (*n*=255,453 rounds) and for the 128 golfers that played a minimum of 50 rounds on both tours (*n*=81,403 rounds).

Scoring per Round

There was a significant difference between rounds (*p*<0.0001), with third round scores lowest (71.43±0.07) followed by second (71.63±0.06), first (71.78±0.06) and then final round (71.79±0.07). When considering the more experienced players that had played a minimum of 50 rounds on both tours, the scoring pattern remained the same with the lowest scores also occurring in Round 3. (Figure 3) This scoring pattern across rounds reflects a combination of two factors. As only the better golfers for each tournament go on to play the final two rounds, it is realistic to expect higher scores in the first two rounds. Secondly, the location of the hole on the green has a big impact on scoring. Lower scores in the third round probably reflect easier pin placement whilst higher scoring in the final round would reflect a policy to make the holes harder in the ultimate round.

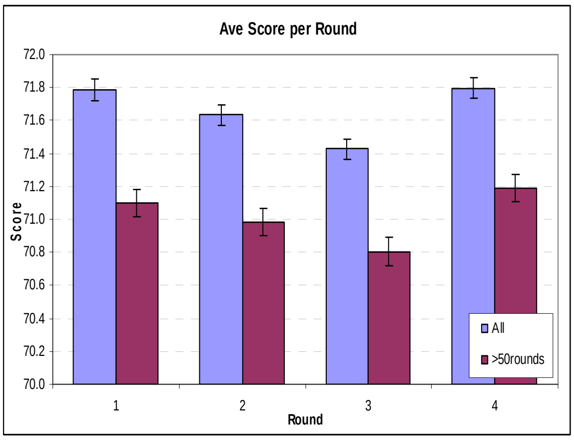


Figure 3: Scores per round for all 860 golfers that played on both tours (n=255,453 rounds) and for the 128 golfers that played a minimum of 50 rounds on both tours (n=81,403 rounds).

Scoring per year

There was a fairly consistent increase in scoring over the 10 year period of data collection across both continents, with scores peaking in 2008 before a slight fall in 2009. (Figure 4) Whilst the true reason for this pattern remains unknown, it could reflect either a drop in the standard of golfers on tour, or an increase in the difficulty of golf courses.

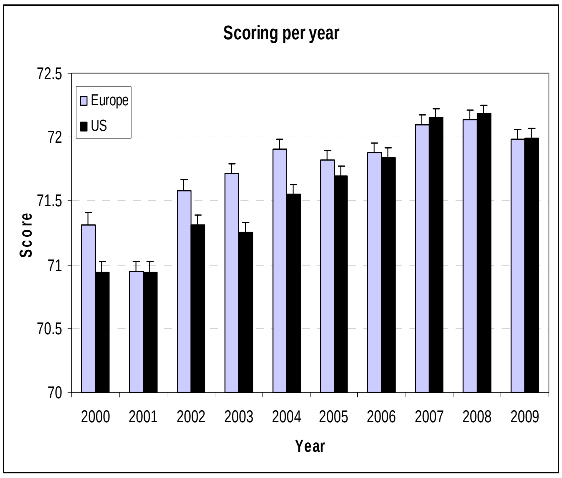


Figure 4: Average scores per year for both European and US tours

US versus European golfers

Golfers that played more often on the US tour were over half a stroke consistently better than golfers that played more regularly on the European tour (71.39±0.07 vs 71.93±0.08 p<0.0001). This difference was less pronounced for the more experienced golfers with US golfers about a quarter of a stroke better off. (70.89±0.09 vs 71.15±0.11 p=0.04). When considering all rounds, there was a significant interaction between continent and home tour (p<0.0001). US golfers averaged 71.21±0.08 on the US tour and 71.58±0.08 on the European tour, while European golfers averaged 71.91±0.08 on the European tour and 71.95±0.08 on the US tour. When considering those that had played a minimum of 50 rounds on both tours the message was the same with experienced US golfers scoring about 0.4 strokes lower in the US than in Europe, while European golfers were only 0.05 strokes worse off when playing in the US.

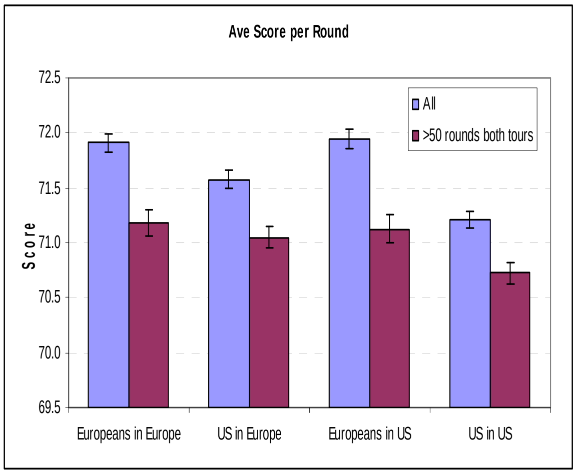


Figure 5: Overall scoring for all 860 golfers that played on both tours (n=255,453 rounds) and for the 128 golfers that played a minimum of 50 rounds on both tours (n=81,403 rounds).

Experience

When considering all players, experience is a highly significant predictor of scoring (p<0.0001) with scores improving at a rate of about 3/4 of a stroke per 1000 rounds (0.00074±0.0001 p<0.0001) This improvement remains similar for those who have played a minimum of 50 rounds on each tour (0.00075±0.0001 p<0.0001). As the data set includes a high percentage of players that commenced their career prior to the year 2000, this is unlikely to be an accurate reflection of the true effect of experience. To derive a less biased estimate, all players that commenced their career prior to 2001 were removed from analysis leaving a database of 62,722 players. When considering this subset of players, the effect of experience was closer to a half a stroke for every 1000 rounds. (0.00048±0.0002 p<0.0001).

Familiarity

When considering all players, there was evidence to show that improvement occurs with each round played at a tournament by a rate of 0.012±0.0015 strokes per round (p<0.0001). This rate is slightly less for those who have played a minimum of 50 rounds on each tour (0.0079±0.002 p=0.0004), and slightly more when considering a subset of players who commenced after the year 2000 (0.016±0.004 p=0.0001).

Individual golfers

Tiger Woods dominance of world golf can clearly be seen by individual player averages over the past 10 years. (Table 1) Woods averages 68.0±0.15 strokes per round overall, 68.0±0.16 in US and 68.1±0.27 in Europe. The difference in scoring average from Woods to the next golfer is 1.4 strokes; amazingly, only 0.8 of a stroke then separates the remaining 19 players that rank in the top 20.

5. DISCUSSION

There is statistical evidence to support HA in professional golf. Whilst US golfers perform better in both the US and Europe, they also have the greatest discrepancies between performances at home versus performances away. Whilst the US tour has more prize money and lower scoring suggesting a higher standard in the US, the reason why the magnitude of HA differs between US and European golfers is open to speculation.

This retrospective analysis of home continent advantage has both strengths and weaknesses. With over a quarter of a million rounds of golf, it represents one of the most comprehensive reports on professional golf in publication. Unfortunately however there was a host of relevant information that could readily affect results that were not collected. Course name, par score, golfer age, golfer nationality, weather conditions, pin placement and travel schedules are just some of the many parameters that could affect scoring. Given that familiarly is an established component of HA, it was also necessary to assign the home continent to the tour in which the most rounds were played. This occasionally resulted in golfers that were born and raised in Europe paradoxically being assigned a home continent of the USA as they had played more professional golf in the US than in Europe (and vice versa). This essentially meant that continent familiarity has been used as a surrogate marker for HA.

As some tournaments such as the US and British Open are regularly played on different courses, our current marker of the familiarity, namely rounds played per tournament can only be considered as a surrogate marker for true familiarity. A further pitfall of measuring player and tournament experience from a 10 year snapshot of data arises because the true level of experience for players commencing prior to 2000 will not be known. To accommodate for this we have considered a subset of players for whom complete career information was known.

There is further evidence to suggest that it may not be appropriate to treat both experience and familiarity as having linear relationships with scoring as both may tend to have an exponential relationship with a reduction in scoring reducing over time. Because of the previously mentioned pitfalls, it was considered beyond the scope of this manuscript to explore familiarity and experience in depth.

It is interesting to note that player scoring has increased by over a stroke per round over the last 10 years. This may be attributed to the fact that more tournaments are played now than ten years ago, but, given the continued improvements in golfing technology, it would be realistic to expect that scoring would go down over time rather than up. Perhaps we can conclude that tournament organisers and course managers are finding ways to modify courses and place pins so that increased length and accuracy do not equate to reduced scoring.

One potential reason for this increase in scoring may in part be due to the Tiger Woods' effect. Tiger is clearly head and shoulders above the rest of the golfing fraternity. The difference in averages (1.4 strokes fewer per round) from Tiger to the second ranked player (Phil Mickelson) is the same magnitude as the difference from Phil Mickelson to the 68th ranked golfer! With Tiger continually shooting low scores, measures may have been made to make courses and scoring harder which has resulted in increased scoring for most others.

Finally, whilst this manuscript has been able to show highly statistically significant effects (p<0.0001), the magnitude of this database results in an imbalance between statistical and practical significance. As any aspiring golfer will tell you, the difference between a good and bad round can come down to a matter of millimetres, in which case home continent advantage may count for very little!

6. CONCLUSIONS

Scores shot on the US tour are significantly lower than scores shot in Europe. Whilst there is a home continent advantage of approximately 0.4 strokes per round for golfers playing primarily on the US tour, the advantage is not present for golfers that play primarily on the European Tour. There is statistical evidence to show that differences exist between rounds, years and across continents and that familiarity and experience are both significant predictors of scoring. Furthermore, there is clear statistical evidence to support the argument that Tiger Woods is a class above all other golfers.

Overall Rank	Golfer	All	All SE	Euro	Euro SE	Euro Rank	US	US SE	US Rank	Diff Euro-US
1	Tiger Woods (US)	68.0	0.15	68.1	0.27	1	68.0	0.16	1	0.1
2	Phil Mickelson (US)	69.4	0.15	69.8	0.30	7	69.3	0.16	2	0.5
3	Vijay Singh (US)	69.4	0.14	69.6	0.28	3	69.3	0.15	3	0.3
4	Ernie Els (US)	69.4	0.14	69.3	0.19	2	69.5	0.17	5	-0.2
5	Jim Furyk (US)	69.5	0.14	69.9	0.33	11	69.4	0.15	4	0.4
6	Sergio Garcia (US)	69.7	0.14	69.6	0.20	4	69.7	0.17	8	0.0
7	Anthony Kim (US)	69.7	0.24	69.7	0.54	6	69.6	0.26	6	0.2
8	Retief Goosen (US)	69.7	0.14	69.8	0.17	10	69.7	0.18	7	0.2
9	Padraig Harrington (US)	69.8	0.14	69.8	0.17	8	69.8	0.19	10	0.0
10	Luke Donald (US)	69.8	0.16	69.7	0.24	5	69.8	0.18	9	-0.2
11	Davis Love III (US)	69.9	0.15	70.0	0.38	12	69.9	0.15	13	0.0
12	Adam Scott (US)	70.0	0.15	70.2	0.19	17	69.9	0.19	11	0.3
13	David Toms (US)	70.0	0.14	70.9	0.39	59	69.9	0.15	12	1.0
14	Camilo Villegas (US)	70.0	0.20	69.8	0.44	9	69.9	0.21	14	-0.1
15	Zach Johnson (US)	70.0	0.17	70.1	0.43	15	69.9	0.17	15	0.2
16	Kenny Perry (US)	70.0	0.14	70.2	0.43	16	70.0	0.15	18	0.2
17	Stewart Cink (US)	70.1	0.14	70.7	0.34	39	70.0	0.15	19	0.7
18	Scott Verplank (US)	70.1	0.14	71.4	0.36	93	70.0	0.15	17	1.4
19	Mike Weir (US)	70.1	0.14	70.6	0.35	34	70.1	0.15	20	0.6
20	Robert Allenby (US)	70.2	0.14	70.3	0.28	24	70.2	0.15	24	0.2

*SE=standard error

Table 1: Top 20 golfers with lowest overall scoring average (>50 rounds both tours)

References

Bailey, M., and Clarke, S. R. (2004). Deriving a profit from Australian Rules football: A statistical approach. *Proceedings of the Seventh Australian conference on Mathematics and Computers in Sport*. Palmerston North, Massey University, R Hugh Morton & S Ganesalingam: 48-56.

Barnett, V. and S. Hilditch (1993). The effect of an artificial pitch surface on home team performance in football (soccer). *Journal of the Royal Statistical Society (Series A)* 156(1): 39-50.

Clarke, S. R. (2005). *Home advantage in the AFL. Journal of Sports Sciences* 23 (4): 375-385.

Clarke, S. R. and J. M. Norman (1995). Home ground advantage of individual clubs in English soccer. *The Statistician* 44, No 4, p509-521.

Courneya, K. S. and A. V. Carron (1992). The home advantage in sport competitions: A literature review. *Journal of Sport & Exercise Psychology* 14: 13-27.

Holder, R. L. and A. M. Nevill (1997). Modelling performance at international tennis and golf tournaments: is there a home advantage? *The Statistician* 46(4): 551-559.

Lee, A. (1999). Modelling rugby league data via bivariate negative binomial regression. *Australian and NZ Journal of Statistics* 41(2): 141-152.

Neave N, Wolfson S. 2003. Testosterone, territoriality, and the 'home advantage'. Physiol Behav 78:269-275.

Neville, A.M, Holder, R.L, et al, (1997) Identifying home advantage in international tennis and golf tournaments, *Journal of Sports Sciences*, 15, 437-443

Nevill, A. M. and R. L. Holder (1999) Home advantage in sport: an overview of studies on the advantage of playing at home. *Sports Medicine* 28 (4) 221-236

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sport Sciences* 4: 237-248.

Schwartz, B. and S. E. Barsky (1977). The home advantage. *Social Forces* 55: 641-661.

Stefani, R. T. and S. R. Clarke (1992). Predictions and home advantage for Australian rules football. *Journal of Applied Statistics* 19(2): 251-261.

Wright, E.F, Christie, S.D, et al, (1991). The home-course disadvantage in golf championships: Further evidence for the undermining effect of supportive audiences on performance under pressure. *Journal of Sport Behaviour*, 14, 51-60.

STROKER: SOFTWARE FOR PRESENTING ROWING DATA

Bob Buckley, Conny Draper

Australian Institute of Sport, Australia
Corresponding author: robert.buckley@ausport.gov.au

Abstract

Data logging devices are used increasingly during competition and training in elite sport. Stroker is a small program used to present data collected using the MinimaxX and Weba devices in Australian elite rowing competition and training (Draper C, 2010).

Rowing and the data collection processes used have some specific features that affect the design of the program and the data presentation. A 2000m race takes over 220 strokes.

Elite rowers in Australia routinely attach a MinimaxX device to their boat. The device is set in rowing mode. The MinimaxX device collects tri-axial accelerometer, gyro- and magnetometer data at 100Hz as well as GPS information at 5Hz. After the race or training session the device is taken from the boat and put in the MinimaxX cradle. The Logan software downloads the data stored on the device and performs basic transforms to create the data file.

The Stroker software was created as a prototype (or proof of concept) to explore ways to display data from the AIS's Logan software program. It aims to display the available data so analysts can investigate the nature of rowing performance. Stroker might be described as a simple data visualisation tool.

Stroker was quickly adopted to provide reporting to coaches at the elite level. It is a work in progress. It shows that some simple approaches still need to be explored in the context of sport science.

Colour-coded multi-dimensional graphical displays were found to be a powerful analysis and visualisation tool to profile comprehensive stroke by stroke (sbs) changes of characteristic boat-related curve patterns in relation to the sbs average boat velocity and stroke rate throughout a rowing race (Draper C, 2009). Data from 114 races of 43 Australian boats in 18 classes were analysed, the sbs data detected, statistical tests and graphical sbs colour-coded technique & performance-related time scatter plots were used to profile boat performance and assess stroke consistency patterns in relation to boat velocity (Draper C, 2008).

Keywords: elite rowing, data visualisation, instrumentation, GPS, accelerometer

References

Draper C, Rice T, et al (2008). Characteristic curve patterns of 3D boat motion in international rowing races in all boat categories. *13th Annual Congress of the European College of Sport Science*, Estoril/ Portugal .p.558.

Draper C, Ting KM, et al (2009). Profiling Rowing races and crews using visualisation techniques to assess multi-dimensional boat motion. *14th Annual Congress of the European College of Sport Science*, Oslo/ Norway.

Draper C, (2010). Identifying critical determinants of rowing. *High Performance Sport Research Workshop*, Australian Institute of Sport, Canberra.

SKILLS ACQUISITION IN BADMINTON: A VISUAL BASED APPROACH TO TRAINING

Minh Vu Huynh [a,b] , Anthony Bedford [a]
[a]School of Mathematical and Geospatial Sciences, RMIT University
[b] Corresponding author: minh.huynh@student.rmil.edu.au

Abstract

Currently, there are a number of training programs that attempt to improve decision making and awareness in badminton. However, these programs are extremely limited, and do not provide athletes with the necessary improvements needed to optimise their in-game performance. In developing and improving decision making, the ideal strategy would be to expose the athlete to all possible situations and scenarios that they may face. This allows them to retain certain responses in their subconsciousness; leading their bodies to instantaneously select the appropriate action to take in similar situations. This paper provides an overview of the electronic training program currently being developed to improve reaction time and awareness in badminton players. Particular emphasis is placed on a player's ability to estimate and predict shuttle location. Using this program, we will be able to identify the player's awareness and attempt to improve their in-game performance and decision making. These findings will not be limited to badminton, and applicability to other sports will be discussed.

Keywords: Skills acquisition, badminton, visual training, decision making

1. INTRODUCTION

In any fast paced sport or activity, the outcome of a match may be defined by the ability to make decisions quickly and accurately (Blomqvist, Luhtanen and Laakso, 2000). Considering the importance of swift decision making, it has become imperative that athletes train and improve their ability to instantaneously determine the best course of action. However, improving an individual's capacity for decision making is more complex and detailed (Macquet and Fleurance, 2007) than improving physical abilities such as strength or agility. Currently, coaches and trainers attempt to maximise athletes' physical skills and competencies (Chin, Wong, So, Siu, Steininger and Lo, 1995; Fahlstrom, Lorentzon and Alfredson, 2002), rather than their decision making awareness. In attempting to optimise an athlete's competency, the development of a training program that incorporates the improvement of reaction time and awareness in juxtaposition with physical performance would be ideal.

The traditional approach to training badminton players follows a three step sequential training process: perception, decision-making and movement execution (Abernethy, 1996; Blomqvist, Luhtanen and Laakso, 2001) training. Typically, coaches and trainers place heavy emphasis on the movement execution component of the traditional training method (Blomqvist, Luhtanen and Laakso, 2001) and tend to overlook the significance of the cognitive processes of perception and decision-making. This is unfortunate however, considering the quality of decision-making in a game situation is often as important as the execution of the motor skills (Blomqvist, Luhtanen and Laakso, 2001; Thomas, 1994).

Currently, the majority of training methods focus on improving a badminton player's physical capabilities. Minimal research has been carried out to examine the use of other training methods (Blomqvist, Luhtanen and Laakso, 2001) such as creative problem solving and video-based methods in badminton, despite numerous studies suggesting that video-based training methods can improve perceptual skills (decision accuracy and decision-making speed) in athletes (e.g., Abernethy, Woods and Parks, 1999; Christina, Barresi and Shaffner, 1990; Farrow, Chivers, Hardingham and Sachse, 1998; Starkes and Lindley, 1994). The small amount of training programs that do attempt to improve decision making and awareness in badminton players are however extremely limited (Macquet and Fleurance, 2007), and do not provide athletes with the necessary improvements needed to optimise their in-game performance.

While it is essential that athletes continuously train and improve their physical capabilities (Chin, Wong, So, Siu, Steininger and Lo, 1995; Fahlstrom, Lorentzon and Alfredson, 2002), it seems evident that cognitive components of badminton must not be underemphasised when training athletes (Blomqvist, Luhtanen and Laakso, 2001). In developing and improving decision making, the ideal strategy would be to expose the athlete to all possible situations and scenarios that they may face. This allows them to retain certain responses in their subconsciousness; leading their bodies to instantaneously select the appropriate action to take when similar situations arise (Hall, Schmidt, Durand and Buckolz, 1994). With the traditional training method, it is extremely difficult to expose athletes to all possible situations and scenarios they may face as they acquire a regular standard of play from training with the same athletes and coaches. As such, video based training methods expose players to an abundance of different scenarios and situations, preparing them for in-game utilisation (Blomqvist, Luhtanen and Laakso, 2001).

The purpose of this study was to develop a visual based training program that aims to improve reaction time and awareness in badminton players. Particular emphasis is placed on a player's ability to estimate and predict shuttle location. Using this program, we will be able to identify the player's awareness and attempt to improve their in-game performance and decision making.

2. METHODS

Participants

The participants for this study were collected from two separate groups: (i) athletes from the Australian Badminton Olympic (junior division) team (n = 3) and (ii) college students (n = 13) from RMIT University. All college participants had prior experience in playing badminton, with most participants having played in high school round robins. Athletes from the Australian Olympic team in conjunction with some members from RMIT served as the experimental group (n = 8) and were assigned to the treatment group (age mean = 23, SD = 13.46). College students were randomly assigned to either the treatment group (with the Olympic team) or the Control group (n = 8, age mean = 18.63, SD = 3.78).

Measures

To effectively improve and evaluate decision-making and reaction time in badminton players, we developed a visual based training (VBT) program using Microsoft's Visual Basic for Applications (VBA) software. Since particular emphasis was placed on improving player's ability to estimate and predict shuttle location using skills acquisition, the program was named Skills Acquisition Trainer for Badminton (SATB).

Knowledge Measures Test

Prior to administering the treatment program (SATB) participants were given a Knowledge Measures Test (KMT) to evaluate their comprehension and awareness of badminton rules, strategies and techniques. The KMT was a simplified version of McGee and Farrow's (1987) book "Test questions for physical education activities," and was designed to test participant's knowledge of different shot types used in badminton. Participants were timed on how quickly they could match up a shot type with a description of a shot from eight choices. Figure 1 gives an example of the questions in the KMT.

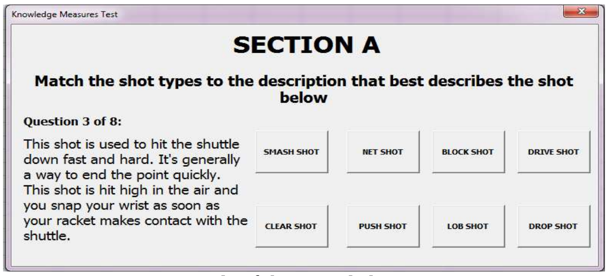


Figure 1: An example of the Knowledge Measures Test

Following the KMT, the researcher would administer the SATB. The SATB is a VBT program that consists of ten visual questions. Participants would watch different clips of badminton rallies being played, with sequences running for 2 – 30 seconds and was followed by a still frame for 1 second with which participants would be asked to answer (on screen) what type of shot was about to take place (e.g. drop shot) as well as the location that shot will be played (e.g. middle right). An example of this is shown in Figure 2 with the corresponding answer options.

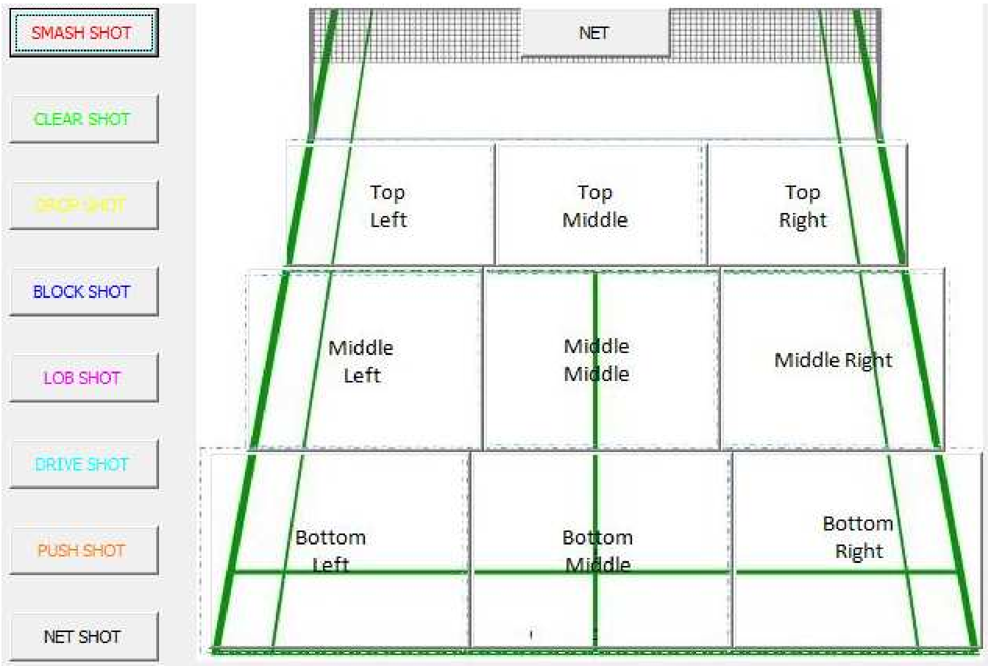


Figure 2: An example of the SATB answer selection screen

Testing Procedure

Participants were tested before and after the treatment period for shot type knowledge and speed using the KMT. Participants in the treatment group were also administered the SATB once a week for four weeks to examine their awareness, decision making, and response time. Participants in the Control group were only given the KMT to complete (in week one and week four). Participants' performance in-game was also recorded on a weekly basis to examine if they could apply the skills acquired from the SATB in a live match. These video recordings were filmed at the Melbourne Sports and Aquatic Center (MSAC) for the Olympic team and the Aqualink Leisure Centre for the college students.

The SATB was based on a weighted system, with the assistance of experts' judgement and opinion (coaches and trainers who have played and taught for many years). Two points were awarded for the correct shot type response, one point for other possible shot types in that situation, and no points for any other shot types. Similarly, two points were given if participants chose the correct location the shuttle will land, one point if it is adjacent to the shuttle location, and no points for any other location selection. Participants are also timed from the point the rally sequence finishes to the point they had inputted their response to shot type and location to examine response time. With ten questions per session, the maximum score a participant could acquire was 40, with an optimal time of 11.9 seconds. Hence, the equation for the SATB score is given by:

$$SATB = \frac{11.9}{TIME} \times SCORE \qquad (1)$$

From equation 1, *TIME* is the combined time it took participants to answer all ten questions regarding shot type and shuttle location. The value of 11.9 was based on Jorgensen, Garde, Laursens and Jensen (2002) study "*Using mouse and keyboard under time pressure: preferences, strategies and learning*" (a click response time = 1.1±0.08 s), in conjunction with experts opinion that it would take two seconds to select both location and shot type. *SCORE* is the combined score of each correct response from the ten questions. Therefore the maximum score an individual can acquire on the SATB is 40. Upon completion of the final SATB session (week eight) participants are given the KMT once again to examine the change in badminton knowledge and awareness that the SATB has on participants.

3. RESULTS

A repeated measures ANOVA for the SATB was carried out to examine the change in scores over the six weeks of experimentation. Assumptions of normality, homogeneity of variance and sphericity ($\chi^2(5) = 9.294$, $p = 0.102$) were met. Results showed that differences between conditions were unlikely to have arisen by sampling error ($F(3,21) = 17.23$, $p < .001$); an overall effect size of 0.77 (partial η^2) showed that 77% of the variation in score can be accounted for by improvement over time.

Similarly, a repeated measures ANOVA was carried out to examine the change in scores on the KMT across the experimental and control groups. Results revealed a significant interaction, $F(1,7) = 143.36$, $p < .001$ between the two groups. Figure 3 shows that the experimental group was able to improve their badminton knowledge significantly ($p < .001$) compared to the control group.

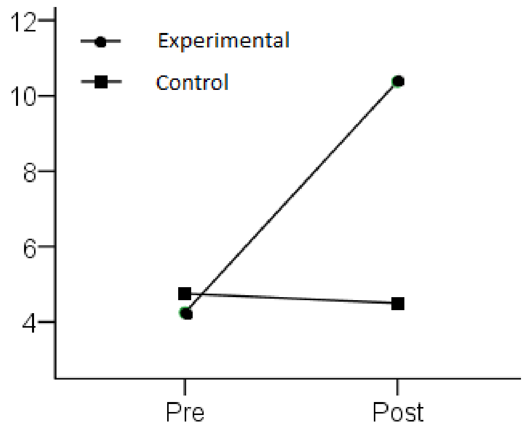


Figure 3: Pre and post test scores on the KMT for the experimental and control groups

Also, one way repeated measures ANOVAs were carried out to examine significance between time and score for both the KMT (refer to table 1) and SATB (refer to table 2). Results revealed a significant improvement in time on the KMT ($F(1,7) = 9.42$, $p = 0.02$) as well as the SATB ($F(3,21) = 15.12$, $p < .001$).

	Pre		Post	
	M	*SD*	*M*	*SD*
Experimental	276.12	94.64	192.21	31.35
Control	284.47	65.49	279.55	71.12

Table 1: Times on the KMT (in seconds) for both groups

	Pre		Post	
	M	*SD*	*M*	*SD*
SATB	91.44	37.04	32.77	7.96

Table 2: Times on the SATB (in seconds) for both groups

4. DISCUSSION

The results from this study showed that VBT methods helped to improve participants' reaction time and awareness in badminton. The significant interaction on the KMT was caused by an increase in the scores of the experimental group on the post-test, while the scores of the control group either remained the same, or decreased. Furthermore, participants' response time on the SATB improved significantly over the eight week period which was further supported by their improved response time on the KMT and in-game matches. The overall results from this study revealed that participants were able to improve their badminton knowledge (awareness and decision making) in conjunction with their reaction time, utilising the SATB.

These results are consistent with other studies that have utilised similar methods in attempting to improve athletes' skill and performance (Blomqvist, Luhtanen and Laakso, 2001; Christina, Barresi and Shaffner, 1990) via a VBT method. Blomqvist and colleagues (2001) argue that although general tactics will develop in athletes automatically from just playing the game, good decision making skills will only develop if taught extensively. The authors suggest that visual-based learning tasks encourage participants to develop their tactical awareness, bringing cognitive aspects of their game to a conscious level. Similarly, the present study found that participants who used the SATB showed a consistent improvement in their ability to predict and react to the visual-based sequences over the eight week period.

Interestingly, in sessions with sequences involving athletes they have already seen play, their score on the SATB was higher than it was if they were seeing those athletes play for the first time. Tong and Hong (2000) suggest that there are many different playing styles in badminton, varying from strength types to speed types. The authors suggest that knowing your opponents play style can help you predict what type of shot they will play and where they will hit the shuttle. The SATB supported this notion, with participants scoring significantly higher in sessions where they have acquired an athlete's playing style from viewing them play in previous sessions.

Further points of interest regarding style of play was noted by the researchers after viewing the video recordings of both the treatment and control groups' in-game matches and training. Firstly, participants in the treatment group were able to improve their overall badminton skills as well as respond more efficiently (with greater speed and reaction time) to their opponents' actions. Secondly, participants in the control group were able to improve their overall badminton skills and showed slight improvement in their ability to predict their opponents' next action. Both groups played/trained in badminton on a weekly basis so it was expected that both groups would improve their in-game skills but it was surprising to find that the Control group had also improved in predicting their opponents' actions. A possible explanation for this might be that the Control group only consisted of eight participants and after eight weeks of playing/training together, they had already played with each member in that group numerous times. This essentially leads the players to become used to each others' playing styles, hence the increase in predicting their opponents' actions.

A number of points about the methodology should be noted. Blomqvist et al. (2001) suggest that age and experience affects the outcomes of the experiment. The present study supports this argument with the results being skewed by participants of a higher skill level. Future studies should use larger sample sizes with non-random groups (perhaps a beginner, intermediate, advanced system) to avoid distorted outcomes. Secondly, although this study can be extended to other sports, it is not applicable to invasion games (e.g. soccer, hockey) due to differences in the number of players and the tactical aspects of evasion games (Blomqvist et al., 2001). The sports that would be applicable to this type of study would ideally be racket sports such as tennis and squash. The present study would need to be recoded (SATB and KMT) for application to invasion games.

5. CONCLUSIONS

With the majority of coaches and trainers placing little emphasis on the cognitive aspects of training, these findings will introduce a new form of preparation that is both effective and successful in developing the optimal athlete. The researchers found that VBT methods were effective in improving reaction time, awareness and decision making in badminton players. As such, it can be an issue for future studies to apply VBT to other sports for coaches and trainers to develop the ideal training method.

Acknowledgments

We wish to thank the coaches and athletes from Badminton Australia for their participation in the study. Furthermore, we would also like to thank the students of RMIT University for their participation in the study.

References

Abernethy, B. (1996). Training the visual-perceptual skills of athletes. *The American Journal of Sports Medicine, 24, 89-92*.

Abernethy, B., Wood, J.M., & Parks, S. (1999). Can anticipatory skills of experts be learned by novices? *Research Quarterly for Exercise and Sport, 70*, 313-318.

Blomqvist, M., Luhtanen, P., & Laakso, L. (2000). Expert-novice differences in game performance and game understanding of youth badminton players. *European Journal of Physical Education, 5*, 208-219.

Blomqvist, M., Luhtanen, P., & Laakso, L. (2001). Comparison of two types of instructions in badminton. *European Journal of Physical Education*, *6*, 139-155.

Chin, M., Wong, A.S., So, R.C.H., Siu, O.T., Steininger, K., & Lo, D.T.L. (1995). Sports specific fitness testing of elite badminton players. *British Journal of Sports Medicine, 29,* 153-157.

Christina, R.W., Barresi, J.V., & Shaffner, P. (1990). The development of response selection accuracy in football linebacker using video training. *The Sport Psychologist, 4,* 11-17.

Fahlstrom, M., Lorentzon., & Alfredson, H. (2002). Painful conditions in the Achilles tendon region in elite badminton players. *American Journal of Sports Medicine*, *30*, 50-54.

Farrow, D., Chivers, P., Hardingham, C., & Sachse, S. (1998). The effect of video-based perceptual training on the tennis return of serve. *International Journal of Sport Psychology*, *29*, 231-242.

Hall, C., Schmidt, D., Durand, M.C., & Buckolz, E. (1994). Imagery and motor skills acquisition. In Sheikh, A.A., & Korn, E.R. (Eds.) *Imagery in sports and physical performance* (pp. 121-133). *Baywood Publishing,* NY: Amityville.

Jorgensen, A.H., Garde, A.H., Laurens, B., & Jensen, B.R. (2002). Using mouse and keyboard under time pressure: Preferences, strategies and learning. *Behaviour & Information Technology*, *21*, 317-319.

Macquet, A.C., & Fleurance, P. (2007). Naturalistic decision-making in expert badminton players. *Ergonomics*, *50*, 1433-1450.

McGee, R., & Farrow, A. (1987). Test questions for Physical Education Activities. Champaign, II., Human Kinetics.

McMorris, T. (1998). Teaching games for understanding: Its contribution to the knowledge of skill acquisition from a motor learning perspective. *European Journal of Physical Education*, *3*, 65-74.

Starkes, J.L., & Lindley, S. (1994). Can we hasten expertise by video simulations? *Quest*, *46*, 211-222.

Thomas, K.T. (1994). The development of sport expertise: From Leeds to MVP legend. *Quest*, *46*, 199-210.

Tong, Y.M., & Hong, Y. (2000). The playing pattern of world's top single badminton players. *Journal of Human Movement Studies*, *38*, 185-200.

TALENT ID IN 100M SPRINTING: HOW THE WORLD'S BEST SPRINTERS SOLVE THE PROBLEM BASED ON MATHEMATICAL MODELS AND MULTIVARIATE STATISTICS

Ian Tim Heazlewood[a]

[a] *Charles Darwin University*
Corresponding author: *Ian.Heazlewood@cdu.edu.au*

Abstract

Talent identification is now common practice in Australian sport to identify future Olympians and Paralympians. Mathematical models and multivariate statistical approaches to predicting performance in sport are frequently recommended, however rarely implemented, to identify those factor that are associated with high performance athletes. Australia had difficulty finding talent to complete in the 100m at the recent 2008 Beijing Olympics and at the 2009 World Athletics Championships in Berlin. This research is focussed on how the best 100m sprinters, both male and female, actually run the race and to derive models for high performance sprinting. The research analysed data (IAAF, 2009) from the recent 2009 World Athletics Championships in Berlin. The data set consisted of 23 male performances from the 100m final and semi-finals and 36 female performances from the 100m final and the fastest athletes in the 100m sprint rounds. The data analysed consisted of reaction time to the gun, race segment time for 20m, 40m, 60m, 80m and 100m and conversion of race segments times to average velocity per 20m. Comparisons were conducted for male and female sprinters in terms of multivariate factor analysis of times per segment to assess if the race segments represented distinct race constructs or were representing the underlying factor of sprint ability, nonlinear regression (curve estimation) of average velocity with distance, and linear regression analysis in terms of predicting 100m time from race segments times. Individual (male=.844-.994, 88.9% common variance; female=.906-.993, 92.1% common variance) and pooled (0.985-0.999, 98.7% common variance) factor loadings indicated that sprinting ability is common construct across all race segments. The best fit nonlinear regression mathematical functions for the relationship of average velocity per distance segment were cubic functions for both males (R=.987, R^2=.975) and females (R=.989, R^2=.977), where the transition from positive to negative acceleration occurring at 58-59m for both genders. The regression analysis using pooled data indicated that 60m time was an excellent predictor of 100m time (R=0.994, R^2=0.982, p<0.001) indicating both males and females solve the problem of sprinting the 100m in almost an identical manner in terms of mathematical and statistical models.

Keywords: Curve estimation, nonlinear regression, factor analysis, mathematical modelling

1. INTRODUCTION

Talent identification is now common practice in Australian sport to identify future Olympians and Paralympians. Mathematical models and multivariate statistical approaches to predicting performance in sport are frequently recommended, however rarely implemented, to identify those factor that are associated with high performance athletes. Australia had difficulty finding talent, both males and female to complete in the 100m and 200m at the recent 2008 Beijing Olympics and at the 2009 World Athletics Championships in Berlin as the one female athlete who made the 100m is primarily a 100m hurdler, and no other athletes had sufficiently good qualifying times to be considered for selection.

The Australian National Talent ID program (National Talent ID & Development, 2009) for the athletic sprint events, the 100m and 200m, is based on a simple sports specific motor fitness test. This is the time for the 60m sprint from a standing start and the vertical jump height, a measure of the standing anaerobic power of the legs. Explosive power in the legs is thought to be related to performance in sprint events, such as the 100m and 200m sprints. Kinanthropometric data such as standing height and body mass are also recorded to assess if the potential athlete is within the usual ranges of heights and mass for sprinters. Anthropometric and sports specific motor fitness are related to both age and gender constructs.

Other research (Heazlewood, 1998) has indicated approximately 56% of 100m and 400m sprint performance can be predicted by power/weight ratio using isokinetic leg extension at 300^0s^{-1} using CYBEX 340 technology. However, such information does not indicate how the athletes actually sprint the entire race in terms of race distance, velocity segments, velocity curve characteristics or the factor structure of the race segments. Coaching and biomechanical recommendations indicate the 100m sprint can be divided into acceleration (0m - 50m), maximum running speed (50m - 80m) and speed endurance or negative acceleration (80m - 100m) stages based on examining the curves of individual athletes such as Usain Bolt (IAAF, 2009). However, the data is descriptive in nature and does not indicate which race segment is the best predictor of total race performance (100m time), which type of mathematical curve best fits the data or indicates the factor structure, if any, that exists between the acceleration, maximum running speed and speed endurance stages. Siegel (2009) has used past trends, world records, over time to predict future times.

This predominant research focus was on how the best 100m sprinters, both male and female, actually run the race and to derive models for high performance sprinting.

Models that will:

1. Predict the total race time based on the race segment that is the most predictive of total race time. In this context to test the model developed by the Australian National Talent ID program for Athletics to select 100m sprinters based on a 60m time trial.
2. Discover the mathematical curve and function that best fits the data and to use this function as a sprint model in talent identification as well as to evaluate changes in athlete performance based on training programs to fit the high performance sprint model.
3. Evaluate the factor structure that exists between the acceleration, maximum running speed and speed endurance stages based on evaluating the factor structure of sprint segments in the 100m.

2. METHODS

The research analysed data provided by the International Associations of Athletics Federations (IAAF, 2009) from the recent 2009 World Athletics Championships in Berlin. The data set consisted of 59 athletes where 23 male performances from the 100m final and semi-finals and 36 female performances from the 100m final and the fastest athletes in the 100m sprint rounds. The data analysed consisted of reaction time to the gun, race segment times for 0-20m, 20-40m, 40-60m, 60-80m and 80-100m, as well as the conversion of race segments times to average velocity per 20m segment. Descriptive statistics were derived to provide an overall understanding of the data set. Tests to assess normal distribution characteristics were also conducted to evaluate compliance with statistical assumptions using the following statistical methods. The race segment data was then applied to linear regression analysis to derive a regression equation based on which 20m race segment was the most accurate predictor of total race time. The intention of linear regression analysis is to; develop an equation that summarizes the relationship between a dependent variable and a set of independent variables or variable; identify the subset of independent variables or variable that are most useful for predicting the dependent variable; and finally to predict values for a dependent variable from the values of the independent variable(s) (Hair et al., 2006). Gender specific models were derived as well as models based on the pooled data.

The times were then converted average velocities for the 20m segments, which enabled the derivation of a velocity (y-axis) and distance (x-axis) curve. As the derived curves were definitely nonlinear in shape the statistical method of curve estimation regression was applied. The curve estimation procedure produces curve estimation regression statistics and related plots for 11 different curve estimation regression possibilities. These regression models are based on linear, logarithmic, inverse, quadratic, cubic, power, compound, S-curve, logistic, growth and exponential fits. In terms of evaluating model fit the following statistics are generated, these are, regression coefficients, multiple R, R^2, adjusted R^2, standard error of the estimate, analysis-of-variance table with appropriate levels of significance, predicted values, residuals and prediction intervals. Gender specific models were derived as well as models based on the pooled data.

Factor analysis was applied to assess if the 100m sprint is represented by unique factors or reflects underlying human motor fitness ability, such as sprint ability expressed across all race segments. A number of comparisons were conducted for male and female sprinters in terms of multivariate factor analysis to assess if the male and female sprinters displayed any gender specific models. A number of factor solutions were applied, such as principle component analysis, maximum likelihood and principal axis factoring to develop the most interpretable and pragmatic factor solution.

The goals of factor analysis (Norusis, 1985; Hair el at., 2006) are:

1. To identify underlying constructs or factors that explain the correlation's among a set of variables.
2. To test hypotheses about the structure of the variables.
3. To summarize a large number of variables with a smaller number of derived variables.
4. To determine the number of dimensions required to represent a set of variables.

All calculations were conducted with SPSS Software Version 17.0 (SPSS - Version 17.0, 2009).

3. RESULTS

The tests on normally distributed data (normal Q-Q plot, detrended normal Q-Q, normally distributed histograms, stem and leaf plots, Kolmogorov Smirnov test and Shapiro Wilk test) indicated in the majority of cases the variables satisfied normal distribution assumptions. The descriptive statics for each race segment are displayed in table 1.

The minimum values in the range for the race segments are for Usain Bolt and 9.58 seconds represents the current World Record for the men’s 100m. The regression analysis using pooled data indicated that 60m time was an excellent predictor of 100m time (R=0.994, R^2=0.982, p,0.001; standard error predicted value, mean=.014s, SD=.002s) and illustrated in figure 1.

Variable	Mean	SD	Min.	Max.	Range
Reaction time (s)	.15	.02	.12	.21	.09
Time 0 - 20m (s)	3.10	.13	2.89	3.32	.43
Time 20 - 40m (s)	5.05	.25	4.46	5.39	.93
Time 40 - 60m (s)	6.96	.37	6.31	7.42	1.11
Time 60 - 80m (s)	8.84	.51	7.92	9.45	1.53
Total time100m (s)	10.80	.67	9.58	11.57	1.99

Table 1: Descriptive statistics for pooled data for reaction time and each race segment time in seconds.

This indicates both males and females solve the problem of sprinting the 100m in almost an identical manner in terms of mathematical and statistical models. The predictive regression equation using unstandardised beta coefficients is:

$$100\text{m Time(s)} = 1.791(60\text{m Time(s)}) - 1.66\text{s} \qquad (1)$$

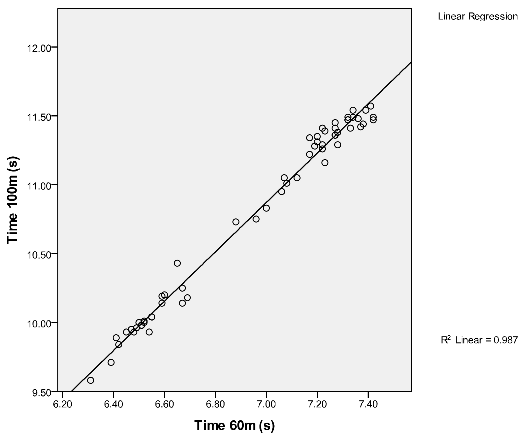


Figure 1. Line of best fit for 60m time predicting 100m time.

The best fit nonlinear regression mathematical functions for the relationship of average velocity per distance segment were cubic functions for both men (R=.987, R^2=.975) and women (R=.989, R^2=.977), where the transition from positive to negative acceleration occurring at 58-59m for both genders. The graphs for the individual velocity-distance curves for both men and women 100m are displayed in figure 2.

It is interesting to note that the two curves are almost identical, the main construct is the men sprinters run faster than the women sprinters, however the shape of the curves reflect underpinning capacities required for 100m sprinting at the highest level for men and women. Pooled data is represented by figure 3.

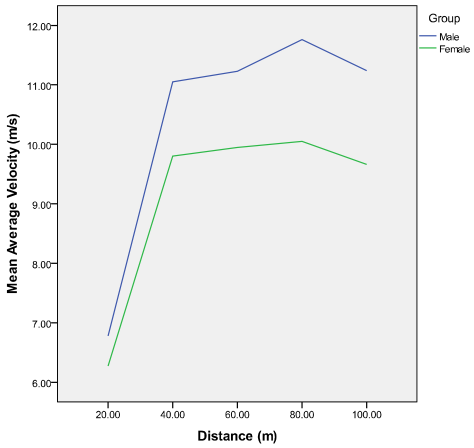


Figure 2. Graph of men's and women's velocity-distance graph (top line men's data).

The pooled data equation 2 for the cubic function is:

$$v = 4.139 + .144d - .001d^2 + .0000055d^3 \quad (2)$$

Where v = velocity and d = distance segment (cubic function R^2 = .99; p=.029).

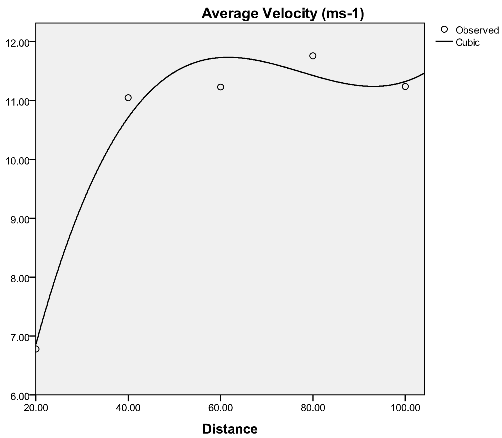


Figure 3. The cubic data curve fit with observed data points based on pooled data.

Individual cubic equations are provided for both men and women based on the velocity-distance 100m data and illustrated in equation 3 and equation 4.

Men

$$v = 4.208 + .159d - .002d^2 + .000006d^3 \quad (3)$$

The almost perfect nonlinear curve fits for both men and women are illustrated in figure 4 and figure 5.

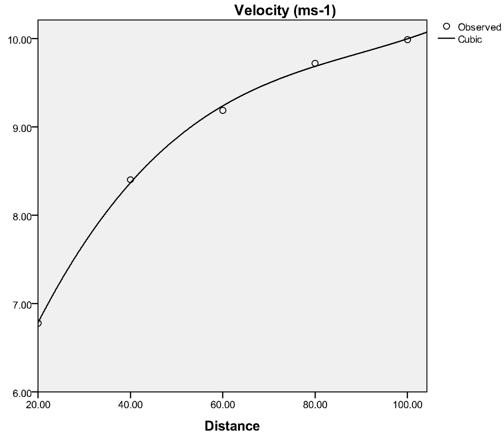


Figure 4. The cubic data curve fit with observed data points for men 100m.

Women

$$v = 4.007 + .141d - .001d^2 + .0000056d^3 \quad (4)$$

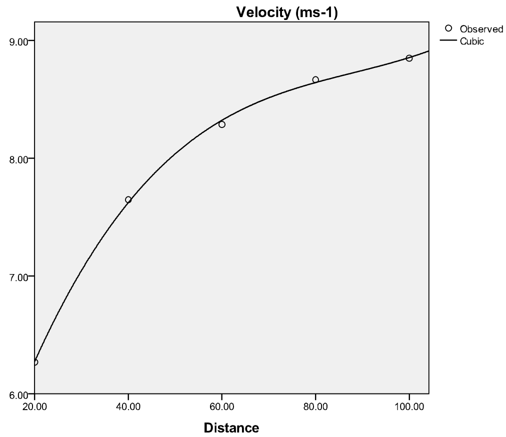


Figure 5. The cubic data curve fit with observed data points for women 100m.

Both curve fits are almost an identical match in terms of best nonlinear solution (cubic function), extremely high R^2 (> .975), coefficients for the two cubic equations (equations 3 and 4) and the curve fitting the data, which is observed cases plotted against line of best fit. The data indicate that the world's best men and women 100m sprinters solve the problem in an essentially identical manner.

Comparisons were conducted for male and female sprinters in terms of multivariate factor analysis of times per segment to assess if the race segments represented distinct race constructs or were representing the underlying factor of sprint ability as compared to individual race segments, such as reaction time, acceleration phase, maximum sprint running phase and speed endurance. The individual race segment model would suggest different factors.

Segments	Component 1 Factor Loading
t20m	.985
t40m	.995
t60m	.999
t80m	.997
t100m	.993

Table 2. Pooled Data Component Matrix with Factor Loadings for Race Segments for One Significant Factor.

Segments	Component 1 Factor Loading
t20m	.844
t40m	.959
t60m	.994
t80m	.978
t100m	.926

Table 3. Men's Data Component Matrix with Factor Loadings for Race Segments for One Significant Factor.

Segments	Component 1 Factor Loading
t20m	.906
t40m	.975
t60m	.993
t80m	.979
t100m	.943

Table 4. Women's Data Component Matrix with Factor Loadings for Race Segments for One Significant Factor.

Factor solutions using principal component analysis produced the most interpretable and parsimonious solution based on men's data (loadings = .844-.994, 88.9% common variance); women's data (loadings = .906-.994, 92.1% common variance) and pooled data (loadings = 0.985-0.999, 98.7% common variance) and factor loadings indicated that sprinting ability is a common construct across all race segments. Tables 2, 3 and 4 indicate the factor loadings for the men's, women's and pooled data based on the derivation of a single significant factor explaining the relationships in the correlation matrices.

A final factor analysis was conducted with reaction time included with 20m race segments to assess if reaction time is a unique constructs and different from actual sprint running. Pooled data were used to increase sample size and generalisability of the factor solution.

Variables	Component 1 Loading	Component 2 Loading
t20m	.933	.322
t40m	.964	.248
t60m	.969	.244
t80m	.970	.233
t100m	.966	.231
reaction (s)	.252	.968

Table 5. Pooled Data Component Matrix with a Two Factor Solution, Loadings for Race Segments on Factor 1 and Reaction Time Loaded Uniquely on Factor 2.

A two factor solution was achieved by principal component analysis with varimax rotation and Kaiser normalisation. This solution indicated that reaction time and the ability to sprint 20m race segments were very different constructs, as the factor loading for reaction time was very low on factor 1 (.252) and very high on factor 2 (.968). The reverse situation applied for the sprint segments, which are indicated clearly with the factor loadings in table 5 and highlighting the simple factor structure of these constructs.

4. DISCUSSION

The data indicate clearly that both men and women world's best sprinters solve the problem of running the 100m in essentially an identical manner. The linear regression solution in predicting final 100m time from race segment time indicated the 60m time could predict accurately final 100m time. It is interesting to note the Australian talent identification Athletics program (National Talent ID and Development, 2010) for sprinting uses the 60m time trial as a selection criterion combined with vertical jump, standing height and body mass (weight). Based on these data alone it would appear that the

60m time explains most of the performance variance (98.2%) in the 100m sprint for both men and women and figure 1 indicates that both men and women sprinters fit the linear regression line almost perfectly when using 60m time to predict 100m time. The Australian talent ID process for the selection of sprinters based on 60m time has some support based on these findings. However, the factors of vertical jump, standing height and body mass may contribute minimally to the predictive model as most of the performance variance has been explained by just using race segments.

The nonlinear regression indicated clearly the cubic function to describe the velocity distance graph, as the data and the model were an almost perfect fit for both men and women sprinters. The underlying reason may reflect that both men and women use the same energy systems to solve the problem when competing at maximal effort. It definitely indicates that the velocity-distance curves are relatively fixed in shape and which describes the positive acceleration phase, the maximum sprinting speed phase (approximately to 58-59m), and the significant negative acceleration phase from 80m to 100m. These data indicate how athletes distribute speed across the entire race and indicate that all sprinters solve the problem in the same manner, which hints of finite, but like capacities, when men and women use maximal sprinting in competition. The major issue is, the male sprinters just sprint faster at any stage in the race, even though the velocity-distance distributions are the same for men and women.

The findings probably indicate that both men and women should train using very similar training methods to develop the underpinning energy systems and biomechanics (force, power, work and speed) although force, torque, power, power/weight ratios are different and favour male sprinters (Heazlewood, 1998: Richmond, 2009). However, Richmond's research was based on one athlete when describing forces involved in sprinting as compared with the best 59 men and women sprinters in 2009. The model also indicates that individual sprinters can be modelled against the World's best sprinters to see how other sprinters of sub high performance ability (next level below World Championship representatives) fit the curve. This level of sprinter would want to fit the curve as closely as possible and training should be directed towards getting the velocity-distance characteristics to match the high performance curves, especially the positive acceleration, maximum sprinting speed and negative acceleration components. This is the approach Valerie Borzov's coach applied back in the late 1960's and early 1970's (Borzov, 2009), when he discovered Borzov's acceleration and maximum sprinting speed lagged behind the best male sprinters in the world at that time. Borzov went on the win the 100m and 200m in athletics at the 1972 Munich Olympic Games based on this training principle.

The factor analysis provided further confirmation the top men and women 100m sprinters solve the problem of sprinting the 100m using almost identical methods and identified a common underpinning factor that explained the majority of common variance (98% plus) for pooled data among the 20m race segments. Using the data for men, women and pooled data in separate factor analyses all derived a single factor solution, where each race segment loaded significantly on this one factor. This suggests that underpinning sprint ability is expressed across all race segments for this level of World Championship sprint athlete, irrespective of gender and does not indicate independence of constructs such as positive acceleration, maximum sprinting speed or negative acceleration (speed endurance).

The inclusion of reaction time with sprint segment times indicated that the reaction time construct is unique to sprinting and resulted in a two factor solution, one factor for 20m sprint segments and one factor for reaction time from the starting gun. The two factor solution was an orthogonal solution indicating no correlation between the reaction time to gun factor and the 20m sprint segment factor.

This implies the reaction ability construct appears to be based on another perceptual motor ability, which is independent of actual sprint running ability and must be trained independently from actual sprint running.

It would be of heuristic value to evaluate other ability sprinters, such as sub high performance men and women and sports where sprinting up to 100m is important to corroborate or refute the general mathematical models and factors structures that describe high performance sprinters. Most codes of football actually play on fields that are 100m from goal line to goal line.

5. CONCLUSIONS

Sprinting running ability at the highest level is an almost identical construct for both men and women sprinters as they solve the problem of sprinting the 100m in competition in an almost identical manner. Predictive equations based on linear and non linear regression methods and factor analytic models using exploratory and confirmatory factor analysis indicate identical problem solving approaches using velocity-distance graphs and times of 20m race segments. In terms of training both genders should train essentially using the identical training methods. The only real difference being women sprinters do not sprint quite as fast as men. As well, the information can be applied in other sports where the dimensions of playing fields in terms of length are 100m or approximate 100m, such as Rugby League, Rugby Union, AFL, Touch (70m) and Football.

References

Borzov, V. (2009). Training procedures in sprinting. http://speedendurance.com/2009/01/12/valeri-borzov-training-procedures-in-sprinting/

Hair, J. E., Block, W., Babin, B.J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate Data Analysis. (6th Ed.) Upper Saddle River: Pearson - Prentice Hall.

Heazlewood, I. (1998). The influence of power, work and torque acceleration in predicting long jump, 100m, 400m high jump performance. Conference proceedings Australian Track and Field Coaches Congress 1998. University of Western Sydney. 23 - 26th October.

IAAF (2009). Biomechanics research project: Men 100m Final – Usain Bolt. IAAF Berlin 15-23 August, 2009.

National Talent ID and Development. (2010). Talent ID Athletics – sprints. ausport.gov.au/eTID.

Norusis, M. J. (1985). SPSSX. Sydney: McGraw-Hill Book Company.

Richmond, J. (2009). Newtonian model of an elite sprinter: How much force do athletes need to produce each step to be World class? http://www.elitetrack.com/article_files/newtonian-sprinting.pdf

Siegel, E. (2009). The maths of the fastest man alive. http://scienceblogs.com/startswithabang/2009/08/the_math_of_the-fastest_human.php.

SPSS Inc. (2009). SPSS Statistics Base 17.0 User’s Guide. Chicago: SPSS Inc.

A COMPARISON OF CLASSIFICATION ACCURACY FOR KARATE ABILITY USING NEURAL NETWORKS AND DISCRIMINANT FUNCTION ANALYSIS BASED ON PHYSIOLOGICAL AND BIOMECHANICAL MEASURES OF KARATE ATHLETES

Ian Tim Heazlewood[a], Hovik Keshishian[b]

[a] *Charles Darwin University*
[b] *Australian Catholic University*
Corresponding author: *Ian.Heazlewood@cdu.edu.au*

Abstract

Neural networks can be applied to many predictive data mining applications due to their power, flexibility and relatively easy operations. Predictive neural networks are very useful for applications where the underlying process is complex, such as in classification using a mix of nominal and ratio level variables and for predictive validity based on classification modelling. A neural network can approximate a wide range of statistical models without requiring the researcher to hypothesize in advance certain relationships between the dependent and independent variables. Neural networks and discriminant function analysis (a more traditional statistical approach), based on physiological and biomechanical measures of karate ability and collected within the ACU exercise physiology laboratory, were compared for there classification accuracy. Twenty four karate athletes were assessed, 12 were classified as high performance athletes with black belt or higher and 12 were classified as non-high performance athletes, green belt and lower. Ability level served as the classification variable. The dependent variables were height, weight, age; motor fitness variables were Margaria power test, standing long jump, isometric grip strength, sit-reach flexibility, arm crank, peak aerobic power, anaerobic Wingate power test for peak power, time to peak power, mean power and power/weight; and Karate specific motor fitness tests were karate agility, power punch, speed punch, reaction time, balance and lower limb bilateral flexibility. ANOVA indicated the general motor fitness constructs of Margaria power test, sit-reach flexibility, arm crank and Wingate power test for peak power; and karate specific motor fitness tests for karate agility test, power punch, speed punch, balance and lower limb bilateral flexibility or lateral split were significantly different ($p<0.05$ to 0.001). These two data sets were used in the multilayer perceptron (MLP) neural networks and method enter discriminant function analysis. The neural network solution based on the training data set and testing (holdout) data set classified at 100% accuracy karate ability (high and non-high) for the karate specific tests, as well as general motor fitness tests. Discriminant analysis was marginally less effective in classifying ability level. The karate specific tests produced a 95.8% and general motor fitness tests 91.3% correct classifications, respectively. Neural networks, specifically the multilayer perceptron (MLP) networks, were more effective in predicting group membership and displayed higher predictive validity when compared to discriminant analysis.

Keywords: Neural networks, multilayer perceptron, multivariate classification, talent identification, classification accuracy

1. INTRODUCTION

Neural networks can be applied to many predictive data mining applications due to their power, flexibility and relatively easy operations. Predictive neural networks (Fausett, 1994; SPSS Inc., 2007a) are very useful for applications where the underlying process is complex, such as in classification using a mix of nominal and ratio level variables and for predictive validity based on classification modelling. A neural network can approximate a wide range of statistical models without requiring the researcher to hypothesize in advance certain relationships between the dependent and independent variables.

Neural networks are the preferred tool for many predictive data mining applications because of their power, flexibility, relevance and ease of use. Predictive neural networks are particularly useful in applications where the underlying process is complex, especially pattern recognition and classification problems that are based on predictive and concurrent validity.

Neural networks used in predictive applications, such as the multilayer perceptron (MLP) and radial basis function (RBF) networks, are supervised in the sense that the model-predicted results can be compared against known values of the target variables. These target variables are identified on *a priori* criteria by the researcher.

The term neural network applies to a loosely related family of models, characterized by a large parameter space and flexible structure, descending from studies of brain functioning. As the family grew, most of the new models were designed for non biological applications, though much of the associated terminology reflects its origin in biology.

A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use and is analogous to human brain function. Specifically, it resembles the brain in two respects:

- Knowledge is acquired by the network through a learning process.
- 'Interneuron connection' strengths known as synaptic weights, analogous to human synapses, are used to store the knowledge.

A neural network can approximate a wide range of statistical models without requiring that you hypothesize in advance certain relationships between the dependent and independent variables, a non *a priori* model. Instead the form of the relationships is determined during the learning process. A type of neural processing phenomenology in this context.

The trade-off for this flexibility is that the synaptic weights of a neural network are not easily interpretable. Thus, if you are trying to explain an underlying process that produces the relationships between the dependent and independent variables, it would be better to use a more traditional statistical model, such as discriminant analysis or logistic regression. However, if model interpretability is not important, you can often obtain good model results more quickly using a neural network.

Although neural networks impose minimal demands on model structure and assumptions, unlike inferential statistics, it is useful to understand the general neural architecture or neural network structure. The multilayer perceptron (MLP) and radial basis function (RBF) networks are functions of predictors (also called inputs or independent variables) that minimize the prediction error of target variables (also called outputs).

Discriminant analysis (or discriminant function analysis) based on classification modelling is applied to classify cases into the values of a categorical dependent variable, usually a dichotomy (SPSS Inc., 2007b). In sport this could be males compared to females on different motor fitness tests or different player grades using the same principles. If discriminant function analysis is effective for a set of data, the classification table of correct and incorrect estimates will yield a high percentage correctly classified cases and maybe useful in such processes as sport talent identification, such as in the Olympic sport of karate.

The major foci of discriminant analysis (Hair et al., 2006: Norusis, 1985: StatSoft Inc., 2010) are to:

- Classify cases into groups using a discriminant prediction equation.
- Test theory by observing whether cases are classified correctly as predicted.
- Investigate differences between or among groups.
- Determine the most parsimonious way to distinguish among groups.
- Determine the percent of variance in the dependent variable explained by the independents.
- Determine the percent of variance in the dependent variable explained by the independents over and above the variance accounted for by control variables, using sequential discriminant analysis.

- Assess the relative importance of the independent variables in classifying the dependent variable.
- Discard variables which are little related to group distinctions.

The aim of this research was to apply both neural networks and discriminant function analysis (a more traditional statistical approach under the general linear model) and compare their ability as statistical techniques to classify different ability karate groups (high versus non high performance) correctly, based on karate specific and general motor fitness test using physiological and biomechanical measures.

2. METHODS

Twenty four karate athletes, 12 classified as high performance athletes with black belt or higher and 12 classified as non-high performance athletes, below green belt, participated in the ethics approved study. The data were collected in an environmentally controlled exercise physiology laboratory at ACU.

Ability level served as the dichotomous classification variable. The dependent variables in both the neural network and discriminant analyses represented anthropometric factors, which were height (cm), weight (kg), age (years); motor fitness variables of Margaria power test (W), standing long jump (cm), isometric grip strength (kg.wt), sit-reach flexibility (cm), arm crank (W), peak aerobic power (W), anaerobic Wingate power test for peak power (W), time to peak power (s), mean power (W) and power/weight (W/kg). As well, karate specific motor fitness tests, as selected by a panel of experts with backgrounds in karate and motor fitness assessment, were included as dependent variables in the analysis. These were karate agility (s), 'power punch' (kg.wt and actually measure of force), speed punch (s), reaction time (s), balance (s) and lower limb bilateral flexibility or lateral split (cm in seated lateral split). The 'power punch' was actually measured in kg.wt force, but was defined as a power punch by the 6^{th} dan black belt karate expert.

Statistical Analysis

An initial two group ANOVA was applied to assess which general motor fitness and karate specific variables were most discriminatory at the univariate level. Subsets of these significant variables were then included in the discriminant analyses using general motor fitness and karate specific variables separately and finally merging the two dependent variable sets, in an attempt to achieve improved classification accuracy.

Means and standard deviations of the significantly different general motor fitness and karate specific motor fitness tests were derived based on the ANOVA findings.

Neural network analysis applied the multilayer perceptron (MLP) procedure, which produces a predictive model for one or more dependent (target) variables based on the values of the predictor variables. The nominal or classification variable in the analysis was once again karate ability level and in neural network jargon these are defined as the dependent or target variables. The covariates or predictor variables (neural network jargon) were the significantly different general motor fitness and karate specific tests as identified by ANOVA analysis. It must be emphasised that the comparison of discriminant analysis with neural network analysis were based on the identical subsets of data in both analyses.

In the analysis rescaling is applied where scale-dependent variables and covariates are rescaled by default to improve network training. All rescaling is performed based on the training data, even if a testing or holdout sample is defined (SPSS Inc., 2007a). The mean, standard deviation, minimum value, or maximum value of a covariate or dependent variable is computed using only the training data. The neural network multilayer perceptron architecture was based on:

- Selecting one hidden layer where the hidden layer contains unobservable network nodes (units). Each hidden unit is a function of the weighted sum of the inputs. The function is the activation function, and the values of the weights are determined by the estimation algorithm.
- The selected activation function was the hyperbolic tangent, where the activation function links the weighted sums of units in a layer to the values of units in the succeeding layer.
- Hyperbolic tangent function has the form
  $$\gamma(c) = \tanh(c) = (e^{c}-e^{-c}) / (e^{c}+e^{-c}). \qquad (1)$$
- It takes real-valued arguments and transforms them to the range (–1, 1). When automatic architecture selection is used in SPSS, this is the activation function for all units in the hidden layers.

The identity function was selected and this function has the form: $\gamma(c) = c$. It takes real-valued arguments and returns them unchanged. When automatic architecture selection is used, this is the selected activation function for units in the output layer if there are any scale-dependent variables. Training the network was based on the batch method. This method updates the synaptic weights only after passing all training data records, which means batch training uses information from all records in the training dataset. Batch training is often preferred because it directly minimizes the total error and is most useful for smaller datasets, such as the one used in this research.

Discriminant analysis was based on using the significantly different general motor fitness and the karate specific motor fitness tests separately in the analyses by applying the method enter (all variables included in the discriminant model and determined by the researcher) as compared to stepwise method, which is based on statistical criteria to enter the model at each calculation step. Ability level was used as the independent dichotomous variable in the analysis. The data from both general motor fitness and karate specific tests that were significant at the ANOVA level of analysis were pooled and then subjected to discriminant analysis. Relevant fit statistics, canonical discriminant functions, hierarchy of importance in terms of variables and classification tables were generated.

3. RESULTS

ANOVA indicated the general motor fitness constructs of Margaria power test, sit-reach flexibility, arm crank and Wingate power test for peak power were significantly different between the high and non high ability groups($p<0.05$ to 0.001). Karate specific motor fitness tests that discriminated at the univariate level were karate agility test, power punch, speed punch, balance and lower limb bilateral flexibility and were significantly different ($p<0.05$ to 0.001). These two data sets were used in the multilayer perceptron (MLP) neural networks and method enter discriminant function analysis.

The neural network solution based on the training data set and testing (holdout) data set classified at 100% accuracy karate ability (high and non-high) for the karate specific tests for the training and testing samples (refer to table 1).

Classification

Sample	Observed	Predicted		
		High	Non High	Percent Correct
Training	High	10	0	100.0%
	Non High	0	8	100.0%
	Overall Percent	55.6%	44.4%	100.0%
Testing	High	2	0	100.0%
	Non High	0	4	100.0%
	Overall Percent	33.3%	66.7%	100.0%

Dependent Variable: Participants ability level.

Table 1: The Neural Network Solution Based on the Training Data Set and Testing (Holdout) Data Set Classified at 100% Accuracy for Karate Ability (High and Non High) for the Karate Specific Tests.

Classification

Sample	Observed	Predicted		
		High	Non High	Percent Correct
Training	High	10	1	90.9%
	Non High	1	6	85.7%
	Overall Percent	61.1%	38.9%	88.9%
Testing	High	1	0	100.0%
	Non High	1	3	75.0%
	Overall Percent	40.0%	60.0%	80.0%

Dependent Variable: Participants ability level.

Table 2: The Neural Network Solution Based on the Training Data Set and Testing (Holdout) Data Set Classified at 88.9% and 80.0% Accuracy Respectively for Karate Ability (High and Non High) for the General Motor Fitness Tests.

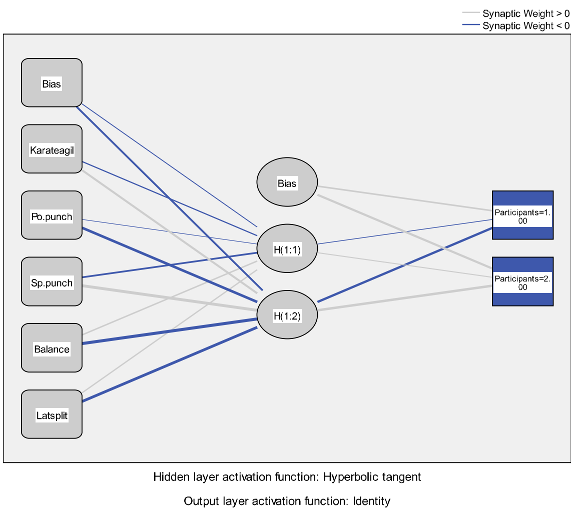


Figure 1: Diagrammatic representation of neural network architecture for karate specific tests with one hidden layer (hyperbolic tangent).

The classification accuracy for the general motor fitness tests were 88.9% for the training and 80.0% for the testing samples respectively (refer to table 2). Diagrammatic representation of neural network architecture for karate specific tests with one hidden layer using a hyperbolic function and the output layer an identity function are represented in figure 1.

Parameter Estimates

Predictor		Predicted: Hidden Layer 1 H(1:1)	Predicted: Hidden Layer 1 H(1:2)	Output Layer: Participants = High	Output Layer: Participants = Non High
Input Layer	(Bias)	.723	-.129		
	Karateagil	-.884	.417		
	Po.punch	1.444	-.549		
	Sp.punch	-2.338	1.123		
	Balance	2.069	-.784		
	Latsplit	1.137	-.653		
Hidden Layer 1	(Bias)			.487	.518
	H(1:1)			.982	-.975
	H(1:2)			.502	-.497

Table 3: Parameter Estimates for Hidden Layer 1 and Output Layer for High and Non High Karate Athletes for the Predicted Outcome (Classification).

Synaptic weights display the coefficient estimates that indicate the relationship between the units in a given layer to the units in the following layer (refer to table 3). The synaptic weights are based on the training sample, even if the active dataset is partitioned into training, testing and holdout data. The number of synaptic weights can become large with large numbers of variables in the analysis and as a consequence these weights are generally not used for interpreting network results.

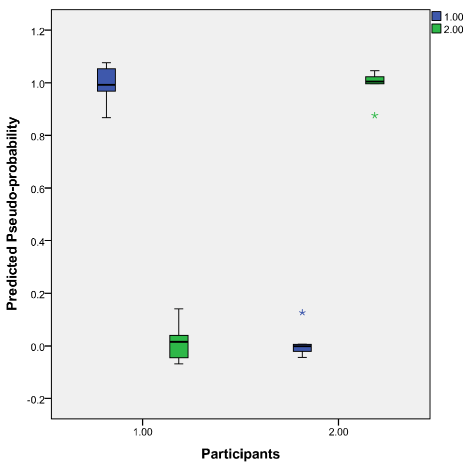


Figure 2: Predicted by observed chart indicating 100% probability of correct classification for ability level.

The predicted by observed chart (figure 2) supports the accuracy of the model. The top left box plot is for the high performers (p=1) and the top right box plot for the non high performers (p=1) and the reverse probabilities apply (p=0).

The cumulative gains chart shows the percentage of the overall number of cases in a given category "gained" by targeting a percentage of the total number of cases (refer to figure 3).

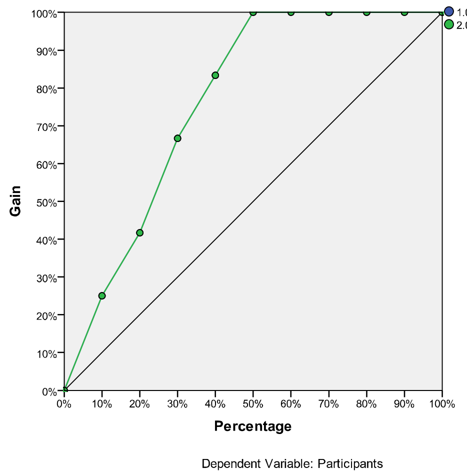


Figure 3: The cumulative gains chart shows increase in percentage against gain, indicting 50% results in 100% gain.

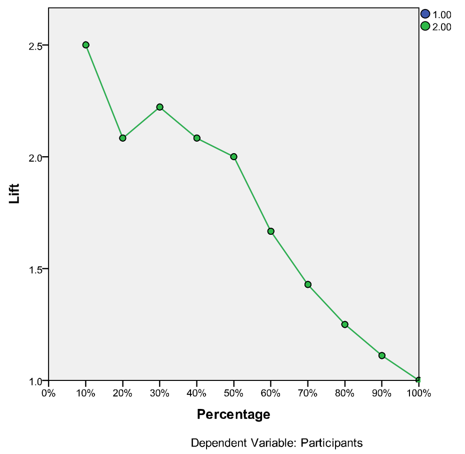


Figure 4: The lift chart displayed is derived from the cumulative gains chart and the values on the *y*- axis correspond to the ratio of the cumulative gain for each curve to the baseline.

For example in figure 3, the point on the curve for the non high performer category is at 50% on the x-axis and indicates gain of 100%, meaning that if you score a dataset with the network and sort all of the cases by predicted pseudo-probability of non high performer, you would expect the top 50% to contain approximately 100% of all of the cases that actually take the category non high performers. Once again, indicating the accuracy of the model.

The lift chart displayed in figure 4 is derived from the cumulative gains chart and the values on the *y*-axis correspond to the ratio of the cumulative gain for each curve to the baseline. Thus, the lift at 50% for the category non high performers is 100% gain and gives the ratio of 100%/50% = 2.0. It provides another way of looking at the information in the cumulative gains chart and minimal lift occurs beyond 50% (refer to constant negative slope from 50% to 100% on lift graph). The cumulative gains and lift charts are derived from the combined training and testing samples. Table 4 indicates the importance and normalized importance of the variables in the neural network analysis.

Independent Variable Importance

	Importance	Normalized Importance
Karateagil	.122	39.6%
Po.punch	.148	47.8%
Sp.punch	.267	86.4%
Balance	.309	100.0%
Latsplit	.154	49.8%

Table 4: Displays Independent Variable Importance and Normalized Values.

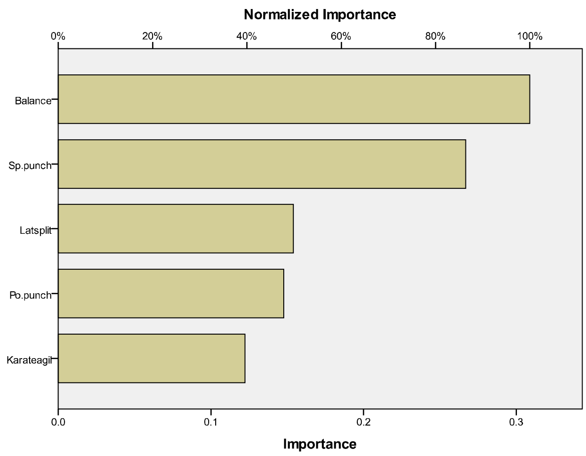


Figure 5: Normalized importance as a histogram.

Table 4 and figure 5 indicate the most important discriminating variables in the neural network analysis are balance and karate speed punch.

Similar analysis using the general motor fitness data, which classified 88.9% correctly from both ability groups indicated arm cranking (100% normalized) and Margaria power test (85.6% normalized and a leg power test) were the most important discriminators for the general motor fitness tests.

Discriminant analysis was marginally less effective in classifying ability level when using the karate specific tests, however slightly more accurate when using the general motor fitness tests. The general motor fitness tests produced 91.3% (Wilks' Lambda = .425, p=.003) and karate specific tests produced 95.8% (Wilks' Lambda = .362, p<.001) correct classifications, respectively. The means and standard deviations for sit and reach flexibility, arm crank, Wingate peak power and Margaria power test are displayed in table 5 and high performance athletes display higher scores than the non high performance athletes on all tests.

	Variables	Mean	Std. Deviation
High (n=12)	Sit-Reach (cm)	38.16	8.93
	Arm Crank (W)	230	9.34
	Wing. Peak (W)	1092	119.54
	Margaria (W)	919.7	134.50
Non High (n=11)	Sit-Reach (cm)	29.18	11.23
	Arm Crank (W)	206.64	18.30
	Wing. Peak (W)	976.18	125.59
	Margaria (W)	748.9	167.14

Table 5: Means and Standard Deviations for High and Novice Karate Athletes Based on Significantly Different General Motor Fitness Variables (p<0.05 to 0.001).

Variables	Function
	1
Sit-Reach (cm)	.041
Arm Crank (W)	.047
Wing. Peak (W)	.000
Margaria (W)	.004
(Constant)	-15.525

Table 6: Canonical Discriminant Function Coefficients with Unstandardized Coefficients to Derive Equation.

Table 6 displays the unstandardised canonical discriminant function coefficients and table 7 the classification function coefficients for the high and novice groups enabling the derivation of the discriminant function scores for both groups.

Variables	Participants	
	High	Non High
Sit-Reach (cm)	.206	.115
Arm Crank (W)	1.056	.951
Wing. Peak (W)	.022	.021
Margaria (W)	.045	.035
(Constant)	-158.826	-124.452

Table 7: Classification Function Coefficients for Each Ability Group with Fisher's Linear Discriminant Functions.

Table 8 indicates the accuracy of classification based on general motor fitness constructs. In this model 91.3 % of both ability groups were classified correctly. In this context the classification was marginally better (2.4%) than the neural network solution.

		Participants	Predicted Group Membership		Total
			High	Novice	
Original	Count	High	12	0	12
		Novice	2	9	11
	%	High	100.0	.0	100.0
		Novice	18.2	81.8	100.0

Table 8: Classification Results where 91.3% of Original Grouped Cases Correctly Classified.

The means and standard deviations for the karate agility test, power punch, speed punch, balance and lower limb bilateral flexibility or lateral split are displayed in table 9. One again, the scores are higher for the high performance athletes when compared to non high performance athletes. The timed scores of karate agility and speed punch tests (referred to as speed in the tables 9, 10 and 11) reflect that a lower score is equated with higher ability on these tests or faster speed means lower score.

Table 10 displays the unstandardised canonical discriminant function coefficients and table 11 the classification function coefficients for the high and non high groups, which enable the derivation of the discriminant function scores for each ability group.

The values in table 11 represent unstandardized classification coefficients.

Variables		Mean	Std. Deviation
High	Agility (s)	7.52	2.37
	Power (kg)	71.16	7.27
	Speed (s)	.33	.11
	Balance (s)	3.46	1.32
	Latsplit (cm)	40.50	7.69
Non High	Agility (s)	9.85	2.54
	Power (kg)	63.50	9.23
	Speed (s)	.53	.18
	Balance (s)	1.65	.68
	Latsplit (cm)	27.91	9.18

Table 9: Means and Standard Deviations for High and Novice Karate Athletes Based on Significantly Different Karate Specific Fitness Variables (p<0.05 to 0.001).

Variable	Function
	1
Agility (s)	-.122
Power (kg)	.011
Speed (s)	-3.375
Balance (s)	.439
Latsplit (cm)	.055
(Constant)	-1.252

Table 10: Canonical Discriminant Function Coefficients with Unstandardized Coefficients to Derive Equation.

Variable	Participants	
	High	Non High
Agility (s)	1.306	1.616
Power (kg)	1.093	1.064
Speed (s)	25.986	34.559
Balance (s)	.827	-.289
Latsplit (cm)	.381	.242
(Constant)	-57.938	-54.758

Table 11: Classification Function Coefficients for Each Ability Group with Fisher's Linear Discriminant Functions.

Table 12 indicates the accuracy of classification based on karate specific motor fitness constructs. The classification accuracy was 95.8% (Wilks' Lambda = .362, p<.001) correct classifications.

		Participants	Predicted Group Membership		Total
			High	Non High	
Original	Count	High	12	0	12
		Non	1	11	12
	%	High	100.0	.0	100.0
		Non	8.3	91.7	100.0

a. 95.8% of original grouped cases correctly classified.

Table 12: Classification Results where 95.8% of Original Grouped Cases Correctly Classified.

The discriminant analysis performed marginally lower (4.2%) than the neural network solution using the same subset of variables, however it should be noted that the differences between the two approaches were not that different.

4. DISCUSSION

Neural networks, specifically the multilayer perceptron (MLP) networks, were more effective in predicting group membership and displayed higher predictive validity when compared to discriminant analysis for the karate specific tests consisting of karate agility test, power punch, speed punch, balance and lower limb bilateral flexibility or lateral split. However, the discriminant function analysis using the general motor fitness tests of Margaria power test, sit-reach flexibility, arm crank and Wingate power test for peak power marginally out performed the neural network approach. This presents some what contradictory findings. If the outcomes of this research are utilised in talent identification, using karate specific tests would provide marginally better classification outcomes. This argument was supported by the karate expert used in this research to design the karate specific tests, who believed these types of specific tests should provide a greater differentiation or discrimination between karate athletes of different abilities as an outcome of karate specific training or as a natural propensity for the sport.

The surprising finding was the general motor fitness tests used in the research almost performed as well as the karate specific tests. In this context the general motor fitness tests, especially arm crank power (arm power test) and Wingate power test for peak power (leg power test) may in fact be testing some underlying power constructs that were also measured to some degree by the karate specific tests. In terms of training implications the findings may also provide direction as to what sports specific constructs can be trained, such as karate agility, power punch, speed punch, dynamic balance and lower limb bilateral flexibility or lateral split, which are thought to be sport specific. Many of the anthropometric, general and some of the sports specific tests did not provide discriminant or predictive validity, such as anthropometrics of height, weight, age; general motor fitness of standing long jump, isometric grip strength, peak aerobic power, anaerobic Wingate power for time to peak power and power/weight; and Karate specific motor fitness test of reaction time. These non discriminating tests based on this research should probably not be included within tests attempting to identify karate ability in adult males. It is important to emphasise that both groups in the study were mature males with no significant differences for height, weight and age and so developmental maturation factors can be discounted.

5. CONCLUSION

Karate specific tests using classification and predictive validity methods, such as multilayer perceptron neural networks and discriminant function analysis, both provided accurate classification of high and non high performance karate ability groups. However, the multilayer perceptron neural network method performed marginally better than discriminant function analysis and provides a model for talent ID in karate.

Acknowledgements

We wish to thank the karate participants in the study.

References

Fausett, L. (1994). *Fundamentals of neural networks: Architectures, algorithms and applications.* Upper Saddle River NJ: Prentice Hall.

Norusis, M. (1985). *Advanced statistics guide: SPSSX.* Chicago, IL: SPSS Inc.

SPSS Inc. (2007a). *SPSS statistics base user's guide 17.0.* Users Guide. Chicago, IL: SPSS Inc.

SPSS Inc. (2007b). *SPSS Neural Networks^{TM} 17.0.* Chicago, IL: SPSS Inc.

StatSoft, Inc. (2010). *Electronic statistics textbook.* Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/.

AN INVESTIGATION INTO THE POSSIBILITY OF THEORETICAL MODELLING FOR THE PURPOSES OF EXAMINING THE NON-ASSOCIATIVE NATURE OF PROBABILITIES OF GAME BASED SKILL TASKS WITHIN COMPETITIVE SPORTS MATCHES

Joe Walsh[a,b]

[a] *Australian Catholic University*
[b] *COPAAL: The Centre for Physical Activity Across the Lifespan*
Corresponding author: joewalsh@acu.edu.au

Abstract

Mathematical modelling has been used in many sports to predict game winners based on past performances. This has been done for individual as well as team sports. There is also the possibility of determining probabilistic expectations for results of matches. As an alternative to examining this construct from game result data, this paper examines prospects for such predictions based on execution of game based skill tasks. This methodology is directed towards selecting an example sport and investigating solving the problem of predicting the outcome of individual elements of the competitive match. The possibility of doing this via the probabilities of successful play execution, by an attacking side, is examined. The ability of the defending team to execute the appropriate strategy for defence is also examined. This rudimentary theoretical model is then used to investigate the associative nature of winning probabilities involving matches between different teams in the example sport. These probabilities are shown to be non-associative, therefore in some example cases the use of predictions based on game dependant tasks will be shown to be superior for individual match prediction than those probabilities derived from results of past performances. There are also implications for the structure of tournaments from the results of this paper.

Keywords: Mathematical model, match analysis, performance prediction, touch football

1. INTRODUCTION

This is a hypothetical exploration of the possibility of investigating attacking and defending options for different teams within a sport and evaluating the best options for these teams. Certainly attempting to divulge the probable outcome of a sports match or ranking teams by looking at offensive and defensive abilities of a team has been considered as a viable avenue to pursue (recently, for example Govan, Langville, & Meyer, 2009). Similarly the potential for prediction via examination of a team's ability to execute certain relevant skills has also been preliminarily investigated (for example Zetou, Moustakidis, Tsigilis, & Komninakidou, 2007, in volleyball).

In touch football most defensive structures will be forced to give an attacking option to the offensive side due to there being less onside defensive players than available attackers. Therefore the option for defensive structure most usually selected is typically the one that is hardest for the attackers to execute and defenders must take note of the weaknesses of the attacking side in designing their strategies. Various types of defence present different opportunities for an attacking side to score if they have the personnel to successfully select and execute the available option. For example a "man on man" defence does not adjust for the fact that one of the defenders is offside as this defender effected the touch on the person performing the rollball (the act of bringing the ball into play, following a touch or change of possession). (For further clarification on this and other touch football terminology, the reader is referred to two compatible publications, namely playing rules from Touch Football Australia, 2007 and Federation International Touch Inc., 2003.) This presents opportunities for the attacking player assigned to this defender to score. In this situation, given that these two players are of similar agility and no mistakes are made by the attacking team, a touchdown should be scored. Several variations of defence are therefore initiated. These options include some variation, either instantaneous or delayed, of compressing the defence around the ruck (the area between the attacking player performing the rollball and the half (the player who is to take possession of the ball, behind the player who performs the rollball)) to force passing plays. Some variations of this involve leaving certain players unmarked. An example of this would be leaving a player unmarked, such that an attacking player not known for his/her passing game would have the option to create a scoring play if they could execute a long (e.g. 25-30m) pass to an eligible wing (those two players in the team positioned closest to either sideline) receiver using their non-dominant hand. For this particular player, this would be an unlikely option to be executed successfully and hence a wise selection of defensive structure. Similar to this example there are numerous other defensive options that attempt to capitalise on the strengths and weaknesses of various attacking teams, both focused on individual players within a team and according to how the players coordinate their strengths and weaknesses within a team structure. At a district, state or national representational level, these strengths and weaknesses of opponents will usually be well known by the coaching staff of the defending team.

A few additional terms that will be mentioned include line attack, which will refer to a state of play in which the attacking team has gained sufficient territory that they are close enough to the opponents line in order to launch a successful attack on the following play. Second phase play is used to refer to an attacking play conducted immediately after an unsuccessful attack with the purpose of scoring due to defensive disarray caused by the previous attack. The author has also submitted another paper (Walsh, 2010) to this conference and more extensive definition of these terms is available in this paper on Markov States for touch football. This paper is an extension of the Markov states paper, focusing on one particular set of states.

Another consideration that will be relevant to this paper is the importance placed on achieving a second or third placed position within a contest. At least since Galton 1902, much scientific literature has been published on tournaments. Following from this paper, it has been clearly established that there is an expectation of recognition for other teams that achieve victory, not just the team that comes first. It is clear from the literature on the design of tournaments that a major goal for the organisers is to maximise total effort of all participants by designing a contest that will engineer appropriate results as well as the potential to achieve sufficient recognition according to the achievement obtained. This recognition and reward are the performance incentives necessary for maximising participant effort. A relevant example of this would be Ehrenberg & Bognanno, 1990. In this paper on PGA golfers, it was found that the athletes responded more optimally when the level and structure of tournament prizes was enhanced. Similarly Becker & Huselid, 1992 found that driver performance in racing increased under similar conditions. Incidentally in this latter case however driver safety also decreased. It is also evident that when this recognition was accompanied by participant uncertainty as to whether it would be correctly assigned, the quality of an athlete's performance would be reduced. Certainly there is an equivalency in the work rate shown in many fields, such as for example workplace labour when rewards are attributed to effort and the enhanced performance of athletes at tournaments. Many of the purely theoretical models of both economic labour and work input and athletic tournaments are therefore interchangeable (Green and Stokey 1983). This would best be seen by examining some of the many well established foundation papers within this area, such as Lazear and Rosen (1981), Nalebuff and Stiglitz (1983) and O'Keeffe, Viscusi, and Zeckhauser (1984).

In addition to increasing performance, it is also clear that sometimes incentives can cause teams to deliberately lose games. Studies such as Taylor & Trogdon, 2002 indicate that when incentives are present to lose, teams are more likely to do so. In this particular paper, evidence implied that NBA team performance responded to changes in the underlying tournament structure to adapt to the rule changes allowing preferential choice of draft picks for losing teams.

2. METHODS

Let us start by considering the simplest model for a sport such as touch football

Consider two hypothetical teams, Team A and Team B. Team A has 4 set attacking plays designated as A_A, B_A, C_A, and D_A respectively (with the subscript denoting the play) in this example we are using A or D for denoting an attacking or a defending play. A_A, B_A, C_A, and D_A are the probability that the play will be executed as intended by the attacking team. Team B has a set defensive policy with probabilities A_D,

B_D, C_D, and D_D of successfully defending moves A_A, B_A, C_A, and D_A respectively. We can therefore write this as:

Team A (A_A, B_A, C_A, D_A)
Team B (A_D, B_D, C_D, D_D)

For example:

Team A (0.8, 0.6, 0.3, 0.2)
Team B (0.95, 0.85, 0.1, 0.2)

So here Team B adopts a defensive structure, which gives Team A the option of scoring via move A_C or A_D, but defending moves A_A and A_B more effectively (e.g. A_A and A_B are quick release plays focusing attack on the central area of the field, whilst A_C and A_D focus attack towards the sidelines with for example a long pass).

In order to model this situation, assume that if Team B executes defence correctly, with the probabilities show above, Team A will not score from this particular play. In order to score it is required for the defence to be unsuccessful at blocking all options for this particular attack and given that this attack would be presented with an opportunity to score, that Team A successfully executes the appropriate option for this particular play. To re-iterate, in order to score a touchdown, Team A must execute an attack correctly, given that B does not defend it successfully.

We must therefore consider the probability that Team A executes their attack correctly and Team B fails to defend this. Therefore firstly consider whether A and B can be considered as independent events. They may well not be so. Certainly it is well known in sport that a defence that exerts a great deal of pressure on an attacking team may change (most likely reduce) successful execution of a move. This could be incorporated into the probabilities of the defence successfully stopping the attack, as a probability of either directly stopping the move or applying pressure which results in its failed execution. However some teams may respond to this pressure better than others so if probabilistic descriptors are to be designed for incorporation into team statistics then this is best done as a component of the attacking team statistics. There is the possibility for development of a factor that represents pressure. If the defensive team can pressurise the attacking team it could be incorporated into model design such that it will activate a pressure function built into the attacker's probability of success. This could vary across teams, for some teams reducing the chance of successful execution of a move, while for other teams it would have little or no effect. This however is very hard to quantify. Whilst the average defending and attacking probabilities, or even the time taken to execute stages within an attack or defence can be easily measured using game video data, the pressure a team places on another team and how it affects the probability of executing an attack is difficult to measure purely from game data. It could be considered that the distance of defending players from attackers and their approach velocities/accelerations could be extrapolated in order to develop a play pressure function. This could then be incorporated as an additional factor affecting execution of a particular attacking move. This would be different for different teams as some teams would absorb this pressure poorly and the probability of a successful score would decrease. Additionally others would absorb it well, or at a higher level use the momentum of the defenders to contribute to the successful execution of the first and/or second phase execution of their scoring strategy. This could be examined via analysis of the success ratio of set moves and if this success was dependant on the distance of the defending players from the attackers. Hypothetically this would be an interesting option to examine. However, consider this case for teams with a good level of experience where there is enough game data to be of use to the coaching team. In this case, if there was enough data to extrapolate a displacement function, the condition would be most likely purely academic. This is because most teams at this level would place pressure on the opposition on all plays and all plays would be executed under defensive pressure. It would be most likely that without any pressure from the defensive structure, it would become very easy to execute most moves and this fact would be well known to the higher level performer. For the lower to intermediate level teams with poor technical skills and limited technical coaching the pressure would fluctuate with each defensive action. Excluding defensive mistakes (which would be represented by the defenders failing to execute the correct defence response as per the previously mentioned probability functions), at a higher level all defences would place pressure on the opposition for the key attacking players and plays would therefore be executed under pressure. The degree of pressure would be therefore factored into successful defensive probabilities as all attacking and defending teams at a high level would be used to applying pressure and operating under pressure, so this could be fairly assumed at all plays. There are some plays which (such as players that are to throw a long pass standing deeper and thus further from advancing defenders or giving the ball to an intermediate, unpressured player that is situated in a preferential isolated position to allow undisturbed execution of the pass) are designed to reduce pressure on the key players at the key points in the play. This bonus would however already be represented in probabilities for correct execution of the particular play as it would be a higher value anyway, as it would already be accounting for the fact that this helped for a given team's execution of this play. Importantly some moves are best conducted when the defenders are very close to the attackers in order to reduce the opportunities for the defence to accurately read the play. Therefore use of a displacement from opposition based pressure function would not be appropriate for all plays and at this stage in model development it is best to make an approximation. This approximation would be that at high level of performance, attacking players are used to being pressurised by defence. The defence pressurises attacking teams according to their defensive probability functions and therefore independency within probability functions can be assumed to exist between attack and defence probability functions, but overall probability of scoring is dependant on both of these functions, therefore implying that this latter probability has no degree of independence.

If Team B's defence and Team A's attacking probabilities, as expressed above, are independent then:

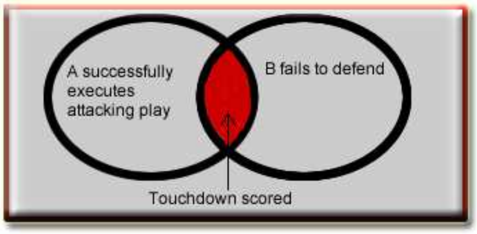


Figure 1: Defensive and Offensive probabilities for an attack by Team A against Team B

$$A \cap B = P(A)(1-P(B))$$

Let us now use Teams A and B from our previous example. The chances of scoring on any set play for set plays A-N are given as follows:

The chance of Team A scoring on any set play (A-N) against opposition B are given below:

$$(P(A_A)(1-P(B_A), P(A_B)(1-P(B_B)), P(A_C)(1-P(B_C)),\ldots\ldots, P(A_N)(1-P(B_N))$$

Giving for our example two teams:

(0.8(0.05), 0.6(0.15), 0.3(0.9), 0.2(0.8))

Which equates to:

(0.04, 0.09, 0.27, 0.16)

This implies that the coaching staff are best to employ moves C and D against this particular defence. These are also the attacking options that the defensive team are making available to the offensive side, this is because these moves are those which the offensive team is least capable of executing successfully, however due to the defensive structure presented, they become the most beneficial attacking options for scoring. Extrapolating this concept, if sufficient data on defending or attacking teams is not available then a probability could be estimated from successful execution of a particular class of defensive or attacking strategy.

Consider now offensive and defensive line attack profiles for Teams A and B

Team A
Attack profile (0.8, 0.6, 0.3, 0.2)
Defence profile (0.8, 0.6, 0.3, 0.2)
Team B
Attack profile (0.95, 0.85, 0.1, 0.2)
Defence profile (0.95, 0.85, 0.1, 0.2)

We already have calculated P(Team A scoring against Team B), we now have P(Team B scoring against Team A) for the various scoring options available

= (0.9(0.2), 0.85(0.4), 0.3(0.7), 0.1(0.8))
= (0.18, 0.34, 0.21, 0.08)

While this hypothetical data is just for successful execution of line attack/line defence, it can be seen that per attack executed, the best attacking options for Team B, when playing against Team A are options B, C, A, D in that order. Different attacking options work well against different defensive patterns and rely on the attacking team being able to adapt to the scoring opportunities presented by a particular defensive structure.

It can be clearly seen that different line attacks are optimal for different teams and when facing different defensive structures. For example Team B may have not employed a man on man defence in order to provide opportunities for Team A to score with a long pass, whilst Team C may structure their defence in a manner that requires Team A to attempt to use their agility in order to score, by adopting a man on man defensive structure with profile:

Team C (0.6,0.5, 0.8, 0.7)

So for Team A:

P(A scoring against C) = (0.8(0.4), 0.6(0.5), 0.3(0.2), 0.2(0.3))
= (0.32, 0.3, 0.06, 0.06)

This would imply that moves A and B are the attacking options with the higher probabilities of a successful score. This differs from those that should be employed against Team B due to the alternative defensive structure.

A factor of interest is to consider whether there can exist teams D, E and F such that

P(D scoring against E)> P(E scoring against D)

And
P(E scoring against F)>P(F scoring against E)
But
P(D scoring against F)<P(F scoring against D)

3. RESULTS

Let us now define the parameters that govern the probabilities that these teams will successfully score or defend a given play. Clearly it is easier to engineer such a situation with more attacking plays and their corresponding probabilities for defence available to choose from for each team. However we will consider if this is possible for a situation with as few as two attacking plays and their corresponding defence for each team. Let Team N have attacking plays α_N and β_N and defensive structure such that the probabilities of stopping these two plays are γ_N and δ_N respectively. We can therefore express each team as a 1x4 probability matrix such that

Team N is given by (α_N, β_N, γ_N, δ_N) giving us

Team D (α_D, β_D, γ_D, δ_D)
Team E (α_E, β_E, γ_E, δ_E)
and
Team F (α_F, β_F, γ_F, δ_F)

Given the assumption that the correct tactical approach is chosen by the coaching staff/players with regard to the attacking play selected against a particular defence (which is a reasonable assumption given at a high level, coaches are well aware of the strengths, weaknesses and methodologies of their opposition):

If P(D scoring against E)> P(E scoring against D)
Then Either $\beta_E(1-\delta_D)<\alpha_D(1-\gamma_E)>\alpha_E(1-\gamma_D)$
Or $\beta_E(1-\delta_D)<\beta_D(1-\delta_E)>\alpha_E(1-\gamma_D)$

If P(E scoring against F)> P(F scoring against E)
Then Either $\beta_F(1-\delta_E)<\alpha_E(1-\gamma_F)>\alpha_F(1-\gamma_E)$
Or $\beta_F(1-\delta_E)<\beta_E(1-\delta_F)>\alpha_F(1-\gamma_E)$

Similarly if P(F scoring against D)> P(D scoring against F)
Then Either $\beta_D(1-\delta_F)<\alpha_F(1-\gamma_D)>\alpha_D(1-\gamma_F)$
Or $\beta_D(1-\delta_F)<\beta_F(1-\delta_D)>\alpha_D(1-\gamma_F)$

Of course if we cannot assume that the correct attacking play is chosen against a particular defence, then for one of these teams to have the better chance of scoring than its opposition obviously we need to change the conditional nature of these equations to a situation where both conditions apply for each set of teams. So for example for teams D and E, if P(D scoring against E)> P(E scoring against D)
Then Either $\beta_E(1-\delta_D)<\alpha_D(1-\gamma_E)>\alpha_E(1-\gamma_D)$
And $\beta_E(1-\delta_D)<\beta_D(1-\delta_E)>\alpha_E(1-\gamma_D)$

Let us however once again return to the situation where the offensive team have chosen the correct attacking option. Consider the following situation:
α and β are two types of attacking play focusing on exploiting different kinds of defence. Usually there will be significant difference in the ability of a defensive team to defend against each of these attacks as there will be a different focus on both plays and they will be designed to exploit very different defensive shortcomings. Let us assume that:
Team D can successfully execute an attacking play α with a higher probability than play β.
Team D can defend attacking play β with higher probability than attacking play α.
Team E can successfully execute an attacking play β with a higher probability than play α.
Team E can defend attacking play β with higher probability than attacking play α.
Team F can successfully execute an attacking play α with a higher probability than play β.
Team F can defend attacking play α with higher probability than attacking play β.
If the teams are described by the following matrices:
Team D (α_D, β_D, γ_D, δ_D)
Team E (α_E, β_E, γ_E, δ_E)
and
Team F (α_F, β_F, γ_F, δ_F)

First we can consider the following example of such a situation:
Team D (0.6, 0.5, 0.5, 0.6)
Team E (0.5, 0.6, 0.5, 0.6)
and
Team F (0.6, 0.5, 0.7, 0.4)
If P(D scoring against E)> P(E scoring against D)
Then Either $\beta_E(1-\delta_D)<\alpha_D(1-\gamma_E)>\alpha_E(1-\gamma_D)$
Or $\beta_E(1-\delta_D)<\beta_D(1-\delta_E)>\alpha_E(1-\gamma_D)$

So
Either $0.6(0.4)<0.6(0.5)>0.5(0.5)$
Or $0.6(0.4)<0.5(0.4)>0.5(0.5)$
Giving
Either $0.24<0.3>0.25$
Or $0.24<0.2>0.25$
The first condition being met, implying that by selecting the correct approach, P(D scoring against E)> P(E scoring against D).
Similarly if P(E scoring against F)> P(F scoring against E)

Then Either $0.5(0.4)<0.5(0.3)>0.6(0.5)$

Or $0.5(0.4)<0.6(0.7)>0.6(0.5)$

The second condition being met, implying that by selecting the correct approach, P(E scoring against F)> P(F scoring against E).

Similarly if P(F scoring against D)> P(D scoring against F)

Then Either $0.5(0.4)<0.6(0.5)>0.6(0.3)$

Or $0.5(0.4)<0.5(0.4)>0.6(0.3)$

The first condition being met, implying that by selecting the correct approach, P(F scoring against D)> P(D scoring against F). If we model these three teams as having an equal ability to gain sufficient field position in order to structure an attacking play, then all teams have an equal probability of making a given number of scoring attacks. Therefore if a team has a higher probability of scoring once it is in this attacking position it will also have a higher probability of winning the game. Since P(D scoring against E)> P(E scoring against D), P(E scoring against F)> P(F scoring against E) but P(F scoring against D)> P(D scoring against F) the probability of one team beating another at a sport with similar play structure to that discussed in this paper has thus been shown to not be of an associative nature.

Assuming the teams are similar in ability with their more successful offensive and defensive options having probability $n+\Delta$, and their less favoured options having probability n, where $0<n<1$, $\Delta<n$, $0<\Delta>1$ and $n+2\Delta<1$, then we can consider this algebraically. Clearly If Team D has an attacking option for which Team E's defence does not optimally defend, but for Team E's favoured attacking option Team D can more optimally defend, Team D would stand more likelihood of prevailing in this encounter. For this situation the matrices become:

Team D $(n+\Delta, n, n, n+\Delta)$

Team E $(n, n+\Delta, n, n+\Delta)$

If P(D scoring against E)> P(E scoring against D)

Then Either

$(n+\Delta)(1-(n+\Delta))< (n+\Delta) (1-n) >n(1-n)$

Or $(n+\Delta)(1-(n+\Delta))< (n) (1-(n+\Delta) >n(1-n)$

Clearly by inspection, given the parameters confining the values of the variables, the first of these equations is correct, whilst the second is not valid. This therefore confirms the previous observation with regard to these two teams.

We can then consider a Team F for which the favoured defensive option is the opposite focus, but of similar standard to teams D and E, however the attacking structure is less specialised, meaning players can execute a range of different attacking options, though none as well as the more specialised teams D and E. A real world example of this would be the situation where Team D has an exceptional, highly agile team or player able to execute a variety of stepping, or diving plays relying on the superior agility of their players. Team E meanwhile might have a player or players that are less agile, but highly proficient in a long range passing game. Team F might have players able to execute attacking players with less proficiency than the other two teams but might able to use both styles in their attack. Therefore we have:

Team F $(n+0.5\Delta, n+0.5\Delta, n+\Delta, n)$

If P(E scoring against F)> P(F scoring against E)

Then Either

$(n+0.5\Delta)(1-(n))< (n+\Delta)(1-n) > (n+0.5\Delta) (1-(n+\Delta))$

Or

$(n+0.5\Delta)(1-(n))<(n)(1-(n+\Delta))>(n+0.5\Delta)(1-(n+\Delta))$

Clearly via inspection, the conditions in the first line are met, but the conditions in the second line are not. Therefore P(E scoring against F)> P(F scoring against E) and assuming other variables governing the performance of the two teams are comparable, Team E should win this encounter.

However these conditions are not met for Team D. In fact, examining this relationship

IF P(F scoring against D)> P(D scoring against F)

Either $(n+\Delta)(1-(n+\Delta)) < (n+0.5\Delta)(1-(n)) > n(1-n)$

Or $(n+\Delta)(1-(n+\Delta)) < (n+0.5\Delta) (1-(n+\Delta)) > n(1-n)$

Then from the first of these two equations, Team F should have the higher probability of scoring, provided $\Delta> 0.5 -1.5n$. This can be displayed graphically as per Figure 2.

Consider Team G $(n, n, n+2\Delta, n)$. Clearly via inspection P(F scoring against G)> P(G scoring against F). Additionally (E scoring against G) > P(G scoring against E) also by inspection. For (D scoring against G) > P(G scoring against D)

Then Either

$n(1-(n+\Delta))< (n+\Delta) (1-(n+2\Delta) >n(1-n)$

Or $n(1-(n+\Delta))< (n) (1-(n)) >n(1-n)$

The success probabilities of D's attacks are therefore $(n+\Delta) (1-(n+2\Delta)$ and $n(1-n)$. While for G these are $n(1-(n+\Delta))$ and $n(1-n)$.

Clearly $(n+\Delta) (1-(n+2\Delta) >n(1-(n+\Delta))< n(1-n)$

Therefore if $(n+\Delta) (1-(n+2\Delta)< n(1-n)$ D has greater probability of attack success. Which is so if $3n+2\Delta>1$. If $3n+2\Delta=1$, the two teams attacks having equal chance of success, whilst if $3n+2\Delta<1$, Team G is more likely to prevail in this contest.

Expanding this model it can be shown that given $n+\Gamma\Delta<1$ and $\Gamma>0$, any team $(n, n, n+\Gamma\Delta, n)$, will be defeated by teams E and F. If $\Gamma=(1-n)/(\Delta+n)$ this team will have an attack as effective as Team A's best attack and the teams (within the confines of this limited model) can be considered equivalent. If however, Γ is less than this value Team A has the most effective attacking option and if Γ is greater than this value, Team A has a less effective options to select in attack than the primary option of this other hypothetical team.

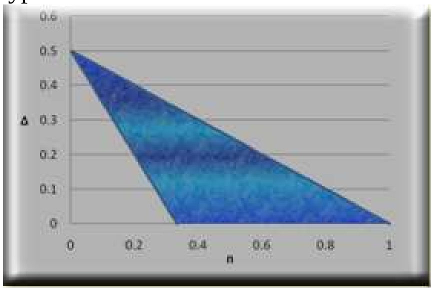


Figure 2: Area indicating Team F's probabilities of scoring. The coloured area meets $\Delta > 0.5 - 1.5n$ and $n+2\Delta<1$ for $0<n<1$, $0 >\Delta<0.5$.

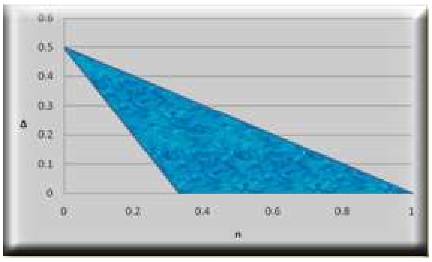


Figure3: Graph of probabilities between Team D and Team G. Coloured area indicates region for which $3n+2\Delta<1$ and $n+2\Delta<1$ for $0<n<1$, $0 >\Delta<0.5$.

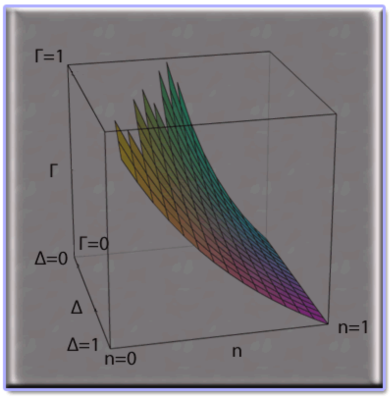


Figure 4: Graph of $\Gamma=(1-n)/(\Delta+n)$, the area below this three dimensional surface satisfies the condition $\Gamma<(1-n)/(\Delta+n)$, a requirement for the hypothetical team defined using parameter Γ, to have a more effective attacking option than any of team A's options.

4. DISCUSSION

Just because a team is better at line attack/defence than another team, it does not mean that they are the better team. The teams must also drive the ball to the position where they can set up an attempt to line attack, something the other team is trying to stop them from doing. For an effective model of a game of touch football we need to consider the different phases of the game and the transition between the different phases. This however must be developed in steps and this paper is focused on the initiation of model development with room for future expansion.

This examination of one component of the sport of touch football needs to be further examined and modelled within the context of models representing the other game elements. Once this is done this model needs to be examined with respect to real game data. If the non-associative nature of this relationship is found to be statistically significant then this will have implications for this category of sport. Provided this significance can be extrapolated to the probabilities of winning/losing a game (e.g. if the teams were equal in other parameters), then these implications would include tournament design, particularly with relevance to the finals structure of tournaments. An example of this would be coaching decisions with regard to purposefully losing games, or failing to win games by a given margin in order to change tournament finals placing and receive a more probabilistically favourable pathway to the final. Certainly the potential for purposefully losing games is not a new concept. We have already seen one example of this in the NBA, where there is a prospect to gain higher honours and greater renown at a later stage, it is clear that purposefully losing games is a path that could feasibly be taken by some teams. This sort of ethical choice would be shown to be necessary if this non-associative nature of winning probabilities was found to be of performance significant. Of course this would be in the context of sheer mathematical probabilities of a winning result, without consideration of moral and psychological issues. It must also be remembered that whilst recognition and reward are known to enhance the effort from competitors, it is also clear that uncertainty as to whether this recognition will correctly be given reduces this effort. A team may be playing in a tournament where they could be eliminated by a generally less effective team, but with a structure that is more effective in this particular situation. According to the literature the possibility and uncertainty generated by this would therefore reduce effort from competitors, undermining a primary goal of tournaments.

The implications of this study can be applied to other sports with similar strategies, which defend one or two particular attacking options specifically at the expense of others. The relevance of this will therefore vary from sport to sport, however in certain sports, such as for example gridiron and touch football team defences will adopt certain formations which present a favourable attacking option, which is tailored to the relative strengths and weaknesses of the two teams. Often as indicated in the introduction the strengths and weaknesses of a given opponent will be well known at a higher level. This situation would however only of course apply if these results are based on a series of league type match results involving many different teams. Of course in the situation in which these predictions were based on a series of results involving events with only the same two teams involved in each event the conditions of the associative nature of predictions between 3 or more teams would not be relevant as only two teams are used to supply data sets.

5. CONCLUSIONS

Many factors are discussed and some only in outline, requiring further evaluation and exploration in future papers. The simple probabilities involved within this paper could be crudely estimated from observing game data. This paper however, allows some ability to look into the statistical significance of any trends based on number of data points selected from and allows, whilst crude, at least some degree of quantifiable basis for results prediction and attacking or defending options chosen by coaches. For example which moves and defensive options will be most appropriate against which opposing teams

This paper simply and logically shows that the most important part of touch football, namely the action of attacking the scoreline to score touchdowns and the defence of this line can be shown to be non-associative between different teams. This was done by consideration of as few as two attacking options and their associated defence. Clearly the likelihood of such a non-associative situation arising in a tournament can be considered more likely when a range of dozen or so plays are well practiced by all times given that such a situation can be engineered with as few as two different attacking plays.

As discussed in the introduction, one of the important aspects of a tournament is that the best team wins when a particular tournament system is used. Logically it is therefore also important that the second best team finishes in second place, whilst the third best team finishes in third place and soon. This is particularly important when medals, trophies or other accolades are awarded for these places. It is therefore important that the correct finals structure is identified to facilitate this. Whether the probability of winning a game is of an associative nature when considering games between several teams ranked as being of different standard is an important issue to consider. This may well in fact be an important consideration not just for touch football, but additionally for other similar invasion type games such as gridiron and rugby league.

Acknowledgements

The author would like to thank those he has played touch football with. Primarily Manu Wakely, coach of numerous teams the author has been fortunate enough to be involved with, for sharing highly insightful and informative technical analysis of the sport of touch football.

References

Becker, B., & Huselid, M., (1992). The incentive effects of tournament compensation systems. *Administrative Science Quarterly*, 37

Ehrenberg, G., & Bognanno, L., (1990). Do tournaments have incentive effects? *The Journal of Political Economy*, *98*(6), 1307-1324.

Federation International Touch Inc. (2003). *Playing rules with explanations and interpretations 3rd Edition.* Curtin, Australia: F.I.T.

Galton, F., (1902). The most suitable proportion between the values of first and second prizes. *Biometrika*, *1*(4), 385-389.

Govan, A., Langville, A., & Meyer, C. (2009). Offense-Defense Approach to Ranking Team Sports. *Journal of Quantitative Analysis in Sports*, *5*(1).

Green, J., & Stokey, N., (1983). A comparison of tournaments and contracts. *Journal of Political Economy*, *91*, 349–64.

Lazear, E., & Rosen, S., (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, *89*(5), 841-864.

Nalebuff, B., & Stiglitz, J., (1983). Prizes and incentives: toward a general theory of compensation and competition. *Bell Journal of Economics*, *14*(1) 21-43.

O'Keeffe, M., Viscusi, W., & Zeckhauser, R., (1984). Economic contests: comparative revenue schemes. *Journal of Labor Economics*, *2*, 27-56.

Taylor, B., & Trogdon, J. (2002). Losing to Win: Tournament Incentives in the National Basketball Association. *Journal of Labor Economics*, *20*(1), 23-41.

Touch Football Australia (2007). *Playing rules and referees signals 7th Edition*. Australia: Touch Football Australia.

Walsh, J. (2010). *A Markov States model for touch football: Proceedings of the tenth Australasian Conference on Mathematics and Computers in Sport, Charles Darwin University* (pp259-267). Darwin, Australia.

Zetou, E., Moustakidis, A., Tsigilis, N., & Komninakidou, A. (2007). Does effectiveness of skill in complex I predict win in men's Olympic volleyball games? *Journal of Quantitative Analysis in Sports*, *3*(4).

A TWO-STAGE SIMULATION TO PREDICT MEDALISTS IN PISTOL SHOOTING

Rebecca Shepherd, Anthony Bedford [a] and Adrian Schembri

School of Mathematical and Geospatial Sciences
RMIT University
[a] *Corresponding author: anthony.bedford@rmit.edu.au*

Abstract

The International Sports Shooting Federation (ISSF) World Cup Series for air pistol consists of four events each year. Each event comprises two stages: four qualifying rounds and a final round for the top eight competitors. This paper considers best predicting the outcome of both qualifying and final rounds throughout the event, as well as how this information can be utilised to develop a guide for shooters wanting to improve their standing during competition. Due to the differing scoring methods used during an event a two staged approach was required and a variety of methods and models examined to best evaluate this, including multiple and logistic regression, discriminant analysis and simulation. The 2005-2008 World Cup Series results in the Women's 10m Air Pistol were used with a simulation derived from distribution and probabilities determined to be the most appropriate and accurate method of prediction. This was confirmed in practice using data from a single World Cup competition held in 2009.

Keywords: simulation, modeling, shooting, prediction

1. INTRODUCTION

With its beginnings in war, hunting and archery, the first known competitive shooting event occurred in Bavaria late in the 15th Century. A painting from Switzerland from the early 16th Century depicts a rifle match reminiscent of modern setups, and wooden targets from the same era are somewhat common in German collections.

Since its inception in 1896, the modern Olympic games has hosted shooting events utilising both live and static targets. Although few women participated, competition was technically mixed gender. In the 1970s a female competitor was awarded the silver medal for a rifle event. In 1984 several women's events were introduced, however it was not until the early 1990s that mixed gender competition was abandoned completely. The 10 metre Women's Air Pistol, the subject of this research, was introduced to the Olympic agenda in Seoul in 1988.

The International Shooting Sport Federation (ISSF) has existed in one form or another since 1907 and is now the preeminent governing body for several shooting disciplines. In 1986 the ISSF initiated the World Cup series in response to a request from the International Olympic Committee for a qualification system. World Cup performances can be acknowledged as world records.

Competition ranges from Olympic Games and World Cups, held in Europe, North and South America, Asia and Australia, through to regional contests, with numerous events in shotgun, rifle and pistol. The number of competitors can vary greatly, possibly due to the depth of competition level, location and timing, and has observed to be anywhere between 40 and 140 participants.
The women's 10m air pistol begins with four qualifying rounds consisting of 10 shots each. Interestingly, each shot score is rounded down to the nearest whole number (i.e. using a floor function), resulting in a maximal single shot score of 10,

maximal round score of 100, and maximal qualifying score of 400.

Upon completion of qualifying rounds, the top eight competitors advance to the final, with a shoot-off adopted in the case of a tie for eighth place. It is at this point that the scoring method is unique. Scores are cumulative throughout the tournament and each shot during the final is scored to one decimal place, resulting in a maximum final round score of 109 (i.e. ten shots of score 10.9) yielding a total potential event score of 509.0. The competitor with the highest total is awarded gold, next silver and third bronze with competition often incredibly close. One observed event concluded in a tie breaking shoot-out with the gold medal decided by only 1mm.

A computerised target system is utilised for scoring, feeding back results at the conclusion of each qualifying round and after each final round shot, typically on a monitor. Strict rules apply in relation to timing between shots.

Forecasting the results of competitive sporting events is a much-considered topic in literature; however it is ratings models that prove popular in this regard. Most models utilise the past performance of the relevant individual or team. A noteworthy exception is the method formulated by Duckworth and Lewis (1998). Used to reset targets in interrupted cricket matches, it is a method formulated from past scores and suitable for use at the different stages of play for any teams.

Specific papers on the prediction of sports shooting and multi-staged approaches were not found. Utilising a complete set of data from the women's 10m Air Pistol World Cup events for the years 2005-2008 (excluding World Cup Finals), we aim to construct a dynamic in-competition method of predicting the outcome of both the qualifying and final rounds for all competitors.

2. METHODS

As a precursor to the approaches taken, we need to consider the nature of shooting data utilised in this research.

2.1. The data

The qualification stage (for our purposes, stage one) of the tournament consists of four rounds of ten shots. Let us denote the score at round *i* for player *k* as $R_{i,k}$, which is the sum of each shot at shot *j*, denoted $s_{j,k}$, rounded using the floor function as in (1) below

$$R_{i,k} = \sum_{j=1}^{10} \lfloor s_{j,k} \rfloor \qquad (1)$$

where $0 \le s_{j,k} \le 10.9, i \in \{1,2,3,4\}$,
$\lfloor c \rfloor = \max\{n \in \mathrm{Z}, n \le c\}$.
The cumulative qualifying score at round *i* for player *k* is given by $Q_{i,k}$ which is given as

$$Q_{i,k} = \sum_{m=1}^{i} R_{m,k} \qquad (2)$$

where $i \in \{1,2,3,4\}$.

The use of the floor function in scoring has a bearing on (2) especially for shooters that are shooting scores 'close to the integer'. For instance, suppose shooter *A* scores {9.2, 9.3, 9.2, 9.4, 9.8, 9.6, 8.8, 8.9, 10.1, 9.9} which sums to 94.2 and receives $R_{i,k}$ =89. This exhibits no difference to a shooter *B* scoring {9.6, 9.5, 9.9, 9.8, 9.5, 9.9, 8.9, 8.9, 10.5, 9.9} which sums 96.4 yet also receives $R_{i,k}$ =89.

The finals stage (for our purposes stage two or round five) consists of ten shots measured to the single decimal level. To make this stage shooters must make it into the top eight at round *i*=4, with tie-breaking procedures used if there is equality in scores for eighth.

Scores from (2) are added to the ten final shots to form the grand total, $G_{i,k}$ with ranks deciding the final order, so

$$G_{i,k} = 1_{\{i>0\}} \cdot \sum_{j=1}^{i} s_{j,k} + Q_{4,k} \qquad (3)$$

where $0 \le s_{j,k} \le 10.9, i \in \{0,...,10\}$ and $1_{\{k\}}$ is the indicator function taking value 1 when *k* is true.

The distribution of the data in the qualifying rounds is left-skewed, which is to be expected given the nature of the sport (predominately good shots, very few poor and excellent). In the data for the qualifying rounds, a score of 100 occurs only (approximately) 0.2% of the time, a score of 99 occurs only 1.1% of the time; and 98 around 4% of the time (see Table 1).

Score	R_1	R_2	R_3	R_4	Total
		Frequency			
98	59	53	42	53	207
99	13	21	15	6	55
100	1	2	2	3	8
		% Frequency			
98	4.5%	4.0%	3.2%	4.0%	4.0%
99	1.0%	1.6%	1.1%	0.5%	1.1%
100	0.1%	0.2%	0.2%	0.2%	0.2%

Table 1: Frequency and Percentage Frequency of top three round scores

However we found that in the finals round the data was roughly symmetrical; part of this is due to the fact that each shot is now measured and recorded to one decimal place (which allows for wider dispersion). Another explanation could be that players may 'give up' when out of medal contention.

So, given that there are differing stages, and different scoring systems within, a two-stage model was required to predict qualification and finals place. The how is now for discussion.

2.2. Preliminary Analysis

Prior to arriving at the appropriate method, a variety of models were explored, including multiple and logistic regression; and discriminant analysis. Whilst some interesting results were observed, these were determined to be inappropriate given the dependant nature of the data. Ultimately it was decided that a simulation model derived from the distribution of probabilities of all past scores, rather than grouped past performances of an individual, was the most applicable method for investigation.

2.3. Probabilities and the Model

To calculate the probability of the available outcomes for any given round score R_i it was first necessary to categorise the data into bands. For the first stage this was determined by mapping the total qualifying round score (i.e. Q_k) over three bands, namely scores *always* observed qualifying at the end of round 4 (Q), scores *sometimes* observed (maybe) qualifying at the end of round 4 (MQ) and scores *never* observed (do not) qualifying at the end of round 4 (DNQ). Following this, counts are performed on the bands to determine the likelihood of falling in one of the three bands for each $R_{i,k}$. So essentially, we are determining the conditional probability of Q, MQ and DNQ given the current round score.

Similarly, categorising the data was necessary in stage two, however due to the more precise nature of the scoring, a different approach was required. Banding was again determined by the final outcome, in this case medal and no medal, however counts were performed on the shots observed within each cumulative *rank* rather than *score*. This approach was imperative due to the dependent nature of final stage score on qualifying as seen in (3). Bands were divided into medal (M) and no medal (NM) as defined previously in the paper. Probabilities can again be calculated using these figures.

2.4. Data smoothing

Once the bands were determined and probabilities calculated, a degree of noise was found. Sargent and Bedford (2009) demonstrated that application of a non-linear smoother to AFL data improved forecasting by removing noise. Here several smoothing methods are applied to both the qualifying and final round data. Initially, stage one probabilities were interpolated if there was a need to account for any as yet unobserved scores. Following this, a Tukey T4253H smoother was imposed on the transition probabilities for both stage one and two data (see Figure 1). Tukey (1971) offers a diverse mix of nonlinear smoothers using running medians with which to remove unwanted noise from data sets. A comprehensive discussion on the use of median smoothers is given in Sargent and Bedford (2009). As an example of the need to utilise the T4253H smoother, Table 2 provides the values and effect of the smoother as applied to the round 1 probabilities for band Q.

As seen for a score of 99, the $P(\mathrm{Q} \mid R_{1,k} = 99) < P(\mathrm{Q} \mid R_{1,k} = 98)$ due to sample variation. This is was seen as an unlikely outcome, given that $P(\mathrm{Q} \mid R_{1,k} = 100) = 1$, thus the smoother was used. For notational purposes, we denote the smoothed probabilities as P^*.

R_1	$P(\text{Q} \mid R_1)$	$P^*(\text{Q} \mid R_1)$
91	0	0
92	0.01	0
93	0.00	0.01
94	0.03	0.04
95	0.08	0.07
96	0.12	0.11
97	0.13	0.16
98	0.29	0.25
99	0.23	0.49
100	1	0.87

Table 2: Round 1 scores banded by the chance of qualifying (Q) at the end of Round 4; raw and smoothed likelihoods.

A similar process was applied to date for the finals round. However, due to the variability in Q_k, attempting to band based on this measure yielded farcical scenarios.

Suppose that in tournament 1, the gold medal score was 493.1. In tournament 2, let Q_1 = 370 for the top qualifier. Based on Q_1 in tournament 2 it would be impossible to reach the gold medal score of tournament 1. Thus prediction of a medal based on entering score in poorer tournaments may yield predictions outside of reasonable bounds. Specifically, it would predict that no one would win a medal!

The same process was used for each shot in the final round, however scores were replaced with ranks. The bands were divided into medal (M) and no medal (NM). Table 3 provides the smoothed values for ranks at the end of round 1. Again, for notational purposes, we denote the smoothed probabilities as P^*.

$rank(s_1)$	$P^*(M \mid s_1)$	$P^*(NM \mid s_1)$
1	0.75	0.25
2	0.70	0.30
3	0.58	0.42
4	0.39	0.61
5	0.22	0.78
6	0.13	0.87
7	0.11	0.89
8	0.11	0.89

Table 3: Post final round shot 1 scores banded by the chance of medal and non medal (M, NM); raw and smoothed likelihoods.

2.5. Probability distributions

We define the process results detailed in 2.4 as follows.

Let

$$\Omega_i = P^*(z \mid R_{i,k} = x) \qquad (4)$$

where $z \in \{Q, MQ, DNQ\}, x \in \{0,...,400\}$ and Ω_i denotes the adjusted probability of qualify, maybe qualify or do not qualify for any player k at round i where the round score R is value x.

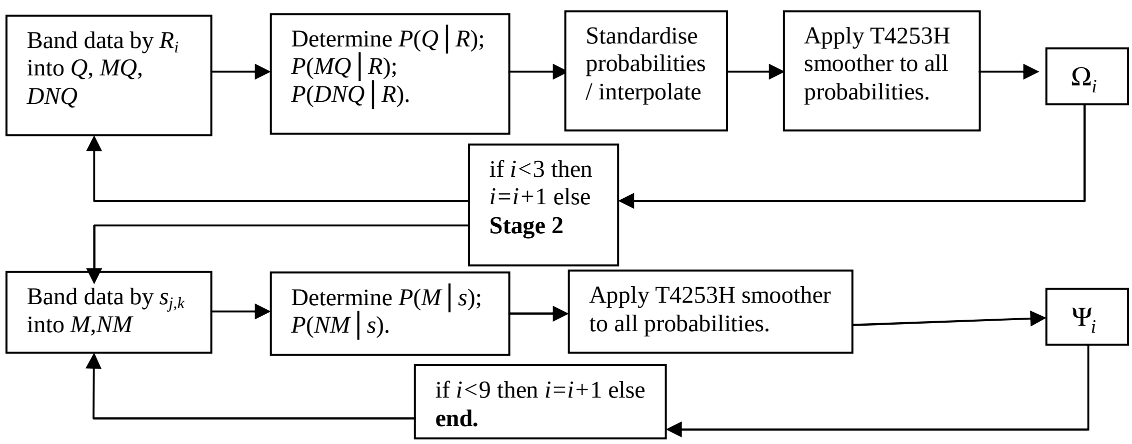


Figure 1: Flowchart of the smoothing process to obtain classification likelihoods.

Let us also denote

$$\Psi_i = P^*(z \mid rank(G_{i,k}) = x) \qquad (5)$$

where $z \in \{M, NM\}, x \in \{1,\ldots,8\}$ and Ψ_i is the adjusted probability of medal or no medal for any player k at round i where the cumulative score G is ranked x. The above set of probabilities yield quasi-transition matrices used in the simulation at each stage of the tournament. For example, for i =1,

$$P^*(z \mid R_{1,k} = x) = \Omega_1 =$$

$$\begin{array}{cc} & z \\ & \begin{array}{ccc} Q & MQ & DNQ \end{array} \\ x & \begin{array}{r} 0 \\ \vdots \\ 90 \\ 91 \\ 92 \\ 93 \\ 94 \\ 95 \\ 96 \\ 97 \\ 98 \\ 99 \\ 100 \end{array} \left(\begin{array}{ccc} 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \\ 0 & 0.02 & 0.98 \\ 0 & 0.06 & 0.94 \\ 0.01 & 0.12 & 0.87 \\ 0.04 & 0.19 & 0.77 \\ 0.07 & 0.25 & 0.68 \\ 0.11 & 0.31 & 0.58 \\ 0.16 & 0.35 & 0.49 \\ 0.27 & 0.37 & 0.36 \\ 0.49 & 0.34 & 0.17 \\ 0.74 & 0.26 & 0 \end{array} \right) \end{array}$$

so a player at score 90 or lower always fits into the *DNQ* band of shooters and a player at score 96 has a 11% chance of being a *Q*, 31% of *MQ* and 58% of *DNQ*. If a player falls into the *DNQ* band this does not imply that the player will not qualify, rather that this shot at this stage typically came from a player that *DNQ*. This logic will be visited in the next section.

In similar vein for shot i =0 (i.e. before the first shot of the finals round)

$$P^*(z \mid rank(G_{0,k}) = x) = \Psi_0 =$$

$$\begin{array}{cc} & z \\ & \begin{array}{cc} M & NM \end{array} \\ x & \begin{array}{r} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \left(\begin{array}{cc} 0.74 & 0.26 \\ 0.61 & 0.39 \\ 0.43 & 0.57 \\ 0.26 & 0.74 \\ 0.17 & 0.83 \\ 0.15 & 0.85 \\ 0.15 & 0.85 \\ 0.15 & 0.85 \end{array} \right) \end{array}$$

so a player ranked sixth to eighth at the end of qualifying won a medal 15% of the time.

So a complete tournament consists of a qualification stage $\{\Omega_1, \Omega_2, \Omega_3\}$ and finals stage $\{\Psi_0, \Psi_1, \Psi_2, \ldots, \Psi_9\}$. These transitions alone do not offer any solution as such; they are simply the likelihood at stage i to either qualify or win a medal. However, coupled with both actual and forecast scores, we can determine an approximate final qualifying score and rank and if in the top eight, medal position at any stage. Let us now set up this process.

2.6. Simulation of R_i and sj,k

The ability to forecast the likelihood of qualifying, and medalling, required extensive distribution fits. Whilst we first had estimates of conditional probabilities of both qualifying (Ω_i) and medalling (Ψ_i), we then needed to fit the distribution of scores for the corresponding predicted states on actual data. That is, we then used the banded scores at each round to determine the distribution of shots at stage i. This was fundamental to realistically simulating tournaments. For example, if a player shoots a round 1 score of 93, they have a 12% chance of *MQ*. So, given this information, what is the distribution of shots for all players at round 1 given that are in the state *MQ*? Players that shot a 93 may well go on to shoot a 100, others maybe a 90, however all players in this range have a distribution of shots. For this we utilised the @Risk package, and found all scores for shots fired within each band followed the binomial distribution.

So our final model consisted of distribution fits for each qualification type (Q, MQ and DNQ) and each stage R_i when i = 2,3 and 4; and distribution fits for two finals outcome (M, NM) of $rank(G_{i,k})$ for each stage i = 0,..,9 in the finals. So, let

$$XQ_{i,m} \sim B(n_{i,m}, p_{i,m}) \qquad (6)$$

be the distribution of shots for the qualification rounds, R_i when i = 2,3 and 4 and player k is in state $m \in \{Q, MQ, DNQ\}$. Also, let

$$XF_{i,m} \sim B(n_{i,m}, p_{i,m}) \qquad (7)$$

be the distribution of shots for the finals rounds, when i = 0,…,9 and player k is in state $m \in \{M, NM\}$. This feeds into the final simulation model, with the process detailed in Figure 2.

3. RESULTS

To begin testing of the simulation, we simulated a hypothetical tournament 10,000 times. We set the tournament length to k = 56 competitors. As a way of validating the simulation, we investigated the distribution of scores throughout the tournament; the type of competitor that wins a medal; and the scores by final rank. Given this was a hypothetical tournament, the first round was also simulated, something which is of no use in practice.

To determine the likelihood of competitors winning a medal, we split the first round shots from excellent to good, i.e., from around 99 to 90. The rank of the competitors after round 1 was matched against all subsequent results. Notably, $P(Gold) \propto rank(R_1)$, with

$P(Gold \mid rank(R_1) = 1) \approx 0.13$

$P(Gold \mid rank(R_1) = 2) \approx 0.10$ and

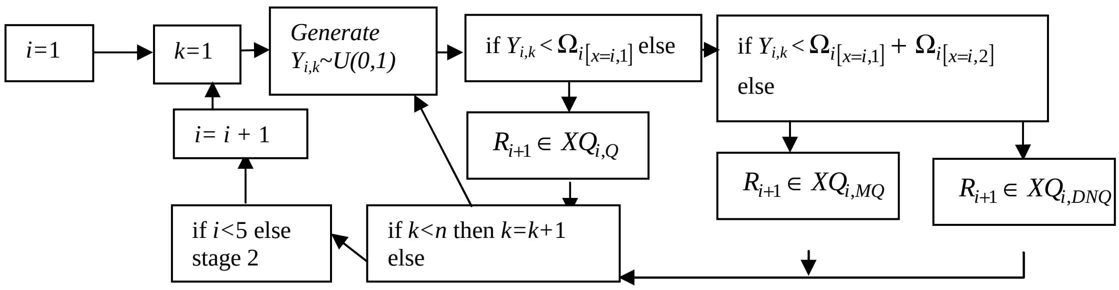
$P(Gold \mid rank(R_1) = 3) \approx 0.08$.

All competitors had won the gold at least once, indicating that it is at least possible to shoot the poorest R_1 score and still win a gold medal. Whilst in many sports this is unlikely, shooting is different in that a comeback (and indeed failure) is still possible with three rounds remaining.

Figure 3 below displays the likelihood of gold by rank in round 1. We also exhibit the distribution of finalists by round 1 rank using boxplots in figure 4. Notably, the dispersion for the IQR is tighter for the gold, and wider as the rank declines.

In figure 5, we plot the distribution of all scores for the medallists. The output reveals an expected outcome, with the mean score for a gold medal at 491.2, silver at 489.1 and bronze 487.7.

Stage 1: Qualifying Simulation Set Up



Stage 2: Finals Stage Simulation

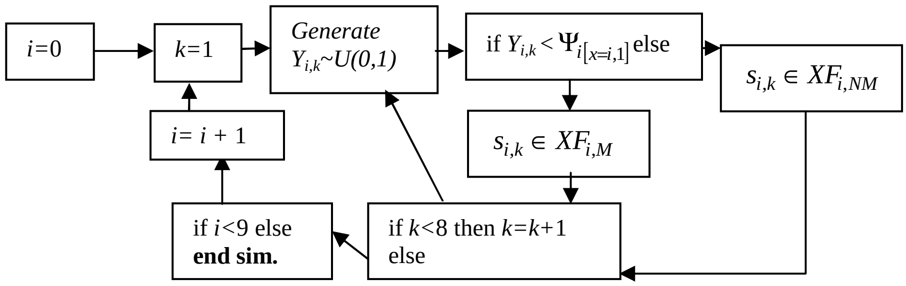


Figure 2: Flowchart of the simulation.

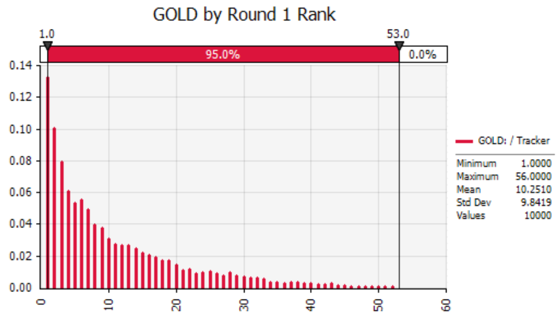


Figure 3. Round 1 rank by P(Gold).

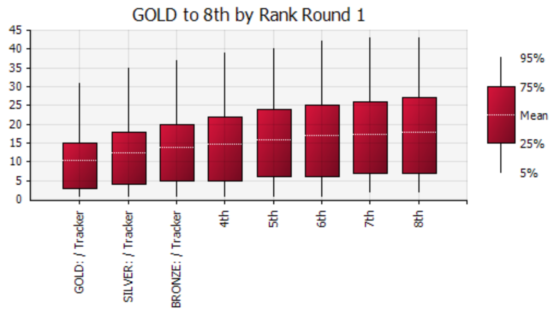


Figure 4. Boxplots of final position by round 1 rank.

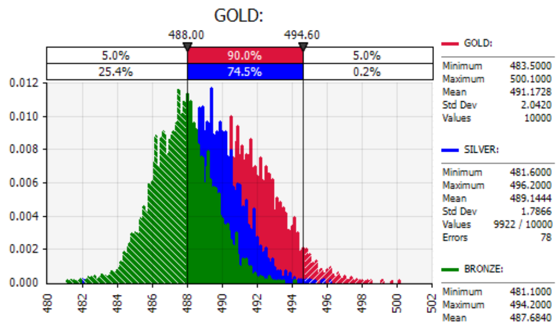


Figure 5. Distribution of scores for medals.

The middle 90% of data for gold spans 488 to 494.6 and contains 74.5% of silver. Only 0.2% of silver scores occur in the top 5% of gold scores.

As a test case, we used the 2009 Munich World Cup. In this event, we had a interesting scenario whereby one competitor shot an R1 of 100, yet did not qualify for the finals. This meant the simulation would certainly be tested in terms of prediction ability, given this unusual outcome.

First, we input all $R_{1,k}$ scores then simulated 20,000 times to determine probability of medals, and top eight finish. We then re-simulated each stage, re-entering results as if live.

Table 4 exhibits the predicted mean round score for the top ranked position, and the eighth ranked position. Arguably it is the likely eighth that is most informative, as this provides an estimate of the score needed to qualify for the finals. After the second round of shooting, the simulation estimated the actual result and did not deviate from this prediction after the third round.

	Predicted R_4 at end			Actual
Position	R_1	R_2	R_3	R_4
1st	392	390	388	390
8th	385	384	384	384

Table 4: Predicted and actual results: Munich 09

Using Table 4, we inspected the scores of all competitors and ascertained the required scores for round 4 for all competitors (shown in Table 5). Notably, two competitors needed 'easy' rounds of at least 93 (65.2% of all shots at this point were >=93), one competitor needed a 94 and shot a 92 to miss out. No competitor requiring 98 or more qualified.

Score to Q	Freq	# qualify
>100	66	0
100	2	0
99	5	0
98	4	0
97	5	1
96	9	3
95	3	2
94	1	0
93	2	2

Table 5: Final round shots by estimated score of 384

At the end of round 3, there were 15 competitors 'qualified' due to a large number tied for seventh. At the end of round four, the 16th ranked round three competitor qualified with a final round of 97.

Finally, we looked at the predicted medal score at each point throughout the tournament. Again, the bronze estimate provides a reasonable medal estimate, and was typically underestimated throughout the tournament.

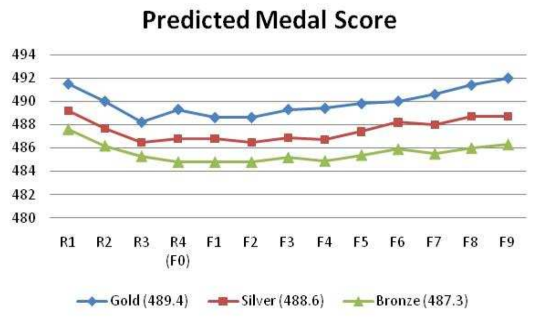


Figure 6: Predicted medal scores by Rounds

Interestingly, the medal was decided by the last shot. Xin Tong (479.3) took the lead for the first time from Olena Kostevych (479.0) after the ninth shot in the final. Olena went into the final round with a huge five point lead, however shot several poor shots. On the final shot, Xin shot her worst for the final at 9.3, and Olena shot an impressive 10.4 to take gold.

4. DISCUSSION

The simulation time is of great importance given that the in-game use of the results. The simulation run time varied markedly by the number of competitors, given that for each competitor we had to simulate 22 distributions of data.

For the Munich simulation, 20,000 simulations for 91 players took around 20 minutes. As live data was added, the simulation run marginally faster, due to a reduced load on simulated data. For example, at 20,000 simulations when R_4 was unknown, runtime was around 10 minutes. Plans are underway to improve the runtime.

The usefulness of the simulation becomes apparent when competitors need to know how they are performing, and particularly, an estimate of qualification requirements (or medals). As seen in the example given for Munich 2009, the score required to qualify for the finals converged to that observed by the end of round 2. This could simply be converted into an array of values needed for each competitor.

Estimating the required number of iterations is important. For this work, we simply ran 20,000 simulations, however utilising convergence of the estimates for the medals may shorten the number of simulations. Furthermore, we did not validate the bounds of error, and this will required more simulation work.

The implemented simulation would need to be leaner, and the simulation can be cut down through the removal of multiple distribution simulations.

Estimating effect size also remains important-we would like to know what it takes to win a gold medal with greater certainty.

A set back might be that there is no dependency in shooting form; a ratings model might assist in improving prediction accuracy.

The trial of the work with competitors is planned. The psychological aspect of such predictions is also very important. The knowledge of underperformance may have a negative rather than positive effect. This is an investigation of its own.

5. CONCLUSION

We have demonstrated the constraints and challenges faced in simulating a shooting tournament. The framework provided, whilst requiring further testing, is certainly capable of running live. We aim to further determine the specifics of winning medals and making the qualification stages therein.

ACKNOWLEDGEMENTS

The authors would like to thank Nick Sanders of the Victorian Institute of Sport and Jan Ketzel of the International Shooting Sport Federation for the provision of data used in this analysis.

References

Duckworth, F. C. and Lewis, A. J. (1998). A fair method for resetting the target in one-day cricket matches, *J. Oper. Res. Soc.*, 49, 220-227.

Sargent, J. and Bedford, A. (2009). Improving Australian Football League player performance forecasts using optimized nonlinear smoothing. *International Journal of Forecasting*, *17*, 30-36.

Tukey, J.W. (1971). *Exploratory Data Analysis*. Addison-Wesley, Reading.

BIAS IN SPORTING MATCH STATISTICS

Pollard, Graham [a, d], Pollard, Geoff [b], Lisle, Ian [a], Cross, Rod [c]
[a] *Faculty of Information Sciences and Engineering, University of Canberra, Australia*
[b] *Faculty of life and Social Sciences, Swinburne University of Technology, Melbourne, Australia*
[c] *Physics Department, University of Sydney, Australia*
[d] *Corresponding author: graham@foulsham.com.au*

Abstract

Tennis is typically modeled using two parameters, pa, the probability player A wins a point on his service, and pb, the probability B wins a point on service (Carter and Crews, 1974; Pollard, 1983). Sports where serving is an advantage (e.g. tennis) or a disadvantage (e.g. volleyball), are called 'bipoints' (Miles, 1984), whereas sports where serving is 'neutral' (e.g. squash) are called 'unipoints'. Although our primary interest in this research is related to tennis, the results are applicable to many sports, particularly 'racquet family sports'.
For many tennis matches the proportion of points won by each player on service is recorded. As the winner must have won the last point, last game and last set, the winner's service statistics can have an upwards bias, and the loser's service statistics a downwards one. We address several questions. For example, in a game, set, and match, what are the typical sizes of these biases? What is the size of the negative correlation between player A's proportion of points won on serve, and player B's? How does nesting (points within games, games within sets, and sets within the match) affect the bias? In volleyball for example where serving is generally a disadvantage rather than an advantage as in tennis, does the bias operate in a 'reverse manner' to tennis? The commonly used sports scoring systems, B_{2n-1} and W_n (n = 2), are studied. How does the bias in these systems change as n increases? Some of the examples in this paper might be of particular interest to teachers and students of probability and statistics.

Keywords: Bias, tennis, squash, volleyball, best of 2n-1 scoring, win-by-two scoring, teaching probability and correlation/dependence, gambling in sport

1. INTRODUCTION

For many years tennis has been modelled using two parameters, pa, the probability player A wins a point on service, and pb, the probability player B wins a point on service (Carter & Crews, 1974, and Pollard, 1983). Thus, games such as tennis and volleyball involving two types of points are called bipoints games, whilst games requiring only one type of point for accurate modelling, such as squash (Pollard, 1985), are called unipoints games (Miles, 1984). Although our primary interest in this work is related to tennis, many of the results are applicable to the scoring systems used in other sports, particularly those in the 'racquet family of sports'.
In the first five subsections of the Methods section we consider various unipoints scoring systems. For these systems the outcome of a point does not depend on who is serving. In particular we consider the B_{2n-1} system, the convolution of such systems, the W_n system, a single advantage game of tennis, and a 'nested' system.
In the next two subsections of the Methods section, we consider the situation in which a player's probability of winning a point does depend on whether he is serving or not. This is the bipoints case. In particular, we consider B_2 (B_3), B_4 (B_3) and a (modified) tiebreak game.

In the last three subsections of the Methods section several practical bipoints cases are considered…a single set of tennis, best-of-three sets of tennis, and best-of-five sets of tennis.

2. METHODS

Unipoints: The bias in B_{2n-1}

We consider unipoints within the best-of-2n-1 points (B_{2n-1}) scoring system (n = 2, 3, 4, …). Player A has a probability p of winning each point, and points are independent. We start with a simple example. When n = 2, A can win or lose in 2 or 3 points, and the proportion of points he wins in one play of B_3 is equal to 2/2, 2/3, 1/3 or 0/2 with probabilities p^2, $2p^2q$, $2pq^2$ and q^2, where q = 1 - p. The expected value of the proportion of points won by A in B_3, E(Prop), is equal to the above proportions weighted by their respective probabilities. As shown in Table 1, when p = 0 6, E(Prop) = 0.616, so the bias in this case when using Prop to estimate p is equal to 0.016. Some comments for B_3 (n = 2) and p = 0.6 are

1. P(2 points are played) = 0.36 + 0.16 = 0.52, and P(3 points are played) = 0.288 + 0.192 = 0.48.
2. E(Number of points) = 2*0.52 + 3*0 48 = 2.48.
3. P(A wins given 2 points) = 0.36/0.52 = 9/13.
4. P(A wins given 3 points) = 0.288/0.48 = 0.6.
5. The better player has a higher probability of winning when the match is shorter.
6. Given 2 points are played, E(Prop) = ((2/2)*0.36 + (0/2)*0.16)/0.52 = 9/13 = 0.6923.
7. Given 3 points are played, E(Prop) = ((2/3)*0.288 + (1/3)*0.192)/0.48 = 8/15 = 0.5333.
8. The expected proportion of points won by the better player is higher when the match is shorter.
9. Proportion of points won by the winner = (2/2)* 0.52 + (2/3)*0.48 = 0.84.
10. Proportion of points won by the loser = 0.16.
11. If A wins, E(Prop) = ((2/2)*0.36 + (2/3)*0.288) /(0.36+0.288) = 0.851851.
12. If B wins, E(Prop won by B) = ((2/2)*0.16 +(2/3)*0.192)/(0.16+0.192) =0.818818.
13. P(A wins) = 0.648, and P(B wins) = 0.352.
14. E(Prop points won by winner) = 0.648*0.851851 +0.352*0.818818 = 0.84, as in 9 above.
15. If X_1 = # points played and X_2 = # points won by A (when p=0.6), then, $P(X_1=2, X_2= 0) = 0.16$, $E(X_1) = 2.48$, $E(X_2) = 1.488$, $E(X_1*X_2) = 3.744$, $cov(X_1, X_2) = 0.5376$, and the $corr(X_1, X_2) = 0.1425 > 0$.
16. If Y_2 = proportion of points won by A when p=0.6, $cov(X_1, Y_2) = -0.03968$. When estimating p, we are interested in using Y_2. But $E(Y_2) = (E(X_2) - cov(X_1, Y_2))/E(X_1)$. That is, $E(Y_2)$ is not equal to $E(X_2)/E(X_1)$ as the relevant covariance is not zero.

E(Prop)	p = 0.6	p = 0.7	p = 0.3
B_3	0.616	0.728	0.272
B_5	0.6189	0.7316	0.2684
B_7	0.6189	0.7303	0.2697
B_9	0.6182	0.7280	0.2720
B_{11}	0.6173	0.7257	0.2743

Table 1: E(Prop) for B_{2n-1} (unipoints)

For B_{2n-1} the probability A wins by n points to i points, P(i), and the probability A loses n/i , Q(i), are $P(i) = {}^{n-1+i}C_i\, p^n q^i$ and $Q(i) = {}^{n-1+i}C_i\, q^n p^i$, (i = 0, 1, .., n-1). Thus, an expression for E(Prop) can be derived. For p = 0.6, the bias has its maximum value for B_7 (which is just the no ad game in tennis). For p = 0.7, the bias has its maximum for B_5. *Note that as n tends to infinity, the bias tends to zero, and the bias is negative when p < 0.5.*

Unipoints: The bias in several B_{2n-1} systems

For two B_3s and p = 0.6, E(Prop) = 0.6081, a bias of 0.0081, approximately half the bias of 0.016 for a single B_3. For three B_3s and p = 0.6 E(Prop) = 0.6053, a bias of approximately one-third of 0.016. *Thus, the biases in a full set of tennis resulting from the biases within each player's service games are reduced considerably from the single-game biases.*

Unipoints: Bias in the win-by-n points system, W_n

In an advantage game of tennis, the win-by-two or W_2 system is used to determine the winner of the game if the game score reaches deuce. For W_2 the probability A wins by (i+2) points to i, P(i), and the probability A loses by (i+2) to i, Q(i), are given by $P(i) = (2pq)^i p^2$ and $Q(i) = (2pq)^i q^2$, (i = 0, 1, 2, …), and so an expression for E(Prop) can be written down. For p = 0.6, E(Prop) = 0.6362 (Table 2), a bias of 0.0362 which is greater than that for B_3. The bias increases as n increases, and then decreases. *When p = 0.7, the bias is greatest for W_3. Clearly, the bias in W_n converges to zero (n large).*

n	P(A wins), p = 0.6	E(Prop), p = 0.6	P(A wins), p = 0.7	E(Prop), p = 0.7
2	0.6923	0.6362	0.8448	0.7594
3	0.7714	0.6502	0.9270	0.7727
4	0.8351	0.6549	0.9674	0.7709
5	0.883	0.6554	0.9857	0.7646

Table 2: Values for W_n scoring system (unipoints)

Unipoints: Bias in an advantage game of tennis

For an advantage game of tennis, when p = 0.6, E(Prop) = 0.6264, a bias of 0.0264, bigger than the bias in any B_{2n-1} when p = 0.6. E(Prop) = 0.7391 when p = 0.7, and when p = 0.5, E(Prop) = 0.5.

Unipoints: Bias in a nested scoring system, $B_3(B_3)$

We can consider $B_3(B_3)$ using first principles, but it is perhaps a little easier using probability generating functions. Let X_1 = the number of points won by A, Y_1 = the number of points played, and Z_1 = an indicator variable that equals 1 if player A wins the first B_3 'game' played. Then, the probability generating function (PGF) for (X_1, Y_1, Z_1) (with respective variables (s, t, u) is

$$F(s,t,u) = t^2(q^2(1+2pst) + up^2s^2(1+2qt))$$

If these variables for the second B_3 played are denoted by (X_2, Y_2, Z_2), then the PGF for the sums $(X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2)$ is given by the square of the above PGF, and this is equal to

$$t^4(q^4(1+2pst)^2 + u^2p^4s^4(1+2qt)^2)$$
$$+ 2uq^2p^2s^2(1+2pst)(1+2qst)t^4$$

where these three terms refer respectively to A losing, A winning and A drawing a $B_2(B_3)$ nested match. In order to get the relevant PGF for winning a $B_3(B_3)$ match, it is necessary to multiply the draw component by a win component for the third B_3 game. Thus, the PGF for the case where A wins the $B_3(B_3)$ match can be shown to equal the product of

$(t^4p^4s^4 + 2q^2p^4s^4t^6 + 4q^2p^5s^5t^7)$ and

$(1+4qt+4q^2t^2)$.

Noting the coefficients of $s^i * t^j$ in this expression for this case where A wins the $B_3(B_3)$ match, the proportion of points won by player A
equals 4/4 with probability p^4 (here i = 4 and j = 4),
equals 4/6 with probability $2p^4q^2$, (i = 4, j = 6),
equals 5/7 with probability $4p^5q^2$,
equals 4/5 with probability $4p^4q$,
equals 4/7 with probability $8p^4q^3$,
equals 5/8 with probability $16p^5q^3$,
equals 4/6 with probability $4p^4q^2$,
equals 4/8 with probability $8p^4q^4$, and
equals 5/9 with probability $16p^5q^4$ (i = 5 and j = 9).

The above analysis can be repeated for the case where A loses $B_3(B_3)$, giving 9 corresponding expressions. E(Prop) for this $B_3(B_3)$ match is equal to the sum of these 18 values weighted by their respective probabilities, and equals 0.6225 when p = 0.6 (verified by first principles). *Note that this bias is greater than that of B_{2n-1} for any n (when p = 0.6), and certainly greater than one with a similar expected duration. Nesting has increased bias here.*

Bipoints: Bias in two $B_{2n}(B_3)$ systems

(i) Consider $B_2(B_3)$ with A serving in the first 'game' (a B_3 game), and B serving in the second 'game'. E(Prop A), the expected proportion of points won by player A on service, and E(Prop B) are the same as above (as each player serves exactly once).

(pa, pb)	P(A wins)	P(B wins)	P(Draw)	E(Prop A)	E(Prop B)
(0.5, 0.5)	0.25	0.25	0.5	0.5	0.5
(0.6, 0.6)	0.2281	0.2281	0.5438	0.616	0.616
(0.7, 0.6)	0.2760	0.1400	0.5841	0.728	0.616
(0.6, 0.7)	0.1400	0.2760	0.5841	0.616	0.728
(0.3, 0.4)	0.1400	0.2760	0.5841	0.272	0.384

Table 3: Some characteristics of the system $B_2(B_3)$

(ii) Consider $B_4(B_3)$ with A serving first with probability pa of winning each point, and B serves in the second game (B_3) with probability pb. A serves 3rd and B 4th, if necessary. The 'set' is won by the first player to win 3 'games', or the set is drawn at 2-2. Each 'inner nest' can result in 4 outcomes…A wins in 2 points, A wins in 3 points, A loses in 3 points or A loses in 2 points. Thus, A can win the match in 3 games (AAA) in 8 ways (keeping track of number of points played in the first, second and third games). He can win each of the ways BAAA, ABAA and AABA in 16 ways, etc. Thus, A can win in 56 ways, and he can lose is 56 ways and there are 96 draw possibilities. We have the following results.

(a) When pa = pb = 0.5, P(A wins) = P(A loses) = 0.3125, P(Draw) = 0.375. Denoting the proportion of points won by player A on service by Prop A, correspondingly for B, E(Prop A) = E(Prop B) = 0.5.
(b) If pa = pb = 0.6, P(A wins) = P(A loses) = 0.3001, P(Draw) = 0.3998. E(Prop A) =0.6081 (see above), and E(Prop B) = 0.5863. The result AAA (with only one service break) is more likely to occur than the result BBB with two service breaks. In the first case AAA, B never gets the chance to improve his service proportion having lost his first serve, and hence his lower proportion of points won on serve.
(c) If pa = pb = 0.4 (as in 'volleyball' where serving is a disadvantage), P(A wins) = P(A loses) = 0.3001, P(Draw) = 0.3998 (as in (b) above), E(Prop A) = 0.3919 and E(Prop B) = 0.4137 (each can be obtained by subtracting the relevant numbers in (b) from unity). The result AAA is now less likely than the result BBB. In the second case BBB, player B, having 'broken the odds' and held service in the second game, does not have to risk lowering his service proportion with a second service game.

(d) If pa = 0.7 and pb = 0.6, P(A wins) = 0.3985, P(A loses) = 0.1831, P(Draw) = 0.4184, E(Prop A) =0.7144, and E(Prop B) = 0.5647. The result AAA is more likely to occur than the result BBB. Given AAA, player B never gets the chance to improve his service proportion having lost his first serve.

(e) If pa = 0.6 and pb = 0.7, P(A loses) = 0.3985 (see (d) above), P(A wins) = 0.1831, P(Draw) = 0.4184. E(Prop A) =0.6081 as earlier (as A always has 2 serves), and E(Prop B) = 0.6995. The difference between this example and (d) is striking. The bias for the second server is again negative (-0.0005), in this 'tennis context' (where pa > 0.5 and pb > 0.5).

(f) If pa = 0.6 and pb = 0.0, P(A wins) = 0.8761, P(A loses) = 0.0, P(Draw) = 0.1239, E(Prop A) = 0.6081 as earlier, and E(Prop B) = 0.0.

(g) If pa = 0.0 and pb = 0.6, P(A wins) = 0.0, P(A loses) = 0.8761, P(Draw) = 0.1239, E(Prop A) = 0.0, and interestingly E(Prop B) = 0.6929. Thus, unlike above, the bias is strongly positive for this second server. Hence, this is considered more closely. BBB occurs with probability 0.648, and B's expected proportion of points won on service given BBB is 0.851851 (see earlier). BABB and BABA contribute to E(Prop B) , and have values (i.e. corresponding to 0.648 and 0.851851) of (0.2281 and 0.510101) and (0.1239 and 0.198347) respectively. Considering BABB, there are 2 ways the first service game of B could have occurred (loss in 2 points or loss in 3) and there are 2 ways the second service game of B could have occurred (win in 2 points, or win in 3). Thus, there are 4 ways the two service games of B could have occurred, with associated proportion of points won on service and probabilities given by (2/4 and 0.16*0.36), (2/5 and 0.16*0.288), (3/5 and 0.192*0.36), and (3/6 and 0.192*0.288). These 4 have a total probability of 0.2281, and a contribution to E(Prop B) of 0.1164. (0.1164 is more exactly 0.116352, and 0.2281 is more exactly 0.228096, so the ratio 0.1164 divided by 0.2281 is more exactly 0.510101.) The draw BABA is now considered in detail. It has a probability of 0.1239 and contributes 0.0.024576 to E(Prop B). The ratio (0.0.024576) /(0.123904) is equal to 0.198347). Adding all the contributions to E(Prop B), we have 0.648*0.851851 + 0.116352 + 0.024576 = 0.6929, as above.

(h) If pa = 0.6 and pb = 1.0, P(A wins) = 0.0, P(A loses) = 0.5801, P(Draw) = 0.4199, E(Prop A) = 0.6081 as earlier and E(Prop B) = 1.0.

(i) If pa = 1.0 and pb = 0.6, P(A wins) = 0.5801, P(A loses) = 0.0, P(Draw) = 0.4199, E(Prop A) = 1.0, and E(Prop B) = 0.5311, a negative bias for the second server in this tennis context (pa, pb > 0.5).

Concluding, the interdependency of E(Prop A) and E(Prop B) can be quite complex as it depends on who is the first server, who is the better player, and the size and sign of the difference pa – pb.

Bipoints: The bias in the tiebreak game

The joint distribution of (Prop A, Prop B) for the tiebreak game is outlined. Consider, for simplicity, the tiebreak with the modification of a draw at 6-6. Suppose A serves first, and the outcome is A wins 7-3. The last point (10th) was a b-point lost by B (with probability qb). The first 9 points were 5 a-points and 4 b-points. A won 6 of these 9 points (i.e. won 2 a-points and 4 b-points, or 3 a-points and 3 b-points, or 4 and 2, or 5 and 1). Correspondingly, the pair of variables (Prop A, Prop B) were (2/5, 0/5), (3/5, 1/5), (4/5, 2/5) and (5/5, 3/5), after including the 10th point. The probabilities of these four outcomes are $({}^5C_2pa^2qa^3)(qb^4)(qb)$, $({}^5C_3pa^3qa^2)({}^4C_1pbqb^3)(qb)$, $({}^5C_4pa^4qa)({}^4C_2pb^2qb^2)(qb)$, and $(pa^5)({}^4C_3pb^3qb)(qb)$. The (Prop A, Prop B) distribution can be derived.

Observations from Table 4 include

(a) Prop A and Prop B are negatively correlated. The 'stopping rule' for this modified tiebreak game (B12) often stops play when one player is by chance doing ordinarily or relatively well, and the other player is doing ordinarily or not doing so well (Rows 2, 3).

(b) In the tennis context (ie. pa > 0.5 and pb > 0.5), for two equal players, the bias in both Prop A and Prop B is slightly negative, and slightly more for the player who serves second (see rows 2 and 3).

(c) In the tennis context, for two unequal players, the bias for the better player is increased (relative to two equal players). That is, it is less negative, or positive. Correspondingly, the bias for the weaker player is decreased i.e. more negative. Thus, if pa, pb > 0.5, the bias in Prop A (or Prop B) has an extra positive component when playing a weaker player, and an extra negative component against a better player.

The correlation between Prop A and Prop B is +1.0 if the duration of the tiebreak is exactly 12 points, and it equals -1 if the duration is exactly 7 points. *The stopping rule of this modified tiebreak game has lead to an overall negative correlation, even though points have been assumed to be independent.*

1	(pa, pb)	E(PropA)	E(PropB)	Correlation
2	(0.6,0.6)	0.5972	0.5969	-0.1751
3	(0.7,0.7)	0.6951	0.6943	-0.1610
4	(0.7,0.6)	0.7026	0.5871	-0.1619
5	(0.6,0.7)	0.5889	0.7034	-0.1607
6	(0.7,0.5)	0.7096	0.4814	-0.1447
7	(0.5,0.7)	0.4831	0.7113	-0.1455

Table 4: Modified tiebreak game characteristics

The bias in a single set of tennis

An exact analysis of a full set is possible, but very tedious. Simulation (4,000,000 sets), using the software developed by Brown, Barnett, Pollard, Lisle and Pollard (2008), was used to get estimates of the bias (Table 5). SS1 is an advantage set and SS2 is a tiebreak set. Estimates of pa and pb are pa^ and pb^. The proportion of points won by the winner on service is p(W) and the proportion of points won by the loser on service is p(L). Although all values in Table 5 have 4 decimal points, accuracy is not quite to that level. This has been done so that the reader can get a feel for various differences in the table, including some on which no comment is made.

Some observations when (pa, pb) = (0.65, 0.65) are

(a) p^(first server) – p^(second server) equals 0.0033 or 0.0034. Also, (pa^ - pb^) given A serves first minus (pa^ - pb^) given B serves first is 0.0067 for SS1 and 0.0068 for SS2. This difference is a measure of the differing bias between that when serving first and that when receiving first.

(b) The correlation coefficients between pa^ and pb^ are smaller in absolute value for SS2 than for SS1, and they are all negative.

(c) The proportion of points won on service by the winner is considerably larger than the proportion of points won on service by the loser, even though the players are equal.

Some observations when (pa, pb) = (0.7, 0.6) are

(d) The biases are positive for the better player and negative for the weaker player.

(e) (pa^ - pb^) – (pa – pb) is 0.0234 and 0.0177 for SS1, and 0.0203 and 0.0147 for SS2. Also, (pa^ - pb^) given A serves first minus (pa^ - pb^) given B serves first equals 0.0057 for SS1 and 0.0056 for SS2. These differences are slightly smaller than in (a).

(f) The correlation coefficients between pa^ and pb^ are smaller in absolute value for SS2 than for SS1, and are smaller in absolute value than those in (b) above.

(g) The proportion of points won on service by the winner minus the proportion won on service by the loser, minus (pa – pb), is greater than zero, as expected.

The bias in a best-of-three sets match of tennis

The results for best-of-three sets matches are given in Table 6 (4,000,000 matches). System SS3 is three advantage sets, SS4 is two TB sets followed by a third advantage set, SS5 is three TB sets, and SS6 is two TB sets followed by a 10 point TB match decider (as used in some doubles events).

Some observations when pa = pb = 0.65 are

(h) The value of p^(first server) – p^(second server) equals 0.0006, 0.0007 or 0.0008, and (pa^ - pb^) given A serves first minus (pa^ - pb^) given B serves first equals 0.0014, 0.0015 or 0.0016 for SS3 to SS6. As expected, these values are smaller than the corresponding ones from Table 5, as the effect on bias resulting from serving first decreases as the matches are 'longer'.

(i) The correlation coefficients get smaller in absolute value as we go from SS3 to SS6.

(j) The proportion of points won on service by the winner is larger than the proportion of points won on service by the loser, but the differences are smaller than in Table 5.

Some observations when (pa, pb) = (0.7, 0.6) are

(k) The biases are positive for the better player and negative for the weaker player.

(l) (pa^ - pb^) – (pa – pb) ranges from 0.0177 or 0.0162 for SS3 down to 0.0111 or 0.0097 for SS6. Also, similarly to (e), (pa^ - pb^) given A serves first minus (pa^ - pb^) given B serves first equals 0.0015, 0.0015, 0.0016 and 0.0014 for SS3, SS4, SS5 and SS6. These are smaller than in (e).

(m) The correlation coefficients are smaller in absolute value than those in (f).

Scoring System	First Server	(pa, pb)	Pa^	pb^	(pa^ - pb^) -(pa – pb)	Correlation (pa^, pb^)	P(W)	p(L)	P(W) - p(L) - (pa – pb)
SS1	A	0.65,0.65	0.6512	0.6478	0.0034	-0.2218	0.7047	0.5944	0.1103
SS1	B	0.65,0.65	0.6479	0.6512	-0.0033	-0.2210	0.7047	0.5944	0.1103
SS2	A	0.65,0.65	0.6535	0.6501	0.0034	-0.1747	0.7053	0.5984	0.1069
SS2	B	0.65,0.65	0.6501	0.6535	-0.0034	-0.1753	0.7053	0.5983	0.1070
SS1	A	0.7, 0.6	0.7101	0.5867	0.0234	-0.1415	0.7227	0.5741	0.0486
SS1	B	0.7, 0.6	0.7094	0.5917	0.0177	-0.1599	0.7236	0.5774	0.0462
SS2	A	0.7, 0.6	0.7103	0.5900	0.0203	-0.1195	0.7225	0.5779	0.0446
SS2	B	0.7, 0.6	0.7096	0.5949	0.0147	-0.1365	0.7234	0.5812	0.0422

Table 5: The bias results for a single set of tennis when (pa, pb) = (0.65, 0.65) and (0.7, 0.6)

(n) The proportion of points won on service by the winner minus the proportion won on service by the loser is slightly bigger than (pa^ - pb^) for SS3 to SS6.

The bias in a best-of-five sets match of tennis

Corresponding simulations of best-of-five sets matches were also carried out, but are not reported here due to space limitations. Suffice it to report that the various trends observed in going from one set to three sets continued. A few figures of interest however, are reported in the results and conclusions.

3. RESULTS

For B_{2n-1} (unipoints), the proportion of points won by player A, p^, is a biased estimator of p, the probability A wins a point. If $p > 0.5$, the bias is positive. It increases initially as n increases from 2, reaches a maximum, then decreases and converges to zero. When $p < 0.5$, the bias is negative.

When two or more B_{2n-1} unipoints systems are used, the bias was seen to decrease (in absolute value).

The bias when B_{2n-1} systems are nested was shown to be typically greater than without nesting.

The bias for the W_n system was seen to be greater in absolute value than for the B_{2n-1} system, but otherwise it had similar characteristics.

The bias in a single advantage game of tennis is greater than that of the B_{2n-1} system.

In bipoints systems the interdependency of pa^ and pb^ can be quite complex, as it can depend on who serves first, who is the better player, and the size of pa – pb. Stopping rules for sports typically lead to negative correlations for pa^ and pb^.

In a single tiebreak set of tennis when pa = 0.7 and pb = 0.6, the total bias, (pa^ – pb^) – (pa – pb), was equal to 0.020 when A served first, and 0.015 when B served first. The correlation coefficient between pa^ and pb^ was -0.120 (A first) and -0.137 (B first). P(W) – P(L) was 0.145 (A first) and 0.142 (B first).

For a single tiebreak set of tennis when pa = 0.65 and pb = 0.65, the total bias, (pa^ – pb^) – (pa – pb), was equal to 0.003 (A first) and -0.003 (B first). The correlation coefficient was -0.175 (A or B first). Also, P(W) – P(L) was 0.107 (A or B first), a substantial amount given that the players are equal.

For a best-of-three tiebreak sets match when pa = 0.7 and pb = 0.6, the total bias, (pa^ – pb^) – (pa – pb), was equal to 0.016 (A first) and 0.014 (B first).

The correlation coefficient between pa^ and pb^ was -0.071 (A first) and -0.074 (B first). P(W) – P(L) was 0.119 (A first) and 0.118 (B first).

For a best-of-five tiebreak sets match when pa = 0.7 and pb = 0.6, the total bias, (pa^ – pb^) – (pa – pb), was 0.012 (A or B first). The correlation coefficient between pa^ and pb^ was -0.040 (A or B first). Also, P(W) – P(L) was 0.112 (A or B first).

The various results for the best-of-three tiebreak sets match are less than the corresponding results for the single set, and all of the corresponding values for the best-of-five tiebreak sets match were smaller again.

4. DISCUSSION

The stopping rule in bipoints games such as tennis (and volleyball) has been shown to lead to an overall negative correlation between the performances of the two players/teams. The interdependent number of a-points and b-points played leads to this negative correlation. This negative correlation would not exist if the duration of the match involved a fixed number of points of each type. An example of this might be in the penalty shoot-out in soccer, where (at least initially) each team has a fixed number of penalty shots. Here the independence (no correlation) bipoints model would appear to be a reasonable first

Scoring System	First Server	(pa, pb)	pa^	pb^	(pa^-pb^) -(pa – pb)	Correln (pa^, pb^)	p(W)	p(L)	P(W) – p(L) - (pa – pb)
SS3	A	(0.65, 0.65)	0.6499	0.6491	0.0008	-0.2250	0.6826	0.6164	0.0662
SS3	B	(0.65, 0.65)	0.6491	0.6499	-0.0008	-0.2249	0.6826	0.6164	0.0662
SS4	A	(0.65, 0.65)	0.6512	0.6504	0.0008	-0.1844	0.6839	0.6177	0.0662
SS4	B	(0.65, 0.65)	0.6504	0.6511	-0.0007	-0.1846	0.6838	0.6177	0.0661
SS5	A	(0.65, 0.65)	0.6513	0.6507	0.0006	-0.1763	0.6836	0.6185	0.0651
SS5	B	(0.65, 0.65)	0.6506	0.6514	-0.0008	-0.1759	0.6836	0.6185	0.0651
SS6	A	(0.65, 0.65)	0.6514	0.6507	0.0007	-0.1180	0.6819	0.6203	0.0616
SS6	B	(0.65, 0.65)	0.6507	0.6515	-0.0008	-0.1177	0.6819	0.6203	0.0616
SS3	A	(0.7, 0.6)	0.7079	0.5902	0.0177	-0.0834	0.7099	0.5883	0.0216
SS3	B	(0.7, 0.6)	0.7079	0.5917	0.0162	-0.0901	0.7100	0.5896	0.0204
SS4	A	(0.7, 0.6)	0.7081	0.5921	0.0160	-0.0704	0.7101	0.5901	0.0200
SS4	B	(0.7, 0.6)	0.7081	0.5936	0.0145	-0.0732	0.7102	0.5914	0.0188
SS5	A	(0.7, 0.6)	0.7079	0.5924	0.0155	-0.0706	0.7097	0.5907	0.0190
SS5	B	(0.7, 0.6)	0.7079	0.5940	0.0139	-0.0736	0.7098	0.5921	0.0177
SS6	A	(0.7, 0.6)	0.7061	0.5950	0.0111	-0.0709	0.7071	0.5940	0.0131
SS6	B	(0.7, 0.6)	0.7060	0.5963	0.0097	-0.0727	0.7071	0.5952	0.0119

Table 6: The bias results for a best-of-three sets match when (pa, pb) equals (0.65, 0.65) and (0.7, 0.6)

approximation to the practical situation. Another example might be in rifle shooting when two (or more) shooters have a fixed number of shots.

5. CONCLUSIONS

The statistics for many sports matches are biased.

In sports such as squash it is widely accepted that only one type of point is really necessary for accurately modelling the sport. In such unipoints sports the expected value of the proportion of points won by the better player, E(p^), is greater than the probability that the better player wins a point, p, for all the commonly used scoring systems. Thus, if the proportion of points won by the better player, p^, is used as an estimator of p, it will tend to overestimate it. Further, if the proportion of points won by the winner (not necessarily the better player) is used as an estimate of the winner's probability of winning a point against that opponent, it will tend to be even more of an overestimate.

Sports where serving is an advantage and two types of point are considered necessary for modelling purposes (eg. tennis) are called bipoints sports. In bipoints sports the expected value of the proportion .of points won by the better player on service, E(pa^), is greater than the probability that the better player wins a point on service, pa, for the commonly used scoring systems. Also, the expected value of the proportion of points won by the weaker player on service, E(pb^), is less than the probability that the weaker player wins a point on service, pb. Thus, if pa^ is used as an estimator of pa, it will tend to overestimate it, and if pb^ is used as an estimator of pb, it will tend to underestimate it.

In bipoints sports such as tennis and volleyball, pa^ and pb^ are negatively correlated variables for many scoring systems. For a given sport and parameter values, the bias in a scoring system with a large expected duration is typically smaller than that in a scoring system with a smaller expected duration.

More specifically, in tennis the proportion of points won on service is typically biased upwards for the better player and downwards for the weaker player. The total effect of these two biases is about 0.014 for a best-of-three tiebreak sets match (when pa = 0.7 and pb = 0.6), and about 0.012 for a best-of-five tiebreak sets match (when pa = 0.7 and pb = 0.6). Although these are not substantial amounts, it would appear to be useful to know the typical sizes of these biases. It could be of marginal relevance in the gambling context.

In a best-of-three tiebreak sets match between two equal players (say pa = pb = 0.65), the proportion of points won on service by the eventual winner of the match is shown to be about 0.065 on average greater than the proportion of points won on service by the loser. For a best-of-five tiebreak sets match between these two equal players, this difference is shown to average about 0.049. It would appear that these are quite substantial amounts given that the two players are actually equal.

References

Brown, A., Barnett, T., Pollard, G. N., Lisle, I., & Pollard, G. H. (2008). The characteristics of various men's doubles scoring systems, *Proceedings of the ninth Australasian Conference on Mathematics and Computers in Sport*, Ed. J. Hammond, Tweed Heads, Australia, 52-59.

Carter, W. H. & Crews, S. L. (1974). An analysis of the game of tennis. *The American Statistician, 28(4)*, 130-134.

Miles, R. (1984). Symmetric sequential analysis: the efficiencies of sports scoring systems (with particular reference to those of tennis). *Journal of the Royal Statistical Society B. 46(1)*, 93-108.

Pollard, G. H. (1983). An analysis of classical and tie-breaker tennis. *Australian Journal Statistics, 25(3)*, 496-505.

Pollard, G. H. (1985). A statistical investigation of squash. *Research Quarterly for Exercise and Sport, 56(2)*, 144-150.

MODELLING MATCH OUTCOMES IN TEST CRICKET

Sohail Akhtar [a,b], Philip Scarf [a]

[a] *Centre for Operations Management, Management Science and Statistics, Salford Business School, University of Salford, Salford, M5 4WT, UK*
[b] *Corresponding author: s.akhtar@pgr.salford.ac.uk*

Abstract

This paper analyses match outcomes in test cricket. Our analysis is based on 146 international test matches, from November-2005 to April-2010. We model match outcome given the position at the end of each session. Match outcome probabilities are determined using multinomial logistic regression. Theses probabilities can facilitate a team captain or management to consider an aggressive or defensive batting strategy for the next session. We investigate how the outcome probabilities (win, draw, loss) vary session by session and how the covariate effects vary session by session. The covariates in the models include the score or lead, overs-remaining, run-rate, the pre-match strength of teams, batting resources, and home factor and toss outcome indicators. Our analyses suggest that lead has a small effect on the match outcome early on but dominates later. On other hand batting resources dominate throughout the match.

Keywords: Test cricket, multinomial logistic regression, strategy

WHAT WILL THEY SCORE? IN-GAME PROJECTION OF FIRST INNINGS FINAL SCORES IN LIMITED-OVERS CRICKET

Stern, Steven Edward [a,b]

[a] *The Australian National University*
[b] *Corresponding author: Steven.Stern@anu.edu.au*

Abstract

In this paper, we build upon the work of Lewis (2008) and combine the Duckworth-Lewis methodology, used primarily in setting revised targets in interrupted limited-overs cricket matches, with a credibility theory model to obtain a dynamic approach to estimating a team's final first innings total based upon the current state of play. The approach is valid for limited-overs matches of any duration, and can even be applied in circumstances where the first innings has been shortened during play due to weather or other interruptions. Finally, we use a database of international limited-overs matches to assess both appropriate values and structure for the credibility factor as well as the accuracy of the proposed projection method.

Keywords: Credibility theory, credibility factor, Duckworth-Lewis method, dynamic estimation

1. INTRODUCTION

During play in the first innings of a limited-over cricket match, it is natural to speculate on what the likely final total will be based on the current score. Accurate estimates in these circumstances, available dynamically as play progresses, can have beneficial impacts both on playing strategy as well as the entertainment value of the match for those watching. Indeed, as Bailey & Clarke (2006) demonstrate, predictions of first innings totals, as well as eventual margins of victory, are of wide interest; for example, as an aid in sports betting markets. However, current practice for broadcasters of limited-overs cricket matches is based simply on assuming various common, fixed run-rates per over for the remainder of the innings. However, as Lewis (2008) indicated, such an approach is unrealistic and unsatisfactory for a number of reasons, and he suggested ways in which the Duckworth-Lewis (D/L) methodology, currently used primarily in setting revised targets for interrupted matches, might also be used for various other calculations of interest, including first innings score projection.

Specifically, suppose that the team batting first in a limited-overs cricket match has scored s runs and lost w wickets with v out of their maximum of M overs still remaining. In order to project the likely total score at the completion of the innings, we first need to assess how much of the innings has been completed. This may initially seem a trivial question, as the "obvious" answer is proportion $p = (M-v)/M$ of the innings has been completed and, thus, proportion $1-p$ remains. This naïve and overly simplistic definition of the remaining portion of the innings, based only on number of overs, is the basis for the simplest and most common current score projection technique, which simply posits a (fixed) rate, r, of runs scored per over for the remaining portion of the innings, so that the projected score is:

$$P_{run-rate} = s + (1-p)Mr = s + vr\ .$$

However, Duckworth & Lewis (1998, 2004) demonstrated clearly that focusing solely on the number of overs played to determine the effective proportion of the innings that has been completed is not appropriate. Instead, the fundamental principle of the Duckworth-Lewis methodology is based around "scoring resources", which take into account not only the proportion of overs remaining to a batting side, but also the number of wickets still available. Indeed, it is quite clear that a batting side that has already lost 9 of their available 10 wickets is nearly finished with their innings, no matter how many overs they may have remaining. Accordingly, we define $R(v,w)$ to be the Duckworth-Lewis scoring resources remaining to a batting side when they have v overs of their innings remaining and have lost w wickets (for details regarding the calculations of Duckworth-Lewis resource values, see Duckworth & Lewis, 2004). We can then drastically improve our projections over those based on the naïve runs per over approach by considering instead a rate of runs *per unit of resource*, u, over the remaining scoring resources as

$$P = s + \rho u R(M,0) = s + uR(v,w),$$

where $R(M,0)$ represents the scoring resources available to the batting side at the beginning of their innings (when the maximum M overs are still available and no wickets have been lost) and $\rho = R(v,w)/R(M,0)$ is the proportion of the original available scoring resources that still remain to the batting side when the projection is to be calculated. In fact, once we have recognised the problem as one pertaining to scoring resources rather than simply to overs, we can replace $R(v,w)$ and $R(M,0)$ by R_{rem} and R_{max}, the resources currently remaining to the batting side and the maximal resources available to the batting side, which now may have been calculated according to the Duckworth-Lewis methodology to reflect any shortenings to the innings made necessary by weather or other interruptions. In this case, the score projection becomes

$$P = s + \rho u R_{max} = s + uR_{rem}.$$

2. A CREDIBILITY THEORY BASED PROJECTION METHOD

One natural choice for u is the observed scoring rate, $u_{obs} = s/(R_{max} - R_{rem})$. Using this choice produces the so-called "direct scaling" projection:

$$\begin{aligned} P_{direct} &= s + u_{obs}R_{rem} = s + \frac{sR_{rem}}{R_{max} - R_{rem}} \\ &= s\left(1 + \frac{R_{rem}}{R_{max} - R_{rem}}\right) = \frac{sR_{max}}{R_{max} - R_{rem}} \\ &= u_{obs}R_{max}. \end{aligned}$$

This direct scaling approach often performs quite well in accurately predicting the final score, especially if it is applied after a reasonable proportion of the innings has been completed. However, the direct scaling approach is highly sensitive to extreme scoring rates over short periods. That is, if a batting team scores a very high (or very low) number of runs during the early portion of their innings, corresponding to the use of a relatively small proportion of their available scoring resources, the resulting direct scaling projection of their final score can be wildly unrealistic. Indeed, as a simple example, if a batting team has scored no runs in the first couple of overs, the direct scaling score projection will necessarily be zero, which is nonsensical. Moreover, this issue has been magnified in recent times with the explosion of Twenty20 cricket, a game in which short, extremely high bursts of scoring are commonplace.

Alternatively, "expert opinion" might be used to determine an appropriate choice of u, potentially taking into account a range of match-specific conditions. While this approach may seem appealing, it has the clear disadvantage of requiring unique, sometimes almost intangible inputs for each match. Moreover, "expert opinion" is notoriously variable, and may well be locally influenced (if only sub-consciously) by the observed scoring rates. The vagaries of "expert opinion" might be removed by a more systematic modeling of first innings totals using match-specific factors, such as that carried out by Bailey & Clarke (2006). However, while such an approach certainly reduces the subjectivity in the projections, it adds a degree of complexity which makes it difficult to implement generically. Moreover, such an approach may fail to adequately account for the highly relevant information available in the current score of the actual innings for which a projection is being made. Therefore, in what follows, we choose to set u according to a credibility theory model, which takes the form

$$u = (1-\pi)y + \pi z,$$

where $0 \le \pi \le 1$ is a weighting factor, y is an estimate of the future scoring rate of the innings which is based on a relatively small amount of directly relevant information and z is an estimate of the future scoring rate that is based on a relatively large amount of perhaps only indirectly relevant information (see Bulmann & Gisler, 2005, among others, for an introduction to credibility theory).

Employing this approach for first innings score projection, it is straightforward to set $y = u_{obs}$, as this is plainly the most directly relevant information regarding the plausible scoring rate for the remainder of the innings. The choices for π and z are somewhat more equivocal. As noted above, choosing a value for z based on expert opinion or on models including match-specific characteristics has merit, but will not facilitate a simple, automatic process and is subject to the substantial problem of disagreement among "experts" and various model structures (nevertheless, a potential area of further research would be the inclusion of these approaches into the determination of z). So, in what follows, we shall set z equal to the overall average of completed first innings scores from recent international 50-over matches. This value was denoted G_{50} by Duckworth & Lewis (1998), and is currently set at 245. [NOTE: the use of the average score in 50-over matches is necessitated by the fact that D/L scoring resources are calibrated so that $R(50,0) = 1$; that is, a full unit of scoring resources corresponds to the resources associated with a 50-over innings.] Finally, based on a heuristic argument outlined in the Appendix, we choose a weighting factor of the form

$$\pi = (R_{rem}/R_{max})^{k-1},$$

for some appropriately chosen value of $k \geq 1$. This form of the weighting factor ensures that as the resources remaining, R_{rem}, diminish, so does the weighting factor, with the result that u becomes increasingly weighted to the observed scoring rate. This aspect of the weighting factor appropriately reflects the fact that as an innings draws to a close, the observed scoring rate is an increasingly accurate reflection of the actual final scoring rate.

Using this approach, we combine the above choices to construct a credibility estimate for the scoring rate during the remainder of the innings:

$$u = \left\{1 - \left(\frac{R_{rem}}{R_{max}}\right)^{k-1}\right\} u_{obs} + \left(\frac{R_{rem}}{R_{max}}\right)^{k-1} G_{50},$$

resulting in the first innings projection formula:

$$P = P_{direct}\left\{1 - \left(\frac{R_{rem}}{R_{max}}\right)^{k}\left(1 - \frac{G_{50}}{u_{obs}}\right)\right\}, \qquad (1)$$

provided $u_{obs} > 0$ [otherwise, the projection formula becomes $P = (R_{rem}/R_{max})^k(R_{max}G_{50})$].

The projection formula (1) shows that the approach proposed here constitutes either a damping (if $u_{obs} > G_{50}$) or an enhancement (if $u_{obs} < G_{50}$) of the direct scaling target. Moreover, the degree of damping or enhancement is decreased as the proportion of available resources already used increases. Similarly, the degree of the damping or enhancement decreases as the observed scoring rate approaches the overall average score of G_{50}. Figure 1 shows the degree of damping or enhancement for various scoring rates and resources remaining using $k = 2$. The damping and enhancement occur in the upper corners, where the proportion of scoring resources remaining is high. Damping occurs in the upper right corner, where scoring rates are very high, and enhancement in the upper left corner, where scoring rates are well below average.

Of course, the degree of damping and enhancement shown in Figure 1 will vary according to the value of the tuning constant, k, and so an appropriate value of this parameter must be determined. Before discussing that choice in detail, we note that choosing $k = 1$ amounts to projecting the final score by assuming the remainder of the innings will see the batting team score at the overall average resource usage rate (i.e., a usage rate corresponding to an overall score of G_{50}), while as k increases, the projection approaches the direct scaling method. Thus, as k increases, the damping and enhancement plot would become increasingly flat.

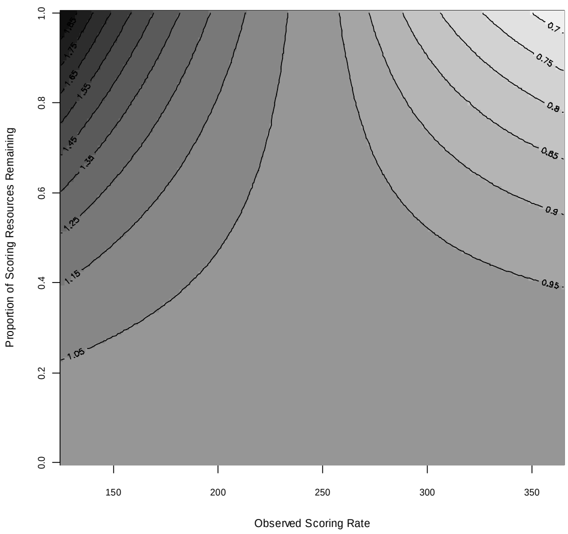


Figure 1: Contour plot of the degree of damping or enhancement of P_{direct} (using $k = 2$).

3. CHOOSING THE TUNING PARAMETER

In order to assess the most appropriate choice for the tuning parameter, k, we investigate a database of 574 international cricket matches played between June 2005 and December 2009. The collection consists of a mixture of 50-over One-Day International (ODI) matches and 20-over Twenty20 International (T20I) matches. For each match, the final first-innings score is recorded, as well as the scores when any multiple of 5 overs was remaining. For example, for a standard ODI, the scores at 5, 10, 15, 20, 25, 30, 35, 40 and 45 overs were recorded, while for a T20I, the scores at 5, 10 and 15 overs were recorded.

Overall, the collection included 101 full T20I first innings, 451 full ODI first innings and 22 ODI matches which were shortened to between 22 and 49 overs at their outset. In total, this created a set of 4,981 partial-innings scores (including the 0 runs for 0 wickets scores at the start of each match) from which projections could be made and subsequently assessed for predictive accuracy. One reason for using partial-innings scores at 5-over intervals is to reduce any dependency between scores from the same match. In the analysis that follows, we proceed as if all 4,981 partial-innings scores represent independent pieces of information.

We start by presenting some descriptive statistics regarding the database of matches. Table 1a shows the average (standard deviation in brackets) scores at various stages of the T20I first innings, and Table 1b shows the same for full-length ODI first innings.

Overs Played	Wickets Down	Runs Scored
5	1.25 (1.10)	38.62 (10.80)
10	2.59 (1.46)	74.58 (17.32)
15	4.32 (1.59)	113.06 (25.96)
20	N/A	157.76 (35.04)

Table 1a: Average scores (standard deviations in brackets) at various stages of 101 T20Is. [NOTE: Wickets down at completion of innings not available in dataset.]

Overs Played	Wickets Down	Runs Scored
5	0.67 (0.77)	19.83 (8.64)
10	1.28 (1.02)	43.23 (14.55)
15	1.97 (1.30)	66.34 (19.11)
20	2.62 (1.47)	88.57 (23.57)
25	3.17 (1.59)	109.92 (27.14)
30	3.64 (1.74)	132.51 (30.76)
35	4.11 (1.79)	156.73 (34.21)
40	4.68 (1.70)	184.08 (38.68)
45	5.60 (1.61)	216.72 (45.02)
50	N/A	244.09 (63.87)

Table 1b: Average scores (standard deviations in brackets) at various stages of 451 full ODIs. [NOTE: Wickets down at completion of innings not available in dataset.]

Initially, we estimate the value of k using a simple least-squares approach. Specifically, for each of the 4,981 partial-innings scores in the dataset, we calculate the difference between the actual final score, S, and the projected final score as a function of k, say $P(k)$. That is, for the i^{th} partial-innings score in the dataset, calculate $P_i(k)$, the projected score based on the choice of value k, according to the projection formula (1). Then choose the tuning parameter as the value of k that minimises

$$d(k) = \sum_{i=1}^{4981} \{P_i(k) - S_i\}^2 .$$

This approach yields an estimate of k = 1.6046. (Solutions to this and all subsequent minimization problems were carried out using the `nlmin` function within the computer package S-Plus). However, as Figure 2 demonstrates, the accuracy of a projection will be inversely related to the amount of scoring resources remaining at the time the projection was made.

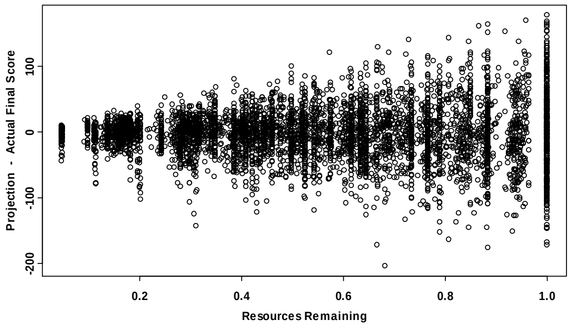


Figure 2: Difference between projections, $P_i(1.6046)$, and actual final score versus resources remaining at time of projection.

To account for this heterogeneity, we use a weighted approach to estimating the value of k, appropriately accounting for differential accuracy of predictions. So, we use the alternative objective function

$$d_w(k,\gamma,\sigma) = \sum_{i=1}^{4981} \left[\left\{ \frac{P_i(k) - S_i}{\sigma R_{rem,i}^{\gamma}} \right\}^2 + \ln(\sigma^2 R_{rem,i}^{2\gamma}) \right],$$

where $R_{rem,i}$ is the resources remaining associated with the projection $P_i(k)$. The form of this function is based on a weighted normal log-likelihood applied to the runs remaining to be scored; that is, the values $S_i - s_i$ are assumed to be normally distributed with mean $P_i(k) - s_i$ and standard deviation $\sigma R_{rem,i}^{\gamma}$.

The chosen structure for the standard deviation is based on the increasing spread observed in Figure 2, and also allows the estimation of a generic spread parameter, σ, which equates to the accuracy of a score projection at the start of a 50-over innings, $P = G_{50} = 245$. Minimising $d_w(k,\gamma,\sigma)$ yields the normal-based estimates and standard errors of k, γ and σ shown in Table 2 below. In addition, Figure 3 presents a plot of weighted differences between projections and actual scores

$$e_i = \frac{P_i(k) - S_i}{\sigma R_{rem,i}^{\gamma}}$$

using the estimates of k, γ and σ derived, and suggests that the weighting scheme adopted has remedied the heterogeneity problem arising when un-weighted least squares was used. However, while Figure 3 appears more homoscedastic, there is still an issue when resources remaining are small.

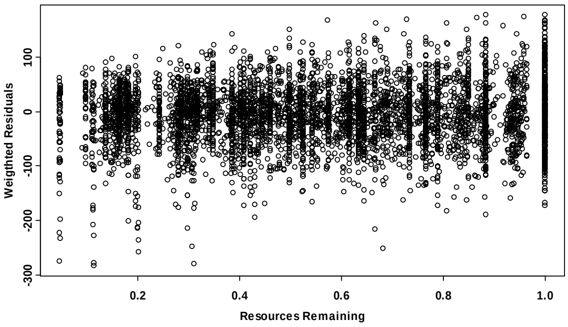


Figure 3: Weighted difference between projections and actual final score, e_i, versus resources remaining at time of projection.

An investigation of the partial-innings scores associated with those matches with weighted residuals in the lower left corner of Figure 3 revealed that they were associated with matches in which either the ninth or tenth wicket partnership produced an unusually high number of runs. This circumstance highlights a key problem with a normal-based approach. While the runs yet to be scored in a match will likely be reasonably symmetrically distributed around its average value when a reasonable amount of the innings still remains to be played, when the innings is nearly complete, the distribution of runs yet to be scored tends to be skewed. This phenomenon arises because runs yet to be scored cannot be negative, and when 8 or 9 wickets have already been lost, it is likely that only a few more runs will be scored, yet in a small number of cases the last batting few partnerships will produce a large number of runs.

In response to this problem, we consider an alternative distribution for the values $S_i - s_i$, using a gamma distribution rather than a normal distribution, but maintaining the same mean and standard deviation structure. Doing so yielded the maximum likelihood estimates and standard errors for the parameters shown in the final column of Table 2.

Parameter	Normal-based Estimates	Gamma-based Estimates
k	1.4858 (0.0360)	1.6781 (0.0389)
γ	0.5917 (0.0141)	0.7059 (0.0138)
σ	55.49 (0.80)	59.95 (0.85)

Table 2: Parameter estimates (standard errors in brackets) using normal- and gamma-based models. [NOTE: Standard errors calculated as the square-root of the diagonal entries of the inverse of the Hessian matrix of the log-likelihood function evaluated at its maximum; that is, the observed Fisher information.]

The estimates from the two different methods are reasonably similar (although, the differences are statistically significant). However, the gamma-based estimate for σ is more closely in line with the observed standard deviation of 63.87 for first innings scores in full 50-over matches (see Table 1b). As such, we prefer this method.

4. CONCLUSION

The proposed projection formula fits the observed data well, and using our database, we have arrived at a projection formula of the form:

$$P = P_{direct}\left\{1-\left(\frac{R_{rem}}{R_{max}}\right)^{1.6781}\left(1-\frac{245}{u_{obs}}\right)\right\}$$
$$= u_{obs}R_{max}\left\{1-\left(\frac{R_{rem}}{R_{max}}\right)^{1.6781}\left(1-\frac{245}{u_{obs}}\right)\right\}$$

This method is based on the direct projection approach, which works well in a wide range of circumstances, but also behaves more stably in situations where the observed scoring rate is extreme over a relatively short period.

As an independent assessment of the accuracy of the method, projected first innings totals were calculated for a new set of 44 matches (8 T20Is and 36 ODIs) played between January and April 2010. Projections were again made from each partial-innings score at the end of overs which were multiples of 5. This yielded 345 predictions which could be assessed for accuracy. Figure 4 shows the histogram of the differences in the projections and the actual final first innings scores for these 345 predictions. In addition, Figure 4 shows the histogram for projections based on direct scaling, P_{direct}.

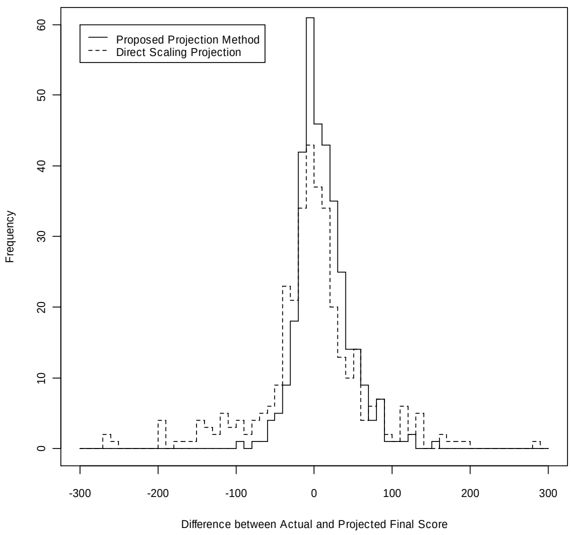


Figure 4: Histograms of differences between 345 actual and projected final first innings scores.

As can be seen from Figure 4 and from Table 3, the projection method proposed here outperforms the direct scaling method. On average, the projection method proposed is within 25 runs of the actual final score, nearly twice as accurate as direct scaling (the actual average final scores for the 8 T20I matches was 131.88 with a standard deviation of 37.28 and for the 36 ODI matches was 258.36 with a standard deviation of 52.53).

Projection Method	Average Absolute Deviation	Standard Deviation of Differences
P	24.45	32.37
P_{direct}	44.46	67.07

Table 3: Summary of 345 projections from 44 new matches.

Finally, the likelihood framework outlined in the previous section provides a ready methodology for attaching a likely range to our projected total. In particular, the standard deviation of runs yet to be scored is estimated as $59.95R_{rem}^{0.7059}$, so there is an approximate 70% chance that the final score lies in the range $P \pm 59.95R_{rem}^{0.7059}$. Alternatively, we can use appropriate gamma-based quantiles to derive more precise upper and lower confidence bounds for the final score projection. For the 345 projections from the 44 matches played between January and April 2010, 76.81% of the ranges so calculated actually covered the true final first innings score. This is in close agreement with the theoretical value of 70%, the excess coverage in part explained by the inter-dependence of projections from the same match.

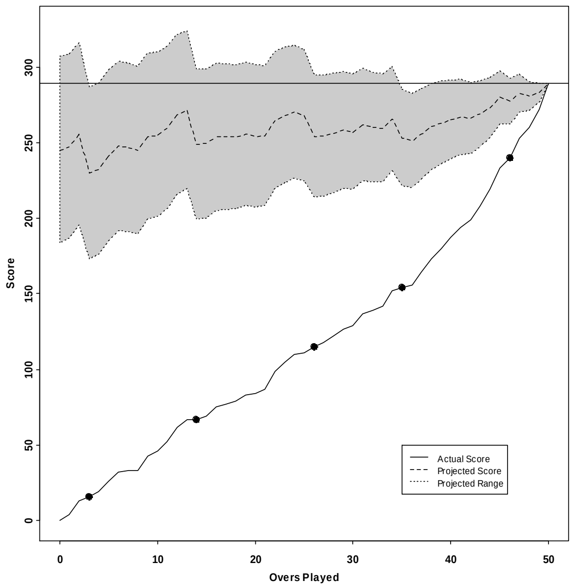


Figure 5a: Progression of actual and projected score throughout the Sri Lankan innings during their 2007 Cricket World Cup semi-final match against New Zealand.

We conclude by indicating one way our projection procedure can be used to enhance the entertainment and information value of live limited-overs cricket broadcasts. Figures 5a and 5b show the run-scoring progression from the semi-finals of the 2007 Cricket World Cup, and are based on plots shown during live cricket broadcasts, whereby the score through an innings is depicted as an increasing "worm" with dots at the fall of wickets. Figure 5a shows the first innings of the match between Sri Lanka and New Zealand, in which Sri Lanka batted first, Figure 5b the first innings of the match between Australia and South Africa, in which South Africa batted first. The addition to the common display in these plots is the inclusion of the projected score and likely range, based on our proposed methodology, above the run-scoring "worm". The display shown here is for the entire innings; however, at any stage of play, the plot can be produced up to the current over.

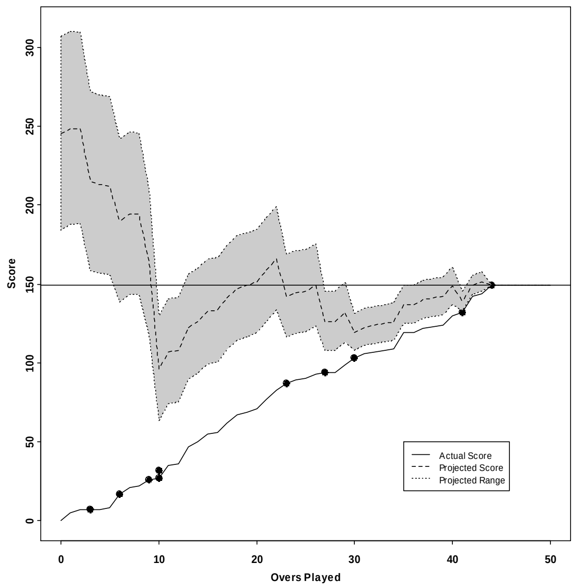


Figure 5b: Progression of actual and projected score throughout the South African innings during their 2007 Cricket World Cup semi-final match against Australia.

Note that the variation in the projections can also serve as an indication of which team is currently performing better in the game; that is, whether the batting side has begun to improve their performance or the bowling side has begun to rein them in.

Acknowledgement

I wish to thank Frank Duckworth and Tony Lewis for their support and ongoing conversations regarding their methodology.

Appendix

A simple heuristic argument using geometric series expansion shows why choosing

$$\pi = (R_{rem}/R_{max})^{k-1}$$

is appropriate. The direct scaling projection is:

$$P_{direct} = s\left(\frac{R_{max}}{R_{max}-R_{rem}}\right) = s\sum_{i=0}^{\infty}\left(\frac{R_{rem}}{R_{max}}\right)^{i}$$

$$= s + s\left(\frac{R_{rem}}{R_{max}}\right) + s\left(\frac{R_{rem}}{R_{max}}\right)^{2} + \cdots$$

Thus, a simple way to adjust the direct scaling method to reduce the effect of extreme scoring rates is to replace s by $(R_{max} - R_{rem})G_{50}$, which is equivalent to replacing u_{obs} by G_{50}, after a certain point in the infinite series expansion:

$$P = s + s\left(\frac{R_{rem}}{R_{max}}\right) + \cdots + s\left(\frac{R_{rem}}{R_{max}}\right)^{k-1}$$

$$+ (R_{max} - R_{rem})G_{50}\left(\frac{R_{rem}}{R_{max}}\right)^{k} + \cdots$$

$$= s + s\sum_{i=1}^{k-1}\left(\frac{R_{rem}}{R_{max}}\right)^{i}$$

$$+ (R_{max} - R_{rem})G_{50}\sum_{i=k}^{\infty}\left(\frac{R_{rem}}{R_{max}}\right)^{i}$$

$$= s + s\left\{\frac{\frac{R_{rem}}{R_{max}} - \left(\frac{R_{rem}}{R_{max}}\right)^{k}}{1-\frac{R_{rem}}{R_{max}}}\right\}$$

$$+ (R_{max} - R_{rem})G_{50}\left\{\frac{\left(\frac{R_{rem}}{R_{max}}\right)^{k}}{1-\frac{R_{rem}}{R_{max}}}\right\}$$

$$= s + \frac{s}{R_{max}-R_{rem}}\left\{R_{rem} - R_{rem}\left(\frac{R_{rem}}{R_{max}}\right)^{k-1}\right\}$$

$$+ R_{rem}G_{50}\left(\frac{R_{rem}}{R_{max}}\right)^{k-1}$$

$$= s + R_{rem}u_{obs}\left\{1-\left(\frac{R_{rem}}{R_{max}}\right)^{k-1}\right\}$$

$$+ R_{rem}G_{50}\left(\frac{R_{rem}}{R_{max}}\right)^{k-1}$$

$$= s + R_{rem}\{(1-\pi)u_{obs} + \pi G_{50}\}$$

Clearly, the later in the expansion the replacement takes place, the closer the projection is to the direct scaling projection.

References

Bailey, M. & Clarke, S.R. (2006). Predicting the match outcome in one day international cricket matches, while the game is in progress. *Journal of Sports Science and Medicine, 5,* 480-487.

Bulmann, H. & Gisler, A. (2005). *A course in credibility theory and its applications*. Springer-Verlag.

Duckworth, F. C. & Lewis, A. J. (1998). A fair method for resetting the target in interrupted one-day cricket matches. *The Journal of the Operational Research Society, 49,* 220-227.

Duckworth, F. C. & Lewis, A. J. (2004). A successful operational research intervention in one-day cricket. *The Journal of the Operational Research Society, 55,* 749-759.

Lewis, A. J. (2008). Use of the Duckworth/Lewis methodology to provide additional benefits for one-day cricket. *Proceedings of the 9th Annual Australasian Conference on Mathematics and Computers in Sport*. (ed. John Hammond). ANZIAM.

DATA MINING IN PROFESSIONAL CRICKET FOR PERFORMANCE ENHANCEMENT

Abid Ghaffar [a,b] and Ehtesham Taha [a]

[a] *University of Engineering & Technology Lahore,*
Alkhawarzmi Institute of Computer Sciences
[b] *Corresponding author: abidghaffar@gmail.com*

Abstract

Data mining can be viewed as the process of automatically extracting previously unknown information from large databases and utilizing this information to make crucial business decisions. In Data mining data is looked for patterns and relationships which later on lead to some previously unknown facts, i.e. knowledge. The sports world is known for the vast amounts of statistics that are collected for each player, team, game, and season. Since the processing of data is much slower than its generation, this information is not useful in most of the cases; especially for the players or even teams of limited resources. It is virtually impracticable to get any advantage from that data. Using data mining, this raw data could be converted into wealth of knowledge. A study was conducted wherein Cricket, which is also known as game of records, was taken as the model sport. Ball to ball data of a batsman was mined to prepare a decision tree. This decision tree highlighted the strong & weak areas of that batsman. Using this decision tree, that batsman could be suggested to improve his weak areas. On the other hand, this decision tree could also be used for selection of a batsman against any particular team or situation. The results of this study were presented to stakeholders in Pakistan Cricket Board (PCB), who validated and appreciated the results.

Keywords: Data Mining, Classification, Decision Tree, Cricket, Performance Enhancement

1. INTRODUCTION

Data mining is the blend of some tools and techniques for the exploration of knowledge from huge amount of data (Han, Jiawei and Micheline Kamber, 2006). In Data mining data is looked for patterns and relationships which later on lead to some previously unknown facts, i.e. knowledge. Moreover the future events can be predicted from these patterns and relationships. Areas where the presence of Data mining can be felt includes industrial applications, criminal investigations, bio-medicine (Chen et al, 1998), sports (K. Solieman, Osama 2006) and counter-terrorism (Bhavani Thuraisingham).

Most customers when they go out for shopping normally follow a similar pattern. It has been seen that when a customer purchases milk & bread, he is likely to purchase butter. It’s the data mining which disclose these patterns. On the basis of this knowledge recommendations to the vendor/customer can be given.

Another area where loads of data is generated and gathered is the world of sports, where data is collected for players, their teams, season and venue where the matches have been played. The statistics that are gathered for a cricket player might be – runs scored, wickets taken, centuries made, batting average maintained, batting strike rate maintained, bowling average accomplish, bowling strike rate achieved etc for each match.

The level of competition is high in the world of sports, the difference of skills and temperament is very low between the international teams so a slight edge normally goes a long way. By tradition it was considered that the experts of the game are the repository of all the knowledge regarding that game, experts might be scouts, coaches or even managers.

Considering the significance of historical data in the field of sports and games, it is imperative to find the patterns in that data which can lead to new knowledge. This could extremely helpful not only for the player and their coaches but also for managers and the contractors of the teams who offer expensive contracts to the players.

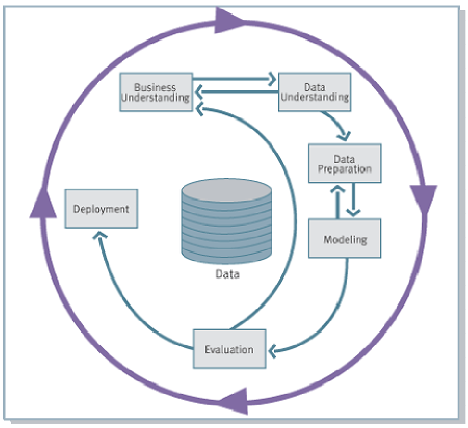
In this study Classification technique is proposed for mining and knowledge extraction from sports data.

2. METHODOLOGY

Since data mining is a growing area, the techniques are constantly changing, as new improved methods are discovered. Decision tree classification can find out interesting trends from historical data of any cricket player. Decision tree classification has certain distinct qualities which are listed in (Borisov, A., Chikalov, I., Eruhimov, V., & Tuv, E. (2005), (Rose Quinlan 1996) and (SK Murthy (1998)).

The Life Cycle of any data mining application consists of following six phases (Crisp).

1. Business understanding
2. Data understanding
3. Data preparation
4. Modelling
5. Evaluation
6. Deployment



Business understanding

Major concern of this study is towards the performance enhancement of batsman using data mining. So ball to ball data about a batsman batting will be collected.

Data understanding

Before processing it's imperative to understand the data. There must be a domain expert there, who would be able to set direction of data mining expert by pointing out what are most important pieces of data and how does they relate to each other. Because selection of relevant attributes is crucial to Data mining success, without adequate understanding of data, the return on the resources invested in data mining would certainly be disappointing. In quest of finding the most important attributes that deals with the performance of a batsman we consulted experts of the game and finally concluded the following attributes

Bowler Style

To keep things simple and easy to understand, bowlers a batsman faced were categorized as fast, medium and spinner. This categorization is important because some batsmen have good technique against faster bowlers but they are vulnerable against the spinner and vice versa.

Foot

There are distinct shots which can be played on front foot as well as back foot. It can be a good measure of players technique as it is seen that some batsman hesitate to play on the front foot and others fails because they commit themselves on back foot or on front foot too early.

Ball Length

All the balls were categorized on the length. This attribute could have the values like short pitch, good length, fuller length or even full toss. Choice of attribute is obvious as there are only few players who can play the balls of every length with ease.

Ball Line

Ball line is another judge of batsman's technique. Some players are strong on the leg side and others feel it comfortable to play on the off. Earlier days of the Indian former captain Sourav Ganguly is an example, who played almost every ball on the off side, conversely former kiwi Captain Stephen Fleming when he came to bat used to play every ball on the leg side.

Shot Played

What shot the player is playing. This variable is categorized as following

- Beat
- Cover Drive
- Cut, Edge
- Flick
- Glance
- Left
- Lifted Drive
- Off Drive
- On Drive
- Pull
- Push
- Reverse Sweep
- Smashed
- Square Drive
- Stop
- Straight Drive
- Sweep
- Thump

Result

A variable named 'Result' was added to the above data. This variable is the output (dependent) variable. This variable has only two values i.e. 'convincing' or 'not convincing'. If the batsman has scored a run, it is marked as 'convincing' and if he hasn't scored any 'not convincing'.

As it is not possible to infer from the available data (Cricinfo) that whetter or not batsman has played a convincing shot or not, runs scored was used to infer that whether or not the shot played was 'convincing' or 'not convincing'

Data preparation

Data collection and preparation for the evaluation of the proposed model was not an easy task. Ball to ball data of a batsman was required. Keeping this issue in mind major sports organizations were consulted and data from multiple sources especially from (CricInfo) was obtained and integrated. A Pakistan test cricketer Imran Nazir was selected as a model batsman. The batting data of Imran Nazir for season 2006-07 was collected, transform & integrated.

The shape of our final data will look like.

Bowler style	foot	Ball length	Ball Line	shot	Result
FAST	BACK	Full Length	OFF	LEFT	Convincing
FAST	LEFT	Full Length	OFF	LEFT	Convincing
MED	LEFT	Good Length	ON DRIVE	LEG	Not Convincing
IUM					
FAST	LEFT	Short Pitch	OFF	LEFT	Convincing
FAST	LEFT	Full Length	LEG	LEFT	Not Convincing
FAST	LEFT	Full Length	OFF	LEFT	Convincing
FAST	LEFT	Short Pitch	LEG	PULL	Convincing
FAST	LEFT	Good Length	LEG	ON DRIVE	Convincing

Modelling

In this phase, many data mining modelling techniques are selected and applied, and their parameters are calibrated to most favourable values. Typically, for the same data mining problem, several techniques might work. Stepping back to data preparation phase is sometimes required because of the difference in the format of data used in different techniques.

For this particular study, we selected decision tree modelling approach.

Evaluation and Validation

How well your mining models have performed against real data is accessed in the process of Validation. It is imperative that you authenticate your mining models by considering their quality and characteristics before their deployment into the production environment (MSDN). Following metrics are used to access the quality of a data mining model.

Accuracy is a measure of how well the model correlates an outcome with the attributes in the data that has been provided. The accuracy of these rules was checked with the help of confusion matrix (Kohavi and Provost, 1998). A confusion matrix is built as a comparison of actual values that exist in the testing dataset against the values that the mining model predicts.

Usefulness deals with a variety of metrics that tell you whether the model provides useful information.

To evaluate the proposed methodology, dataset was divided into two partitions, training subset and test subset. First a decision tree was build using training subset and its accuracy was verified using test subset. After this the results are discussed with a domain expert to gauge the usefulness of results.

Deployment

Creation of the model is usually not the end of the project. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process (Crisp).

In this particular case, we were only interested in getting a decision tree which would highlight the strong and weak areas of a batsman.

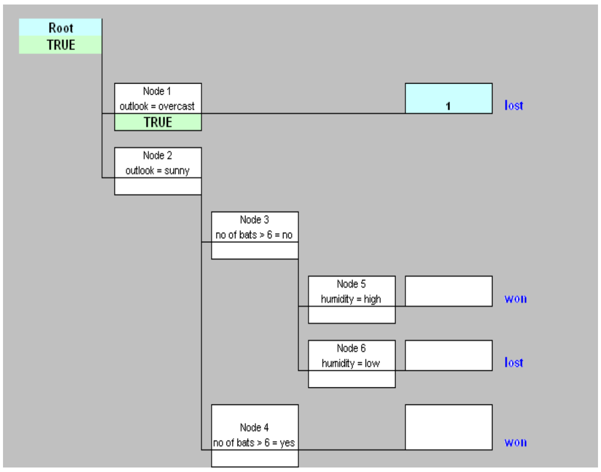
Example

To understand the usefulness of decision trees in cricket world, consider the start of cricket match where two captains are involved in a toss. The winning captain has the option to decide whether to bat or field first. It's a difficult decision on his part as sometimes this decision plays a major role in the outcome of a match. Let's suppose he decides to collect the data of last 10 matches where the captains decide to bat first after winning the toss. The data set that has been chosen for this table comprises of Humidity, Outlook and the number of regular batsmen in the team.

Following Table showing the hypothetical data of last ten games where the winning captain decided to bat first after winning the toss.

Independent Variables			**Dependent Variable**
Outlook	Humidity	Number of batsmen > 6	Final Outcome
Sunny	High	Yes	Won
Overcast	High	No	Lost
Sunny	Low	No	Lost
Sunny	High	No	Won
Overcast	Low	Yes	Lost
Sunny	Low	Yes	Won
Sunny	Low	No	Lost
Sunny	High	No	Won
Sunny	Low	Yes	Won
Sunny	Low	Yes	Won

Though it's an example containing very few records but still it is apparently confusing for captain to figure out what to do on seeing it in the above format. Let's see how a decision tree model can help the captain in understanding the situation. To make the decision tree of the above data, we proceed as follows:



An equivalent rule solution is:

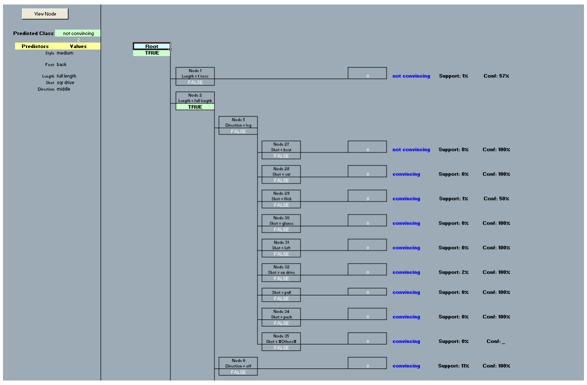
If outlook = Overcast then
Lost
Else if outlook = sunny then
If humidity = high then
Won
Else if humidity = low then
If batsmen > 6 then
Won
Else
Lost
End if
End if
End if

From this decision tree the captain and his other fellows come to know historically and statistically that it is better to put opposition in for bat if the day is overcast and conversely If the day is sunny and humidity is high batting first would have been a better option and finally. If the day is sunny, with low humidity, and the team contains at least six regular batsmen, batting first is safe; otherwise they

should field. Hence, the captain can now make a knowledgeable decision after winning the toss.
This is basically a very simple example to illustrate the use of data mining in sports. By looking at the proposed example, a captain can easily predict that the result of batting first is not going to be favourable if the Outlook is overcast. On the other hand, there is fair chance of winning if you opt to bat first after winning the toss on a day when humidity is high and the outlook is sunny.

3. IMPLEMENTATION AND CASE STUDY

Classification Software was designed in Excel that utilized the standard C4.5 algorithm to generate decision tree. Excel was selected due to its intuitive user interface and ease of use. Its very easy for a novice user to input, edit & manipulate data in excel, while simultaneously able to mine it. Ball to ball detailed data of the Imran Nazir was feed to this software which generated the following tree.



From the detailed study of the tree which is shown above is pruned and made on C4.5 model, we can come across the following conclusion about the technique of the batsman ,which in our case is Imran Nazir.

If length = "full toss"
    Not convincing
Elseif length = "full length"
    If direction = "off side"
        Convincing
    Elseif direction = "middle stump"
        Not convincing
    End if
Elseif length= "good length"
    If bowler = fast and direction = "off stump" and shot tried = "cover drive"
        Convincing
    Else
        Not convincing
    End if;
    If shot = "cover drive" and bowler = "medium fast"
        Not convincing
    Else if shot = "cover drive" and bowler = "spin"
        Convincing
    End if
    If shot = "cover drive"
        Convincing
    End if
End if

If we go through above generated rules, we will find discover following.

- Imran Nazir feels comfortable while playing
    - "Full Length" balls that come on "off side"
    - "Cover Drive" to "Good Length" balls especially for "slow" bowlers.

- Imran Nazir feels difficulty while playing
    - "Full toss" balls.
    - "Full Length" balls that come on "middle stamp"
    - "Cover Drive" to "Good Length" balls "Medium fast" bowlers.

These rules can help in following ways.

- Coach can use these rules while planning practise session, training for batsman. He can give more attention to grey areas and resultantly the performance of batsman would improve.
- The captain of the team can use these rules to select the best batsman for the current situation during a match.
- The opponent team can use these rules to exploit the weak areas of a batsman.
- Keeping in view the weak and strong areas of other team batsman, selection of best bowlers can be made.

These rules were presented to Imran Nazir & other stake holders in Pakistan Cricket Board (PCB), who validated and appreciated the results.

Confusion Metrics for the Case Study are as under:

Training Data

True Class	Predicted Class: convincing	Predicted Class: not convincing	
convincing	3156	24	3180
not convincing	38	1003	1041
	3194	1027	4221

Test Data

True Class	Predicted Class: convincing	Predicted Class: not convincing	
convincing	35	1	36
not convincing	1	13	14
	36	14	50

- On Training Data the class "Convincing" is miss-Classified only 24 times and class "Not Convincing" is miss-classified 38 times. So miss-classification is 1.47% on Training Data.

- On Test / Validation Data both of the classes have been miss-classified once and the miss-classification is mere 4.00 %.

4. CONCLUSIONS

Although traditional business organizations has immensely benefited from the use of data mining, its application in the world of sports is still in infancy. But the situation is changing rapidly. Data mining has started showing great value in many dimensions to organizations in the sports world and the recent future will undoubtedly bring increased research as well as commercial opportunities in the area of Sports Data Mining.

In this paper, it has been shown that data mining could successfully be used to analyse and enhance the performance of any batsman. In the future this technique could also be used to analyse the performance of bowlers, fielders, wicket keepers.

Acknowledgements

I wish to thank my supervisor Dr. Waqar Mahmood who is always very inspirational and supportive and gave me the opportunity to pursue my research ideas.

References

Borisov, A., Chikalov, I., Eruhimov, V., & Tuv, E. (2005). Performance and scalability analysis of tree-based models in Large-Scale Data-Mining Problems. International Technology Journal, 9(2), 143-151.

Chen et al, 1998 Internet browsing and searching: User evaluations of category map and concept space techniques. Journal of the American Society for Information Sciences (JASIS), 49(7).

Bhavani Thuraisingham, "Web Data Mining: Technologies and Their Applications to Counter-terrorism," CRC Press, FL, 2003.

Cricinfo www.cricinfo.com Feb. 2010

Crisp www.crisp-dm.org/Process/index.htm Feb. 2010

Han, Jiawei and Micheline Kamber (2006) Data Mining: Concepts and Techniques

K. Solieman, Osama. (2006) Data mining in Sports

Kohavi Ron , Foster J. Provost : ,1998 Guest Editors' Introduction: On Applied Research in Machine Learning. Machine Learning 30(2-3): 127-132 (1998)

MSDN Validating Data Mining Models msdn.microsoft.com/en-us/library/ms174493.aspx Feb. 2010

Rose Quinlan (1996): Top Ten Algorithms in Data mining

SK Murthy (1998)" Automatic construction of decision trees from Data: A multi-disciplinary survey" Data mining and knowledge Discovery. Vol.2 pp. 345-389

SPORT SIMULATION BY USING SIX DEGREE OF FREEDOM CABLE SUSPENDED ROBOT

M. H. Korayem [a, b], M. Bamdad [a], A. H. Korayem

[a] Robotic Research Laboratory, College of Mechanical Engineering,
Iran University of Science and Technology, Tehran, Iran

[b] *Corresponding author:* hkorayem@iust.ac.ir

Abstract

This paper introduced a novel sport simulation system which provides an actual physical trail of the sport performance to be enacted. The case of spatial parallel cable suspended mechanism is addressed. The main specifications for this parallel manipulator have been the prescribed workspace for each of the 6 degree of freedoms. For such training, there are some facilities to move the end-effector according to the kind of simulation including athlete in the workspace. This paper describes the study of the reachable workspace of a cable-suspended parallel robot and the workspace under study is defined as the set of all end-effector poses satisfying tension-ability condition. Since there are some limitations on the external wrench and the dynamic motion of the end-effector, such a workspace is the most desirable workspace for the intended application simulating athlete dynamics. Finally, in order to validate the proposed mechanism dynamic investigation is considered in workspace determination and the cable forces are computed to observe the relevance result to the athlete's position during the simulated trajectory.

Keywords: Sport simulation, cable suspended robot, Dynamic workspace

1. INTRODUCTION

Some kinds of sports are popular pastime for many people. However it requires a large amount of space and/or specialized facilities. In some urban area it may be difficult to find the necessary space to participate in these sports. Moreover, specialized facilities may be expensive to build and maintain. Additionally, Sports almost is subject to the weather, daylight and direct participation in the sport may be dangerous and cause injuries.

Accordingly, there is a need for an alternative to traditional sport that is not subject to the drawbacks noted above. To explore these aspects, we focus on robotic mechanisms as sport simulators. This class of systems allows people to practice their hobby by using real clubs into a screen projection of a sport course. Naturally, building such a system will take on a different character than the more commonly observed cases of developing systems for PC and laptop- based modes of interaction.

The flight simulation is probably the most representative application of the interactive visual simulator (Roman, 1999). This simulator was built with a complex and effective motion system to generate the realistic feeling of takeoff, landing and in-flight turbulence. This manipulator was first presented by Stewart in 1965 and called Stewart Platform Based Manipulator thereafter. This parallel robot provides several distinct advantages over conventional serial robot. Parallel machines are characterized by having multiple closed kinematic chains and actuate a subset of their joints. Due to this general arrangement, a parallel manipulator holds the potential for greater stiffness and higher speed and payload for a given weight. However, workspace is typically substantially smaller than that of a similar serial machine.

Cable array robots are a class of parallel manipulators which utilize multiple actuated cables to manipulate objects. They usually have a lower moving inertia when compared to its rigid link counterpart. Moreover, it has large workspace. For sport simulators, it is necessary to attain enough motion space. The flexibility of a cable allows the replacement of mechanical joints in the design.

They are ideally suited for large scale applications. This robot consist of a fixed base and a centrally-located end-effector, attached to moving payload, connect to cables whose tension is maintained along the tracked trajectory. Dealing with this mechanism SPIDAR is designed to investigate the effect of the force feedback from the environment to an operator, on task performance. However, the performance of the mechanism itself was not of their interests (Ishii, Nakata & Sato,1994).

The cable array mechanisms have been studied in the context of applications such as a virtual sports machine. A virtual tennis system which uses radial cable drive mechanism to enlarge the motion area is designed (Kawamura, Ida, Wada, & Wu, 1995). A versatile cable robot as a haptic interface was developed for sport simulation (Zitzewitz, Rauter, Steiner, Brunschweiler, & Riener, 2009).

There are many papers in the literature which discuss the defining characteristics and the important issues relating to this class of robots. Some of them are interested in the case of sport simulation. The computing load carrying capacity of cable parallel robotic manipulators (Korayem, Bamdad, 2009), workspace analysis (Barrette, Gosselin, 2005) are seen in the theoretical literature. The ideal model includes rigid elements. Elastic cable will tend to sag under its own weight. The effect of cable sag could be reduced by requiring non-zero minimum tension values (Korayem, Bamdad, & Saadat, 2007).

Determining the workspace of cable actuated robots can be more complex than for conventional parallel robots, as the cable actuators can only apply forces when in tension. In this paper, the reachable workspace is studied. One particular type of cable array robot with six cables is presented in detail including kinematic relations and static modelling. The shape, boundary, dimensions, and volume of the workspace of this cable robot are displayed. A simulation system for simulating sport virtual reality experiences is done. These results are discussed in detail based on the sport simulator.

2. MULTIPLE DOF POSITIONING MECHANISM IN SPORT SIMULATION

The paper is to present an architecture and mechanism to design a simulator on sport area and provides sport cases of designing for interactive settings where embodied performance is a central property, and how this influences the activity of programming. We examine a sport simulator setting, which is a case where the target activity is heavily based on physical action and full body interaction. The operational simulation is an operational assistance system, which is to train the operation of manipulating safely. Such systems include sports arcade games, simulators, and training applications for different settings requires users to engage in extensive bodily engagement to interact with a computational system.

The aim is to explore the multitude of use settings that developers may need to relate to robotics. This parallel manipulator meets the above needs and avoids the disadvantages and drawbacks of the prior art. The novel sport simulation system provides an actual physical trail of the sport performance to be enacted. These parallel manipulators relate to apparatus for simulating a sport activity and more particularly an interactive sports simulator system which provides an actual physical trial of the sports performance to be enacted.

The motion platform is an important sensory device for the trainee to fully immerse himself into the training scenario. The platform is able to deliver 6 DOF postures to the trainee, including rotations and shifts with respect to X-, Y- and Z-axis. The system can simulate the situation of acceleration, vibration, and turn over etc.

Rather than documenting the practices in detail, we have chosen to describe specific themes of the cable suspended robot, pointing at how they illuminate certain aspects and challenges that designers have to face when designing for such settings. For example it will be easy to provide a frame including adjustment means to adjust the inclination of the frame relative to a horizontal plane. Through this paper we attempt to contribute to an understanding of some of the dimensions involved in sport simulators, and in particular, the interplay between designing practice and its resources, the physical experiences, and the motion style contexts of the interactive devices.

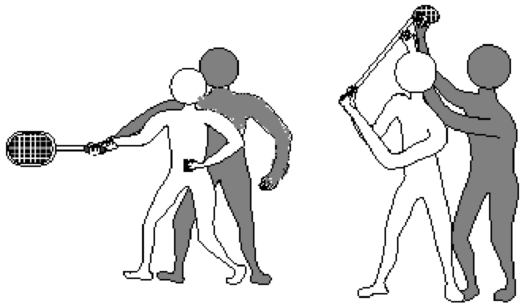


Figure 1: Cable Robot plays a role of trainer

3 CLASSIFICATION OF CABLE ROBOTS

The cable robot is a parallel manipulator consisting of an upper base plate and a fixed moving plate, which are connected together by six cables like adjustable links. This manipulator can be expanded and contracted individually to control the gesture of the platform.
Based on the number of cables (m) and the number of degrees of freedom (n), the cable actuated robots were classified into four categories:

1) Incompletely Restrained Positioning Mechanism :

$m < n+1$

2) Completely Restrained Positioning Mechanism :

$m = n+1$

3) Redundantly Restrained Positioning Mechanism :

$m > n+1$

4) Fully Restrained Positioning Mechanism :

$m \geq n+1$

Here, an aluminium plate is supported by six steel cables whose ends are rolled by pulleys connected to D.C. motors. Seven actuators are necessary and sufficient to produce forces and torques with six degrees of freedom. A fully constrained cable robot possesses more cables, and therefore presents greater risk of cable interference, and therefore limits the usable workspace. The mechanical architecture and computations also increase in complexity for full constraint.
The fewer cables are used to help avoid cable-interference problems by taking advantage of the end-effector's weight, which acts similarly to a "cable" pulling downward with constant tension .the cable robots with fewer cables offer greater simplicity and larger workspaces. Overall, IRPM robots are often preferable. The objectives of this paper are to develop a general algorithm to generate the workspace. Here, the research is focused on IRPM cable robots.

4 CABLE ROBOT KINEMATICS

Cable robots are relatively simple in form, with multiple cables attached to a mobile platform as illustrated in Figure. 2.
The Inverse Jacobian matrix is a position and orientation dependent matrix and for parallel mechanisms it is found to be

$$J = \begin{bmatrix} u_1 & \cdots & u_m \\ [r_1 \times]u_1 & \cdots & [r_m \times]u_m \end{bmatrix}^t \tag{1}$$

where $u_1, u_2, ..., u_m$ are unit vectors along the cable lengths $l_1, l_2, ..., l_m$.

A feature of parallel robots is that it is difficult to obtain Jacobian, where at the inverse Jacobian is common. Inverse kinematics of cable-suspended robot is computationally simpler to solve compared to forward kinematics. The forward kinematics equations for parallel mechanism are usually derived as a set of implicit equations, or a set of relatively complex explicit expressions. Due to the nature of parallel mechanisms, a complete analysis of the forward kinematics model may be impractical.
The position and orientation of the moving platform is known in inverse kinematics and the problem is to determine the cable lengths. The architecture of a six-cable robot and reference frames are described in Figure. 2. All six cables connect from the base to the moving platform. As is common knowledge:

$$\boldsymbol{l}_i^o = \boldsymbol{H} + \boldsymbol{a}_i^o - \boldsymbol{b}_i^o \qquad (i = 1, ..., m) \tag{2}$$

Where $\boldsymbol{H}$ is the position vector of the centroid moving platform with respect to center of the base platform. $\boldsymbol{a}_i^o$, $\boldsymbol{b}_i^o$ are the position vectors of connection points of the cable i on the moving platform and base platform relative to the reference frame O_B .

$$\boldsymbol{a}_i^o = \boldsymbol{R}(\psi,\theta,\varphi)\boldsymbol{r}_i \tag{3}$$

where $R(\psi,\theta,\varphi)$ present the rotation matrix about base coordinate frame and $\boldsymbol{r_i}$ is the moving platform fixed vector to the connecting points. In order to develop the kinematics of six-cable robot, consider that:

$$\boldsymbol{a}_i^o = [x_i, y_i, 0]^T,\ \boldsymbol{b}_i^o = [b\sin(\alpha), b\cos(\alpha), 0]^T \\ \boldsymbol{H} = [0,0,h]^T \tag{4}$$

A. INVERSE KINEMATICS

The equations for the analytic cable lengths can be written in the following form:

$$\begin{aligned}
&(x_1 - b\sin(\alpha))^2 + (y_1 - b\cos(\alpha))^2 + h^2 = l_1^2 \\
&(x_1 + b\sin(\alpha))^2 + (y_1 - b\cos(\alpha))^2 + h^2 = l_2^2 \\
&(x_2 - b)^2 + (y_2)^2 + h^2 = l_3^2 \\
&(x_2 + b)^2 + (y_2)^2 + h^2 = l_4^2 \\
&(x_3 - b\sin(\alpha))^2 + (y_3 + b\cos(\alpha))^2 + h^2 = l_5^2 \\
&(x_3 + b\sin(\alpha))^2 + (y_3 + b\cos(\alpha))^2 + h^2 = l_6^2
\end{aligned} \tag{5}$$

B. FORWARD KINEMATICS

From (5), $x_i\ (i = 1-3)$ for this robot is found as:

$$\begin{aligned}
x_1 &= \frac{(l_2^2 - l_1^2)}{4b\sin(\alpha)} \\
x_2 &= \frac{l_4^2 - l_3^2}{4b} \\
x_3 &= \frac{l_6^2 - l_5^2}{4b\sin(\alpha)}
\end{aligned} \tag{6}$$

The coordinates should be satisfied by the platform geometry constraints:

$(x_1, y_1, h), (x_2, y_2, h), (x_3, y_3, h)$

The position of the connection points on the manipulator shouldn't be varied. The moving platform equilateral triangle side with cable connections is considered as extra equations.

With substituting (6) into (5), nonlinear equations will be appeared. This system of equations can be solved numerically using Newton-Raphson method for finding the position of the robot attachment. All numerical calculations were performed in MATLAB.

5 WORKSPACE ANALYSIS

All the end effector position of the simulator according to its current position, velocity, acceleration, altitude, and gravity must calculate during the operation. The most essential issue in cable-suspended robots is maintaining the cables tensions. Thus, wrench exertion and cable length constraints limit the workspace.

Workspace for parallel robots is typically generated computationally, due to the high complexity of the geometry and for cable-suspended robots the workspace calculation is related to static equilibrium because of the fact that cables can not sustain compressive forces.

The workspace of a robotic system is defined as the volume that the moving platform reference point can reach. In the literature, with classification of cable actuated robots, a number of different workspaces had been studied. The reachable workspace is the largest workspace definable, and is the set of all points that can be reached in at least one orientation, regardless of the required orientation of the platform at that point, or ability to change orientations.

The less dexterous regions of the workspace corresponded to poses of the manipulator that are near singular configurations. As a result, there are additional constraints that must be satisfied in order for the robot to be in a non-singular pose.

Generally the moving platform may not be in equilibrium at a point in the workspace, whereas for a cable-suspended robotic system, the moving platform for all points in the workspace is equilibrated. The following conditions must be satisfied in order for a point to be within the workspace.

1- Tension forces must be above the cable pretension, also the maximum tension is considered for all cables.

2- The cables must be capable of exerting a positive wrench on the platform. All cable tensions must be non-negative to equilibrate the moving platform for an applied force.

3- All active cables must remain in tension to be effective for equilibrium or dynamically motions.

The end-effector may not be in equilibrium at a point in the reachable workspace because of the fact

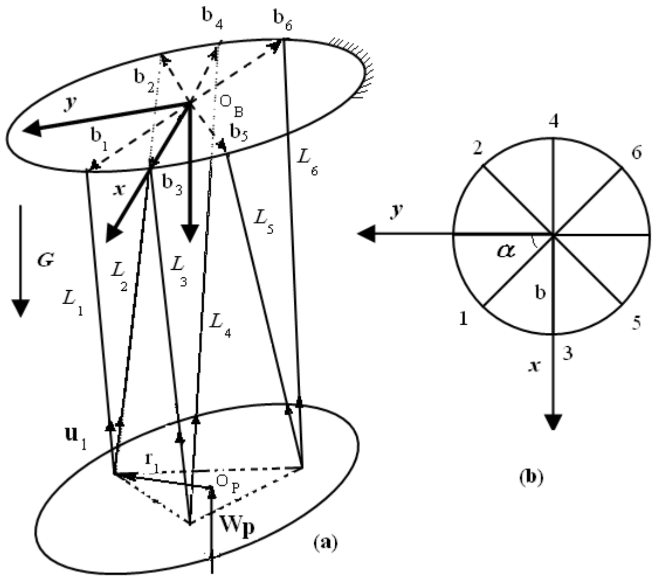


Figure. 2. (a) A schematic of the 6-cable robot, (b) A schematic of the base platform

that cables can not sustain compressive forces. For this reason, there is a need for defining and calculating the statically reachable workspace.
All cable tensions must be non-negative to equilibrate the end-effector for an applied force. The volume and boundaries of the statically reachable workspace depend on the direction of the forces acting on the end-effector. Statically reachable workspace is directly related with static equilibrium of the system given in next section.

6 STATIC ANALYSIS

In this case the target activity is heavily based on physical action and full body interaction. For example an interactive golf simulator environment is used by individuals. A golf course in combination with a sensor-based device for capturing the physical properties of the golf shot, such as ball speed, direction, and spin, to calculate the trajectory and indicate where in the simulated environment the ball will end up. The setup of the system provides an interactive context that involves two distinctly separate modes of interaction, first, hitting golf shots to actually play the game, and second, interacting with the simulation to carry out the actions necessary to play in the manner one wants. The first being a highly physical mode of interaction while the second being more of an information seeking mode of interaction.
It is possible to equilibrate the end-effector for all points in the reachable workspace when we define a statically reachable workspace.
This section presents statics modeling for the robot. Principle of virtual work will be applied to find the tension values caused by the wrench on the platform. Since the virtual displacements are related to the Jacobian matrix. The static balance equation is:

$$J^t \tau + W_p = 0 \tag{7}$$

When the external wrench $W_p = \left(F_p, M_p\right)^t$ is available to tighten the cables, the cable tensions are denoted by $\tau = (T_1, T_2, ..., T_m)^t$.
The static equations include of wrench balance on platform is used to find the force of each cable for planar model. The force equations for the platform can be easily expressed as:

$$\sum_{i=1}^{m} T_i + F_p = 0 \tag{8}$$

$$\sum_{i=1}^{m} R(\varphi)\boldsymbol{r}_i \times T_i + M_p = 0 \tag{9}$$

7 SIMULATION

In this section, a three-dimensional manipulator with six cables is considered. For each cable, there is a separate winch and pulley. The end effector carries a member of body or sport device like hand, rocket. It is kinematically constrained by maintaining tension in all six supporting cables. The suspended movable platform and the overhead support are typically two equilateral triangles.
The mechanism is shown in a virtual reality environment in Figure. 3. In this environment the effects of degrees of freedom and the position of rocket are described with cable forces. Graphically displays articulated geometric figures and allows them to be placed and manipulated in a virtual environment. This virtual world contains a fully functional model of robot, a 3-D tracking device and other inputs sensors in sport simulation. The simulated environment will have the advantage of a knowledge based reactive planner. The knowledge based planner allows for flexible and reactive planning in an unstructured real world domain.

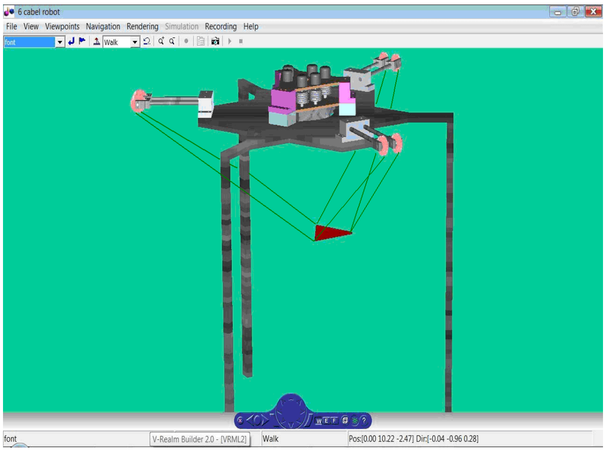


Figure 3: Virtual Reality Toolbox viewer

One type of static workspace can be defined by the task required at the set of all points within the workspace. The constant orientation workspace is defined for a particular orientation of the platform and is the set of points that can be reached by the centroid of the moving platform when the orientation of the moving platform is kept constant.
Here, all rotation angles are assumed to be zero in order to make a stable situation for trainee. Moreover the elevation of end effector (rocket) can be varied during the motion.
For a complete description of static workspace, it is divided to some layers. Each Z level is considered one layer in workspace. The tension norm of cables is an important parameter to determine the suitable operation elevation.

Z(m)	Tension Norm (N)
0.5	15.22
1	11.62
1.5	10.87

Table 1: Tension norm in different levels Z

Each cable is tensioned by a motor-pulley direct-drive system placed on the base platform. According to Table 1, motor efforts can be selected.
The kinematics and the statics of cable-suspended robots have been analyzed. Our theoretical study of a particular cable-suspended robot has been undertaken and to confirm the theory with experiments, an experiment test-bed of a planar cable suspended robot is fabricated and experimental tests have experienced the feasibility of the cable system design and proposed models (Korayem, Bamdad & Zehtab, 2010).
The robot workspace refers to the region in which the robot can operate, which is a defined volume for spatial manipulators. Workspace was computed using a numerical method. To calculate the workspace, one can either use the idea of null vectors or static equilibrium equations. Idea of null vectors is very useful when the system is fully constrained.
The side lengths of equilateral triangle shaped base and moving platforms are respectively 1.19 m and 0.17m. In this case, the workspace area is characterized as the set of points where the centroid of the moving platform O_p can reach with tensions in all suspension cables. By changing the coordinates of O_p the workspace shape can be expressed in 2D space by a series of dots (Figure 4).

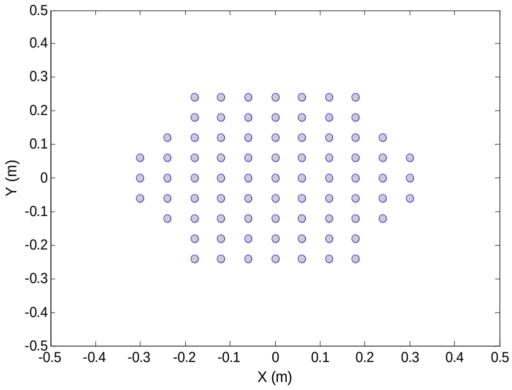


Figure 4: The planar workspace for Z=1 m

The raising degrees of the yaw roll and pitch will cause hazard situation if the angle is over some safety value. The calculation of cable tension by equations in Section 3 is applied. In this way, the cables tensions are depicted in Figures 5-11.

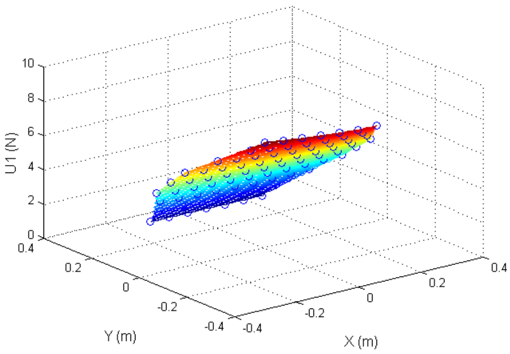


Figure 5: Cable force 1

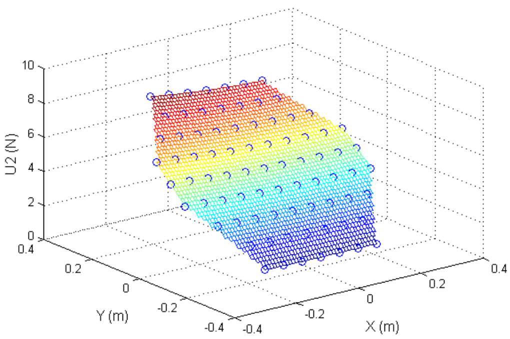


Figure 6: Cable force 2

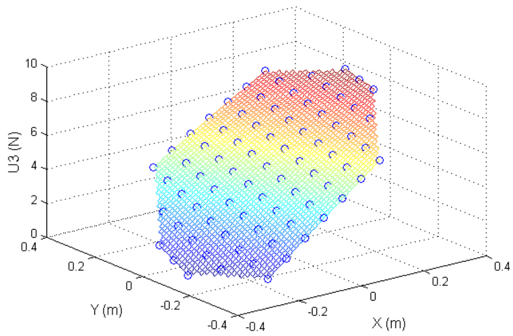


Figure 7: Cable force 3

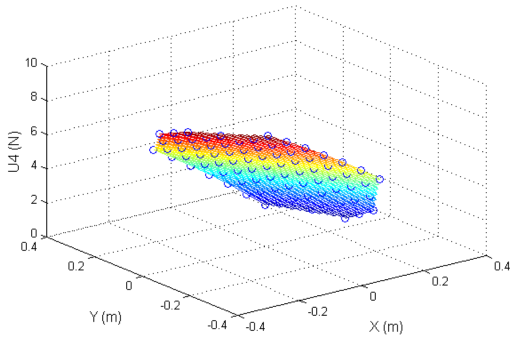


Figure 8: Cable force 4

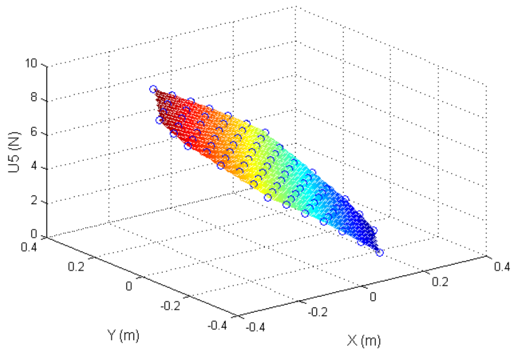


Figure 9: Cable force 5

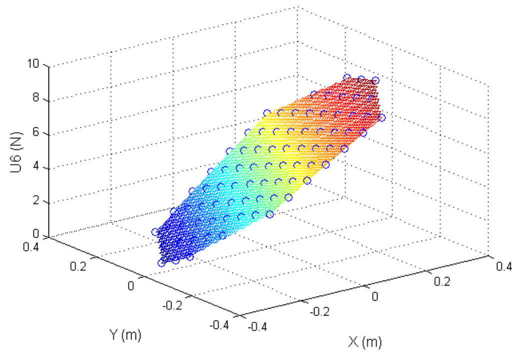


Figure 10: Cable force 6

The investigations conducted earlier, including an analytical solution and the search for solutions of the problem of synthesizing trajectories, planning and implementation of the motion of a cable suspended robot are combined.

For a cable-suspended robot, the optimization problem for minimum effort can be posed in the tension norm calculation (Figure 11). The position of end effector which is in high values of cable force or motors efforts can be deleted in path determination for sport simulator. The motion planner in sport simulation can decide based on the cable force contour.

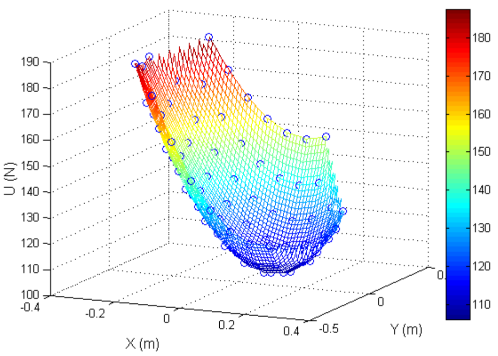


Figure 11: The norm of cable forces

8 CONCLUSIONS

The participation of parallel cable robots in sport events is considered. The planning mechanism and developing a virtual representation allows for both autonomous planning as well as planning through human-machine interaction.

According to the actual physical trail of the sport performance, the target activity is heavily based on physical action and full body interaction. These machines have a large workspace and enough space for sport player swings is guarantied. The workspace determination plays a very important role in the sport simulation.

Acknowledgements

This paper describes a research done in Robotic Research Laboratory, College of Mechanical Engineering, Iran University of Science and Technology (IUST). Authors thank IUST for support the program during our research activity.

References

Barrette, G., Gosselin, C. M., (2005). Determination of the dynamic workspace of cable-driven planar parallel mechanisms. *ASME J. Mech. Des.*, 127, 242–248.

Kawamura, S., Ida, M., Wada, T., & Wu, J. L. (1995). Development of a virtual sports machine using a wire drive system: a trial of virtual Tennis. *Proceedings IEEE International Conference on Intelligent Robots and Systems*, 111–116.

Korayem, M. H., Bamdad, M., (2009). Dynamic load carrying capacity of cable-suspended parallel manipulators. *International Journal of Advanced Manufacturing Technology*, 44, 829-840.

Korayem, M. H., Bamdad, & M., Zehtab, R. M. (2010). First Experimental results of load carrying capacity for a planar cable-suspended manipulator. *Journal Mechanic IAU Majlesi*, in press.

Korayem, M. H., Bamdad, M., & Saadat, M. (2007) Workspace analysis of cable-suspended robots with elastic cable. *Proceedings IEEE International Conference on Robotics and Biomimetics*, 1942-1947.

Ishii, M., Nakata, M. and Sato, M. (1994) Networked SPIDAR: A Networked Virtual Environment with Visual, Auditory, and Haptic Interactions. *Presence*, 1.3, No.4, 351-359.

Zitzewitz, J. v., Rauter, G., Steiner, R., Brunschweiler, A., & Riener, R. (2009). A versatile wire robot concept as a haptic interface for sport simulation. *Proceeding IEEE International Conference on Robotics and Automation*, 313-318.

Roman, M. (1999). Flight Simulators: A Look at Linux in the Aerospace Training Industry. *Linux Journal, Vol. 19*, Available online at http://www.linuxjournal.com/.

TOURNAMENT ROULETTE

Neville de Mestre

Bond University
Corresponding author: margnev2@bigpond.com

Abstract

Tournament roulette is a special type of roulette competition run by some casinos. Competitors play in a series of roulette games in which some competitors are eliminated at the end of each game. A game consists of 25 spins of the wheel followed by an unspecified number of spins until one of seven key numbers on the wheel occurs to end the game. The essential strategies to consider are what to do in the end game depending on whether or not a competitor is in the possible elimination group.

Keywords: roulette, probability

1. INTRODUCTION

Roulette is a gambling game in which a wheel is spun about a vertical axis. In Australia and Europe there are 37 slots in the wheel for a white ball to fall into indicating the winning number from 0 to 36. The lay-out for the wheel is shown in Figure 1.

Figure 1: The Australian and European wheel

Up until recently the wheel was situated at one end of a long table with the remainder of the table being used by players to place their bets using special roulette chips. The lay-out for the area to place bets is shown in Figure 2.

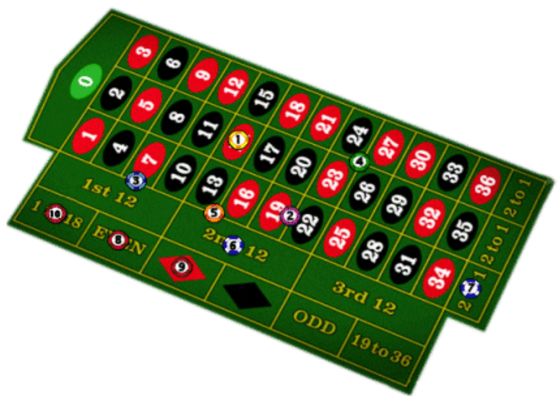


Figure 2: The betting lay-out

At my local casino the roulette tables have recently been removed and replaced by computer screens for placing bets and observing the wheel. Tournament roulette is no longer available.

Tournament roulette is a competition organised usually by a casino over a number of elimination games so that six competitors will meet at a roulette table for the final. Each elimination game involves six players with the top two progressing after the first round of games and the top three progressing from there on.

Players pay the same fee ($250) to enter, and at the beginning of each game are each given chips to the non-redeemable nominal value of $2000. With 144 players entering, all the entry fees are distributed as prize money in the following way after the end of the final.

First place	$20000
Second place	$7000
Third place	$5000
Fourth place	$2000
Fifth place	$1500
Sixth place	$500

The casino takes zero percentage and uses the event as a marketing tool.

2. THE TOURNAMENT

Each round of the tournament consists of a number of games where the field of competitors is reduced in size. Each game has a basic section of 25 spins. Players must bet at least $5 on each spin and no more than $1000. At the end of the 25 spins the amount left with each player is totalled and announced to all. Then follows the end-game section.

In the end game, the competition continues until one of seven key numbers (0, 5, 8, 17, 20, 29, 32) comes up and then the game terminates. The game also terminates when no key number arises after 20 minutes of the end game. The players with the top two amounts in each game of the first round heats, and the top three amounts in each game of subsequent rounds, progress to the next round where they again start with $2000 in chips.

3. THE ODDS

The Australian roulette wheel has 18 black numbers, 18 red numbers and only one zero. Therefore the probability of a particular number coming up is 1 in 37 or approximately 2.7%. This is the house margin for normal roulette.

Bets may be made on various combinations of numbers (singles, doubles, triples, fours, sixes, twelves and eighteens) and some of these are given roulette names such as split, street, corner, etc. The pay-out for each betting category is as follows:

Straight up: A bet placed on any single number including zero pays 35 to 1.
Split: A bet on any two adjacent numbers on the table pays 17 to 1.
Street: A bet covering a row of three adjacent numbers pays 11 to 1.
Corner: A bet covering a square of four adjacent numbers pays 8 to 1.
Six Line: A bet covering any rectangle of two adjacent rows pays 5 to 1.
Column: A bet covering any column of 12 numbers pays 2 to 1.
Dozen: A bet covering a series of 12 consecutive numbers starting with either 1, 13 or 25 pays 2 to 1.
Even chances: A bet covering a group of 18 numbers such as red or black, even or odd, top 18 or bottom 18 (not zero) pays even money.

For each spin, the wheel is spun one way and the ball is projected in the opposite direction until it comes to rest in a numbered slot, thus determining the winning number. All losing bets are first of all removed from the betting table by the croupier, and then the winning bets are paid out in chips.

4. STRATEGIES FOR THE FIRST 25 SPINS

By calculating the expected gain for each type of bet it is seen that, no matter how bets are placed, the expected loss will always be exactly 1/37 or 2.7% approximately in the long run. Now 25 spins is not a particularly long run, but it is long enough to suggest that some competitors who place big bets may wipe themselves out early. Therefore a good strategy for the first 10 spins is to place only one $5 bet on an even-money chance each time, and watch to see if some players drop out. If a number do drop out then the number of opponents to watch in the end game is reduced significantly. Using this strategy, the total chip value for the $5-bet player should stay near $2000. However if two players in a first round heat, or three players in a subsequent round heat, make lucky big bets early on that take them well past $2000, the $5–bet player must now take some action to try to catch up gradually before 25 spins have passed.

To help decide on possible actions a computer program was written so that the Roulette tournament could be simulated. There were six places available for players, but three of these slots were used for pre-determined types of players. One player doubled his bets when he lost and only went back to his original bet when he won, another used a less-risky Martingale system, while a third bet wildly at

random. The remaining slots were used by my colleagues and me to try out our various strategies.
A possible catch–up strategy is to place $540 on numbers 1 to 18, and $360 on numbers 25 to 36. The expected gain is still the same as betting $30 on 30 different numbers, that is -1/37 x $900. However it is far easier and quicker to place bets on two spaces rather than 30. If any of these 30 numbers comes up the gain is $180, and the high probability of this occurring is 30/37 or approximately 81%. Of course, if any of the other seven numbers (0, 19 to 24) come up then the loss is $900. In this case put $1000 on an even-money chance and pray or hope.

5. STRATEGIES FOR THE END GAME

A clever player notes that there are four black key numbers, only two red key numbers, zero, and six of the key numbers are in the central column of the betting matrix in Figure 2. These facts may be useful when trying to decide on betting strategies before each spin of the end game.

At each stage of the end game players must try to ascertain whether they are within or outside the qualifying group and by how much.

When they are within the qualifying group, they should gauge their lead over the nearest current non-qualifier and place $5 on each key number. This will cost them only $35 if a key number doesn't come up but if a key number comes up to end the game they will profit by $145.

When they are outside the qualifying group there are a number of strategies to use depending on how far they are outside. Betting on the key numbers will produce profits which are $(35 times N), where N is the amount bet on each key number. Players must add extra amounts to match the amounts placed by the nearest current qualifier on any particular key number. For example, suppose a current non-qualifying player is $650 behind the nearest current qualifier, who places $5 on number 20 together with other bets spread across the table. The non-qualifying player could place $30 on number 20 and $25 on each of the other key numbers. If a key number comes up the gain will be $725. The loss will be only $180 if a key number does not come up, and the game is still on.

When the deficit to the nearest current qualifier is too big, the non-qualifiers could place $500 on each outside column and hope that a key number doesn't come up (or one of those other pesky numbers in the central column). This will gain $500 with a probability 24/37 of occurring. On the other hand a bet of $100 on each key number gives a return of $2900, but the probability for this is only 7/37. Note that the expected gain in all cases is still (-1/37) x Amount Bet.

6. CONCLUSIONS

When playing in a roulette tournament the aim of any player is to progress through each round to eventually participate in the final.
The simplest strategy in the early part of the game is to wait for other players to wipe themselves out, and this is a definite possibility. Approaching the end of the first 25 spins, players should endeavour to be a current qualifier or near to the lowest current qualifier.

Strategies in the end game should take account of the key numbers, and bets should be made to try to stay within the qualifiers or reach that position on the next spin.

Although luck plays a primary pivotal role in the game of roulette, we believe that mathematical strategies can play a secondary role, and this sometimes tips the balance in close decision-making situations.

7. ACKNOWLEDGEMENTS

The author acknowledges the co-operation of his colleague, Graham McMahon, in developing the computer simulation and suggesting some of the strategies outlined in this paper.

A MARKOV STATES MODEL FOR TOUCH FOOTBALL

Joe Walsh [a,b]
[a] *Australian Catholic University*
[b] *COPAAL: The Centre for Physical Activity Across the Lifespan*
Corresponding author: joe.walsh@acu.edu.au

Abstract

The game of touch football was modelled using a series of modified Markov states. The aim of this model is provide a framework allowing model expansion to account for additional game states, time dependence and displacement dependence. The model was designed to allow adaptation by application of alternative displacement and time dimensional distributions, displacement and time dimensional distribution parameters, and state change probabilities. This would allow adaption to the model according to other various styles of team play and associated levels and composition of teams. Due to the state based nature of the Markov processes involved, the model will be more intensively analysed in following papers focusing on more effectively analysing each state in turn. The goal of this model however is to produce a starting framework to allow for future development in modelling the sport of touch football.

Keywords: Touch Football, Sport, Modelling, Markov States

1. INTRODUCTION

Touch football is one of the most extensively played sports in Australasia. In Australia touch is played by over a quarter of a million registered touch players, half a million school children and up to 100,000 casual players (Australian Human Rights Commission, 2007 and Touch Football Australia, retrieved 2010). Touch is mainly popular in those states in which rugby league is also popular (Queensland, New South Wales and ACT), which is not surprising given its origins as a training activity for this sport (Mathesius & Strand 1994). While New Zealand has less touch players than Australia, it is still the largest participation sport within this country with over 230,000 New Zealanders playing the game. This includes 70,000 under 17 year olds and 160,000 adult players (Touch New Zealand, retrieved 2010). The running of this sport also involves a vast number of hours devoted by support personnel such as team coaches, team managers, referees, selectors and administrators. Touch is also played in other Australasian countries such as Samoa, Fiji, The Cook Islands, and Papua New Guinea. Countries such as South Africa, England, Wales, Scotland, USA, Japan, amongst others also regularly enter teams in international tournaments. Much of the touch football within these countries is of a more casual nature or used primarily as a training tool for other sports. Touch as a sport has grown rapidly within Australia (Philips 2000). This growth seems to be replicated to some extent in these other countries; however the greatest density of participants is still in Australasia.

Despite the large involvement in touch, very little Australasian sports research has been conducted on this sport when compared to the main professional sports played in Australia and New Zealand. Much work has been done on developing theoretical and practical modelling systems in sports such as AFL, rugby league, rugby union, hockey, cricket, basketball, squash, tennis, golf and others. The majority of published research related to touch football is based on injuries associated with the sport, or use of touch players as a subject group to test some experimental intervention in fairly standard study design. This is not surprising considering the large number of touch players available. However limited published research has been done on the sport itself. One paper that has been published Mathesius and Strand 1994 (when touch football was called touch rugby in Australia, a name still used for it in many other countries) is now sadly outdated as the game has changed considerably. Due to the immense volume of time devoted to this sport, particularly by the people of Australasia, there is a need for a basic framework for model development for the sport.

Many invasion type games, in particular sports such as hockey, basketball and football codes can be split into set phases of game play. This is particularly easy for a sport such as gridiron where the very sport is designed around set plays. Most other invasion games can also be split into several distinct components. For example in the game of rugby union different set pieces as well as broken play (i.e. non-set piece play) will have different outcome probabilities for different offensive-defensive team combinations dependant on various factors at any given point in the game, such as for example game score, field position, fatigue and time remaining. In fact a myriad of different factors make modelling of most team sports exceedingly complex and this is particularly so for invasion type sports. Touch football is such a sport and may be modelled as having several components. The first of these components would be a driving phase in which a team aims, amongst other goals (such as gaining a repeat set of touches via an infringement leading to a penalty against the opposition), to gain field position from which to enable both execution of scoring plays and ensuring less probability of their opposition scoring as play is at a larger distance from the scoreline. During this phase, the defensive team has their own corresponding phase of play in which they attempt to both reduce the field possession gained by the attacking team in order to reduce the attacking options of this team and additionally ensure that when they are given their limited touches they are deeper in opposition territory. A secondary, but nethertheless important goal, for the defending team, within this phase is also to defend against any potential scoring plays the attacking team will attempt to execute, even if far from the try line. The next phase of the game to consider would be line attack. In this phase of the game the attacking team is no longer focused on gaining field possession. They have already reached an appropriate distance from the defensive try line to orchestrate a scoring play. For the defenders this phase of the game would be referred to as line defence. There is also the ability to manage player fatigue and optimal application of personnel via substituting players while simultaneously completing driving plays or the defence to these plays. While touch football is a game where players may be freely substituted during the stoppage time after the scoring of a try, it is also a game where unlimited substitutions (from within a team consisting of a limited number of players and sometimes limited by gender) are allowed. In order to maintain optimal intensity within a game, it is usually necessary to substitute players during various phases of the game whilst minimising disruption of the purposes of these phases. Therefore some phases will represent the switching of players on a team in either attack or defence. In some versions of the game of touch, player fatigue is also managed with a break at half time, while in other tournaments the teams switch ends of the field and restart play at halfway with every touch down and there is no half time period (Federation International Touch Inc. 2003, Touch Football Australia 2007). The version played would affect fatigue management and the importance of substitutions. The half-time break is usually used at higher level tournaments and this half-time period can be considered as another phase of the game. There is also a phase of play following a tap-off to restart the game at half time and a phase of play following a penalty.

The author hopes to start with a basic system for the various stages of the game of touch and expand on this systems. If it can be shown to be appropriate to model a stage of the game in this way, the prospect of expanding the model with investigation of different stages and looking at other eventualities that present themselves within the sport can be realised. If a working model for the stages of this invasion sport can be developed and represented via a Markov Chain then future research can expand this model to other invasion sports mentioned. Many characteristics of the other invasion sports are shared by touch football and this will ease adaption. Certainly these vary across the different sports, but can include limited tackle (touch) counts/plays, a single points scoring method, and a game that is played over a limited time span. Additionally the different offensive and defensive alignments and their preferential defence of certain attacking options or choice of certain offensive methodologies in gridiron closely parallel preferential options and patterns in line attack/line defence in touch football (see Walsh 2010). Other factors include importance of field territory, use of and a limit (either strictly numerical or using conjoint numerical and displacement based systems) to the amount of plays/downs, both games being conducted over a set period of time, scoring by crossing a certain line, requirements of choice of defenders to preferentially defend certain players/zones in man on man, uneven or a compressed defence. By its very nature the sport of Gridiron (and 'American Touch Football') is highly suited to adaption as a Markov Chain as by its very nature it is split into discrete stand alone plays. Markov Chains have been used to model many different systems in many different academic fields. In sports, they have had particular success for many years in net games, such as squash, badminton and tennis (for example Schutz 1970, Schutz & Kinsey 1977, Clarke 1979, Croucher 1982, Pollard 1985, Pollard 1987 and Wright 1988). This research has continued progressing over the last 30 years (e.g. Newton & Islam 2009), however touch football has never been looked at in this manner.

If a Markov model for touch can be designed and shown to represent the probability distribution of any future state given the current state then this could have several practical uses in the future. If the model is sufficiently complex to fit game data well, probabilities of relevance can be simulated as a probability distribution function. This would potentially allow for in-game and pre-game strategic questions to be resolved. There would be many potential applications for such a system. Numerous applications could be used advantageously in order to assist with team coaching and selecting optimal plays. While the application to coaching will be apparent from use of match analysis prevalent in many professional sports, in touch football there are a vast number of participants and potential assistance in selecting representative players at various level could in fact be highly beneficial. In fact teams of selectors must watch all major tournaments in order to establish men's, women's and mixed representative teams at various levels and many different age groups. The potential for future modelling assistance in this talent identification or selection processes is therefore also a consideration. This could be extended in many directions, such as selecting particular players for facing particular opposition as well as for executing or nullifying particular plays. Other interesting uses include estimating in game win and event probabilities as well as points expectation for certain games. These particular points could have potential application for such features as pre-game as well as in-game sports betting, deciding games appropriate for television coverage (issues such as the probability of which games are more likely to be high scoring or exciting games), or allocation of playing areas to different teams during a tournament (tournament control allocating fields with the best spectator viewing areas to those games with higher probability of being both higher scoring and highly competitive games). These latter issues would have particular relevance if such a model was expanded to other sports. Rugby league in particular is similar to touch football in its basic framework of a tackle, as opposed to a touch count. Touch football in fact originally evolved from rugby league (Mathesius and Strand 1994). With its high commercial television exposure, in the future it is not inconceivable that some form of statistical modelling or Markov based probability distribution function process will be used for such allocations.

Terminology used in this paper is based around FIT (Federation International Touch Inc., 2003) and Touch Football Australia playing rules 7^{th} edition (Touch Football Australia, 2007). For the reader unfamiliar with the game the terms: toss, tap, rollball, touch count, ruck, half (player who takes possession of the ball behind the player who performs the rollball), 'period of time' dismissal or 'remainder of the match' dismissal (player sent off temporarily or permanently), drop off, scoreline, touchdown, penalty touchdown, onside, offside, interception, penalty, attacking or defending set and the field dimensions, particularly 5m line and halfway line are defined in these publications and the reader is referred to these compatible publications if definition of these terms is required. These publications also make mention of the variety of different rules used in different competitive environments. The lack of one set system of rules is an interesting feature from a modelling perspective. The term "park touch" is described to refer to the less elite level of touch football, where more rule variations are permitted.

Usually the most common long range attack conducted as a scoring play is the scoop. This is when driving plays (or some other effect) have caused defenders to be in an offside position. The half runs with the ball towards the defending scoreline. This may attract opposing player(s), or attacking players may use agility to escape from their marking defender and the half will attempt to pass to an unmarked player in order that they can score. Both long range and short range scoring plays may be preceded by a play, the emphasis of which is not the scoring of points or advancing the ball towards the opposition line, but engineering circumstances such as particular opposition movements or ball placement to enhance the probability of scoring on the follow play. In order to simplify the nomenclature (for this and future papers), the author wishes to borrow a phrase first heard used by the Sydney Rebels Touch Technical

Director (NSW, Australia) and refer to this play as the “strike dump”. Another concept to consider is second phase attacks. If an attack is conducted, but is unsuccessful at scoring, it may still have forced the defending players to be in a poor position for defence. A situation that can be capitalised on by launching a quick attack designed to take advantage of this misalignment. Such a rapid attack following, but dependant on, the misalignment caused by the previous unsuccessful attack is referred to as a second phase attack. A further attack aimed at exploiting the defensive structure adopted as a result of this secondary attack would be referred to as a third phase attack. Hypothetically there could exist fourth and fifth phase attacks, though the difficultly in maintaining momentum and coordinating these across the attacking team according to the hypothetical structure changes of the defensive teams, as well as the chance of a touchdown or stoppage on an earlier play makes these highly unlikely (P~0).

2. METHODS

We first need to define the game states required in order to model the game. State 0 is self explanatory, namely the determination of the team to start off via a coin toss. The winning team captain receives possession for the commencement of the first half, the choice of direction for the first half and the choice of interchange areas for the duration of the game. Many assumptions are made in this model and the first of these is that the choice of interchange areas and choice of direction for the first half do not adversely influence the states model. If a situation arises for which some environmental or psychological factor influences the model due to these choices, then this can later be factored in as an additional parameter. At a competitive level, fields should be situated and both marked and illuminated in a manner that such a choice should not influence play for reasons of field location. However certain environmental factors such as for example sun and wind are more difficult to control, so it is feasible that such an arrangement could have a small influence on the game in certain circumstances. Notwithstanding this approximation is still a reasonable one for the purposes of creating a skeletal framework from which to develop a working model for the sport. Additionally similar assumptions are commonly made in other many examples of sports modelling and yet produce reliable working models (e.g. Noubary 2007).

State 1 will be one of two possible states of equal likelihood with one of the two teams taking a tap off play. This can have several outcomes, the most likely being that the defending players effect a touch on an attacking player in possession and a rollball must be affected with a touch count advanced by 1. This is the most likely state transition, however other possibilities are that one of the teams might make a error leading to the touch count being restarted with a rollball for the attacking team, a penalty for either team, a turnover resulting in a rollball with a restarted touch count for the team that was defending, an interception by the team that was defending or a try for the attacking team. While many of these state changes are unlikely they must still be considered. In fact we can consider that during the driving phase of play, any of these options can occur as well as after a tap off or penalty play. The driving phase of play was therefore modelled as having 6 states within a Markov process representing the touch count for all six touches. The first driving state “rollball T0” (with the number next to T indicating that zero touches had been made by the opposition in this particular touch count) was considered replaceable by one of two alternative states representing a tap from either a penalty or restart from the halfway line. Both of these states differ from a conventional rollball in the distance the defence must retreat from which the position at which play is restarted, therefore the metres play is advanced may be distributed differently. Both of these states have the same outcome options as the first rollball state, though the outcome probabilities may slightly differ.

As the touch count progresses the attacking team has two primary effective options namely to attempt to continue to focus on territorial advantage or attempt to score a touchdown (which may also lead to some advancement of territorial position even if unsuccessful). This leads to the introduction of several additional possible states, namely states representing an attempt at a long range attack, a short range or line attack provided field position is appropriate, a touch deliberately allowed to be executed by the defending players in order to set up a more promising line or long range attacking option (the strike dump) and lastly a second, or later phase attack after an initial unsuccessful line attack (the first phase attacking option). This leads to two categories of line attack, those based from a deliberately engineered platform and those not set up in this way. The probabilities of successful execution will depend on the particular attack used and the appropriate nature of the platform used for any particular combination of attacking and defending teams. We have also explained the concept of second (and later) phase attacking play. As these later phase options are conducted under alternative circumstances, logically the probabilities of success will be different to a conventionally attacking option. Therefore this stage of the model will require a unique state and such a logical argument follows for the prospective extended stage attacks if these are attempted (or even possible). These probabilities will also be affected by the actual attacking option chosen requiring either further sub-states or additional parameters in the probability density function for this particular state as the model expands.

For each team there is an optimal position to initiate a particular line attack. While short range attacks will need to be conducted 5-7m or very close to this distance from the touch line, some attacks such as a scoop can potentially be initiated from anywhere on the field. The probabilities of success depend on acceleration, top end speed, avoidance of deceleration after attaining maximum speed by the scooper and support players (if sufficient distance is covered to attain top end speed) and coordination of the attack across the team. The primary determinant however will be the distance from the opposition scoreline and the point when the attack is initiated. As the distance increases, the probabilities of success will converge towards zero, as the likelihood of successful execution reduces dramatically with significant distance from the scoreline. This particular category of attack could be executed as a long range attack as opposed to those that must only be initiated closer to the opponents score line, namely short range attacks.

In this paper an approximation was made. During the driving phase teams have the ability to attempt to score at any time in their attacking set, though most usually, as field possession is not optimal and a successful score is improbable without field position, this option will be taken on the last, or sometimes second last, touch of the attacking set. These attacks were classified as long range line attacks. Teams also have a range of short range line attacks. In order to initiate these attacks certain conditions must be met and these will vary from team to team. The successful execution of a line attack will be a function of the distance from the scoreline at both the touch it is initiated on and the previous touch. As the model evolves, the time and distance between these touches being executed will also need to be factored into the distribution function. This will reflect the ability of the defence to return to an onside position and make any other necessary adjustments. Both the time dependence and displacement dependence of this function will have a different effect on the probability density distribution for different attacking plays. For some teams and some moves a second phase attack may be launched in some circumstances provided possession is retained and the touch is made with the attacking players in a suitable position. Some defending teams may be more susceptible to such an attack than others.

Each driving or line attack state requires 6 sub-states to represent the touch count. Each strike dump or second phase state requires 5 sub-states to represent the touch count. The need for only 5 is due to the need for a following line attack or a preceding line attack state. Those states that result in the resetting of the touch count or a change of possession do not require additional sub-states within the current model.

The state model in the results section was thus developed (Figure 1). The transition between states is represented by a arbitrary probability or a probability density function according to the transition involved.

Assuming an error is made, we can assume possible results are: a recount (i.e. play continues and the touch count is restarted), penalty, turn over, intercept, send off or player sent for time (we assumed the probability of these last two options as being equal to zero in this simple model, but this can be expanded in future work). Other than these changes of state, teams have the option to execute an attacking play such as a long range attack. We must also consider the possibility of entering into this state at any phase of play and consider that for tactical reasons this is most likely after the 4th or 5th touches have been made (a play such as a scoop is a high risk play, but there is less at risk as possession will most likely be lost anyway).

In the event that the line attack is unsuccessful and assuming the touch count has not expired nor some other event requiring a turnover of possession, there is a possibility that the attacking team will use the following play to execute a second phase attacking option. The alternative to this is to set up an attacking option for the following play or alternatively attempt an attack that does not capitalize on the disarray in defence caused by the preceding attack; however this is a low probability option and therefore an unlikely choice. This last option, while a poor choice earlier in the touch count may however be the only option if no second phase option is available and the 5^{th} touch has been made.

This model makes the following assumptions: no substitution is conducted, no players are sent off or sent for time, there is no drop-off, there are no injuries, fatigue is not considered (i.e. no players fatigue) and all players are identical within a team (which is certainly not the case). While this is imperfect, it is necessary first to generate a basic framework for the game. This framework is designed such that it can easily be expanded in future work.

3. RESULTS

The state transition diagram (Figure 1) was obtained. Arrows and diodes (selected due to the intuitive nature of the symbol) were used to indicate direction

Previous State in Team A's Markov State System	Future State in Team B's Markov State System	Notes
Touchdown Scored Offense	Tap off Restart Team B	
Interception	Turnover of Possession T1	Other possibilities for state change exist for interception
T5 Team A	T0 Team B	Touch count expired, not shown in diagram.
Stoppage due to Error by Offense	Turnover of Possession T0	Other Possible State "Penalty Awarded to Defense"
Penalty Awarded to Defense	Turnover of Possession Penalty	
State zero, the coin toss	Tap off Restart Team B	P=0.5
Touch count expired	Roll ball T0	Not shown on the states diagram, transfer after state T5 if touch is made.

Table 1: States which result in a change of possession from Team A to Team B.

of possible state change. This diagram relies on the assumptions detailed in the methods section. The model is a framework allowing easy addition or modification of states. State transfer probability is in most cases dependant on displacement functions that will be expressed in a future paper. Some state transfers were deemed to have near zero probability. States that were not included in the diagram (For clarity as these states could be transferred into from most of the other states) were a half-time (where applicable with rule variations) and a full-time state. Similarly a touch count expired state was also kept as a hidden state due to the number of states from which this state could be transitioned to. A symmetrical state system exists for the opposing team, for presentation purposes, only one team's states were shown. The teams were labelled Team A and Team B. The state diagram shown is for state transitions when Team A is in possession of the ball. Certain states would lead to a change in possession as indicated in Table 1. In this case the states based model would continue, but in a symmetrical set of states for Team B.

State transfer was dependant on the touch count T0-T5, displacement from the opposition score line and time. Provisional transfer probability functions were developed for state transfers and these will be discussed in future work.

4. DISCUSSION

This is an theoretical paper and is designed to form the backbone for further, progressively more realistic, work modelling invasion type games. Certainly there are a myriad of different effects which need to be acknowledged, but are approximated as being zero, as being implicitly included in the probabilities within the model or as being equal for all teams involved. For the purpose of future expansions to this model consideration of whether a team is winning or losing and its effect on morale and motivation could be added. If one team is winning or losing, either by a great deal or a small amount or the game is close the probabilities of success may well change due to psychological reasons. These probabilities may also change due to the importance of the game. This is hard to represent using this simple base model; however it is an area in which the model could be further expanded. Hypothetically this could be represented using an extra construct representing the effect motivation and morale of the different teams have to alter performance in each of these: situations. It would be recommended to use an adaptation of the surrender parameter, λ (Rump 2008) for psychological issues such as these. Player fatigue is also a factor that this model has not considered. It is something that is very relevant to the coach who wishes to optimally manage fatigue while ensuring the best players are in the key field positions at the appropriate time and for as long as they will continue to perform more optimally than their non-fatigued replacements. The differences between different players themselves and their abilities to execute certain key plays represent another factor for future expansions to the model.

The model can be adapted for a "park touch", or another tournament if rule variations are used by adding additional states. The adaption to the basic framework of the model may not however be necessary for most rule variations however as they do not change the categorization of states. The change would most likely be implemented by minor alteration to the probability function representing transfer between states. Potentially this could also be used for modelling the implications of rule variations, should changes be considered in the future.

A match of many sports is not strictly a Markov Chain. This is because in executing a particular move or action, the reaction/response of the opposition will be influenced by the previous actions of both teams in that position and their outcomes. This is something attacking teams are aware of and purposefully incorporate into their tactics. An example would be conditioning a defence to react to a particular action with a particular response and to begin anticipating that they must respond in a certain way. Once this is achieved the same action is executed by the attacker, but altered at the critical moment in the hope that it is too late for the defender to change his anticipated response. For teams that have played extensively against each other or reviewed each other's performances extensively, this may then not apply to the same extent as in the situation where they have no prior experience of the way the other team plays. In the situation where the team has a set structure that must be followed at all times by the players and provided that this system is not altered in the game as a response to the actions of the opponents or their own failed actions, then this game can be modelled as a Markov Chain. This would also entitle modelling as a Markov Chain a game for a team which had a set structural policy to follow, a policy, which would only be reviewed after the game or a game conducted amongst players with too limited games sense (Piltz, 2003) in order to make tactical changes throughout the game as play progressed. At a higher level it is presumed that leading players and/or coaching staff would possess sufficient ability to make such changes, however if the optimal offensive and defensive patterns had been selected against a particular opponent then a run of unlikely plays, resulting in a sequence of play which was a low probability outcome would not result in changing the patterns for a highly astute coach may identify this for what it is, namely a highly unlikely sequence of events. Therefore, in this case the use of a Markov Chain based methodology would still hold as valid.

Further consideration needs to be given to the distance required for a successful attack. We need to consider scoops, which are longer range attacks and one might model these using the probability of gaining a given number of metres and successfully off-loading which can be treated similarly to a new type of scoring play with a corresponding defence score for each team. For an attack from a dynamic platform this requires far less distance from the score line. Further work needs to evaluate the probability changes caused by preceding attacks with strike dumps as a component of the attacking structure. Particularly focus should be placed on attack types as there is the possibility of attacking several times from some sets of six provided that the first attacks retain the ball if unsuccessful (e.g. quick release plays as opposed to some failed long ball plays which may result in a turn-over due to a failure to deliver the ball appropriately to the receiving player.). Therefore different teams, but particularly different attack types will have different probability outcome functions requiring further state divisions to account for several attacks a team possesses. At the moment however, the model views the outcomes as probabilities given a short or long range attack, so this is not incorrect, just a more macroscopic view. As field position develops, teams can structure an attack using one of these three methods as is relevant: a scoop, a strike dump or primary attack followed by a second phase attack, or an attack from a dynamic driving platform.

Stationarity is challenged under certain conditions in the game of touch. For example a team that has just been scored against at the start of a game is unlikely to change its game plan, however a team losing by 1 touchdown with one minute left of the game would likely play a more high risk form of game in order to try and score on its next set of touches. This singularity is also challenged due to the half and full time state changes, however by consideration of these as special situations and keeping track of time as part of the model (effectively time and displacement are further parallel sub-states) this can be accommodated within the parameters of the model. Eventually it is hoped that GPS data combined with real time video analysis could be used for conjoint theoretical and practical future development of the model.

5. CONCLUSIONS

Due to the large number of touch players in Australasia, the potential international growth of the sport and the relative absence of research, especially mathematical modelling of this sport, there are numerous reasons for developing this particular study. This paper provides a basic framework for modelling touch football as a Markov Chain. Due to the absence of previous work in this area, the focus has been entirely on developing a framework for this model, without specifically detailing transition probabilities. This however is the next stage in the process of building a workable and practical model. A simple score by score workable method could easily be developed, however it is hoped that this model, while requiring further work and less approximations will eventually be more insightful due to more intensive nature of transitions.

Acknowledgements

The author would like to thank those he has played touch football with. Primarily Manu Wakely, coach of numerous teams the author has been fortunate enough to be involved with, for sharing highly insightful and informative technical analysis of the sport of touch football as well as answers to some specific questions relative to this analysis.

References

Australian Human Rights Commission, (2007). *Touch football*. Retrieved from http://www.hreoc.gov.au/racial_discrimination/whats_the_score/pdf/touch_football.pdf

Clarke, S., (1979). Tie point strategy in American and international squash and badminton. *The Research Quarterly*, *50*, 729-734.

Croucher, J., (1982). The effect of the tennis tie-breaker. *Research Quarterly for Exercise and Sport*, *53*, 336-339.

Federation International Touch Inc. (2003). *Playing rules with explanations and interpretations 3rd Edition*. Curtin, Australia: F.I.T.

Mathesius, P., & Strand, B., (1994). Touch rugby: an alternative activity in physical education. *The Journal of Physical Education, Recreation & Dance*, *65*.

Newton, P., Aslam, K., (2009). Monte Carlo tennis: a stochastic Markov Chain Model. *Journal of Quantitative Analysis in Sports*, *5*(3).

Noubary, R., (2007). Probabilistic analysis of a table tennis game. *Journal of Quantitative Analysis in Sport*, *3*(1)

Philips, J., (2000). Sport and future Australasian culture. *International Journal of the History of Sport*. *17*(2&3), 323-332.

Piltz, W., (2003). Teaching and coaching using a 'play practice' approach. In Butler, J., Griffin, L., Lombardo, B., & Nastasi, R., (Eds.), *Teaching games for understanding in physical education and sport*, 189-200. Reston, VA.: National Association for Sport and Physical Education.

Pollard, G., (1985). A statistical investigation of squash. *Research Quarterly for Exercise and Sport*, *56*, 144-150.

Pollard, G., (1987). A new tennis scoring system. *Research Quarterly for Exercise and Sport*, *58*, 229-233.

Rump, M., (2008). Data clustering for fitting parameters of a Markov Chain model of multi-game playoff series. *Journal of Quantitative Analysis in Sports*, *4*(1).

Schutz, R., (1970). A mathematical model for evaluating scoring systems with specific relevance to tennis. *The Research Quarterly*, *41*, 552-561.

Schutz, R., & Kinsey, W., (1977). Comparison of North American and international squash scoring systems – a computer simulation. *The Research Quarterly*, *48*, 248-251.

Touch Football Australia (2007). *Playing rules and referees signals 7th Edition*. Australia: Touch Football Australia.

Touch Football Australia, (2010). *Touch Football Australia Website*. Retrieved from http://www.austouch.com.au./index.php?id=185

Touch New Zealand, (2010). *Touch New Zealand Website*. Retrieved from http://www.touchnz.co.nz/index.php?id=12

Walsh, J. (2010). *An investigation into the possibility of theoretical modelling for the purposes of examining the non-associative nature of probabilities of game based skill tasks within competitive sports matches: Proceedings of the tenth Australasian Conference on Mathematics and Computers in Sport, Charles Darwin University* (pp205-212). Darwin, Australia.

Wright, M., (1988). Probabilities and decision rules for the game of squash rackets. *Journal of the Operational Research Society*, *39*, 91-99.

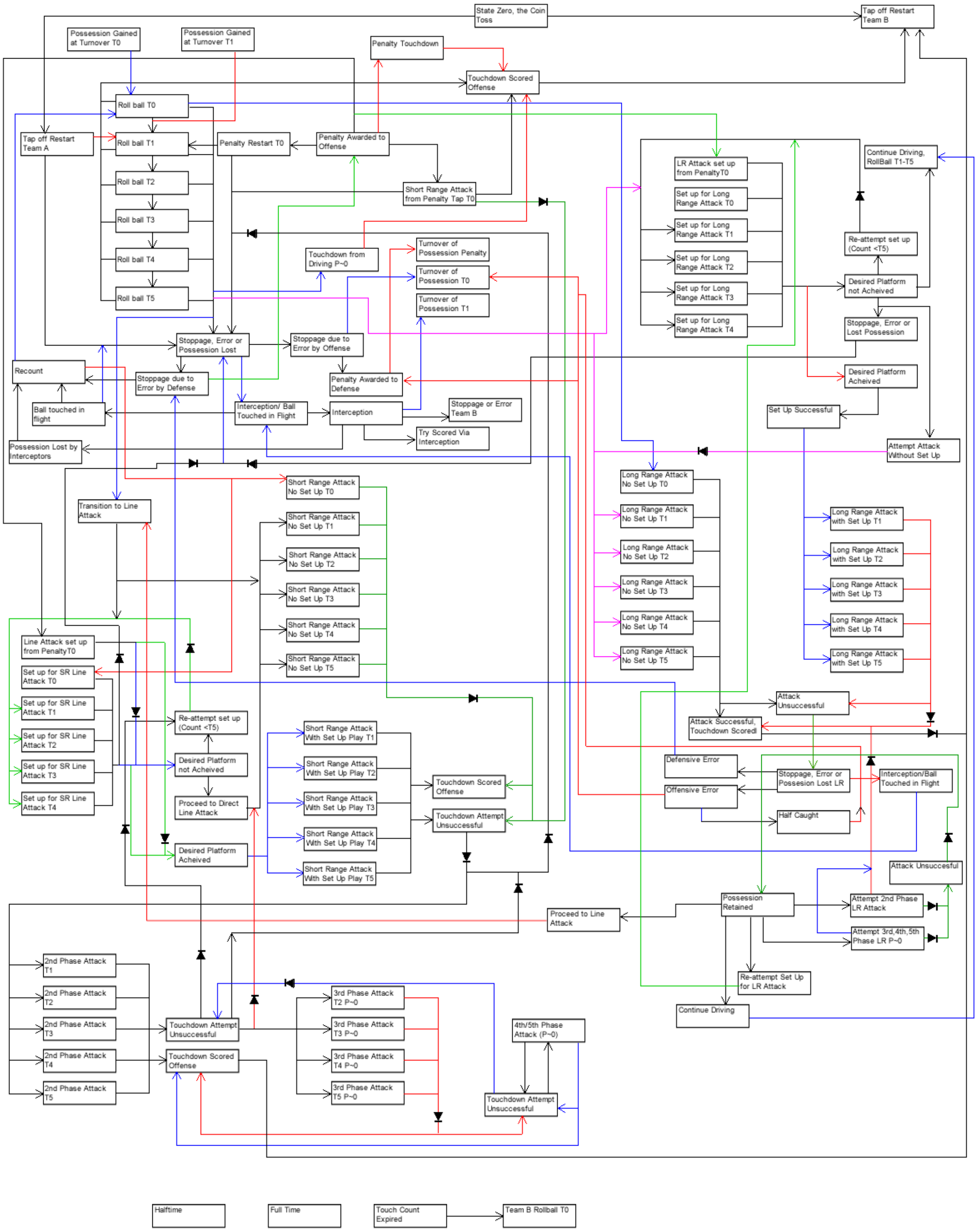


Figure 1: A Markov States Model for the Game of Touch Football. Arrows and Diodes indicate direction of flow. Several state transitions have been coloured in this diagram for clarity.