Proceedings of the Fifth Australian Conference on MATHEMATICS AND COMPUTERS IN SPORT

Edited by

Graeme Cohen and Tim Langtry
Department of Mathematical Sciences
Faculty of Information Technology
University of Technology, Sydney

5M&CS

14 – 16 June 2000

Proceedings of the Fifth Australian Conference on MATHEMATICS AND COMPUTERS IN SPORT

Edited by Graeme Cohen and Tim Langtry
Department of Mathematical Sciences
Faculty of Information Technology
University of Technology, Sydney

5M&CS

14 – 16 June 2000

Typeset by $\LaTeX\,2_{\varepsilon}$

Proceedings of the Fifth Australian Conference on Mathematics and Computers in Sport, 14–16 June 2000, at the University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia.

ISBN 1 86365 535 2

Printed by UTS Printing Services

Contents

Preface v

Participants vii

Email addresses viii

Invited Papers

Mathematics and physics in ball sports
Rod Cross 1

Biomechanics as an ergogenic aid in strength training
Mel C. Siff 9

Contributed Papers

Methods for quantifying performances in one-day cricket
Paul Allsopp and Stephen R. Clarke 28

Identifying arbitrage opportunities in AFL betting markets through mathematical modelling
Michael Bailey 37

A comparison of the ATP ratings with a smoothing method for match prediction
Anthony B. Bedford and Stephen R. Clarke 43

Dynamics of the long jump
Maurice N. Brearley and Neville J. de Mestre 52

Teaching statistics through sport
K. J. Byun 58

Variations in random pairing finals systems
George Christos 66

Home advantage in the Olympic Games
Stephen R. Clarke 76

An annotated bibliography of mathematical articles on sport amenable to elementary university mathematics teaching
G. L. Cohen 86

Player ratings in one-day cricket
John S. Croucher 95

The flight of a cricket ball
Neville de Mestre and Graeme Cohen 107

Applications of the Duckworth–Lewis method
B. M. de Silva, G. R. Pond and T. B. Swartz 113

Curvature effects in running the 200 and 400 metre sprints
Chris Harman 118

Simulation of frisbee flight
M. Hubbard and S. A. Hummel 124

An analysis of the AFL final eight system
J. Lowe and S. Clarke 135

The use of computers to predict the performance of golf clubs
J. S. B. Mather 143

Modelling endurance time at $VO_{2\,\max}$
R. Hugh Morton and Veronique Billat 156

Home ground advantage in the Australian Netball League (Commonwealth Bank Trophy)
Pam Norton and Stephen R. Clarke 168

Optimising downwind sailing
Elliot Tonkes 174

A statistical approach for developing weighted norms based on difficulty ratings on selected fitness parameters for boys in the age group 13–16 years
J. P. Verma 184

Extreme value models in sporting data analysis
J. Z. Wang and P. Cooke 192

Abstracts

Cricket bowling: a two-segment Lagrangian model
R. E. D. Ferdinands, K. A. Broughan and H. Round 197

A greyhound racing betting strategy
L. H. Liyanage, S. Terry and N. Hopkins 198

Seasonal variations and cricket's rule changes
Iain Skinner 199

Preface

The four previous conferences on Mathematics and Computers in Sport (in 1992, 1994, 1996 and 1998) were held at Bond University, on the Gold Coast, Queensland. The Conference Director on each of those occasions was Neville de Mestre, to whom the sporting community as much as the mathematics community owes a huge debt of gratitude.

It was decided in 1998 that the fifth conference should be held in Sydney, to let its participants take in the spirit of the 24th Olympic Games commencing in about three months time. Without doubt, six years of building activity for a host of new stadiums and five-star hotels, the concurrent roadworks, the new railway line and the sprucing up of Sydney have led its residents and visitors to an eager anticipation of the Games' commencement. A tour of the major Olympic Games sites will be a highlight of the conference for many of its participants.

We have dubbed this fifth conference 5M&CS. As on previous occasions, the range of sports and associated analytical techniques to be presented is huge. From frisbees to netball, the long jump to sailing, golf to sprinting, along with the Australian standards of cricket and Australian Rules football, it seems that just about every sport gets a mention. The treatments of those sports include the biomechanical, the physical, the mathematical, the statistical and the educational.

These *Proceedings* begin with the invited papers of our guests, Rod Cross and Mel Siff, followed by twenty contributed papers, arranged in alphabetical order of author, or first author. All these papers have been refereed, and I take this opportunity to thank the referees who have had to work within a very tight time frame. The *Proceedings* end with three further abstracts; the full papers are presented as part of the conference program, but are not reproduced here (mostly because the final submissions were too late for the refereeing process).

For assistance in the organisation of 5M&CS, I am very grateful to the Conference Chair, Neville de Mestre, to Ron Sorli for managing the website, and to Jesu Román for secretarial help. I am particularly grateful also to Tim Langtry, co-editor of the *Proceedings*, for his technical knowledge and skills. As on all previous occasions, the Australian Mathematical Society, through ANZIAM, has again been generous in its financial support.

Graeme Cohen
Director, 5M&CS

Participants

Mr Paul ALLSOPP	Swinburne University
Mr Michael BAILEY	Swinburne University
Mr Anthony BEDFORD	Swinburne University
Emeritus Professor Maurice BREARLEY	Clifton Springs, Victoria
Ms Kyung-Jiu BYUN	Macquarie University
Mr Matt CARRE	University of Sheffield
Mr Steve CHADWICK	University of Sheffield
Dr George CHRISTOS	Curtin University of Technology
Associate Professor Stephen CLARKE	Swinburne University
Associate Professor Graeme COHEN	University of Technology, Sydney
Dr Peter COOKE	University of New South Wales
Dr Peter COUTIS	Queensland University of Technology
Associate Professor Rod CROSS	University of Sydney
Professor John CROUCHER	Macquarie University
Professor Neville DE MESTRE	Bond University
Associate Professor Basil DE SILVA	RMIT University
Mr Richard DIGNALL	University of Sheffield
Mr Rene FERDINANDS	University of Waikato
Mr Jonathan GODFREY	Massey University
Mr Simon GOODWILL	University of Sheffield
Dr Steve HAAKE	University of Sheffield
Associate Professor Chris HARMAN	University of Southern Queensland
Ms Mary HEWETT	University of Canberra
Professor Mont HUBBARD	University of California, Davis
Associate Professor Kuldeep KUMAR	Bond University
Dr Shan LI	University of Canberra
Dr Nick LINTHORNE	University of Sydney
Dr Liwan LIYANAGE	University of Western Sydney
Mr Jonathan LOWE	Swinburne University
Dr Steve MATHER	University of Nottingham
Associate Professor R. Hugh MORTON	Massey University
Ms Pam NORTON	Monash University
Dr Martyn SHORTEN	Biomechanica, Oregon
Dr Mel SIFF	Littleton, Colorado
Dr Iain SKINNER	University of New South Wales
Dr Elliot TONKES	Bond University
Dr J. P. VERMA	Lakshmibai National Institute of Physical Education
Dr Julian WANG	University of Western Sydney

Email addresses

Mr Paul ALLSOPP	paulalls@hotmail.com
Mr Michael BAILEY	mbailey@swin.edu.au
Mr Anthony BEDFORD	abedford@swin.edu.au
Emeritus Professor Maurice BREARLEY	—
Ms Kyung-Jiu BYUN	kbyun@efs.mq.edu.au
Mr Matt CARRE	m.j.carre@sheffield.ac.uk
Mr Steve CHADWICK	mep97sgc@sheffield.ac.uk
Dr George CHRISTOS	christos@cs.curtin.edu.au
Associate Professor Stephen CLARKE	sclarke@swin.edu.au
Associate Professor Graeme COHEN	g.cohen@maths.uts.edu.au
Dr Peter COOKE	pjc@maths.unsw.edu.au
Dr Peter COUTIS	p.coutis@qut.edu.au
Associate Professor Rod CROSS	cross@physics.usyd.edu.au
Professor John CROUCHER	john.croucher@efs.mq.edu.au
Professor Neville DE MESTRE	ndemestr@bond.edu.au
Associate Professor Basil DE SILVA	desilva@rmit.edu.au
Mr Richard DIGNALL	mep97rjd@sheffield.ac.uk
Mr Rene FERDINANDS	redf1@waikato.ac.nz
Mr Jonathan GODFREY	a.j.godfrey@massey.ac.nz
Mr Simon GOODWILL	mep98srg@sheffield.ac.uk
Dr Steve HAAKE	s.j.haake@sheffield.ac.uk
Associate Professor Chris HARMAN	harman@usq.edu.au
Ms Mary HEWETT	maryh@ise.canberra.edu.au
Professor Mont HUBBARD	mhubbard@ucdavis.edu
Associate Professor Kuldeep KUMAR	kkumar@bond.edu.au
Dr Shan LI	shan@ise.canberra.edu.au
Dr Nick LINTHORNE	nick.linthorne@cchs.usyd.edu.au
Dr Liwan LIYANAGE	l.liyanage@uws.edu.au
Mr Jonathan LOWE	jlowe@swin.edu.au
Dr Steve MATHER	stephen.mather@nottingham.ac.uk
Associate Professor R. Hugh MORTON	h.morton@massey.ac.nz
Ms Pam NORTON	pnorton@groupwise.swin.edu.au
Dr Martyn SHORTEN	martyn.shorten@biomechanica.com
Dr Mel SIFF	mcsiff@aol.com
Dr Iain SKINNER	i.skinner@unsw.edu.au
Dr Elliot TONKES	etonkes@bond.edu.au
Dr J. P. VERMA	venpub@gwr1.dot.net.in
Dr Julian WANG	j.wang@nepean.uws.edu.au

MATHEMATICS AND PHYSICS IN BALL SPORTS

Rod Cross
Physics Department
University of Sydney
NSW 2006, Australia

Abstract

The study of sport involves a host of disciplines, including physics and mathematics. Some examples of the latter are presented, including the collision between a bat and a ball, and a study of proposals to slow the serve in tennis.

1 Introduction

The purpose of this paper is to describe a few problems encountered in ball sports, as viewed through the eyes of a physicist. The basic tools available to solve such problems include those commonly used in theoretical and experimental physics, as well as some mathematical and computational tools. The emphasis is therefore on the equipment used in these sports rather than on the person using the equipment, unless the person can be approximated as a spherical or point object. The sports person is more usually studied by biomechanists, psychologists or physiologists. Nevertheless, one can find some interesting physics as well as interesting mathematics in human and animal movement. For example, the centre of mass of a walker rises and falls in a sinusoidal fashion, and describes an arc of radius $R \approx 1\,\mathrm{m}$ about the foot on the ground (Cross [6]). Consequently, the force on the ground, $F = Mg - Mv^2/R$ drops to zero when the walking speed $v \approx 3\,\mathrm{m}^{-1}\mathrm{s}$. If you walk any faster, you will become weightless and start running. This explains why little kids with short legs have to run to keep up with their fast walking parents. The motion of the centre of mass can be analysed using Fourier techniques. Another interesting problem is why walkers and runners travel at almost constant speed, without continuous acceleration, despite the fact that they keep pushing backwards with their back foot in order to move forwards. The reader is challenged to find a simple answer.

A sample of books and papers on the physics of sports is given in the list of references. Armenti [2] includes 303 references. In the remainder of this paper, we will concentrate on (a) the impact of a ball with a bat or racket and (b) some proposals to change the rules of tennis in order to slow down the serve.

2 Impact of a ball with a bat or racket

The collision between a tennis racket (or a cricket or baseball bat) and a ball can be modelled by assuming that the racket is perfectly rigid and that the handle is not subject to any impulsive force during the collision (Brody [4, 5]). The racket and ball speeds after the collision can then be calculated from the conservation equations, in terms of the corresponding speeds before the collision, and an assumed or measured coefficient of restitution. However, the model provides no information on the dynamics during the collision, nor does it provide information on the energy coupled to racket vibrations. These shortcomings can be avoided by modelling the racket as a one-dimensional, flexible beam. The beam model also provides information on the effect of the impulse reflected from each end of the racket back to the ball. For example, if the reflected impulse arrives after the ball leaves the racket, then the ball has no way of knowing if the handle was free or hand-held, or even if it landed at a vibration node.

The equation of motion for a beam subject to an external force, F_o per unit length, has the form (Graff [11], Cross [7])

$$\rho A \frac{\partial^2 y}{\partial t^2} = F_o - \frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 y}{\partial x^2}\right), \tag{1}$$

where ρ is the density of the beam, A is its cross-sectional area, E is Young's modulus, I is the area moment of inertia, and y is the transverse displacement of the beam at coordinate x along the beam. For a uniform beam of mass M and length L, numerical solutions of (1) can be obtained by dividing the beam into N equal segments each of mass $m = M/N$ and separated in the x direction by a distance $s = L/N$. An impacting ball may exert a force acting over several adjacent segments, depending on the ball diameter and the assumed number of segments. For simplicity it is assumed that the ball impacts on only one of the segments, exerting a time-dependent force, F. The equation of motion for that segment (the nth segment) is obtained by multiplying all terms in (1) by s, in which case

$$m\frac{\partial^2 y_n}{\partial t^2} = F - EIs\frac{\partial^4 y_n}{\partial x^4}, \tag{2}$$

assuming that the beam is uniform so that E and I are independent of x. The equation of motion for the other segments is given by (2) with $F = 0$. The boundary conditions at a freely supported end are given by $\partial^2 y/\partial x^2 = 0$ and $\partial^3 y/\partial x^3 = 0$. The boundary conditions at a rigidly clamped end are $y = 0$ and $\partial y/\partial x = 0$. The equation of motion of the ball is given by

$$\frac{d^2 y_b}{dt^2} = -\frac{F}{m_b}, \tag{3}$$

where m_b is the ball mass and y_b is the displacement of the centre of mass of the ball. The elastic properties of the ball can be modelled by assuming that $F = k_1 Y_b$ during the compression phase and $F = k_2 Y_b^p$ during the expansion phase, where Y_b is the compression of the ball and p is a parameter describing the effect of hysteresis in the ball. If k_1 and k_2 are constants and if Y_o is the maximum compression of the ball during any given impact, then $k_1 Y_o = k_2 Y_o^p$ so $k_2 = k_1/Y_o^{p-1}$. The hysteresis loss in the ball is equal to the area enclosed by the F vs Y_b curve for a complete compression and expansion cycle. The parameter p can be chosen to give a total ball loss equal to the experimentally determined loss.

If a soft ball impacts on a hard beam, so that the compression of the beam is negligible, then $Y_b = y_b - y_n$. Alternatively, if the ball impacts on the strings of a racket, and if the string plane is displaced by a distance Y_s relative to the frame during the impact then F is given by $F = k_s Y_s$, where k_s is the spring constant of the strings. In this case, the compression of the ball plus the compression of the strings is given by $Y_T = y_b - y_n$. By equating the forces, it is easy to show that $Y_b = k_s Y_T/(k_1 + k_s)$ during the compression phase, and $Y_b + k_2 Y_b^p/k_s = Y_T$ during the expansion phase.

It is assumed that the ball is incident at right angles to the string plane. It is also assumed that at $t = 0$, $y_b = 0$, $y = 0$ for all beam segments, the beam is initially at rest and that $dy_b/dt = v_1$. The subsequent motion of the ball and the beam is evaluated by numerical solution of equations (2) and (3). These results are used to determine the rebound speed of the ball, v_2, and the apparent coefficient of restitution (ACOR), $e_A = v_2/v_1$. In normal play, a racket is swung towards the ball and is not normally at rest at the moment of impact. The resulting outgoing speed of the ball is easily determined by a simple coordinate transformation from the racket to the court reference frame.

Numerical solutions of the above equations were tested against experimental results. A rectangular cross-section aluminum beam of width 32 mm and thickness 6 mm was supported horizontally, either by a 1.2 m vertical length of string attached to each end or by clamping one end to a rigid support. A 36 mm diameter, 42 gm superball was mounted, as a pendulum bob, at the apex of a V-shaped string support, so that it could impact the beam horizontally and at right angles to the beam. The incident and rebound speeds of the ball were measured using a He–Ne laser beam directed parallel to the aluminium beam. From this data, measurements were obtained of the ACOR as a function of impact location along the beam. The ball was incident at low speed, about $1\,\mathrm{ms}^{-1}$.

Measurements of e_A, and the corresponding theoretical estimates of e_A, are shown in Figure 1 for an aluminum beam of length $L = 110$ cm. The spring constant of the ball was taken as $k_b = 2 \times 10^4\,\mathrm{Nm}^{-1}$ to be consistent with the observed impact duration, about 4.2 ms. Agreement between the theoretical and experimental values of e_A is remarkably good. The results show that, for a sufficiently long beam, (a) the impact of a ball near one end of a beam is not affected by the length of the beam or the method of support at the other end, and (b) e_A for an impact anywhere along the central section of a beam is independent of the impact location and is not affected by the length of the beam or the method of support. The quantity e_A remains constant at $e_A = 0.45 \pm 0.02$ along the beam up to a point about 15 cm from each end. This result implies that the rebound speed is affected only if the impulse reflected off one end arrives back at the impact point within the 4.2 ms period of the impact.

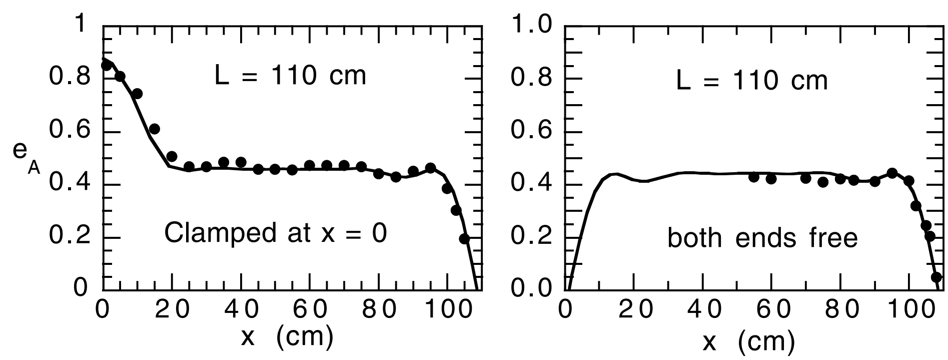

Figure 1: e_A vs x for 110 cm long aluminium beams. Experimental data is shown as black dots, and the solid curves are solutions of (2) and (3).

The dynamics of the situation are illustrated in Figure 2 which shows the theoretical beam deflection, for a freely supported beam, at equal time increments during and shortly after the impact. An impulse propagating towards a free end is reflected without phase reversal, so the beam moves further away from the ball, thereby reducing the rebound speed. A pulse propagating towards a clamped end is reflected with a phase reversal, sending the beam back towards the ball, thereby increasing the rebound speed. The reflected pulse has no effect on the ball if the ball rebounds before the reflected pulse reaches the ball. Figure 2 shows that the beam deflection at the impact point, during the impact, is essentially the same for impacts at $x = 55$ cm or $x = 90$ cm, indicating that the pulse reflected at $x = 110$ cm does not have a significant influence on the impact. However, for an impact at $x = 103$ cm, the reflected pulse acts to deflect the beam away from the ball during the impact, thereby reducing the ACOR significantly.

Calculations are shown in Figure 3 for a graphite/epoxy composite tennis racket of length $L = 71$ cm and mass $M = 340$ gm, modelled as a uniform beam with $EI = 150$ Nm, giving a fundamental vibration frequency of 125 Hz. The ball was modelled with $m_b = 57$ gm, $k_1 = 3 \times 10^4\,\mathrm{Nm}^{-1}$ and $p = 2.55$, corresponding to an impact duration of 4.64 ms on concrete and a COR = 0.751. Figure 3 shows the variation of e_A along the long axis of the racket (passing through the handle) when the handle is (a) rigidly clamped along a 10 cm length at the end of the handle, or (b) pivoted about an axis through the end of the handle, or (c) freely suspended. The results show that e_A is zero at a "dead" spot near the tip of the racket and increases to a broad maximum near the throat of the racket. This behaviour is easily demonstrated experimentally, at least in a qualitative sense, simply by dropping a ball at various spots on the strings and observing the bounce. e_A is independent of the handle support at distances up to 20 cm from the tip of the racket. e_A is affected near the throat of the racket as a result of reflections from a clamped handle, but this is of no consequence in that the hand does not act as a rigid clamp.

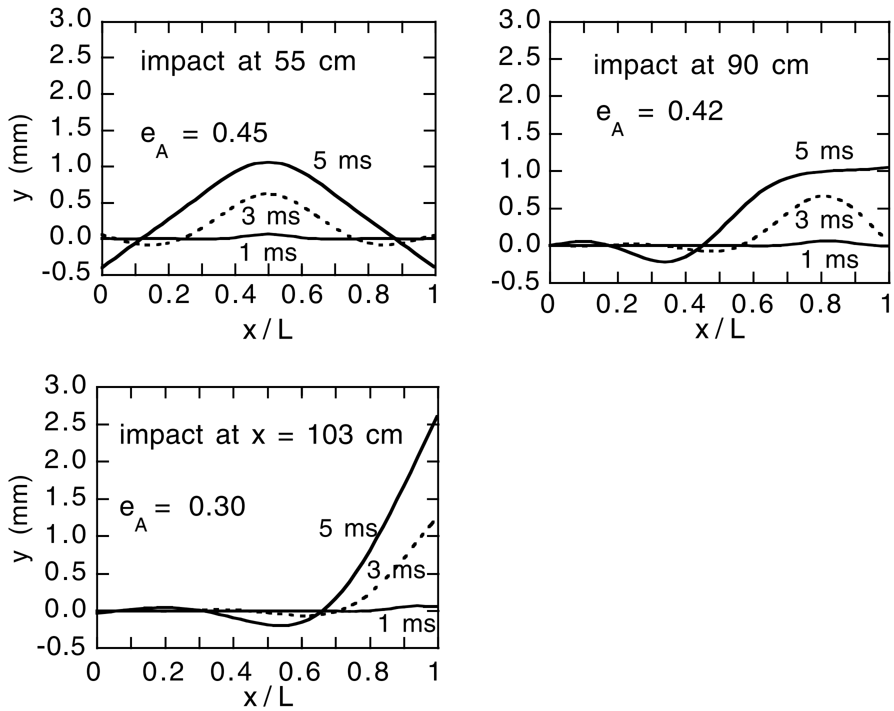

Figure 2: y displacement of a freely suspended beam.

In practice, the racket tends to pivot about an axis through the wrist (Cross [8]).

The change in e_A with string plane stiffness, k_s, is relatively small considering the range in k_s normally used in a tennis racket. For example, for an impact at the centre of the strings, $e_A = 0.44$ when $k_s = 2 \times 10^4\,\mathrm{Nm}^{-1}$, while $e_A = 0.41$ when $k_s = 4 \times 10^4\,\mathrm{Nm}^{-1}$. This corresponds to a 7% change in the ball speed when it bounces off a racket that is initially at rest. Since the ball rebounds at speed $v_2 = e_A v_1$, then in a reference frame where the ball is initially at rest, the racket will be incident at speed v_1 and the ball will come off the strings at speed $v = v_1 + v_2 = (1 + e_A)v_1$. In this case, corresponding to a serve or overhead smash, the above increase in e_A results in only a 2% increase in the ball speed. Given that k_s is proportional to the string tension, then a reduction in string tension from say 60 to 50 lb will increase the ball speed by only 0.7%. A more significant increase in racket power can be achieved by increasing the stiffness of the frame, at least for impacts near the tip of the racket. An increase in frame stiffness has no effect on racket power for an impact at the vibration node near the centre of the strings.

3 Proposals to change the serve speed in tennis

There has been a lot of discussion about slowing down the speed of the serve on fast courts such as the grass courts at Wimbledon. There it has become a first-serve contest, comparable to the quick-draw gun fights in the old west. Some players win 40% of their good first serves as aces. The serve has become so dominant at Wimbledon and service breaks so rare, that some players have over 30% of their sets end in tie-breaks.

The speed and trajectory of a ball can be calculated in terms of measured drag and lift coefficients (Haake *et al.* [12]). A 57 gm, 65 mm diameter standard ball served at 200 km/hr takes 0.594 s to cross the baseline if it bounces off a clay surface, and 0.568 s if it bounces off grass. These figures are shown in Table 1, together with data for a serve speed of 180 km/hr, plus data for larger diameter balls. In

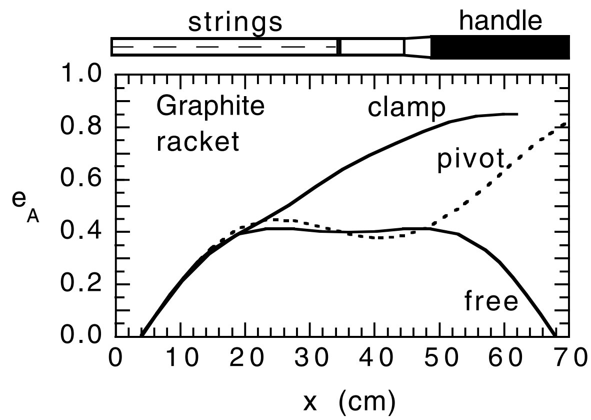

Figure 3: e_A vs x for a tennis racket.

Table 1, θ is the vertical angle of the ball, served down from the horizontal, so that the ball lands on the service line. It is assumed that the ball is served by a tall player from a height of 2.9 m above the centre line and is served down the centre of the court to land on the service line. It was also assumed that clay has a coefficient of friction $\mu = 0.7$, and that $\mu = 0.4$ for grass. The transit time on grass is 0.063 s longer at the lower serve speed, giving the receiver a significantly better chance of returning the ball. If the receiver is running at 6 ms^{-1} to reach the ball, he can cover an extra 38 cm in 0.063 s.

D (mm)	v (km/hr)	θ	Surface	T (sec)
65	200	6.92°	Clay	0.594
65	200	6.92°	Grass	0.568
65	180	6.43°	Grass	0.631
69	200	6.85°	Grass	0.586
72	200	6.80°	Grass	0.600

Table 1: Transit times

Some of the methods proposed to reduce the serve speed, so that the receiver has more time to react to the serve, are as follows:

1. Change the surface. Replacing grass with a slow surface will clearly solve the problem. However, the grass courts at Wimbledon are the heart and soul of "lawn" tennis.

2. Limit the racket power. This would be unpopular, and no-one knows how to do it since it depends more on racket mass than anything else.

3. Eliminate the second serve. It may prove beneficial for a player such as Sampras to hit first serves only as opposed to second serves only. The result would be about 35% of the points on serve would end in a fault and the others would be a one or two hit rally. This could really kill tennis as a spectator sport.

4. Change the footfault rule. A number of years ago the serve footfault rule was changed. At that time, foot contact with the court surface had to be maintained until after the ball was struck. The present rule allows the server to leave the ground and actually be well into the court beyond the baseline. This allows the server to impact the ball about 15 cm higher and up to 60 cm into the court. It also allows the server to get to the net quicker.

5. Change the ball. If the ball is made lighter, it will come off the racket at a higher speed, but it will slow down more in getting to the receiver due to air resistance. If the ball is made heavier, it will come off the racket at a lower speed, but it will slow down less due to air resistance. For a small change in ball mass, the two effects tend to cancel each other out. Reducing the COR of the ball has very little

effect on serve speed since the serve speed depends primarily on the racket speed. An increase in ball diameter is much more effective, due to the increased drag on the ball. With a larger ball, the extra time the receiver has to return a serve on a grass court is comparable to the extra time the receiver has if the ball bounces on clay (see Table 1).

It is difficult to predict with certainty the effect of using a larger ball on grass. A 69 mm ball introduces a delay of 0.018 s for a 200 km/hr serve, which is equivalent to serving a standard ball at 194 km/hr. A 72 mm ball introduces a 0.032 s delay, which is equivalent to a standard ball served at 189 km/hr. Figures 4–6 indicate that on grass, this will make a relatively minor difference to the fraction of unplayable serves. As shown in Table 1, the transit time for a 200 km/hr serve on clay, for a standard ball, is 0.026 s longer than on grass. The larger ball will introduce a comparable delay, but one cannot assume that the same delay on grass will have the same effect as on clay.

6. Shorten the service court. If the service court is shortened, it would force the big servers to reduce their serve speed in order to get in a reasonable fraction of their offerings. Calculations of ball trajectories show that if the service court is reduced by 28 cm, then a 1.93 m (6' 4") player would need to reduce his serve speed from 193 km/hr to 175 km/hr in order to maintain the same serve angle window and hence the same percentage of successful first serves. A 1.75 m (5' 9") player needs to serve at 155 km/hr on a normal size court to maintain the same serve percentage as the tall player. If the service court is shortened by 28 cm, the shorter player could maintain the same serve percentage by reducing his speed to 145 km/hr.

Figure 5 indicates that the percentage of serves won is almost independent of serve speed when the serve speed is low, especially on the second serve.

7. Raise the net. Raising the net height will also force players to reduce their serve speed in order to maintain a reasonable serve percentage. Each 1.0 cm increase in net height is equivalent to a reduction in the length of the service court by 7.6 cm, regardless of the initial ball speed or the initial serve height.

8. Other proposals. The server could be required to be well behind the baseline when serving. If the server stands 1 m behind the service line, then the transit time to the receiver's baseline is increased by 0.02 s at a serve speed of 180 km/hr (50 ms^{-1}). This would also force players to reduce the serve speed to maintain the same serve percentage. However, it could mean the end of the serve and volley game. An interesting proposal has been made to narrow the service box by splitting the centre line. This would also reduce the number of aces.

The effect of these various proposals is difficult to predict with certainty. However, one can make some progress by analysing the available statistical data, shown in Figures 4–6 for the 1999 Wimbledon Open men's singles matches. Figures 4–6 also show data for the 1999 French Open, where matches are played on clay, at a much lower pace. Part of the difference can be attributed to the speed of the surface, which is determined by μ. A 200 km/hr serve on clay takes 0.026 s longer to cross the receiver's baseline. In effect, a 200 km/hr serve on clay is equivalent to a serve speed of about 190 km/hr on grass, in terms of the transit time. However, the main difference between the two surfaces is the rebound angle of the ball. The ball kicks up at a steeper angle on clay, partly because of the larger reduction in the horizontal speed of the ball when it bounces and partly because the coefficient of restitution on clay is larger, meaning that the vertical component of the rebound speed is larger. Most players respond by reducing their serve speed substantially (by about 25 km/hr) at the French Open, in order to apply more spin. A ball with topspin and with a reduced horizontal speed strikes the court at a steeper angle and at a higher vertical speed so it kicks up at an even steeper angle.

The Wimbledon data in Figures 4–6 show that the probability of winning a first serve increases as the serve speed increases, provided the serve is in. However, the probability that a first serve is good decreases as the serve speed increases. The net result, considering all serves, is that players normally win a greater fraction of their second serves than their first serves. Taking an average over all players and all serves, regardless of whether the serve was in or a fault, players who won their matches at Wimbledon won 48% of all their first serves and 56% of all their second serves. Players who lost their matches won 38% of all their first serves and 46% of all their second serves. From this point of view, a reduction in the first serve speed should not be a serious imposition for most players and might even increase the chances of a player winning his first serve, depending on the first serve %. For example, a larger ball will cause the transit time to increase but it will also increase the first serve % due to the larger vertical acceptance angle for good serves. If the net height is raised or if the service court is shortened, the server will be forced to reduce his serve speed in order to maintain the same serve %. Figure 5 shows that the speed of the second serve does not have a significant effect on the chance of winning the second serve, so a reduction in the speed of the second serve should have a relatively small effect.

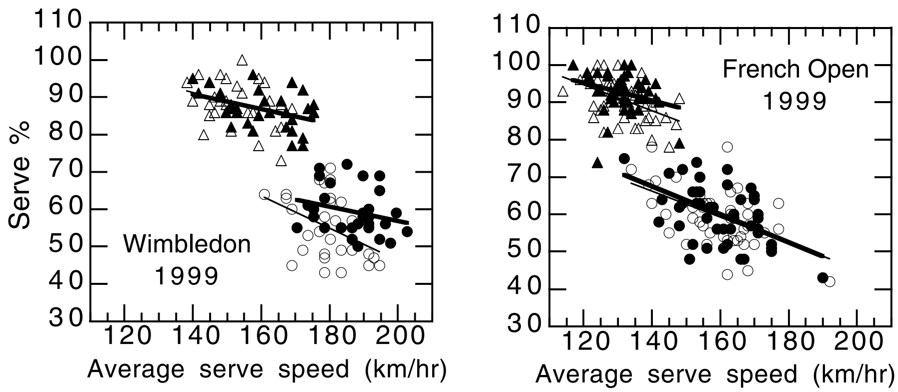

Figure 4: Wimbledon and French Open 1999 men's singles serve % statistics. The serve % is the % of serves that are good (i.e. not a fault). Solid circles = 1st serve (winners), solid triangles = 2nd serve (winners), open circles = 1st serve (losers), open triangles = 2nd serve (losers).

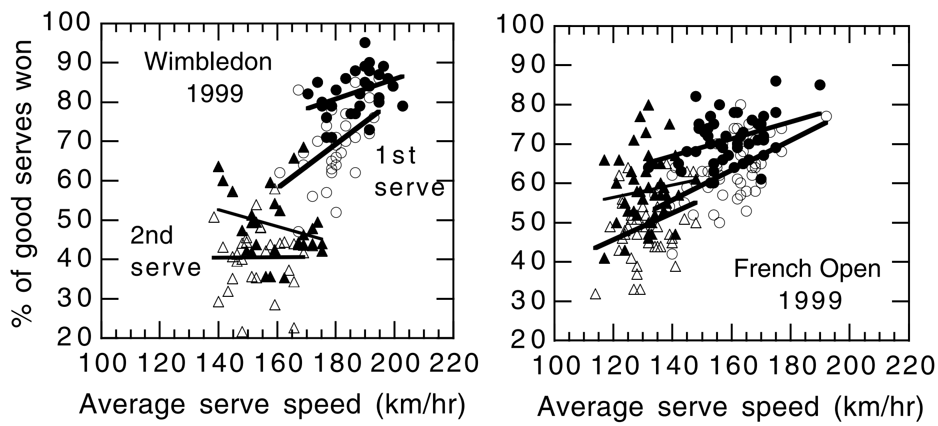

Figure 5: % of good serves won by the server vs serve speed.

References

[1] R. K. Adair, *The Physics of Baseball*, Harper Perennial, New York (1994).

[2] A. Armenti, Jr (editor), *The Physics of Sports*, American Institute of Physics, New York (1992).

[3] H. Brody, *Tennis Science for Tennis Players*, University of Pennsylvania Press (1987).

[4] H. Brody, "The sweet spot of a baseball bat", *Amer. J. Phys.*, **54** (1986), 640–643.

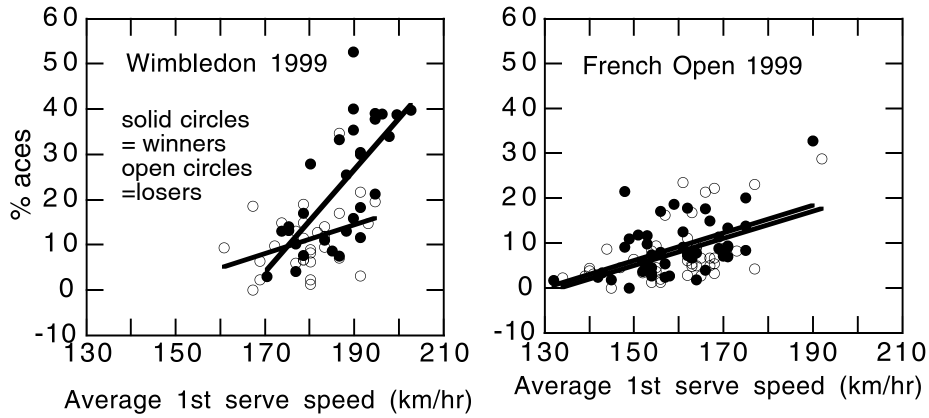

Figure 6: % of good first serves won by aces vs serve speed.

[5] H. Brody, "The physics of tennis III. The ball–racket interaction", *Amer. J. Phys.*, **65** (1997), 981–987.

[6] R. Cross, "Standing, walking, running and jumping on a force plate", *Amer. J. Phys.*, **67** (1999), 304–309.

[7] R. Cross, "Impact of a ball with a bat or racket", *Amer. J. Phys.*, **67** (1999), 692–702.

[8] R. Cross, "The sweet spots of a tennis racket", *Sports Eng.*, **1** (1999), 63–78.

[9] R. Cross, "Flexible beam analysis of the effects of string tension and frame stiffness on racket performance", *Sports Eng.*, to appear.

[10] R. Cross, "The trajectory of a ball in lawn bowls", *Amer. J. Phys.*, **66** (1998), 735–738.

[11] K. F. Graff, *Wave Motion in Elastic Solids*, Oxford University Press, Oxford (1975), 140–210.

[12] S. J. Haake, S. G. Chadwick, R. J. Dignall, S. Goodwill and P. Rose, "Engineering tennis—slowing the game down", *Sports Eng.*, to appear.

[13] T. P. Jorgensen, *The Physics of Golf*, American Institute of Physics, New York (1994).

BIOMECHANICS AS AN ERGOGENIC AID IN STRENGTH TRAINING

Mel C. Siff
7434 S Cherry Ct
Littleton
CO 80122, USA
and
School of Mechanical Engineering
University of the Witwatersrand
South Africa

Abstract

This paper describes how biomechanics may be considered as a type of abstract ergogenic aid that may be applied in sports training to enhance performance through a better understanding and optimal application of the mechanical principles that underlie the production of strength. The importance of the fundamental qualities such as starting-strength, acceleration-strength, explosive strength and reactive ability was established, as was a deeper analysis of the nature of isometric and eccentric muscle action. In particular, it was found that isometric action may be initiated explosively and act as an essential component of stretch-shortening or plyometric actions and that its use need not be confined to maximum strength production. Biomechanical analysis of weightlifting movements emphasised that eccentric prestretch actions may augment strength and power production both under slow and explosive stretch-shortening conditions. Traditional applications of the classical hyperbolic force–velocity curve and Newton's second law were reexamined to establish more appropriate methods of training. Finally, movement against combined free weight–isokinetic and free weight–elastic modes of resistance was examined to assess if such methods could enhance more conventional methods of strength training. The findings that emerged from this work have been successfully applied by a group of elite USA powerlifters, thereby confirming that biomechanics indeed can serve as an effective ergogenic aid in sport.

1 Introduction

Ergogenic aids conventionally are methods or devices such as anabolic steroids or special shoes that enhance sporting performance by their use. Applied mathematics and computers may in a different sense be regarded as similar, though more abstract, ergogenic aids to enhancing performance by enabling the athlete to train and compete more efficiently and safely. All sporting movement depends on the expression of strength in specific forms, patterns, intensities and durations in a given situation. Every movement may be analysed mathematically and computationally to determine the characteristics of efficient and appropriate motion and thereby enable the scientist and coach to devise training methods to enhance the performance of any individual athlete.

These training methods generally are aimed at increasing fitness qualities such as cardiovascular endurance, local muscle endurance, strength, speed and flexibility in as close as possible a balance to what is considered optimal for a given role in a specific sport. In the case of strength development, a great deal of training has been based upon physiological and experiential findings that have focused on increasing muscle mass, on the nature of the different types of fast and slow muscle unit, the process of myofibril contraction and the characteristics of muscle metabolism.

It is the objective of this paper to show that biomechanical analysis can play an equally valuable role in more accurately defining strength as a motor quality, understanding the different forms in which strength may manifest itself, and improving the methods of strength training and injury rehabilitation. The value of this analysis is illustrated by the practical application of some of these findings discussed here being used effectively to enhance the performances of a group of elite athletes.

2 Strength and strength indices

Strength is the ability of a muscle to produce force under specific conditions, where the force of muscle contraction elicited by reflexive or voluntary nervous excitation leads to torque of limbs acting about the various joints as fulcra. All sport involves the production of strength to move the limbs, the whole body or various implements to execute a given motor pattern in a specific type of contact or non-contact play. This strength may be generated at maximal, as is the case in weightlifting and powerlifting, or submaximal levels, as is the case in most other sports, but even in the maximal strength sports just mentioned, the inappropriate generation of high levels of force may be detrimental to the execution of the skilled movements involved in lifting weights competitively.

In other words, it is not simply the production of strength that is important, but also the production of appropriate levels of strength and power at different stages in optimal motor patterns in every sporting movement. Thus, at the very outset, it is vital to emphasise that an increase of strength will not necessarily improve performance in every sport. Certainly in some sports, the quality of absolute strength (strength independent of body mass) may be of relevance, but in most sports, relative strength (relative to body mass) is of far greater consequence.

Another issue of primary concern is that a considerable amount of research into the nature of strength is based on measurement of joint torque using isokinetic dynamometers. This unjustifiably assumes that strength is a generalised motor quality that does not display specificity in sporting movement and that measurement of strength under isokinetic conditions correlates closely with strength production in actual three-dimensional daily movement.

To progress beyond these oversimplified dynamometric analyses of strength and to better understand the role of strength in sport, it is relevant to commence with an analysis of a typical force-time curve and the time derivative of that curve in lifting an unconstrained weight from a given position and returning it to that position (Figure 1).

The time derivative of the $F(t)$ curve, dF/dt, is known as the Rate of Force Development (RFD) or, the Rate of Tension Development (RTD) in the muscle. The graphs enable us to define several fundamental features of strength production, besides maximal strength ($F_{\max}$) given by the peak of the $F(t)$ curve, absolute strength ($F_{\max}$ irrespective of body mass) and relative strength ($F_{\max}$ relative to bodymass, or $F_{\max}/M$). In particular, Russian scientists have used these curves to identify the following (Verkhoshansky [27]):

- starting-strength,
- acceleration-strength,
- strength-endurance.

Starting-strength is the ability to build up working force as rapidly as possible once muscle contraction has begun and it is always produced under isometric conditions. This fact alone has important consequences for strength training, because it dispels the notion that the once-popular method of isometric training should be completely abandoned. On the contrary, the ability to generate starting strength rapidly can exert a profound effect on the dynamics of an entire movement, not only in terms of the magnitude of the impulse, but also regarding the psychophysiological sensation of “lightness” that it creates during the crucial initial stage of a highly resisted movement.

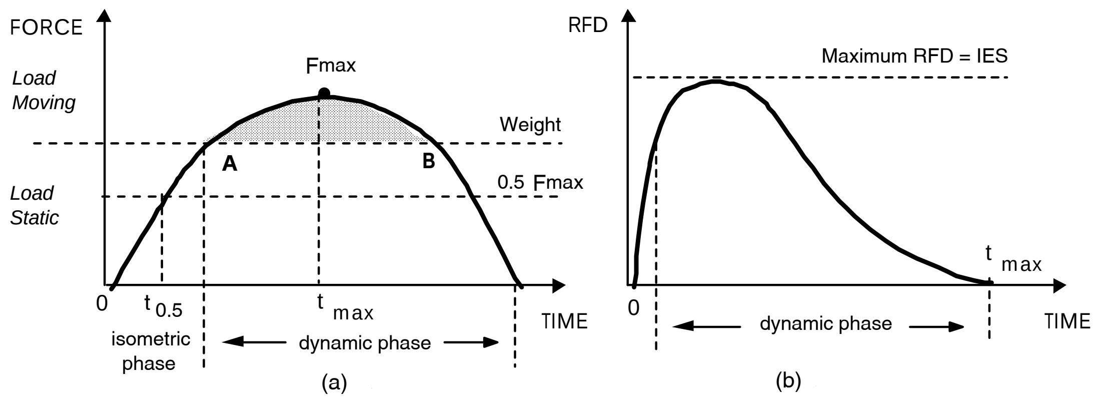

Figure 1: (a) Force–time curve illustrating a method for determining explosive, starting and acceleration strength. W is the weight being overcome by the force $F(t)$. Movement occurs only when the force exceeds the weight W of the object, namely over the shaded portion of the curve. (b) Rate of force development (RFD) curve obtained by plotting the slope of the force–time graph versus time. The maximum rate of force development represents the explosive strength (ES) (Siff and Verkhoshansky [24]).

If the load is near maximal, then the initial slope of the $F(t)$ curve is small and the time taken to produce movement is prolonged. This requires the exhibition of another motor quality known as *static strength-endurance*, as opposed to *dynamic strength-endurance*, which refers to the muscle endurance required to maintain movement over a given time interval. This quality may be involved in carrying out a set of repetitions with a load or by maintaining cyclic work of various intensities One may also identify a property known as *explosive strength-endurance* which involves the repetitive execution of explosive effort.

An index of starting-strength, ISS (the Starting or S-gradient), is estimated by means of the formula:

$$\mathrm{ISS} = \frac{0.5F_{\max}}{t_{0.5}},$$

where $t_{0.5}$ is the time taken to reach one half of $F_{\max}$.

Similarly, an Acceleration or A-gradient index is defined to estimate acceleration-strength, the ability of the muscle to produce force as rapidly as possible under dynamic conditions once the contraction has already begun:

$$\mathrm{IAS} = \frac{0.5F_{\max}}{t_{\max} - t_{0.5}},$$

where $t_{\max}$ is the time taken for the force to reach $F_{\max}$.

The validity of isolating starting-strength and acceleration-strength has been corroborated by electromyography, which reveals differences in their neuromotor motor patterns, the recruitment of motor units and the firing frequency of the motoneurons during explosive force production (Verkhoshansky [27]).

Finally, explosive-strength, the ability to produce maximal force in a minimal time, is estimated by an Index of Explosive Strength (IES), although mathematically it is given by the maximum slope of the $F(t)$ curve, i.e. the peak of the RFD curve:

$$\mathrm{IES} = \frac{F_{\max}}{t_{\max}}.$$

Explosive force production is also described by another index called the Reactivity Coefficient, RC, which is the explosive strength index relative to bodyweight or the weight of the object being moved:

$$\mathrm{RC} = \frac{F_{\max}}{t_{\max}W} = \frac{\mathrm{RFD}_{\max}}{W}.$$

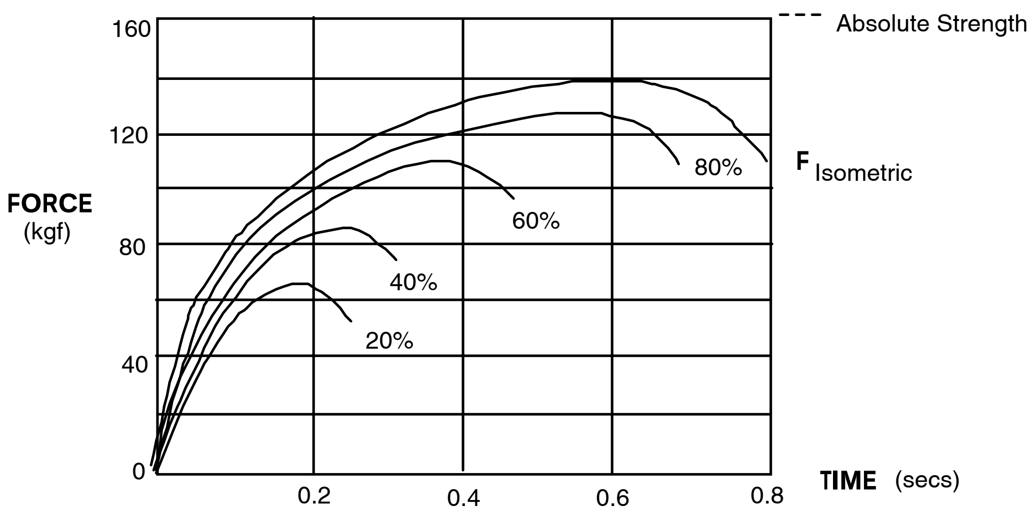

Figure 2: The force–time graph of explosive-isometric muscle tension Fisom and rapid dynamic work with various percentages of maximal strength for a leg press movement (Verkhoshansky [27]).

Explosive strength is most commonly displayed in athletic movements when the contraction of the working muscles in the fundamental phases of the exercise is preceded by mechanical stretching. In this instance, the switch from stretching to active contraction uses the elastic energy of the stretch to increase the power of the subsequent contraction, a process that is central to the so-called stretch-shortening cycle and "plyometric" or rapid rebound action. This specific quality of muscle is called its reactive ability (RA).

The $F(t)$ curve yields even more information if we plot a series of curves produced by examining the dynamics associated with movement against different loads (Figure 2). First of all, it will be noted that the greatest level of force or absolute strength (defined by Zatsiorsky as F_{mm} or maximum maximorum, the maximum of all maxima achieved under any conditions) cannot be produced against loads that are not near maximal (Zatsiorsky [31]).

It will also be noted that it takes more than half a second to produce the greatest levels of force. This has important consequences in sports preparation, because it implies that having great strength is of little consequence unless the athlete can produce a large percentage of this strength when it is most needed. The rapid attainment of peak force for lighter loads also reveals that the impulsive force producing the movement is developed chiefly by starting-strength.

The steepness of the force–time curve and the greater magnitude of maximum isometric than dynamic maximum force for equivalent joint angles show that, if isometric exercises are executed with the accent on RFD, then they can often be as effective for developing explosive strength as dynamic exercises. Because the protective inhibitory effects usually associated with voluntary muscle action are not encountered in reflexive isometric contraction, even greater explosive force can be displayed isometrically than dynamically.

The above has profound implications for the statement that greatest muscle force is produced during eccentric action. While one can lower eccentrically a load that is as much as 35% greater than what one can raise concentrically, it is incorrect to imply that muscle tension is greatest under eccentric conditions. This is corroborated by the finding that superimposed electrical stimulation increases eccentric torque by more than 20% above voluntary levels and electrically evoked torque alone exceeds voluntary torque by about 12% (Westing *et al.* [29]). Interestingly, this research discovered that no corresponding differences were observed between superimposed and voluntary torques under isometric or concentric conditions, so it may be that specific neural mechanisms may protect against potentially damaging muscle tension that could otherwise develop under truly maximal eccentric conditions.

A comparison between EMG recordings during eccentric and concentric exercise, as well as the magnitude of the training-induced changes in the EMG, also suggests that eccentric muscular action may be impaired by mental processes (Handel *et al.* [9]). One method that to counteract this tendency for muscle tension to decrease during eccentric actions is to train against a combination of inertial and elastic or isokinetic loads (see Figure 14).

Eccentric training may have special value in enhancing adaptation to strength training, as suggested by research showing that submaximal eccentric exercise encourages faster initial adaptation to strength training than training with near maximal concentric loading (Hortobágyi *et al.* [13]). Moreover, greatest concentric muscle tension occurs at higher joint velocities, whereas eccentric activity increases as joint velocity decreases (Potvin [21]).

These findings have interesting implications for a method known as Compensatory Acceleration Training (CAT). Proprioceptive feedback makes one aware that the load is changing and enables one to voluntarily accelerate or decelerate the bar to alter muscle tension. It might be preferable to call this method *Compensatory Action Training*, because it does not necessarily alter acceleration significantly, though it might alter the level of neural excitation. This method can be useful in altering muscle tension or movement velocity to achieve a specific training goal, because, as we noted above, neural drive may reflexively attempt to decrease muscle tension during certain phases of movement. Compensatory Action Training may then be seen to increase muscle tension both by accelerative or mechanical means, according to Newton II ($F = ma$), and by increase in neural excitation, especially if the transition phase between the eccentric and concentric phases of the movement involved is made as short as possible.

Thus, we note that muscle tension in strength training may be increased in the following ways:

- voluntary efforts by using Compensatory Action to counteract any decrease in movement velocity,
- involuntary use of reflexive explosive rebound or plyometric actions,
- voluntary use of prestretch to facilitate stronger muscle contraction,
- use of inertial loads combined with elastic or isokinetic loads,
- application of electrical stimulation.

Maximal tension occurs under explosive isometric conditions produced during so-called plyometric actions when an eccentric action is suddenly terminated and immediately followed by a concentric rebound, as was confirmed by research using Siff's tensiometric method (Kirkby [14]). Very often, maximal tension occurs in positions where mechanical leverage is the poorest, in order to facilitate movement from such disadvantageous positions (Siff [23]). The formalisation of exercises based upon this "shock method" of training was carried out during the 1960s by Verkhoshansky and has led to the popular system known as plyometric training. Further research (Verkhoshansky [27]) has determined the correlation coefficient (R) between the different strength indices:

- explosive-strength and acceleration-strength: 0.84,
- explosive-strength and starting-strength: 0.52,
- starting-strength and acceleration-strength: 0.27.

Interestingly, the correlation between absolute strength and the values of the curve $F(t)$ is significant at its starting point for relatively untrained subjects, but this becomes an unreliable measure as strength levels increase. This is one of many reasons why training protocols or research studies may be radically different for elite athletes and novices.

If we return to Figure 1, we will note that the graph involves no information on the muscles generating the force. If we now take into account the fact that the muscle complex can be prestretched during the isometric phase, we will see that this can increase the production of starting-strength both by adding an element of elastic (potential) energy and by activating the stretch reflex more strongly.

3 Different prestart and prestretch states of muscle action

The effects of different starting states of muscle on the subsequent movement were investigated by Verkhoshansky (Siff and Verkhoshansky [25]), who showed that when a movement is begun with the muscles relaxed, they are not optimally ready for work. Under laboratory conditions, a load was projected vertically upwards on a special machine under the following conditions (Figure 3):

1. muscles relaxed,
2. muscles in isometric tension produced by loading with various weights,
3. muscles stretched dynamically, e.g. during the "wave" phase in swimming,
4. muscles stretched by "shock" loading caused e.g. by rapid braking of a load falling from a height.

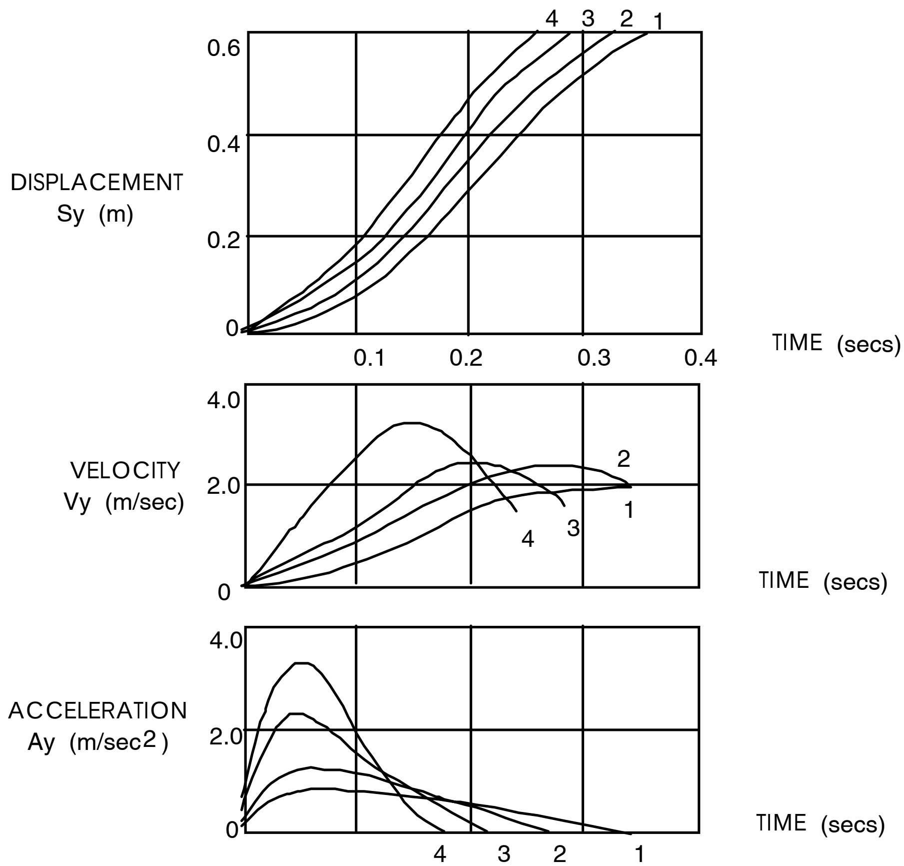

Figure 3: The dynamics of a moving load for different starting states of muscle. The numbers refer to the states detailed above (Verkhoshansky [27]).

This research showed that some form of prestretching of muscle is essential for enhancing force production. Where any prestretch is important, especially in explosive and ballistic actions, its timing of use is critical and its duration must be short, because prolonged or mistimed use can diminish the dynamic functional force.

A deeper analysis of this starting or prestart state of muscle action yields some further interesting insights. For instance, Wilson *et al.* [30] in examining the effects of different delay times in the bench press showed that the benefits of prior stretch may persist for as long as four seconds, at which stage it is suggested that all stored elastic energy is lost (Figure 4 (a)). Chapman and Caldwell [4], on the other hand, found that the benefits of prior stretching during forearm movement are dissipated within 0.25 second, a value which agrees with findings by Siff [24] in an analysis of explosive elbow flexion without additional loading (Figure 4 (b)).

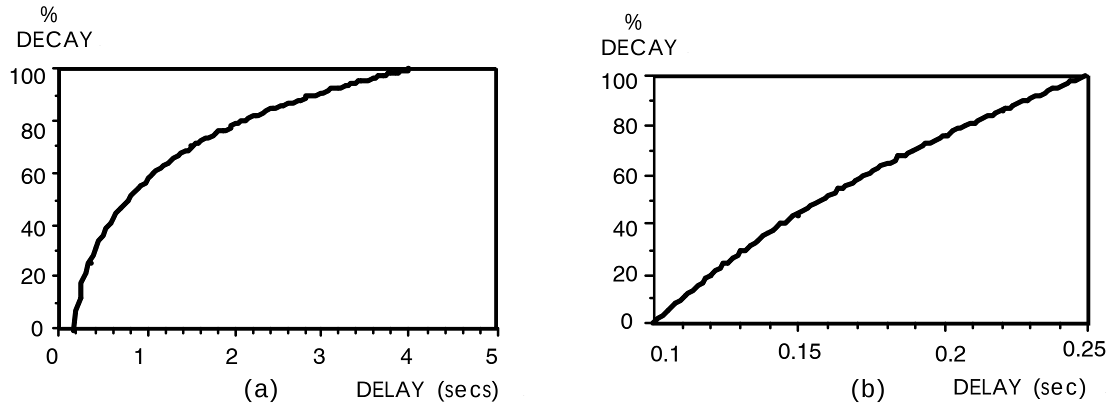

Figure 4: (a) The effect of different time delays on decay of the additional force produced by a preliminary stretch in a bench press (based on Wilson *et al.* [30]). (b) The effect of different time delays on decay of the additional force produced during an explosive preliminary prestretch in unloaded elbow flexion (Siff [24]).

It is curious that muscle prestretch is usually referred to as separate from isometric muscle action, but prestretch occurs under isometric or quasi-isometric conditions, revealing that analysis of isometric training often neglects its role during prestretch and other neuromuscular facilitation processes. As the role of isometric processes in other events such as ballistic and plyometric actions is examined later, their importance in many crucial aspects of human movement will become apparent. Isometrics are not simply a method of muscle action that increases muscle strength at joint angles close to the angle at which isometric training is executed, but can be as effective for developing explosive strength as dynamic exercises if executed explosively with an accent on RFD.

Thus, it must be stressed that isometric training is not simply a matter of holding a static muscle contraction for a given time. Isometric contraction requires a muscle to increase its tension from rest to a maximum or sub-maximal value over a certain time (rise or "attack" time), to sustain this tension for another period (the resistance phase) and to decrease this tension to rest or a lower value (decay time). Consequently, one may distinguish between explosive isometrics, which have a very brief rise time, and slow isometrics, with a much longer rise time.

This research presented in the above graphs suggests that delays of as long as a second or two can still produce significant augmentation of the subsequent concentric phase for some activities, but delays as short as 0.2 second are sufficient to dissipate the benefits of prior stretch during other activities. Research by Bosco *et al.* [2] offers a partial solution to this apparent contradiction. They proposed that individuals with a high percentage of FT fibres in the leg muscles exhibit a maximum plyometric effect when the eccentric phase is short, movement range is small and coupling time between eccentric and concentric action is brief. On the other hand, subjects with a high percentage of ST fibres apparently produce their best jumping performance when the eccentric phase is longer, movement range is greater and the coupling time is longer.

It is also tempting to attribute these differences in coupling times to specific maximum delays for each joint action. While this probably is true for different joint actions, it is also important to note that body exhibits many different reflexes, each of which acts under different conditions and at different rates. In particular, there are tonic (static) and phasic (dynamic) stretch reflexes and very rapid receptors in the joint capsules that detect the rates of movement and allow the nervous system to predict where the extremities will be at any precise moment, thereby facilitating anticipatory modifications in limb position to ensure effective control and stability (Guyton [7]). Loss of this predictive function apparently makes it virtually impossible for one to run, jump, throw or catch.

In addition, weightlifters, powerlifters and sometimes bodybuilders use the so-called prestretch principle to produce a more powerful concentric muscle contraction to enable them to lift heavier loads. In doing so, they begin a movement from a starting position which imposes an intense stretch on the relevant muscles, hold it for a few seconds and then move as strongly as possible from that position. It would seem that this longer delay would implicate a tonic type of reflex with a characteristically greater coupling time between eccentric and subsequent concentric actions. The action cannot be called plyometric, despite the fact that prior stretch had contributed to the subsequent concentric action. Conversely, phasic reflex activity would more likely be implicated in the explosive movements which typify classical plyometrics and the type of activity described in Figure 4.

This explanation also distinguish between plyometric action and plyometric training. One cannot simply separate "plyometric" and "non-plyometric" solely on the basis of coupling times, otherwise one would even have to classify jogging as classical plyometrics, because the time taken for the ground reaction force to reach a maximum can be less than 0.15 second. One also has to take the force–time pattern and the rate of force development (RFD) into account, as is evident from a comparison of different types of vertical jump (Figure 5). Here it can be seen that the RFD during the different prestart conditions have a profound effect on the maximum height reached.

The effect of plyometric or stretch-shortening actions has been attributed to one or more of the following processes, the exact contribution of each still being vigorously debated (van Ingen Schenau *et al.* [26]; Zatsiorsky [33]):

- utilisation of stored elastic energy,
- changes in the mechanical characteristics of the muscle complex,
- eliciting of the various stretch reflexes,
- disinhibition of inhibitory nervous processes,
- reduced electromechanical delay (between stimulus and initiation of muscle action).

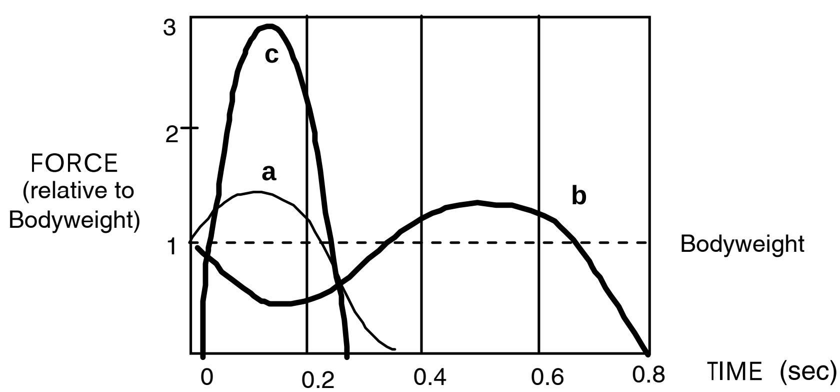

Figure 5: The development of force for various vertical jumps: (a) jumping from rest from a low squat; (b) the usual jump with a controlled dip or amortisation phase; (c) after a depth-jump from 0.4 metre. The jump heights reached were 0.67 m, 0.74 m and 0.80 m respectively (Verkhoshansky [27]).

It is believed that the rapid termination of a powerful eccentric action followed as soon as possible by a rebound under concentric conditions will increase the kinetic energy of the final phase of the propulsive motion.

The apparently contradictory finding that both slower prestretch and explosive ballistic actions both enhance subsequent muscle action emphasises that facilitation of muscle action depends on different neural and biomechanical processes, each of which is implicated under different conditions, a conclusion that is profoundly important for all strength conditioning programmes. Bernstein's comments are highly relevant in this regard: "The movement of the body is more economical, and consequently, more rational, the greater degree to which the body utilises the reactive and external forces and the less he relies on recruiting active muscles" (Zhekov [34]).

It is interesting to note that repeated ballistic prestretches may enhance the subsequent muscle action even further. In studying the different prestart conditions used by Olympic weightlifters, it was learned that a dynamic start using a single or double "pumping" action of raising and lowering the hips produces a greater pulling force than a static start with the knees and hips kept fixed relative to one another before the pull begins (Glyadkovsky and Rodionov [8]). In both of these "pumping" starts, the last pumping action is completed approximately 0.4 second before the eventual pull from the platform.

Then, during the dynamic weightlifting pull from the ground, the downward prestretching action of the knees in the "double knee-bend" phase as the bar passes knee level lasts about 0.08 seconds. It is also relevant that the time taken for the dip during the Olympic jerk of the weight from the shoulders is between 0.2 and 0.5 second, depending on the load being thrust overhead and that the power involved in amortising the downward movement (over a distance of between 8 and 15 cm, depending on the load) is approximately 500 watts (Vorobyev [28]).

Biomechanical findings in normal sporting movement such as these have profound implications for the use of currently popular plyometric or stretch-shortening training, since they emphasise that plyometric actions occur commonly in sport and that plyometric training (a separate system of sports training based upon the use of many jumping and other impulsive plyometric actions) should not be casually prescribed without a thorough understanding of their nature and how they interact with other sporting movements.

4 The force–velocity relationship

One of the best known relationships concerning muscle action is the hyperbolic curve (Figures 6, 7) which describes the dependence of force on velocity of movement (Hill [12]). Although this relationship originally was derived for isolated muscle, it has been confirmed for actual sporting movement, though the interaction between several muscle groups in complex actions changes some aspects of the curve (Zatsiorsky and Matveev [32]; Komi [15]).

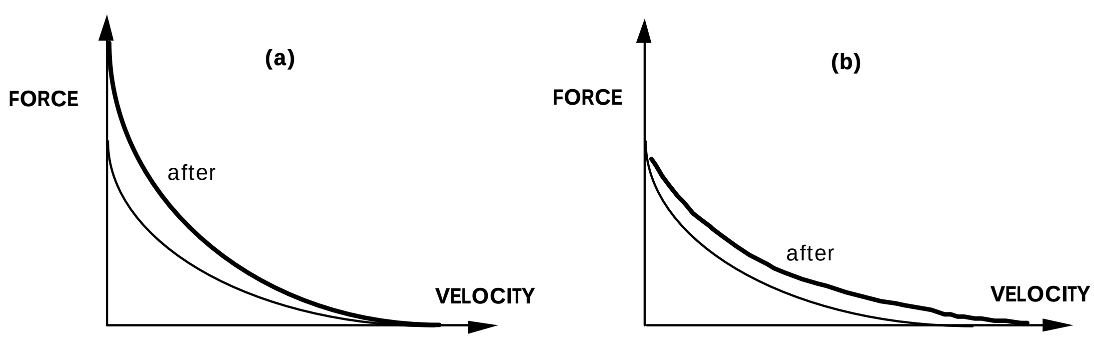

Figure 6: The relationship between force and velocity, based on the work of Hill (1953). (a) The dark curve shows the change produced by heavy strength training (b) the dark curve shows the change produced by low load, high velocity training (after Zatsiorsky, 1995).

This curve implies that velocity of muscle contraction is inversely proportional to the load, that a large force cannot be exerted in very rapid movements, that the greatest velocities are attained under conditions of low loading, and that the intermediate values of force and velocity depend on the maximal isometric force. It is inappropriate to deduce from this that large force can never be produced at higher velocities, because, as discussed earlier, ballistic action implicating stretch-shortening and powerful neural facilitation processes exist primarily to manage such situations.

The influence of maximal isometric strength on dynamic force and velocity is greater in heavily resisted, slow movements, although there is no correlation between maximal velocity and maximal strength (Zatsiorsky [33]). The ability to generate maximum strength and the ability to produce high speeds are different motor abilities, so that it is inappropriate to assume that development of great strength will necessarily enhance sporting speed.

The effect of heavy strength training has been shown to shift the curve upwards, particularly in beginners (Perrine and Edgerton [20]; Lamb [17]; Caiozzo *et al.* [3]) and light, high velocity training to shift the maximum of the velocity curve to the right (Zatsiorsky [31]). Since, in both cases, power = force × velocity, the area under the curve represents power, so that this change in curve profile with strength increase means that power is increased at all points on the curve. The term "strength-speed" is often used as a synonym for power capability in sport, with some authorities preferring to distinguish between strength-speed (the quality being enhanced in Figure 6 (a)) and speed-strength (the quality being enhanced in Figure 6 (b)).

The graph depicting concentric and eccentric muscle action looks like that depicted in Figure 7. Consequently, muscular power is determined by the product of these changes ($P = FV$) and reaches a maximum at approximately one-third of the maximal velocity and one-half of the maximal force (Zatsiorsky [31]). In other words, maximal dynamic muscular power is displayed when the external resistance requires 50% of maximal muscle force.

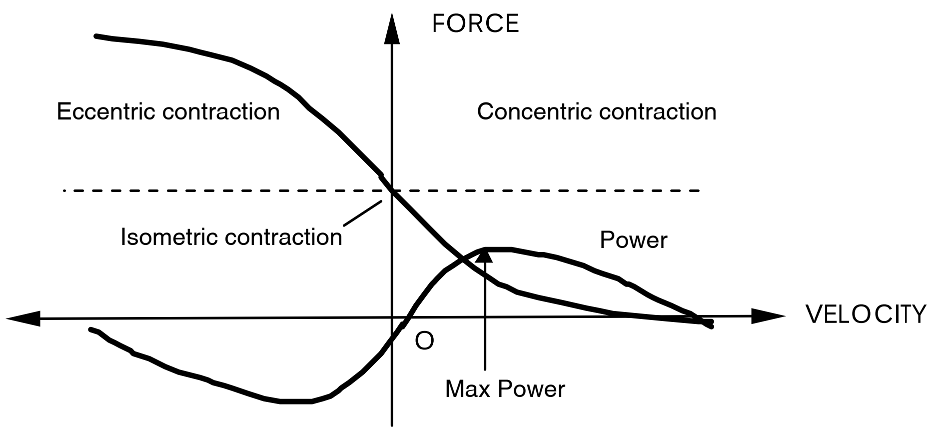

Figure 7: Schematic of the idealised force–velocity curves for concentric and eccentric muscle contraction (not to scale). The change in muscular power with velocity of contraction is also depicted. Note that power is absorbed at negative velocities, i.e. under eccentric conditions.

The pattern of power production in sporting activities can differ significantly from that in the laboratory, just as instantaneous power differs from average power over a given range of movement. For example, maximum power in the powerlifting squat is produced with a load of about two-thirds of maximum. Power drops to 52% of maximum for a squat with maximal load and the time taken to execute the lift increases by 282%. Power output and speed of execution depend on the load; therefore, selection of the appropriate load is vital for developing the required motor quality (e.g. maximal strength, speed-strength or strength-endurance).

It is interesting that Hill's relationship (Figures 6, 7) was modified by more recent research by Perrine and Edgerton [20], who discovered that, for in vivo muscle action, the force-velocity curve is not simply hyperbolic (curve 2 in Figure 8). Instead of progressing rapidly towards an asymptote for low velocities, the force displays a more parabolic shape in this region and reaches a peak for low velocities before dropping to a lower value for isometric contraction ($V = 0$). In other words, maximum force or torque is not displayed under isometric conditions, but at a certain low velocity. For higher velocities (torque greater than about 200 degrees per second), Hill's hyperbolic relation still applies.

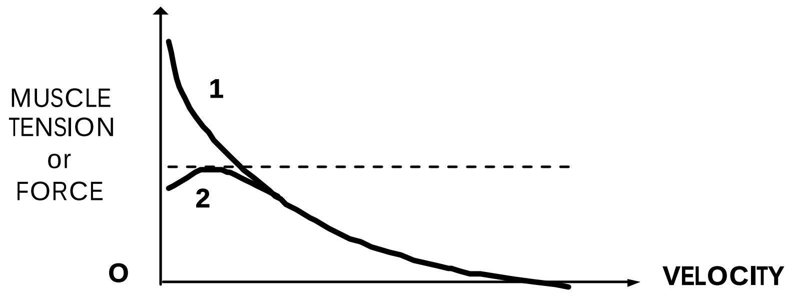

Figure 8: Force–velocity relationship of isolated muscle (1) and in vivo human muscles (2) as determined in two separate experiments under similar loading conditions. The hyperbolic curve is based on the work of Hill, while the other curve is obtained from research by Perrine and Edgerton [20].

In general, therefore, the picture which emerges from the equation of muscle dynamics is that of an inverse interplay between the magnitude of the load and the speed of movement, except under isometric and quasi-isometric conditions. Although this interplay is not important for the development of absolute strength, it is important for the problem of speed–strength.

The above studies of the relationship between strength and speed were performed in single-jointed exercises or on isolated muscles in vitro under conditions which generally excluded the effects of inertia or gravity on the limb involved. Moreover, research has shown that the velocity-time and velocity-strength relations of elementary motor tasks do not correlate with similar relations for complex, multi-jointed movements. In addition, other studies reveal that there is a poor transfer of speed-strength abilities developed with single-jointed exercises to multi-jointed activities carried out under natural conditions involving the forces of gravity and inertia acting on body and apparatus. Consequently, Kuznetsov and Fiskalov [16] studied athletes running or walking at different speeds on a treadmill and exerting force against tensiometers. Their results revealed a force–velocity graph which is very different from the hyperbolic graph obtained by Hill (Figure 9).

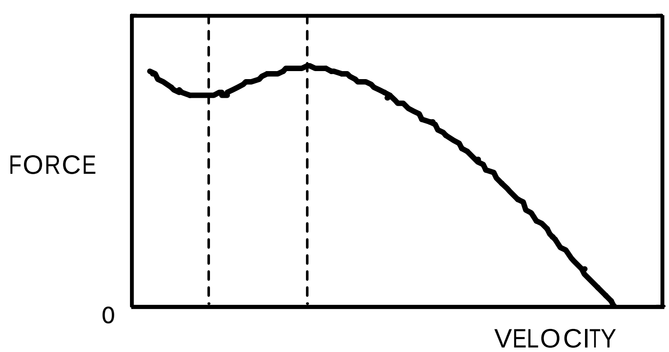

Figure 9: Force–velocity relationship for cyclic activity (based on data of Kuznetsov and Fiskalov [16]).

This figure also shows that jumping with a preliminary dip (or, counter movement) causes the force–velocity curve to shift upward away from the more conventional hyperbola-like force–velocity curve recorded under isokinetically or with squat jumps. For depth jumps, the resulting graph displays a completely different trend where the force is no longer inversely proportional to the velocity of movement. The coordinates describing the more rapid actions of running, high jumping and long jumping also fall very distant from the traditional force–velocity curve (Figure 10).

The reason for these discrepancies lies in the fact that movement under isokinetic and squat jumping

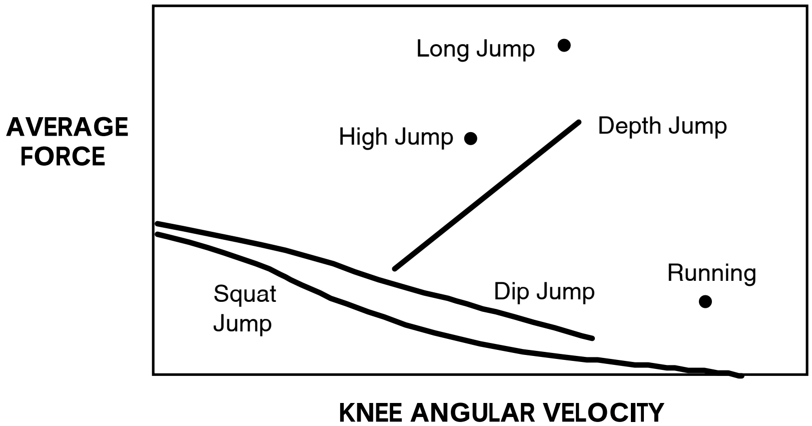

Figure 10: Force–velocity curve for different types of jump (Bosco [1]). In the squat jump, the contractile component of the muscle is primarily responsible for force production, whereas elastic energy, reflexive processes and other muscle changes play additional roles in dip (counter movement) jumps and depth jumping. The calculated values of F and V for high jump, long jump and sprints are also shown.

conditions involves mainly the contractile component of the muscles, whereas the ballistic actions of the other jumps studied apparently are facilitated by the release of elastic energy stored in the SEC and the potentiation of nervous processes during the rapid eccentric movement immediately preceding the concentric movement in each case.

Studies of force–velocity curves under non-ballistic and ballistic conditions (Bosco [1]) further reinforces the above findings that the traditional force–velocity curves (Figures 6, 7) do not even approximately describe the force–velocity relationship for ballistic or plyometric action. The non-applicability of these curves to ballistic motion should be carefully noted, especially if testing or training with isokinetic apparatus is being contemplated for an athlete.

Other work reveals that the jump height reached and the force produced increases after training with depth jumps (Bosco [1]). Whether this is the result of positive changes in the various stretch reflexes, inhibition of the limiting tendon reflex, the structure of the muscle or in all of these factors is not precisely known yet. What is obvious is that the normal protective decrease in muscle tension by the Golgi tendon organs does not occur to the expected extent, so it seems as if plyometric action may raise the threshold at which significant inhibition by the Golgi apparatus takes place. This has important implications for the concept and practical use of plyometrics.

5 Newton's second law and force combinations

Although the biomechanical analysis can yield useful insights into the nature of strength and strength training, it can sometimes be very limited if it is not performed with an understanding of the underlying physiological processes. For instance, if we apply Newton II to examine human force production, we find that this approach fails to solve the fundamental issue of choice of optimal load and acceleration for producing greatest strength or power increase. Since force $F = ma$, let us apply it to produce the same magnitude of force F in several different ways:

(a) $F = Ma$, where the mass M is large and the acceleration a is small;
(b) $F = mA$, where the mass m is small and the acceleration A is large;
(c) $F = ma$, where both mass and acceleration are moderate.

Force production under isometric or isokinetic conditions where acceleration is zero may be added:

(d) $F = kR$, where k is a factor of proportionality and R is the resistance provided by the apparatus.

This might immediately suggest, since the production of an adequate level of muscle tension is necessary for strength training, that all of these methods of "force training" are entirely the same and that is just a matter of one's personal choice which method is used. So, the question arises as to whether or not it makes any real difference which method of strength training is used, as long as adequate muscle tension is produced. If one attempts to answer this question in purely mechanistic terms, one might be tempted to reply in the negative and qualify one's reply with comments about initiating movement against heavy loads with high inertia, possible detrimental effects of sustained loads on the soft tissues of the body, and duration of loading.

Interestingly, experience from three different aspects of strength training, namely Olympic weightlifting, powerlifting and bodybuilding, offers preliminary information. Option (a) with very heavy loads is most commonly encountered in powerlifting, whereas the hypertrophy associated with bodybuilding generally is a product of option (c) training with moderate loads performed for about 8–12 repetitions. Option (b) is characterised by many actions in track and field events. Olympic weightlifting, which involves lifting heavy loads rapidly, appears to contradict evidence that velocity decreases with load, but this is because weightlifting involves ballistic action and relies heavily on quick movement of the lifter under the bar. Isokinetic option (d) occurs only under laboratory conditions using special isokinetic dynamometers to control the motion. Correlation of results obtained under these and actual sporting conditions is low, although it may increase non-specific strength (Rosentzweig and Hinson [22]; Osternig [19]).

Practical evidence shows that the above three ways of generating force do not produce the same results and research reveals that this is because different neural, muscular and metabolic processes are involved in each case. Thus, strength and power training are not simply a matter of using some generalised form of resistance training to produce adequate physical loading and muscle tension; the principle of specificity of training is central to the entire issue.

Research has shown that, to increase muscular strength in the average person, training load must exceed a threshold training stimulus of not less than one-third of one's maximal strength (Hettinger and Müller [11]). The development of strength requires that the stimulus intensity be gradually increased, since it was discovered that every stimulus has a changing strengthening threshold (Hettinger [10]).

Up to this point, strength production has been considered under discrete conditions of inertial, isokinetic and other forms of loading and this has yielded invaluable results for application to training. It is now interesting to examine what happens under conditions which combine different forms of resistance.

6 Combination of free weights and isokinetic resistance

The first combination to be investigated was a combination of auxotonic (free weight, allodynamic or isoinertial) and isokinetic resistance, with the same group of subjects being analysed under separate inertial and isokinetic conditions, then finally under a combination of both types of resistance applied using a cable machine specially designed and constructed for this purpose by one of Siff's mechanical engineering students (Fradd [6]).

The most interesting findings appear in Table 1 and following graphs. In interpreting this information, it is important to note that the shape of the curves varies with every individual, according to their motor characteristics, and, in the combined scenario, with the magnitude of the inertial load and the velocity of isokinetic resistance.

The combined resistance may be made more auxotonic if the added weight is large and more isokinetic if the added load is relatively small. It would be well to note that there may be significant variation among graphs produced by different individuals, because their force output characteristics depend to a large extent on their training history and sporting background.

1. Greatest mean and peak power is produced under auxotonic conditions, reaching a peak approximately midway through the movement (Figure 11).

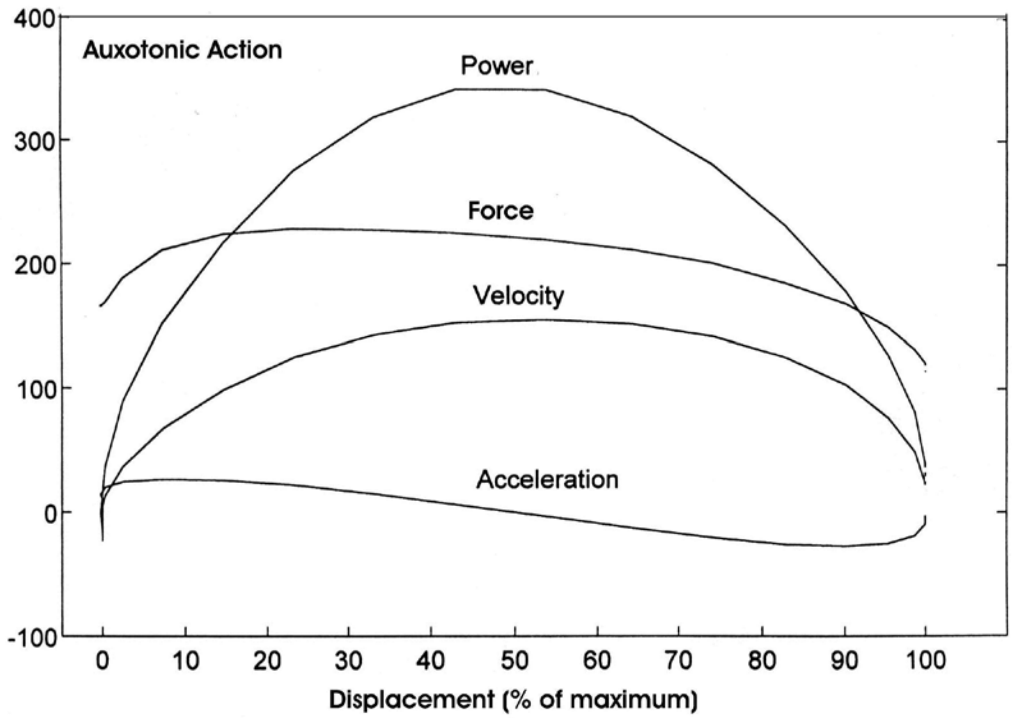

Figure 11: Resisted elbow flexion movement under auxotonic (free weight) conditions.

2. Auxotonic force production is greatest during the earliest phases of joint movement (Figure 11).
3. Power production is considerably lower under isokinetic and combined conditions, but remains near its peak value for a considerable part of the range (Figures 12, 13).
4. Mean and peak force production is greatest during combined conditions, and may peak twice during the movement, most commonly near the beginning and end of range.
5. Isokinetic force tends to plateau about one-third through the range, but commonly increases near end of range.

7 Combination of free weights and elastic resistance

The next resistance hybrid to be examined was a combination of free weights and elastic band resistance, using a box squatting exercise in a power rack with strong elastic bands attached between the ends of the bar and the base of the rack. The length of the elastic bands was adjusted so that elastic resistance varied between set limits at the lowest and the highest positions of the movement. Thus, the powerlifters

Resistance	*Max velocity*	*Mean power*	*Peak power*	*Mean force*	*Peak force*
Isoinertial	1.49	151	306	184	219
Isokinetic	0.42	80	129	247	350
Combined	0.38	102	131	336	380

Table 1: Summary of results obtained from analysis of elbow flexion movement under different resistance conditions (velocity in ms^{-1}, power in watts and force in newtons).

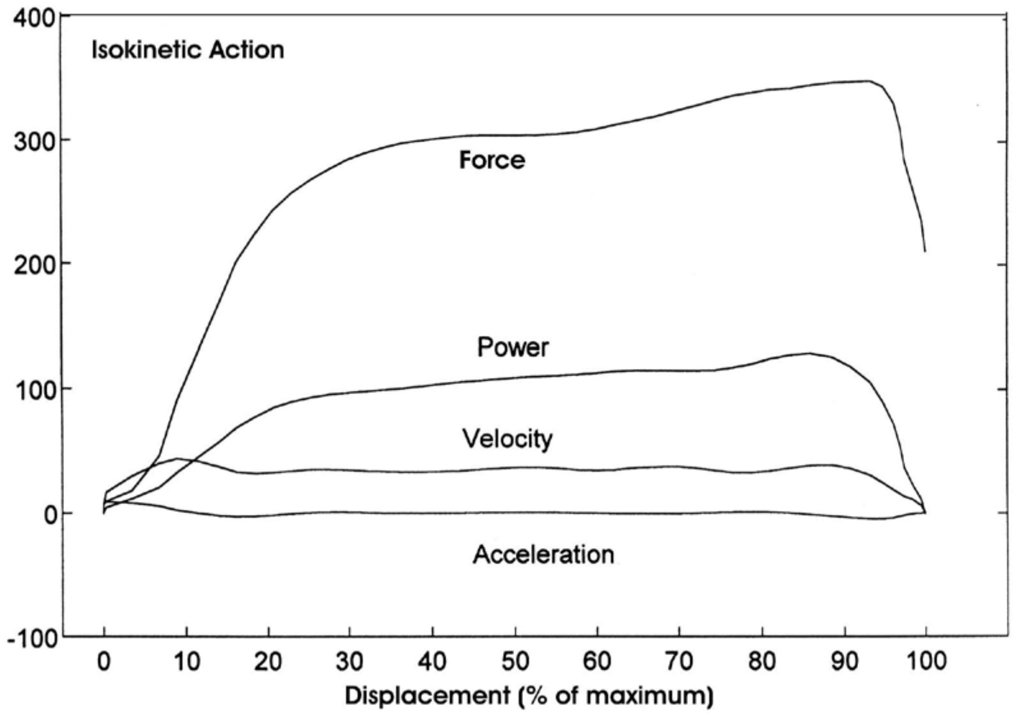

Figure 12: Resisted elbow flexion movement under isokinetic conditions.

involved in this study experienced, in addition to the squat weight of 200 kg, an added elastic resistance of 150 kg at the lowest position while sitting down on the box with knees at approximately right angles and a greatest elastic resistance of 250 kg at the top of the squat with knees fully extended. The results are expressed in the form of percentages of the squat weight used (Figure 14). This method allowed the powerlifters to train under special overload conditions intended to strengthen specific weaknesses in the movement characteristics produced under normal loaded conditions. The main differences between this combined method and free weight training were:

1. A greater mean and peak force were produced throughout the range of movement.
2. The descent onto the box tended to be accelerated above the normal gravitational rate of $9.8\,\mathrm{ms}^{-2}$, so that greater eccentric force had to be generated to control the downward motion.
3. The stronger eccentric loading and the brief transition period involved while sitting before exploding upwards provided neuromuscular stimulation which approximates that usually encountered in popular plyometric training.
4. The force generated during the later stages increased, in strong contrast to the situation of normal squatting in which force production tends to decrease significantly.

As with the other combined method of free weights and isokinetics, it is important to note that there tend to be large individual variations and that the exact shape of the combined resistance curve depends on the magnitude of the weights load relative to the top and bottom resistance provided by the elastic bands.

8 Conclusion

This paper set out to describe how biomechanics may be considered as a type of abstract ergogenic aid that may be applied in sports training to enhance performance through a better understanding and

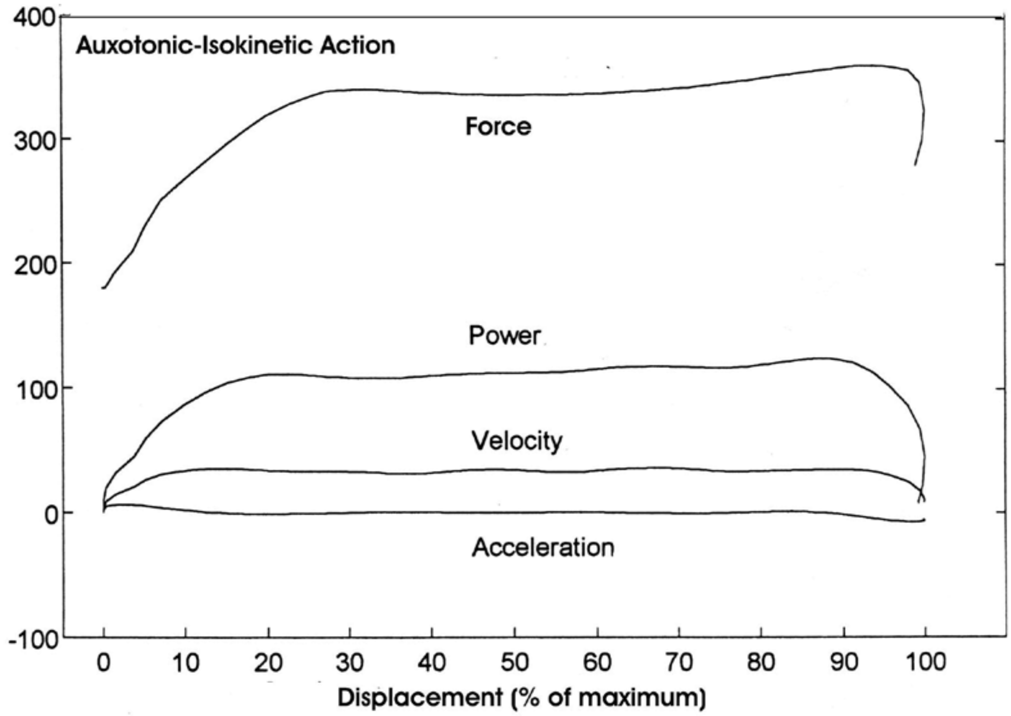

Figure 13: Resisted elbow flexion movement under combined auxotonic and isokinetic conditions. Values of F and V for high jump, long jump and sprints are also shown.

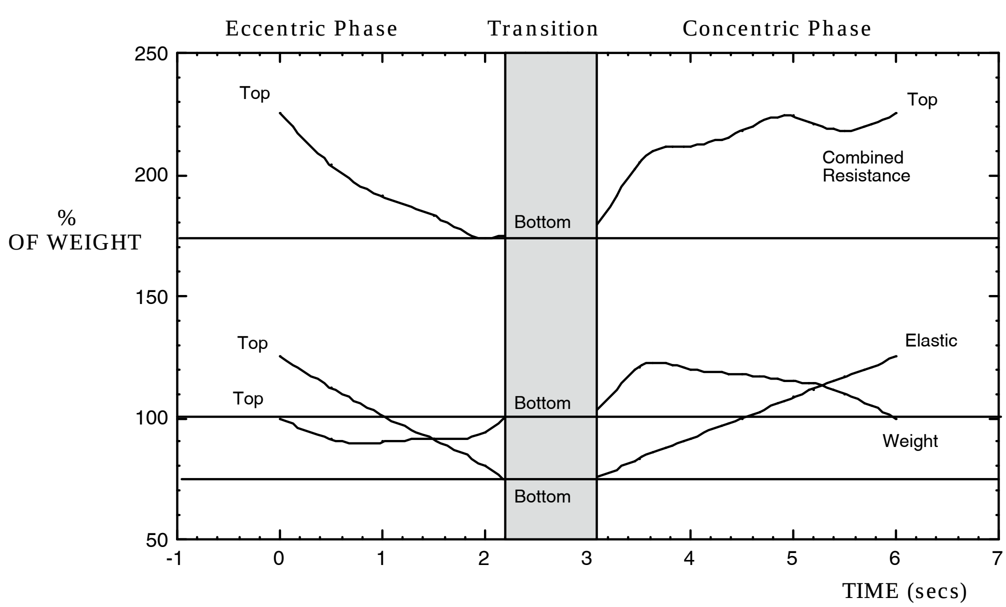

Figure 14: Resisted box squatting under combined free weights and elastically resisted conditions.

optimal application of the mechanical principles that underlie the production of strength. The importance of the fundamental qualities such as starting-strength, acceleration-strength, explosive strength and reactive ability was established, as was a deeper analysis of the nature of isometric and eccentric muscle action. In particular, it was found that isometric action may be initiated explosively and act as an essential component of stretch-shortening or plyometric actions and that its use need not be confined to maximum strength production. Biomechanical analysis of weightlifting movements emphasised that prestretch actions occurring under eccentric conditions may augment strength and power production both under slow and explosive stretch-shortening conditions. Traditional applications of the classical hyperbolic force–velocity curve and Newton's second law were reexamined to show that great force actually can be produced under explosive conditions and that there are more functionally appropriate methods of training than those which are often encountered in training schemes that are derived largely from bodybuilding experience. Finally, movement against combined free weight-isokinetic and free weight-elastic modes of resistance revealed that such methods can enhance more conventional methods of strength training.

Besides the impressive role played over the past 40 years by findings such as the above in guiding the training of many world champions in Russia, all of the findings that emerged from this work have been successfully applied by a group of elite USA powerlifters in a special scheme devised by well-known Ohio coach, Louie Simmons, many of whom have produced USA or World records during the past two years in applying these methods, thereby confirming that biomechanics indeed can serve as an effective ergogenic aid in sport.

References

[1] C. Bosco, "Physiological considerations of strength and explosive power and jumping drills (plyometric exercise)", in *Proceedings of Conference '82: Planning for Elite Performance*, Canadian Track and Field Association, Ottawa 1–5 August 1982 (1982), 27–37.

[2] C. Bosco, P. Komi, J. Thiyani, G. Kekete and P. Apor, "Mechanical power test and fibre composition of human leg extensor muscles", *Eur. J. Appl. Physiol.*, **51**(1) (1983), 129–135.

[3] V. J. Caiozzo, J. J. Perrine and V. R. Edgerton, "Training-induced alterations of the in vivo force-velocity relationship of human muscle", *J. Appl. Physiol.*, **51**(3) (1981), 750–754.

[4] A. Chapman and G. Caldwell, "The use of muscle strength in inertial loading", in *Biomechanics IX-A*, D. Winter, D. Norman and R. Wells (editors), Human Kinetics Publications (1985), 44–49.

[5] R. Enoka, "Eccentric contractions require unique activation strategies by the nervous system", *J. Appl. Physiol.*, **81**(6) (1996), 2339–2346.

[6] D. Fradd, *Analysis of Torque, Velocity and Displacement Relations during Joint Flexion, under Isokinetic-Auxotonic Loading Conditions*, Final Year Project, MECN 446, School of Mechanical Engineering, University of the Witwatersrand, South Africa (1995).

[7] A. Guyton, *Textbook of Medical Physiology*, sixth edition, WB Saunders Co., Philadelphia (1984).

[8] V. S. Glyadkovsky and V. I. Rodionov, "Variations of the dynamic start for the snatch", *Tyazhelaya Atletika*, USSR (1971), 96–102.

[9] M. Handel, T. Horstmann, H. Dickhuth and R. Gulch, "Effects of contract–relax stretching training on muscle performance in athletes", *Eur. J. Appl. Physiol.*, **76**(5) (1997), 400-408.

[10] T. Hettinger, "The physiology of strength", Charles Thomas, Springfield, Illinois (1961), 18–30.

[11] T. Hettinger and E. A. Müller, "Maximal strength development" (in German), *Arbeitsphysiologie*, **15** (1953), 111.

[12] A. V. Hill, "The mechanics of active muscle", *Proc. Roy. Soc. Lond. (Biol.)*, **141** (1953), 104–117.

[13] T. Hortobágyi, J. Barrier, D. Beard, Braspennincx, P. Koens, P. Devita, L. Dempsey and J. Lampert, "Greater initial adaptations to submaximal muscle strengthening than maximal shortening", *J. Appl. Physiol.*, **81**(4) (1996), 1677–1682.

[14] A. Kirkby, *Biomechanical Analysis of the Relation between Joint Torque and Muscle Tension for Different Modes of Muscle Contraction*, MSc Dissertation, University of the Witwatersrand, South Africa (1998).

[15] P. Komi, "Neuromuscular performance: Factors influencing force and speed production", *Scand. J. Sports Sci.*, **1** (1979), 2–15.

[16] V. Kusnetsov and I. Fiskalov, "Correlations between speed and strength in cyclic locomotion", *Teoriya i Praktika Fizischeskoi Kultury (Theory and Practice of Physical Culture)*, **8** (1985), 6.

[17] D. R. Lamb, *The Physiology of Exercise*, McMillan Co. (1984), 82–84.

[18] R. W. Norman and P. V. Komi, "Electromechanical delay in skeletal muscle under normal movement conditions", *Acta Physiol. Scand.*, **106**(3) (1979), 241–248.

[19] L. R. Osternig, "Isokinetic dynamometry: Implications for muscle testing and rehabilitation", *Exercise and Sport Sci. Rev.*, **14**, 45-80.

[20] J. J. Perrine and V. R. Edgerton, "Muscle force–velocity and power–velocity relationships under isokinetic loading", *Med. Sci. Sports*, **10**(3) (1978), 159–166.

[21] J. R. Potvin, "Effects of muscle kinematics on surface EMG amplitude and frequency during fatiguing dynamic contractions", *J. Appl. Physiol.*, **82**(1) (1997), 144–151.

[22] J. Rosentzweig and M. Hinson, "Comparison of isometric, isotonic and isokinetic exercise by electromyography", *Arch. Phys. Med. and Rehab.*, **53** (1972), 249–250.

[23] M. C. Siff, *Ballistic Analysis of Human Knee Stability*, PhD Thesis, University of the Witwatersrand, South Africa (1986), 27–28.

[24] M. C. Siff and Y. V. Verkhoshansky, *Supertraining: Strength Training for Sporting Excellence*, University of Witwatersrand, South Africa (1993).

[25] M. C. Siff and Y. V. Verkhoshansky, *Supertraining: Strength Training for Sporting Excellence*, Supertraining International, Denver (1999).

[26] G. van Ingen Schenau, M. Bobbert and A. de Haan, "Does elastic energy enhance work and efficiency in the stretch-shortening cycle?", *J. Appl. Biomech.*, **13** (1997), 389–415.

[27] Y. V. Verkhoshanky, *Osnovi Spetsialnoi Silovoi Podgotovki i Sporte (Fundamentals of Special Strength Training in Sport)*, Fizkultura i Sport Publishers, Moscow (1977).

[28] A. N. Vorobyev, *Textbook on Weightlifting*, Intern. Weightlifting Federation, Budapest, Hungary (1978), 72–75.

[29] S. H. Westing, J. Seger and A. Thorstensson, "Effects of electrical stimulation on eccentric and concentric torque–velocity relationships during knee extension in man", *Acta Physiol. Scand.*, **140**(1) (1990), 17–22.

[30] G. Wilson, B. Elliot and G. Wood, "The effect on performance of imposing a delay during a stretch-shorten cycle movement", *Med. Sci. Sports Exerc.*, **23**(3) (1991), 364–370.

[31] V. M. Zatsiorsky, *Science and Practice of Strength Training*, Human Kinetics (1995).

[32] V. M. Zatsiorsky and E. N. Matveev, "Force–velocity relationships in throwing", *Teoriya i Praktika Fizischeskoi Kultury (Theory and Practice of Physical Culture)*, **27**(8) (1964), 24–28.

[33] V. M. Zatsiorsky, "The review is nice. I disagree with it", *J. Appl. Biomech.*, **13** (1997), 479–483.

[34] I. P. Zhekov, *Biomechanics of the Weightlifting Exercises*, Fizkultura i Sport Publishers, Moscow, Chapter 3 (1976).

METHODS FOR QUANTIFYING PERFORMANCES IN ONE-DAY CRICKET

Paul Allsopp and Stephen R. Clarke
School of Mathematical Sciences
Swinburne University
PO Box 218
Hawthorn
Victoria 3122, Australia

Abstract

Using the Duckworth and Lewis rain interruption rules for one-day cricket matches, a margin of victory in runs is created for the team batting second using the par, projected and new projected scores. The par score is shown to underestimate the margin of victory, whereas, estimates based on the projected and new scores are essentially equivalent to those obtained by the team batting first. If remaining resources for the team batting second are low, the results show that any differences in the margins of victory using either method are marginal. The resulting margin of victory is also used to model a team's rating and common home advantage in the Australian domestic one-day cricket competition. The differences in team ratings generated by both the projected and new projected scores are shown to be marginal and produce the same ranking order. The application of the projected and new projected scores showed that teams experience a common home advantage of nine and eleven runs, respectively, but these were not significant results. This is supported by application of binary logistic regression techniques. The overall ranking of teams produced by the model is also compared with the ranking based upon each team's mean margin of victory and shown to be in generally strong agreement.

1 Introduction

Duckworth and Lewis [6, 7, 9] have developed innovative rain interruption rules that are extensively used in one-day cricket matches. Their methods differ from previous approaches in that they take into account the available run-scoring resources (overs and wickets) the two competing teams have at their disposal. In summary, the more unutilised run-scoring resources a team has at their disposal at the point of interruption of a match the more runs they would be expected to make if the match was not interrupted.

This paper will adapt methods developed by Duckworth and Lewis [6, 9] and proposed by Clarke [2] and Allsopp and Clarke [1] to estimate a margin of victory (in runs) for the team batting second (i.e. Team 2), after they have gone on to win a match. The estimate will reflect the relative strength of Team 2 and provide a fair appraisal of their performance. We will then use adjusted Team 2 scores to investigate home advantage in the Australian domestic one-day cricket competition (1994–2000).

2 Calculating a margin of victory

2.1 The par score

Clarke [2] suggested that the methods developed by Duckworth and Lewis (D/L) to revise targets in matches interrupted by rain could be used to provide a margin of victory in runs. Using statistical data collected over a long period of time Duckworth and Lewis have developed a method that sets a revised target for Team 2 when overs in either innings have been lost due to a break in play. The target is revised in accordance with the available run-scoring resources the two teams have at their disposal. The adjusted targets ultimately reflect the relative difference in the resource availability of both teams.

In adapting the D/L method to reflect the relative strength of Team 2, when they have gone on to win a match, we treat the completion of Team 2's innings as a break in play. Team 2, in winning the match, has subsequently used up less of their available run-scoring resources in surpassing the target set (unless they win on the last ball). If we denote the revised target for Team 2 by T, Team 1's total score by S, the run-scoring resource percentage remaining by R and the run-scoring resource percentage available to Team 2 by $R2$, then T can be calculated by:

- Scaling Team 1's score downwards in the ratio $R2$ to 100. This is the score to tie.
- Adding one to give the target.

Knowing the number of overs left and the number of wickets Team 2 has lost, Duckworth and Lewis [6, 7] have prepared detailed tables from which the appropriate $R2$ values can be determined. Reading the tables directly provides the resource percentage remaining for Team 2, denoted by R. It follows, $R2 = 100 - R$. As defined by Duckworth and Lewis [6, 7, 9], it follows:

$$T = S\frac{R2}{100} + 1 = S\frac{100-R}{100} + 1. \tag{1}$$

If a match is abandoned during the second innings, $(T-1)$ is defined by Duckworth and Lewis as the par score, or the score that Team 2 will need to have achieved in order to tie the match at this point. If Team 2 is ahead of its target, Duckworth and Lewis [6] quantify the difference between the current and par scores as Team 2's margin of victory. However, if Team 2 is behind the target set at this point the par score is denoted by Team 1's score and the difference between the two scores will be Team 1's margin of victory. At the completion of a match, the par score represents the score that Team 2 will need to have compiled in order to achieve a tied result at the point their innings is completed. If Team 2 wins the match they are obviously ahead of their target and the subsequent difference between the actual and par scores is defined as Team 2's margin of victory. However, if Team 2 is behind the set target at this point, the par score is defined as Team 1's total score and the difference between the two scores in this case will be Team 1's margin of victory.

2.2 The projected and new projected scores

Using the par score to determine the margin of victory gives some indication of how well Team 2 has performed but it does not tell the whole story since we don't know how many more runs Team 2 could go on to make if they batted out their 50 overs. If, for example, Team 2 wins we can only be certain of how far Team 2 is ahead of its target at the completion of their innings irrespective of how many run-scoring resources Team 2 has at its disposal. We will demonstrate that the par score is not a fair indication of how well Team 2 has performed because the margins of victory that are generated will not be equivalent to those obtained by Team 1. We propose that an estimate of Team 2's projected 50-over score will form the basis of a more accurate measure of the margin of victory. This estimate will be based on two methods, namely the projected score and the new projected score. The projected score is the sum of Team 2's current score at the completion of their innings and an estimate of the number of runs they will make in the remaining overs. This estimate is a percentage of 225 runs, which Duckworth and Lewis define as the average score compiled by teams in a 50-over innings. Alternatively, the new projected score assumes Team 2's final score is the par score. This score can then be used to determine the equivalent score Team 1 needed to have achieved in order to tie the match at this point. This in effect provides a measure of how well Team 2 has performed. In both cases the resultant margin of victory is the difference between Team 1's score and Team 2's adjusted score.

If X represents Team 2's current score, to estimate Team 2's projected 50-over score P, we have:

$$P = X + 225\frac{100 - R2}{100} = X + 2.25(100 - R2) = X + 2.25R. \quad (2)$$

If we define Team 2's new projected score as N, then

$$N = \frac{100X}{R2} = \frac{100X}{100 - R}. \quad (3)$$

3 Analysis of the Australian domestic one-day cricket competition (1994–2000)

Tables 1 and 2 provide a summary of results from the Australian domestic one-day cricket competition (1994–2000) which includes results from 117 completed matches. The results are for winning teams only. Figure 1 provides a series of boxplots showing the distribution of the winning scores. The boxplots suggest that, on average, Team 1 posted higher winning scores than Team 2. This is expected, since Team 2, in winning a match, has its innings truncated as soon as they pass Team 1's score. Notably, the winning scores of Team 2 are more variable than those generated by Team 1. The par score also represents a truncated score and is, on average, lower than all listed scores, however, because the par score can result from a relatively wide range of overs, it is more variable. The projected score, on average, is higher than the new projected score, but due to the presence of outliers the new projected score is the more variable. The outliers for the new projected score result from matches in which Team 2 quickly passed Team 1's score with many unutilised run-scoring resources at their disposal.

In comparing the distribution of winning scores the normality assumption (Anderson–Darling test) holds for Team 1 and Team 2's actual winning scores and for both the par and projected scores, but is violated for the new projected score. Using a two-sample t-test the analysis clearly suggests that, on average, Team 1's score is significantly higher than both Team 2's actual score ($p = 0.000$) and the par score ($p = 0.000$). This results from the fact that Team 1, in winning their matches, exhaust available run-scoring resources and so maximise their return. However, Team 2, in winning always has unutilised run-scoring resources at their disposal (unless they win off the last ball) and so is not able to maximise their run scoring potential. Using the non-parametric Mann–Whitney test to compare distribution of scores, both the projected and new projected scores are not significantly different from Team 1's winning score ($p = 0.390$ and $p = 0.535$, respectively). The new projected score is also not significantly different from the projected score ($p = 0.663$).

	Team 1	*Team 2*	*Par score*	*Projected score*	*New projected score*
Mean	248.4	205.2	170.4	243.1	249.9
Median	242.0	208.0	178.3	237.0	238.4
Standard deviation	32.8	38.0	48.9	33.5	47.2

Table 1: Summary of results from the Australian domestic one-day cricket competition (1994–2000).

Tables 3 and 4 provide a respective summary and analysis of the winning margins of victory generated by Teams 1 and 2. Figure 2 provides a series of boxplots showing the distribution of the margins of victory. With reference to the boxplots, the margins of victory (in runs) generated by Team 2's actual score are inconsequential since Team 2's innings is truncated once Team 1's score is surpassed. Notably, Team 1, on average, generated the higher margins of victory, which were also the most variable. The margins of victory generated by the par, projected and new projected scores, on average, were similar, with the new projected score clearly the more variable. Notably, application of the new projected score to generate a margin of victory has produced a relatively high number of outliers. This arises because the method predicts relatively high scores for Team 2 when they have won a match with a high proportion of unutilised run-scoring resources still at their disposal.

	Comparison of scores	*Test*	*p-value*
H_1	Team 1 > Team 2 (actual)	Two-sample t-test	0.000
H_1	Team 1 > Par	Two-sample t-test	0.000
H_0	Team 1 = Projected	Two-sample t-test	0.390
H_0	Team 1 = New projected	Mann–Whitney	0.535
H_0	Projected = New projected	Mann–Whitney	0.663

Table 2: Competition analysis.

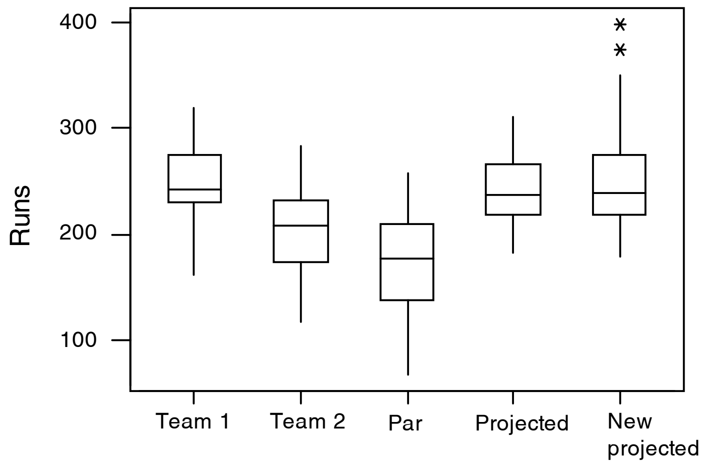

Figure 1: Distribution of winning scores.

Since the normality assumption is violated for all distributions of the margins of victory (Anderson–Darling test), the Mann–Whitney test is used to compare the distributions. The analysis suggests that the margins of victory generated by both the projected score ($p = 0.072$) and the new projected score ($p = 0.190$) are not significantly different from the margins of victory obtained by Team 1. However, the par score generated margins of victory that were significantly less than those generated by Team 1 ($p = 0.005$) and so underestimated the margins estimated by the projected and new projected scores. Both the projected and new projected scores generated margins of victory that were not significantly different from those resulting from adopting the par score ($p = 0.600$ and $p = 0.245$ respectively). Notably, there is no significant difference between the margins of victory generated by the projected and new projected scores ($p = 0.566$).

In relatively few instances the margin of victory generated by the par score exceeds the margin of victory generated by the projected score. This anomaly arises whenever $R(S - 225) > 0$. Since in all cases $R > 0$ (i.e. after at least one ball has been bowled), this situation only arises when Team 2 has gone on to win a match after Team 1 has posted a score in excess of 225. This suggests that Team 1, in losing a match, has performed better than average.

	Team 1	*Team 2*	*Par score*	*Projected score*	*New projected score*
Mean	53.4	2.1	34.8	40.0	46.9
Median	41.0	2.0	30.3	31.7	33.0
Standard deviation	43.2	1.2	21.5	29.1	43.1

Table 3: Margins of victory results.

Figures 3, 4 and 5 represent plots of the differences between the margins of victory generated by

	Comparison of margins of victory	Test	p-value
H_1	Team 1 > Par	Mann–Whitney	0.005
H_0	Team 1 = Projected	Mann–Whitney	0.072
H_0	Team 1 = New projected	Mann–Whitney	0.190
H_0	Par = Projected	Mann–Whitney	0.600
H_0	Par = New projected	Mann–Whitney	0.245
H_0	Projected = New projected	Mann–Whitney	0.566

Table 4: Analysis of margins of victory.

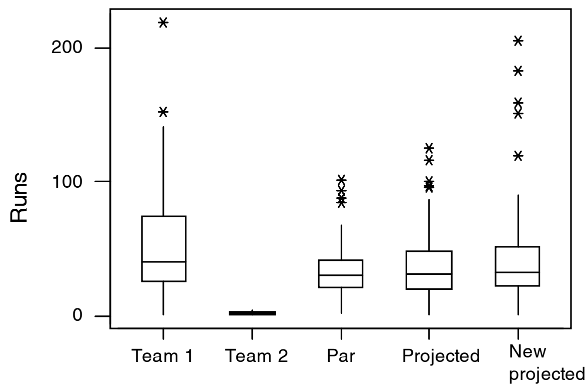

Figure 2: Distributions of the winning margins of victory.

the par, projected and new projected scores against the number of overs remaining. The plots suggest that when the number of overs remaining is relatively small the resulting differences in the margins of victory generated by each representation of Team 2's winning score are also relatively small. However, this difference incr

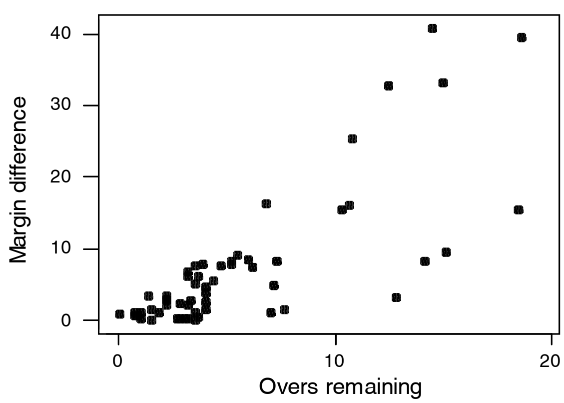

Figure 3: Plot of the differences in margins between the par and projected scores.

Table 5 provides a summary of the correlation coefficients between the margins of victory generated by each representation of Team 2's score. The results suggest that for all winning scores there is a strong positive correlation of 0.949 ($p = 0.000$) between the margins generated by the projected and new projected scores. However there is evidence of only a moderate positive correlation between the margins generated by the par and projected scores and the par and new projected scores (coefficients are 0.636 ($p = 0.000$) and 0.465 ($p = 0.000$), respectively). These observations are deceptive since as is

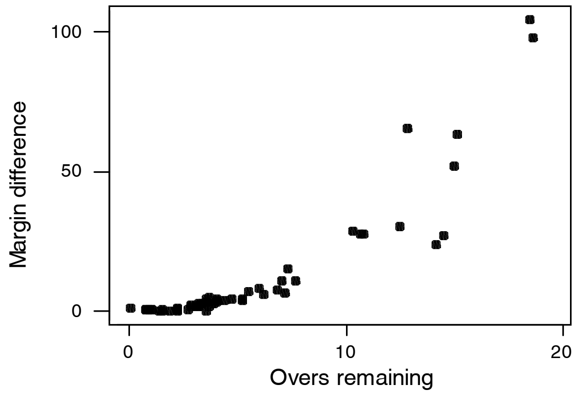

Figure 4: Plot of the differences in margins between the par and new projected scores.

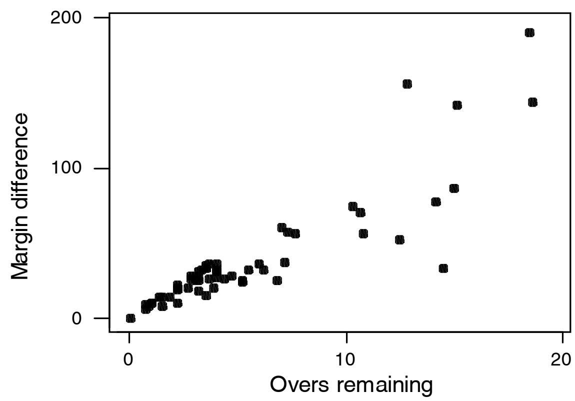

Figure 5: Plot of the differences in margins between the projected and new projected scores.

clear from Figures 3, 4 and 5, in all instances the differences between the margins inflate as the number of overs remaining increases beyond approximately eight. This is confirmed by Table 5, which suggests that the strength of the correlations, in general, diminish as the number of remaining overs increases. This is most apparent when considering the par and projected scores. Notably, when comparing the par and projected, par and new projected and projected and new projected scores the mean difference in the margins of victory in matches with eight or less overs remaining were 26, 28 and 29 runs, respectively. The mean differences increased to 70, 92 and 120 runs, respectively, when the number of overs remaining was nine or more.

Comparisons		*Correlation coefficient*		
		All winning scores	*Less than nine overs remaining*	*More than nine overs remaining*
Par score	Projected score	0.310	0.926	0.730
Par score	New projected score	0.105	0.995	0.960
New projected score	Projected score	0.959	0.946	0.882

Table 5: Summary of correlation coefficients.

4 Analysing home advantage in the Australian domestic one-day competition (1994–1999)

The Australian domestic one-day cricket competition is currently referred to as the Mercantile Mutual Cup and is played between teams representing the six States of Australia and the Australian Capital Territory. The competition is a round-robin tournament with teams gaining two points for a win or one point each for a tie or a "no result". The top four teams at the end of the round robin play off in two semi-finals, with the winners playing each other in the final.

Home advantage is formally defined as the expected difference in score in a game played between two teams (on the home ground of one of the teams) minus the expected difference in score in a game played between the same two teams on a neutral ground. In the context of one-day cricket, home advantage represents the margin of victory in a game played between team i and team j (on the home ground of team i) minus the margin of victory in a game played between team i and team j (on the home ground of team j).

Using techniques adopted by Stefani and Clarke [10], Harville and Smith [8] and Clarke and Norman [4] the winning margin w_{ij} in a match between team i and team j played at the home ground of team i is modelled as:

$$w_{ij} = (u_i + h) - u_j + \epsilon_{ij} = u_i - u_j + h + \epsilon_{ij} \tag{4}$$

where u_i is a measure of the relative ability of team i, h is a measure of the common home advantage and ϵ_{ij} is a zero-mean random error.

A least squares regression model has been fitted to the margins of victory (generated by both the projected and new projected scores) to quantify (a) a team's rating, and (b) any common home advantage. It is assumed that a team's average rating is 100. Table 5 provides a summary of the ratings for each team together with the mean margin of victory over the period 1994 to 2000. Notably, the choice of whether to choose the projected or new projected score to calculate the margin of victory has (a) produced similar ratings and (b) preserved the same ranking order for each team. It is also notable that the teams, on average, did not experience a significant common home advantage under either method, with the advantages generated by the projected and new projected scores being only nine runs ($p = 0.088$) and ten runs ($p = 0.057$), respectively.

In considering the outcome only of each match (i.e. home win/home loss and away win/away loss), we have the home and away teams winning 54% and 46% of matches, respectively. In applying a binary logistic regression model, there is some evidence that the home team experiences an advantage. However, any advantage is not statistically significant ($z = 1.18$, $p = 0.238$), with the odds of winning away being about 1.4 times the odds of winning at home.

Using the mean margin of victory (generated by both the projected and new projected scores) to rank each team shows generally strong agreement with the ranking produced by the model estimates. Home advantage based on estimates generated by the mean margin of victory for each team (i.e. four and six runs for the respective projected and new projected scores) showed some agreement with the model estimates.

5 Conclusions

The use of the D/L method to deal with one-day cricket matches interrupted by rain is well documented and has been used in a number of competitions. The method can also be effectively adapted to provide a relative measure of how well the team batting second has performed by generating a margin of victory in runs equivalent to the team batting first. The margin of victory is a more sensitive measure of the strength of a win or loss.

The par score provides some measure of Team 2's relative performance. However, it does not generate a 50-over based margin of victory. Consequently, use of the par score tends to underestimate Team 2's margin of victory. Using both the projected and new projected scores provides a fairer appraisal of

Team	*Projected score*		*New projected score*	
	Rating	*Mean margin of victory*	*Rating*	*Mean margin of victory*
Western Australia	128	30	132	41
Queensland	124	24	124	25
New South Wales	116	14	118	15
Tasmania	93	−10	97	−6
South Australia	92	−2	89	−7
Victoria	83	0	79	0
ACT	65	−27	61	−26
Common home advantage	9	4	11	6

Table 6: Rating of teams in the 1994–1999 domestic one-day competition.

Team 2's relative performance. In each case the scores generate a margin of victory essentially equivalent to that obtained by Team 1. This suggests that the margins of victory generated by the projected and new projected scores provide a more accurate measure of a team's performance.

In matches won by Team 2, when the number of remaining overs was relatively small, any difference in the margins of victory generated by the par, projected and new projected scores was marginal. This suggests that when the number of overs is low (less than eight) the margin of victory generated by either method will in effect be equivalent to the margin of victory obtained by Team 1. However, as the number of remaining overs increases these differences become statistically significant and so it is more appropriate to use either the projected or new projected scores to generate a margin of victory in these cases.

Using the margin of victory generated by the projected and new projected scores to model team performance in the Australian domestic one-day competition (1994–2000) showed that the team-rating estimates were similar for both methods and the teams were ranked in the same order. Using each team's mean margin of victory to rank the teams showed general agreement with the rankings obtained by the model.

Based on scores estimated by the projected and new projected methods the teams, on average, experienced a common home advantage of eight and nine runs respectively. These results were not statistically significant. Home advantage based on estimates generated by the mean margin of victory showed some agreement with the model estimates. The application of binary logistic regression techniques also support the notion that teams on average did not experience a significant home advantage, with the odds of winning away estimated to be about 1.4 times the odds of winning at home.

References

[1] P. Allsopp and S. R. Clarke, "Fair measures of performance in one-day cricket", in *Mathematics 2000 Festival*, M. Bulmer *et al.* (editors), Melbourne, Victoria, Melbourne University (2000), 1–8.

[2] S. R. Clarke, *Performance Modelling in Sport*, unpublished PhD thesis (1998), Chapter 2.

[3] S. R. Clarke, "Test Statistics", in *Statistics in Sport*, J. Bennett (editor), Arnold, London (1998), 83–101.

[4] S. R. Clarke and J. M. Norman, "Home ground advantage of individual clubs in English soccer", *The Statist.*, **44**(4) (1995), 509–521.

[5] S. R. Clarke and P. Allsopp, "Fair measures of performance: The World Cup of Cricket", submitted.

[6] F. C. Duckworth and A. J. Lewis, “A fair method for resetting the target in interrupted one-day cricket matches”, *J. Oper. Res. Soc.*, **49** (1998), 220–227.

[7] F. Duckworth and T. Lewis, *Your complete guide to the Duckworth and Lewis method for resetting targets in one-day cricket*, University of the West of England, Bristol (1999).

[8] D. A. Harville and M. H. Smith, “The home-court advantage: How large is it, and does it vary from team to team?”, *Amer. Statist.*, **48**(1) (1994), 22–28.

[9] T. Lewis and Duckworth, “A fair method for resetting the target in interrupted one-day cricket matches”, in *Third Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1996), 51–68.

[10] R. T. Stefani and S. R. Clarke, “Predictions and home advantage for Australian Rules football”, *J. Appl. Statist.*, **19**(2) (1992), 251–261.

IDENTIFYING ARBITRAGE OPPORTUNITIES IN AFL BETTING MARKETS THROUGH MATHEMATICAL MODELLING

Michael Bailey
School of Mathematical Sciences
Swinburne University
PO Box 218
Hawthorn
Victoria 3122, Australia

Abstract

Through the use of multiple regression on historical data, it is possible to identify numerically quantifiable factors that can be used to predict the outcomes of AFL games. These predictions can then be compared with the fixed prices offered on the betting market and possible arbitrage opportunities identified. Using complete match results from the 1997 and 1998 AFL seasons a prediction model was developed that, when used on the 1999 season offered positive returns on investment.

1 Introduction

Applying mathematical models to sport is not a new concept. In the past, attempts to model sporting outcomes were always hampered not only by the lack of quality data, but also by the time taken to accurately enter the data for computer modelling. Today, with the growth of the Internet has come a rapid increase in the amount of readily accessible data from which to explore sporting outcomes.

In the early nineties, authors such as Clarke [1] and Stefani and Clarke [6] clearly showed that a certain degree of success in predicting AFL outcomes can be achieved through the use of computer modelling. Similarly, Dixon and Cole [2] have shown that regression techniques can successfully be applied to English football matches in order to identify market inefficiencies. Using a least squares approach, this paper attempts to identify arbitrage opportunities and therefore facilitate a positive return on investment.

By subtracting the away team score from that of the home team, it is possible to model the outcome of AFL football games as a continuous variable. Because the margin of games can be seen to follow an approximately normal distribution (Figure 1), the underlying assumptions needed for least squares regression seem to hold.

Collection and organisation of past match data enables the development and testing of factors that can be used to effectively predict variation in the margin of victory. By testing combinations of predictors using multiple regression, consistent factors that explain variation in the outcome of matches can be identified. The corresponding prediction equation is then applied to future match statistics to produce a predicted margin of victory. By dividing the predicted margin of victory by the calculated standard error and comparing with the standard normal probability distribution it is then possibly to estimate each sides probability of winning. Further comparison with market probability (1/market price) enables detection of market imbalances.

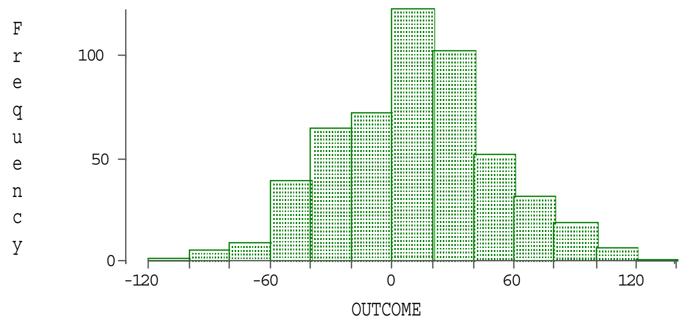

Figure 1: Margin of all AFL home/away games 1997–1999 (home score − away score).

2 Model development

At a univariate level, there are a large number of predictors that can explain variation in the margin of victory. Team differences in age, weight, experience, kicks, marks, disposals, turnovers, times inside 50 and previous winning margins can all explain small but statistically significant portions of variation. The strongest individual predictor of outcome is market price. In the seasons 1997–99 the favourite (shortest price on offer) won 61% of matches. As a continuous predictor, market prices can explain approximately 12% of the variation in the outcome. (see Figure 2.)

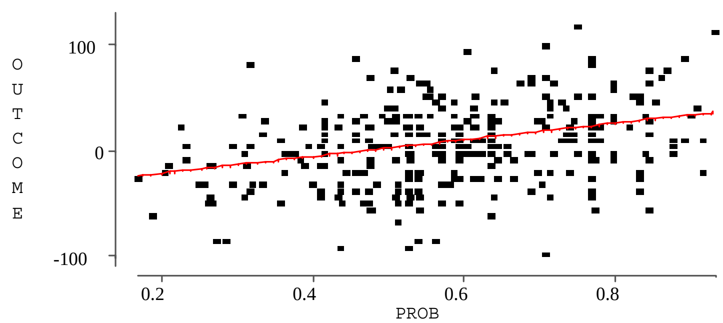

Figure 2: Linear regression of outcome vs market probability (1/price). $R^2 = 12\%$ ($p = .0001$), 1997–1999.

After adjusting for the effects of market price, most of the variables that were significant at a univariate level no longer remained significant. This comes as no surprise, because the market price effectively acts as combination of many variables that seem to effect the outcome of the match. Intuitively, any significant increase in model R^2 over that of the market price represents an increased "knowledge" of the true odds in comparison to the market. Using a combination of factors collected solely at a team level (see Table 1), it was possible to explain an additional 4% of variation giving an overall R^2 of 16%. Note that as outcome is modelled as home score − away score, the home ground advantage (62% of home sides won in 1997–99) is automatically included into the equation in the form of the intercept.

Team data can be collected in two ways. Firstly, the average of the team as a whole, and secondly the team as a sum of the individuals involved. Significant improvement in the amount of variation explained can be achieved by modelling at an individual level. This process also allows a surrogate measurement for injury—a key factor thought by many to influence the outcome of the matches. By modelling individual players on factors known to be predictive at a team level (goals, inside 50, turnovers etc.) and then aggregating the values up to a team score, it is possible to compare sides for who they have playing in each specific game rather than comparing team form. Interestingly, the difference between the team average and the sum of the individuals is a significant predictor. This makes intuitive sense, because, if a team fields a weaker side through injury, it could be expected to perform worse. Table 2

Variable	*Partial* R^2	*Model* R^2	F*-test*[1]
Intercept (home ground adv.)			0.0001
Market price	12%	12%	0.0001
Past margins	2%	14%	0.01
Turnovers	1%	15%	0.01
Inside 50	1%	16%	0.01

Table 1: Variation explained in multiple regression (1997–1998).
[1] Test for significance of partial R^2.

gives an example from a recent pre-season match in which Brisbane was a participant. It shows statistics that can be gathered for each match at an individual player level.

Name	*TK*	*TH*	*DI*	*RE*	*IN50*	*MA*	*HO*	*CL*	*TO*	*FF*	*FA*	*TK*	*G*
Akermanis, J.	9	10	19	1	0	5	1	2	1	1	0	0	1
Ashcroft, M.	10	8	18	1	2	2	0	9	0	1	1	0	0
Bolton, C.	7	2	9	3	2	1	0	1	1	0	0	0	0
Bradshaw, D.	4	1	5	1	0	0	0	0	1	1	3	0	3
Champion, R.	5	2	7	3	0	3	0	0	3	1	2	3	1
Cupido, D.	2	2	4	2	0	0	0	0	2	0	2	2	1
Hart, S.	6	5	11	3	1	0	0	1	0	0	0	1	0
Headland, D.	0	0	0	0	0	0	0	0	1	0	0	0	0
Heuskes, A.	6	9	15	5	3	0	1	1	2	1	2	1	0
Johnson, C.	4	4	8	3	1	2	0	0	2	0	1	1	2
Kenna, S.	6	3	9	1	0	3	0	1	1	1	1	0	1
Kennedy, M.	5	1	6	2	0	2	0	0	0	0	0	0	0

Key:

TK = total kicks, *TH* = total handballs, *DI* = disposals, *RE* = rebounds, *IN50* = times inside 50 zone, *MA* = marks, *HO* = hit outs, *CL* = clearances, *TO* = turnovers, *FF* = frees for, *FA* = frees against, *TK*= tackles, *G* = goals.

Table 2: Example of individual match statistics.

The most difficult part of establishing a model to identify arbitrage opportunities is to ensure that estimates remain free from bias. When fitting data, care must be taken to ensure that predictions are established for the right time frame. For example, Collingwood's average after 22 rounds cannot be reapplied to Round 4. Clearly, if an unbiased estimate was to be established for Round 4, then only data prior to Round 4 can be used. Similarly, due to the nature of the modelling process, it is difficult to avoid "overfitting" past data. It is for this reason that hold-out samples should be used determining the true predictive quality of the models.

Using the model developed on seasons 1997 and 1998 that include past margins, turnovers and inside 50 collected at a team level (Table 1), it was possible to test the validity of the model by applying it to 1999 data. As a predictor this model performed reasonably well, explaining as much variation in 1999 ($R^2 = 16\%$) as it did in 1997/98 ($R^2 = 16\%$). By using the predicted margin as a guide to the actual match winner, this model produced the winning side in 64% of matches.

3 Wagering strategy

In a fixed price market with only two outcomes, the market price becomes a reflection of supply and demand. By multiplying the predicted probability by the market price and subtracting the original unit bet, it is possible to gauge the size of any market imbalance and thus the perceived advantage.

$$A = (P \times M) - 1 \tag{1}$$

where A = advantage, P = predicted probability, and M = market price.

For example, Hawthorn starts favourite (M = \$1.50) against Fremantle (\$2.40). Using a prediction model (Table 1), Hawthorn is given a 75% chance of success ($P = 0.75$), thus $A = (0.75 \times 1.5) - 1 = 0.125$ or a 12.5% advantage. When an interstate side plays in Melbourne, local betting markets sometimes reflect increasing support for the home side—i.e. Hawthorn shortening from \$1.50 (r) \$1.30. This would effectively reduce any betting advantage on Hawthorn: $A = (0.75 \times 1.3) - 1 = -0.025$. With few local betters willing to back the interstate side the price accordingly drifts. If the price drifts too far, an arbitrage opportunity will then exist on the least favoured side. Given Hawthorn's probability of winning at 75%, Fremantle has a 25% chance of winning (excluding a draw). If the price on Fremantle rises to \$4.20, a positive arbitrage opportunity will exist: $A = (0.25 \times 4.2) - 1 = 0.05$, or a 5% advantage. In order to maximise growth of wealth, a betting strategy must incorporate three specific features in determining the size of the wager, namely existing bank size, size of advantage and probability of winning. Kelly [3] developed a betting strategy designed to maximise the growth of wealth by maximising the expected log of wealth. This formula can be effectively simplified to

$$B = \frac{A}{M - 1}, \tag{2}$$

where B = % of total wealth, A = advantage, and M = market price.

Returning to our example above in which Hawthorn was estimated to have a 75% chance of winning, at a price of \$1.50 there would be an advantage of 12.5%, thus the betting fraction $B = 0.125/(1.5 - 1) = 0.25$, represents 25% of total wealth. If Fremantle was paying \$4.20 against a predicted probability of 25%, the advantage and probability of success are both much smaller: hence a bet size equivalent to only 1.5% of total wealth.

Further research into the Kelly wagering strategy by MacLean, Ziemba and Blazenko [4] suggests that the Kelly model in its present form can often be too volatile in nature and a fractional Kelly criterion in which a fraction (e.g. $\frac{1}{2}$) of the recommended bet is placed, offers a greater security. As AFL football matches are often played simultaneously, a full Kelly criterion would not be appropriate, as the investor would not have a complete bank available for each match.

Due to high volatility and the lack of reliable data at the start of each season, it makes sense to restrict betting until sufficient data is available. The predictive power of the model examined was found to significantly increase with the number of rounds played. For the purposes of this paper, the first three rounds of each season were excluded from the analysis.

For each given match it is possible to determine whether either side is undervalued or overvalued by the market. By restricting betting opportunities to those in which the perceived advantage is at least equivalent to a certain amount, it is possible to "insure" against falsely predicting the true match probability. For this paper, a figure of 5% was arbitrarily chosen.

4 Results

By using the model shown in Table 1 that consists of only past margins, turnovers and inside 50 (all measured at a team level), and by restricting betting opportunities to those in which the perceived advantage is at least 5%, 96 arbitrage opportunities were identified after Round 3 of the 1999 season (162 matches). To evaluate the most feasible wagering strategy, two fractional Kelly regimes were considered (half and third) along with a fixed bank strategy in which a fixed bank size of \$100 was used.

In each situation, the starting bank size for each was $100. From Figure 3, it is possible to see that all three methods had positive returns for the season, with the fixed strategy being the most successful ahead of the Half Kelly and Third Kelly models.

Interestingly, similar results could be achieved by considering only matches in which the perceived advantage was greater than 10%. In this case there was only 76 potential arbitrage opportunities available for the season but the final season figures ($380-fixed, $270-half and $210-third) were virtually identical to those seen in Figure 3.

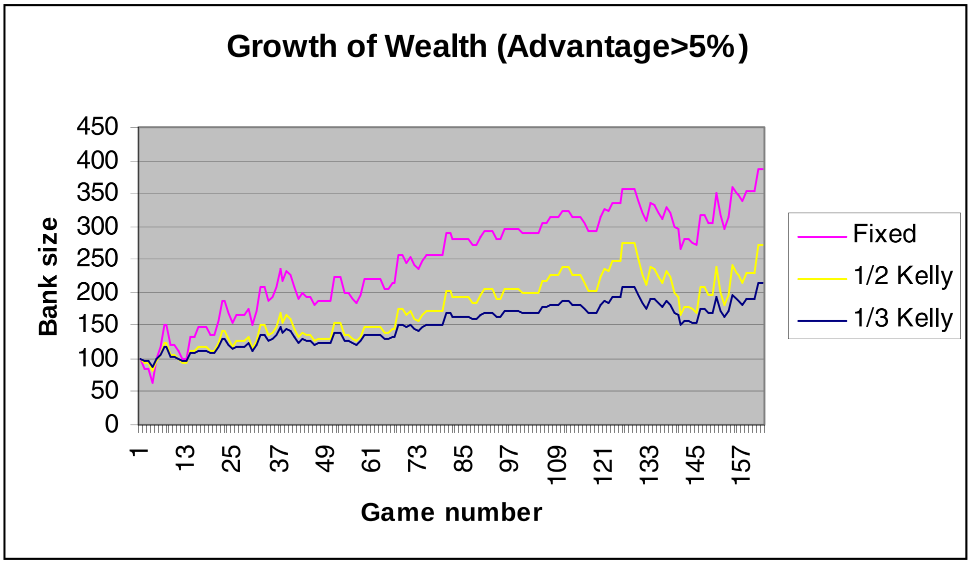

Figure 3: Growth of wealth, season 1999. Initial Bank = $100.

5 Discussion

Further model improvements can be gained by constraining betting opportunities at either end of the market. As prices increase, so too does the variation in return. Inversely, as prices shorten the trade off between return and risk can often be exceeded, especially considering over 80% of variation in the outcome still remains unexplained.

One further cause for bias in the prediction model is the over-inflated importance of large past wins. It could be argued that a team that is ten goals behind won't be trying as hard as a team that is a goal behind, and as such should be taken into consideration.

For the purposes of this paper, market price was taken as the price available through Centrebet on the Friday prior to the weekend round. By comparing the online prices offered by Centrebet, with prices offered in a more local Melbourne market (Sportsbet) it is clear that interstate biases occur, with the differences between bookmakers often being as large as 20%. With the growth of sports betting in general, a rapid increase in the number of bookmakers has led to further gains in wealth by shopping around for the best price.

6 Conclusion

Australian Rules football has evolved dramatically over the past ten years. The introduction of new rules, clubs and venues, have all had major impacts on the predicability of matches. For the purposes of this paper, data has been collected on the years 1997–1999. It could be questioned as to whether this is enough data to accurately construct models with long term predictive powers. For this reason, the author is happy to look upon the results of this study in the same light as that of a pilot study—Is it possible to statistically derive a handicapping model for AFL football that produces positive returns on investment? The answer is surely yes—as to the long-term feasibility of such a process, only more comprehensive longitudinal analysis will answer that question.

References

[1] S. Clarke, "Computer forecasting of Australian Rules football for a daily newspaper", *J. Oper. Res.*, **44**(8) (1993), 753–759.

[2] M. Dixon and S. Coles, "Modelling Association Football scores and inefficiencies in the football betting market", *J. Appl. Statist.*, **46**(2) (1997), 265–280.

[3] J. Kelly, "A new interpretation of information rate", *Bell System Technical J.*, **35** (1956), 917–926.

[4] L. W. MacLean, Ziemba and Blazenko, "Growth versus security in dynamic investment analysis", *Management Sci.*, **38**(11) (1992), 1562–1585.

[5] R. T. Stefani, "Football and basketball predictions using least squares", *IEEE Transactions on Systems, Management and Cybernetics*, **7** (1977), 117–121.

[6] R. Stefani and S. Clarke, "Predictions and home advantage for Australian Rules football", *J. Appl. Statist.*, **19**(2) (1992), 251–261.

A COMPARISON OF THE ATP RATINGS WITH A SMOOTHING METHOD FOR MATCH PREDICTION

Anthony B. Bedford and Stephen R. Clarke
School of Mathematical Sciences
Swinburne University
PO Box 218
Hawthorn
Victoria 3122, Australia

Abstract

A simple method of rating tennis players is investigated. An exponential smoothing method based on margin of victory in terms of sets and games won is used for rating players. The method is optimised over a year's data and compared with the ATP rating for predicting the winner of each match in major tournaments. Results suggest the exponential method performs as well as the more complicated ATP rating method for both match prediction and tournament seeding.

1 Introduction

Methods of rating teams and players across many fields of sport have been the topic of much discussion. The ATP has established two methods of rating tennis players. Now only used for tournament seeding is a moving twelve-month rating method that is based on both tournament and bonus points. In 2000 it was joined by a twelve-month race of point accumulation, called the Champion Race. For both the moving and race systems, point allocation depends upon progression within the tournament and class of tournament. Points are totalled from a maximum of fourteen tournaments in the prior year. Consequently ranking serves as a guide for seeding players within competitions, and is used also in allocation of prize money at the end of the year. As well as wishing to triumph in major opens, being highly ranked is the aim for all players.

Calculating a player's moving ATP rating is a complicated process. Tournament points and bonus points are allocated, with tournament points being those used as a player advances from round to round in a tournament, and bonus points are those a player gains for defeating an opponent based on their ranking. If a player advances to the next round by walkover (match never commenced) then no bonus points are allocated.

Stefani [6] describes how to compute the ATP ratings for players using matrix and vector methods. This provides an alternative way to calculate ratings in a more flexible manner than that described earlier. He extends this procedure further in Stefani [7].

The ATP web site www.atptour.com has a points breakdown for all the players, summarising the best fourteen results in the calendar year. This is very helpful as to calculate from scratch you require the points from all tournaments, including bonus points determined by the ratings at that time of opponents the player has defeated.

Further idiosyncrasies of the ATP ratings include:

- Any challenger tournament player providing hospitality will receive points of the next highest prize money level.
- Regular bonus points are awarded for defeating ranked players but double bonus points are awarded in the case of grand slam and best of five set finals in Super 9 ATP tournaments.
- Player gaining entry to an event through qualifying competition shall receive one half the points awarded to a second round loser in the main draw (in addition to points earned).
- In grand slam qualifying players receive one point for losing in rd1 (Round 1), three rd2 and five in rd3, whereas in ATP Championship Series events, players in qualifying receive one point for losing in rd1 and one quarter of the points a main draw second round loser receives (plus bonus points) for losing in rd2.
- If a player withdraws after 12 noon eastern time USA on the Friday before the start of an event, he will be penalised in the ratings by one event per 12 month period (i.e. best 13 instead of best 14 etc., per violation).

Further explanation of how the ratings work is available on the Web, with prize money and points allocation detailed at www.fortunecity.com/olympia/zola/206/faqp2.html, and at www.stevegtennis.com/tourpts99.txt. This method continued in 2000 for use of seeding, but was not used as the main guide to player ratings.

At the beginning of the 2000 tournaments, the ATP introduced another rating method called the Champion Race ranking, whereby all players start at zero and accumulate points in an annual race from season start to end. The idea is that by season's end the player with most points is the world number one. Every player, regardless of performances in the previous year, counts eighteen performances in their ATP race total. Grand slams and masters must be included in the best, as well as their best five from other international series events. The ATP states the Race gives an idea of the "hottest player at the moment". Further explanation of the Champion Race is available at http://cgi.atptour.com/players.

So there are now two contrasting rating methods, one a cumulative for the calendar year, the other on a cryptic fifty-two rolling week method. Other methods have been discussed in the literature. A method of rating racquetball players in Strauss and Arnold [8] uses maximum likelihood and moment estimation methods to obtain ratings, dependent upon the server probability of victory. Blackman and Casey [1] use game probabilities due to their availability and high probability of picking winners.

Clarke [3] suggested the use of a simple exponential smoothing system similar to the Elo [5] chess rating system. This is aimed to be simplistic in calculation as well as representative of the current form of a player. The method involves smoothing a margin of victory, which can be applied to all standards of players. The beauty of most forms of smoothers is the ability to retain overall trend whilst still updating with most recent occurrences. As tennis ratings should reflect a time related estimation of form and achievement, the use of a smoother seems the obvious choice over the more rigid and specialised ATP rating currently in use.

2 Calculating margin of victory

Clarke [3] suggested a derived margin of victory method based from the points, games, and sets won or combination of the three. The method adopted here is the SPARKS method, discussed in Clarke [3], where the name is derived from Set Point mARKS. The value is dependent upon the games won and sets won in determining a final score. While using points may lead to a more accurate measure, they are not readily available. The SPARKS rating of a player is not only dependent upon the outcome of the match and the quality of the opponent, but the score in sets and games. The formula used to evaluate a specific value for an individual player is given by

$$\text{SPARKS} = (\text{sets won}) \times 6 + \text{games won}. \tag{1}$$

In a game with final score 7–6, 6–4, 4–6, 4–6, 7–5, the SPARKS for the winner is given by (3 sets) $\times$ $6 +$ (28 games won) $= 3 \times 6 + 28 = 46$. The loser gets 2 sets $\times 6 + 27 = 12 + 27 = 39$. So in this close game, the winner has a SPARKS victory of 7. This gives a little more weight than just a games score (28 to 27) which does not account for the fact that a player has won the extra set.

The SPARKS method is also applicable when a game sees the withdrawal of one player through injury, or if the match is not completed. A set is not considered awarded if not completed. At Wimbledon 1999, a semi-final match between Pete Sampras and Mark Philippoussis ended when Philippoussis retired injured. He was leading the match 6–4, 1–2. The associated SPARKS was 13 to 6 in favour of the retiree, leading to the advancing Sampras dropping 10 in rating. It may be argued that the retirement of a player through injury represents a match where form is not of the normal type as one player is not competing up to their best. We have chosen to leave this in, as we wish to be as objective as possible.

Here are some possible test cases and associated SPARKS points:

A. "Triple Bagels" or 18 Love (6–0, 6–0, 6–0). Win: 36, Loser: 0, Margin: 36. This has occurred only once in the 90's: at the French Open, second round in 1993.

B. "Longest Match" 6–3, 6–4, 5–7, 6–7, 15–13. Win: 56, Loser: 46, Margin: 10. Wimbledon semi-final encounter in 1998.

C. "Loser is winner" 0–6, 7–6, 7–6, 0–6, 7–5. Win: 39, Loser: 41, Margin: -2.

There is no set limit to the upper boundary of the scores, but there is on the margin of victory, which ranges from 0 to 36. Case C is somewhat controversial since the margin of victory is in the loser's favour. This could be avoided by altering the multiplier in (1) to 8 for sets won. However, clearly this is a very close match, and the loser can probably be declared very unlucky. Note the ATP method would give this loser no reward, the same as the loser in Case A. This cannot arise in a 3 set match with a set multiplier value of 6. However, lowering this value to 4 gives the possibility of a zero margin, and to 3 an unlucky loser scenario.

3 Using margin of victory for player ratings

The general form for updating the rating of a player as given by Clarke [3] is

$$\text{New Rating} = \text{Old Rating} + \alpha(\text{actual margin} - \text{predicted margin}) \tag{2}$$

for some α.

This method was applied to Australian Rules football prediction by Clarke [2, 4], and the outcome was shown to be as accurate as expert tipsters.

Another advantage of this rating system is that the match results are not required to be kept after a game is completed and the rating is calculated, whereas ATP requires the record of the past twelve months results.

Since we require a large spread of player ratings, but the margin ranges from 0 to 36, we divide the player ratings by 100 to predict the margin in SPARKS. This results in (2) becoming

$$R_t = R_{t-1} + 100\alpha\left(\text{SPARKS}_{\text{player}} - \text{SPARKS}_{\text{opponent}} - \frac{R_{t-1}^{\text{player}} - R_{t-1}^{\text{opponent}}}{100}\right), \quad t > 0, \tag{3}$$

where $\alpha =$ player's rating at period t. (R_0 is the starting point. To use this SPARKS margin through a tournament, and indeed throughout the year, we require a starting rating, and in this case the ATP rating at the start of the year is the obvious choice.)

Repeated application of (3) shows the current rating is a weighted average of the difference in SPARKS margin of victory for all matches played previously with the weights decreasing geometrically. However, there is no requirement of storage of any value other than the player rating prior to the commencement, which is vastly simpler than the ATP method. Furthermore, as we are comparing expected with actual margins, it can occur that a victor of a match does not win by a large enough margin to increase their rating. In this way, it encourages players not to "lie down" in defeat or "take it easy" to victory.

To illustrate the procedure, consider the match between Wayne Black and Pete Sampras in the third round of the Australian Open 2000. The score of the match is given, with the pre-match rating bracketed. Here we will choose $\alpha = 0.06$. Sampras (SPARKS = 3038) d. Black (SPARKS = 748) 6–7, 3–6, 6–3, 7–5, 6–3. Whilst the victory to Sampras is as expected, the margin is not. The expected result is Sampras d. Black: $30.38 - 7.48 = 22.9$. That translates to a SPARKS victory of at least 23 for Sampras to have performed well enough to increase his rating. Sampras wins by 46 SPARKS to 36, translating to a margin of only 10, so his rating changes to $3038 + 100 \times 0.06 \times (10 - 22.9) = 2961$ whilst Black's rating increases to 830, even though he is eliminated from the tournament. Black is rewarded for taking Sampras to five sets in a tight contest. The ATP rating would give him zero.

Calculating the SPARKS and ATP ratings for the calender year is a difficult task to undertake retrospectively, as matching players, matches, scores and current ratings is needed. Obtaining the data required the use of the Internet. A few sites had extensive data stored in text files, which needed some editing before analysis. The best site for obtaining data was Steve G's site at www.stevegtennis.com. These files contained ATP ratings on a week by week basis, and all tournaments played in the calendar year. The ATP site www.atptour.com did not have such information, only the current ratings. The task here involved several packages, including conversion into a library file using SAS then analysis in Excel. Data for 1999 was collated into one large Excel file and manipulated in order to calculate the moving SPARKS ratings. All players that were not in the ATP at the start of the year were given a score of zero. As well as including all ATP tour events, Davis Cup and non-ATP events were also included, increasing the sample size.

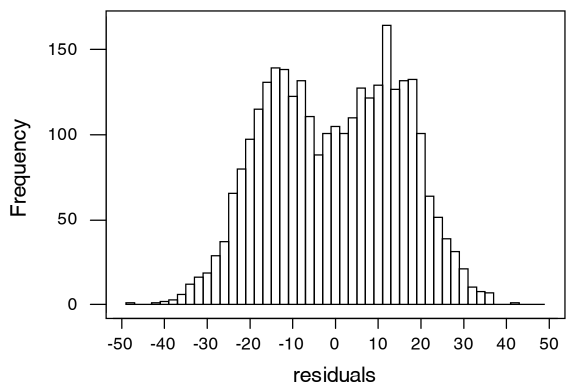

Figure 1: Distribution of SPARKS residuals in 1997.

To identify exactly what values to use in calculating players rating, the Solver function in Excel was used. It allows the optimisation of values subject to constraints. Here we wish to minimise the error by altering α. Furthermore, Solver can be used to adjust both the initial starting value (if the ATP is not used) and the set multiplier which was chosen initially as six. As data were already available for 1997, this was used as a guide for prediction in 1999 and 2000.

The 1997 set of data consisted of 3,306 matches whereby an opponent was ranked in the ATP list. Firstly, the choice was to minimise error by changing the smoother α. The prediction was adjusted where no definitive outcome could be made as both players began the match with equivalent SPARKS ratings. This number decreased as matches with lower rank players were removed, and subsequently the SPARKS method improves. Excel was used to investigate the predictive power for different values of the set multiplier and α. Unfortunately, Solver could not cope with both these parameters changing independently. Consequently, the set multiplier was fixed at 6 first, and Solver was used to minimise absolute errors in SPARKS. Solver converged to an exact solution of $\alpha = 0.063$.

The distribution of the SPARKS residuals is given in Figure 1, illustrating the bimodal distribution. The bimodal distribution is due to the shift caused by the set multiplier at 6.

With the set multiplier set to zero, the residuals are more symmetric, as shown in Figure 2. Whilst a set multiplier of zero may seem to be the best, it may not be acceptable to players, with far more matches having the winner is loser scenario.

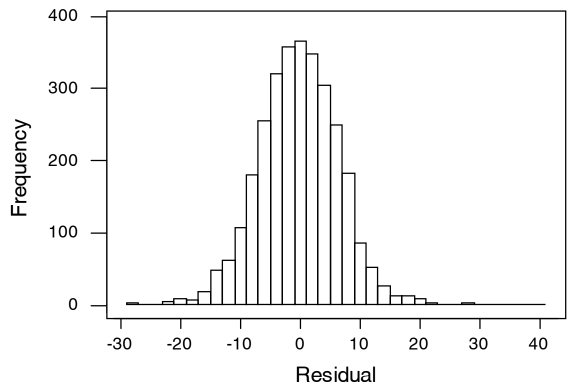

Figure 2: Distribution of SPARKS residuals in 1997 with set multiplier at 0.

After optimising α, different values of the set multiplier were checked against our initial estimate of six. Table 1 illustrates that the set multiplier is indeed best when set at six, returning the highest correct proportion for the 1997 data set. There is some small variation in percentage correct for multipliers 4 through 7. The ATP returns 63.7% correct, which represented 46 less matches correctly predicted than that by SPARKS.

Set multiplier	*% correct*
10	64.0
9	63.9
8	64.0
7	64.2
6	64.6
5	64.5
4	64.4
3	64.0
2	63.5
1	63.5
0	63.2

Table 1: The percentage of all matches correct using SPARKS in 1997 with α fixed at 0.063 and varying set multipliers.

The success rate of both ATP and SPARKS prediction increased when lower ranked players were discarded from the 1997 sample, as can be seen in Figure 3.

This shows the percentage correct in matches where both players are above the cutoff rating. As the cut off increases, the sample size decreases, leading to estimates of a smaller proportion of the total number of matches in the year. For example, in 1997, there were 172 matches where both players' ATP rating exceeded 3,250, and 77% of those players rated higher by the ATP in a match were victorious.

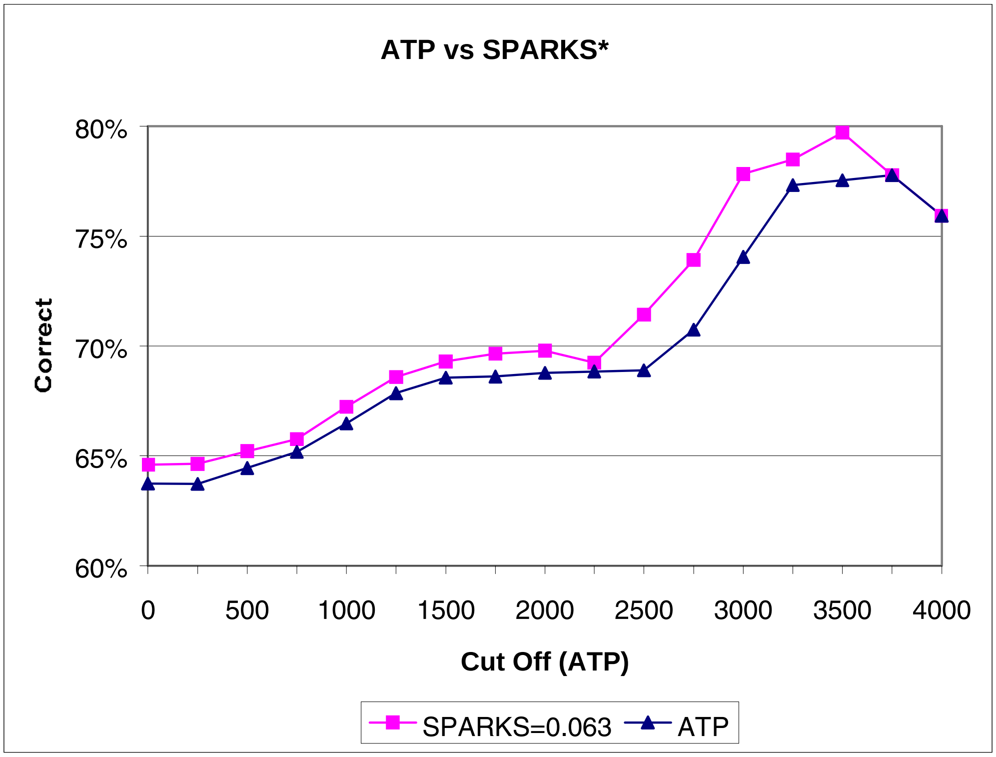

Figure 3: Percentage correct for 1997 for SPARKS and ATP, where matches in which both the players are at or above the cut off are included in the sample.

Now that these values were established, more current data was used for tournament prediction. The 1999 data set resourced from the web contained 3,982 matches, and using this fixed value of α at 0.063 and set multiplier six returned a 62.5% correct prediction using the SPARKS method. This is poorer than the 1997 result, but expected since α is optimised on 1997 data. There are far more lowly ranked players included in the sample. This is highlighted by the fact that 96 matches were not predictable due to equal ratings (3.5% of sample) with no cut off in place. The average absolute error for 1999 data is 14.4. (For interest, Solver was again used to minimise errors for α and arrived at (after 1 hour 51 minutes) 0.102 for a 62.5% success rate. So the errors were decreased minimally with no significant improvement in prediction, further justifying the decisions based on the 1997 data.)

4 Major tournament prediction

Using the data sets available for 1997, 1999 and 2000, the four major tournaments were analysed for each year. A comparison between the seeding and prediction for the pre-existing ATP and SPARKS methods were compared, in the hope that SPARKS was comparable, even superior, to the ATP ratings and rankings. The percentage of matches predicted correctly is given in Table 2. In 1997 SPARKS performs better than the ATP, but this is expected as we optimised on this data set. Two out of four tournaments in 1999 were predicted better by the SPARKS method than using ATP ratings.

There is evidence of surface differences and home advantage influencing prediction when observing the differences in tournaments. The French Open success prediction is well below the other major tournaments, with the surface (clay) the most likely factor in the low results.

An alternative way to assess the merits of the system is to compare player seedings prior to the

Year	*Tournament*	SPARKS	ATP
1997	Australian Open	69.5	69.5
	French Open	60.2	60.9
	Wimbledon	66.4	62.5
	US Open	74.2	64.1
	Average	67.6	64.3
1999	Australian Open	64.1	65.6
	French Open	60.2	60.9
	Wimbledon	68.8	67.2
	US Open	64.1	63.3
	Average	64.3	64.3
2000	Australian Open	68.8	68.8

Table 2: The percentage of all matches correct using SPARKS against ATP by tournament.

tournament. As sixteen players are seeded by the ATP, we will use our best sixteen prior to the commencement of the tournament for comparison. All quantities are the number of players making the stage from the sixteen. Rounds refer to number of players.

Year	*Tournament*	*Round of 16*		*Round of 8*		*Round of 4*		*Round of 2*	
		SPARKS	ATP	SPARKS	ATP	SPARKS	ATP	SPARKS	ATP
1997	Australian Open	9	9	7	7	3	3	1	1
	French Open	7	6	2	2	1	1	1	1
	Wimbledon	8	8	3	3	1	1	1	1
	US Open	11	7	7	4	4	2	2	1
	Total	35	30	19	16	9	7	5	4
1999	Australian Open	6	5	4	3	2	1	2	1
	French Open	6	7	4	4	1	1	1	1
	Wimbledon	8	10	7	7	4	4	2	2
	US Open	9	8	5	5	4	3	2	2
	Total	29	30	20	19	11	9	7	6
2000	Australian Open	7	7	4	5	3	4	1	2
	Overall	71	67	43	40	23	20	13	12

Table 3: Number of players from pre-tournament seeding by SPARKS and ATP methods to reach final rounds of major tournaments.

Clearly, the SPARKS method is an excellent guide to tournament outcomes, with equal or better results in most stages of these tournaments. In particular, for the US Open SPARKS prediction is superior to the ATP. This may be due to the event commencing later in the calendar year, giving the SPARKS ratings a longer period of assessment, as they were initialised in 1997 and 1999 at the beginning of those years using the ATP rating. The good results here are also despite the ATP rankings being used for tournament draws where it often occurs that some of the top sixteen by the SPARKS rankings may meet in the early final rounds. Hence SPARKS performs extremely well considering the draw bias toward the ATP.

SPARKS also has the advantage over the ATP in that recent form is weighted into ratings. After the completion of Wimbledon 1999, the ATP rankings had the victor, Pete Sampras, fall from number one to number three, and his opponent in the final, Andre Agassi, rise from three to one. This highlights the drawback of the ATP system as an indicator of a player's form at the present moment. The SPARKS rating had Andre Agassi at number one prior to the tournament, and he held his ranking but gave significant ground to Pete Sampras who was number three and rose to number two.

At Wimbledon 1999, of the 127 matches, eight times a winning player dropped in rating. These players were all rated above 1400 SPARKS at the time, showing that it is necessary to more than just win a match if you are highly rated. For the Australian Open 2000, this only occurred twice.

The draw in the Australian Open 2000 was the subject of a lot of controversy, with what many considered the two best players, Pete Sampras and Andre Agassi, together in the same half on the draw. Indeed, they went on to clash in the semi-final in what was considered by many to be the best match of the tournament. The implementation of the SPARKS method for the purpose of seeding and subsequent draw would have seen the two in separate halves of the draw, destined to meet in the final. Analysis of the draws using SPARKS ratings reiterated public and media opinion that the lower half of the draw was the easier half. This is highlighted by the fact that the top half of the draw had a sum of 66,220 SPARKS as opposed to 59,182 SPARKS for the bottom half, accounting for the draw bias. Indeed, for each stage of the tournament, the top half clearly exceeded the bottom half in SPARKS ratings. On the other hand, the ATP had a more balanced draw with the top half summing to 45,979 and lower half 46,495 using ATP ratings. The finalist from the lower half, Yevgeny Kafelnikov, was poorly seeded using SPARKS (20th), but his easy draw saw him make the final. Whilst his rating rose from 1,513 prior to the tournament to 2,345 prior to the final, his opponent, Agassi, rose from 2,926 to 3,290, and Agassi defeated Kafelnikov three sets to one in the final with relative ease, reflecting the relative chances from the SPARKS ratings. The semi-final two days earlier between Sampras (3,112) and Agassi (3,234) went to five sets, and left both fans and media believing the final had already been played.

After the Open, in ATP ranking, Kafelnikov dropped from two to three, and Sampras rose from three to two, even though Sampras failed in the semi-final and Kafelnikov made the final! SPARKS had Agassi and Sampras one and two pre-tournament and post-tournament. The Champion Race for the ATP post tournament had Agassi one, Kafelnikov two and Sampras five.

5 Further discussion

Another dilemma is the inclusion of a decay factor for players that do not play for an extended period of time. Two factors can be added to equation (3), a smoothing value for decrease in rating, and an indicator variable dependent upon the number of weeks to begin implementing the smoothing value. Again this could be achieved using Solver. The problem is that there have been instances when returning players of high calibre are poorly rated (e.g. Andre Agassi in 1997), and on other occasions players return ranked too highly (e.g. Mark Philippoussis in 1999). This issue can be extended to include those matches where one player withdraws through injury. A player may be hampered for the entire match or injured only at the point of withdrawal. It may be worth investigating the removal of such matches from the sample.

Unfortunately, the sparse availability of data for the WTA still sees no inclusion of ratings for women. At the time of publishing, 1999 results data for WTA had been archived at

http://neptune/spaceports.com/ lovegame/1999

without WTA player ratings, so restricted analysis could be undertaken. Indeed, the women's draw is considered to be far more consistent, with media speculation that there are far fewer upsets than in the men's draw.

Furthermore, there are no five set tournaments in the WTA, whereas the ATP tour has a mixture of both. This should lead to a difference in the smoothing constant and set multiplier, as for SPARKS analysis of 1997 we chose to include both three and five set encounters. There may be a case for separate constant/multiplier combinations for the type of match (three or five sets), surface, and tour type (women's or men's).

As has been well established, court surface is another factor to be considered. The ground is clearly the most significant external factor, with some players clearly performing poorly on different surfaces. This is explored in detail in a model proposed in Blackman and Casey [1] where their model factors court surface speed and correction factors for the given surface. They use extended Kalman filter measurements. Although this is not factored into SPARKS, it may be possible using a similar technique as another possible rating to be used in conjunction with SPARKS. Indeed, if these ratings are established for all competitors in a tournament, then it is possible to adjust all SPARKS for a particular tournament with respect to the surface.

6 Conclusion

SPARKS has proved to be a valuable predictor of results for both tournament and yearly results. It is comparable to the ATP as a predictor for yearly and tournament results, often surpassing the ATP in successful prediction. It is also easy to update and simple to understand, especially when considering the existing ATP methods, which are both non-systematic and specialist.

SPARKS ratings are also a valuable tool in setting of tournament seeding. SPARKS ratings and seeding reflect the current form of players in a manner easily understood by followers of the game. The outcome of tournament results on a player's ranking are far more reflective of recent form than the ATP method, which has resulted in numerous lead changes inconsistent with immediate results.

The SPARKS method can be easily applied to any tennis tournament, and is a method that is simple to implement by any association. It is an effective way for association officials at all levels to implement seeding for upcoming tournaments, and guide players to their standing within a competition. The current study has shown it to be an effective method of rating players in the top grade of professional tennis. Further investigation of its use for player comparison throughout the full range of tennis levels is warranted.

References

[1] S. S. Blackman and J. W. Casey, "Development of a rating system for all tennis players", *Oper. Res.*, **28**(3) (1978), 489–502.

[2] S. R. Clarke, "Computer forecasting of Australian Rules football for a daily newspaper", *J. Oper. Res. Soc.*, **44**(8) (1993), 753–759.

[3] S. R. Clarke, "An adjustive rating system for tennis and squash players", in *Second Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1994), 43–50.

[4] S. R. Clarke, "Computer and human tipping of AFL football—a comparison of 1991 results", in *Second Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1994), 81–93.

[5] A. E. Elo, *The Rating of Chess Players, Past and Present*, London, Batsford (1978).

[6] R. T. Stefani, "Survey of the major world sports rating systems", *J. Appl. Statist.*, **24**(6) (1997), 635–646.

[7] R. T. Stefani, "A taxonomy and survey of sports rating systems", in *Statistics In Sport*, J. Bennett (editor), Arnold, London (1998), 253–258.

[8] D. Strauss and B. C. Arnold, "The rating of players in racquetball tournaments", *J. Appl. Statist.*, **36**(2) (1987), 163–173.

DYNAMICS OF THE LONG JUMP

Maurice N. Brearley
85 Dandarriga Drive
Clifton Springs
Victoria 3222, Australia

Neville J. de Mestre
School of Information Technology
Bond University
Gold Coast
Queensland 4229, Australia

Abstract

The main dynamical aspects of the long jump are discussed. In particular, it is shown how the length of a jump can be calculated from a knowledge of the approach velocity of an athlete and the forces exerted by the ground on the take-off foot. For this purpose, use is made of measurements taken from a film of an Olympic champion athlete in action.

1 Introduction

The principles involved in the execution of a high standard long jump are well known. A good result depends on optimisation of several aspects of the jump, such as approach, take-off, flight and landing. In this paper, attention will be mainly confined to the take-off and flight, with air resistance ignored.

Use will be made of the magnitude of the forces exerted on the foot of an athlete during contact with the take-off board. This information was obtained from a video [3] showing an Olympic champion athlete in action.

For simplicity the paper will be written in terms of a male athlete, it being understood that a word such as "he" is intended to be read as "he or she".

2 Contributions to the length of a jump

Figure 1 depicts an athlete at the instants of take-off and landing of a long jump.

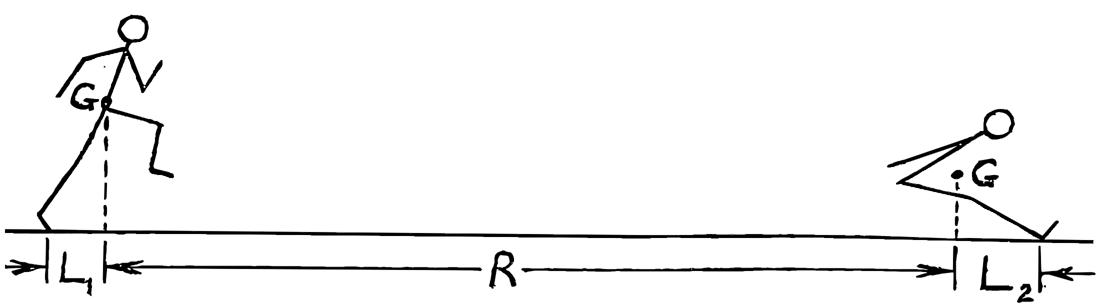

Figure 1: Contributions to the length of a long jump.

The lengths L_1, L_2 and R shown in Figure 1 refer to positions of the centre of mass G of the athlete at the instants of leaving the take-off board and first contacting the landing pit. They will be considered in turn.

(i) *The take-off distance L_1*

L_1 denotes the distance by which G is ahead of the front edge of the take-off board at the instant of take-off. Since the length of the jump is measured from the front of the board it is desirable for the athlete's foot to be as close as possible to this edge. The penalty for overstepping is severe, the jump being ruled invalid if this occurs. The mean value of L_1 for a large sample of male athletes is quoted by Hay [4] as 0.24 metres.

(ii) *The landing distance L_2*

L_2 denotes the distance by which the athlete's feet are ahead of G at the instant when they touch the surface of the sand in the landing pit. The forward momentum of the athlete should be great enough to carry him over the point of contact; if it is not, and he sits down behind this point, the measured length of the jump is reduced accordingly. The mean value of L_2 for a large sample of male athletes is quoted by Hay [4] as 0.53 metres.

(iii) *The flight distance R*

To be in the desirable landing attitude shown in Figure 1 the athlete's body needs to be leaning backwards to a greater degree than it was at the instant of take-off. This can be achieved during passage over the flight distance R by suitable movements of the athlete's body. Moving his feet as if "running in the air" (an action called a "hitch-kick") engenders forward angular momentum in his legs because the bending of each leg during its forward motion greatly reduces its moment of inertia about G. Since there is zero moment of external forces about G during flight, the total angular momentum of his body must remain unchanged, which means his upper body must acquire backward momentum, and it is this which gives his body the extra backward tilt needed for a good landing attitude. If properly executed, a hitch-kick of $1\frac{1}{2}$ or $2\frac{1}{2}$ in-the-air strides will result in the correct landing attitude. Just before landing, the athlete swings both feet forward to maximise his jump length; conservation of angular momentum causes his torso to lean forward, resulting in the "jack-knife" landing position shown in Figure 1.

Calculation of the flight distance R depicted in Figure 1 will be considered in the next section.

3 Calculation of the flight distance

The path followed during flight by the centre of mass G of the athlete is determined by the velocity of G at take-off, and cannot be altered by any in-flight movements of the athlete. The effect of air resistance during flight is very small; it has been shown by Brearley [1] that for a jump of 8.90 m the length reduction caused by air resistance during flight is less than 9 cm at sea-level.

The flight distance R shown in Figure 1 may be determined by means of the well known theory of projectiles (e.g. Bullen [2]). The standard notation is as shown in Figure 2.

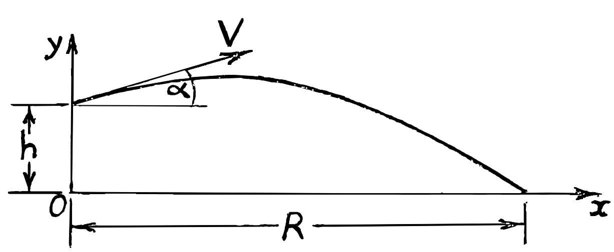

Figure 2: The standard notation in theory of projectiles.

In the case of the long jump, R is the flight distance and h is the difference between the heights of G at the instants of take-off and landing depicted in Figure 1. The theory of projectiles shows that, when h is small compared with R, a close approximation to the value of the flight distance is

$$R = \frac{V^2 \sin 2\alpha}{g} + h \cot \alpha, \tag{1}$$

where α is the angle of projection shown in Figure 2.

In the following section it will be shown how an athlete's flight distance can be found from a knowledge of his approach speed and the force acting on his foot during take-off.

4 Flight distance from approach speed and take-off force

Let u_0, v_0 denote the horizontal and vertical components of the velocity V shown in Figure 2. Then

$$V = \sqrt{u_0^2 + v_0^2}, \tag{2}$$

$$\alpha = \tan^{-1} \frac{v_0}{u_0}. \tag{3}$$

To determine the values of u_0 and v_0 it is necessary to investigate the take-off in some detail. Figure 3 shows the athlete in the positions that correspond to the instants when his foot first makes contact with and finally leaves the take-off board.

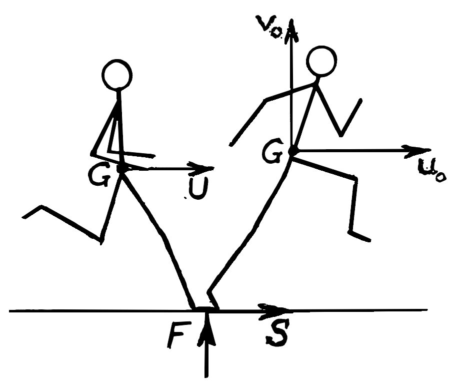

Figure 3: Positions at start and end of take-off.

Figure 3, in addition to the velocity components u_0, v_0 already introduced, shows

U = the velocity of G at the instant of foot plant,

F, S = vertical and horizontal components of the force exerted by the ground on the athlete during take-off.

The direction of U may be taken as horizontal. The values of F and S vary greatly during the brief period for which the athlete's foot is in contact with the take-off board. Experiments have been performed on a number of top-ranked athletes in which their movements have been recorded on video tape, enabling accurate estimates to be made of the approach velocity U. At the same time, by means of an electronic force plate mounted in the take-off board, the values of F and S throughout the take-off have been recorded, as well as the duration τ of the contact with the board. From these data the flight distance R can be calculated in the following way.

The vertical equation of motion of the athlete during the take-off depicted in Figure 3 is

$$m\frac{dv}{dt} = F - mg, \tag{4}$$

where

$$\begin{aligned} v &= \text{the upward vertical velocity of } G \text{ at time } t, \\ m &= \text{the mass of the athlete,} \\ g &= \text{the acceleration due to gravity.} \end{aligned}$$

Integrating (4) with respect to time over the interval τ of the take-off yields the principle of momentum

$$mv_0 = \int_0^\tau F\,dt - mg\tau,$$

the initial vertical velocity of G being taken as zero. Hence

$$v_0 = \frac{1}{m}\int_0^\tau F\,dt - g\tau. \tag{5}$$

The horizontal motion of the athlete during take-off may be treated similarly. In place of (4) we have

$$m\frac{du}{dt} = S, \tag{6}$$

where

$$u = \text{the forward velocity of } G \text{ at time } t.$$

The corresponding principle of momentum is

$$m(u_0 - U) = \int_0^\tau S\,dt,$$

since the initial forward velocity of G is U. Hence

$$u_0 = U + \frac{1}{m}\int_0^\tau S\,dt. \tag{7}$$

In the next section the use of (5), (7), (2), (3) and (1) in calculating flight distance will be illustrated by a particular example.

5 A particular example of a long jump

At the 1980 Olympic Games in Moscow the long jump was dominated by the German athlete Lutz Dombrowski. He won the event with a leap of 8.54 m (28 ft $0\frac{1}{4}$ in); this was the first time that anyone had jumped over 28 feet since Bob Beamon had leaped 8.90 m (29 ft $2\frac{1}{2}$ in) in the 1968 Mexico City Olympics.

A video film [3] of Dombrowski, made for the purpose of training athletes, shows that his approach speed on reaching the take-off board was $U = 9.0\ \text{ms}^{-1}$. He would have been running faster than this earlier in his approach, but some speed reduction occurs during the last stride.

The take-off board used for the training film was fitted with a force plate that measured during the take-off the force components F and S shown in Figure 3. In the film the graphs of F and S as functions

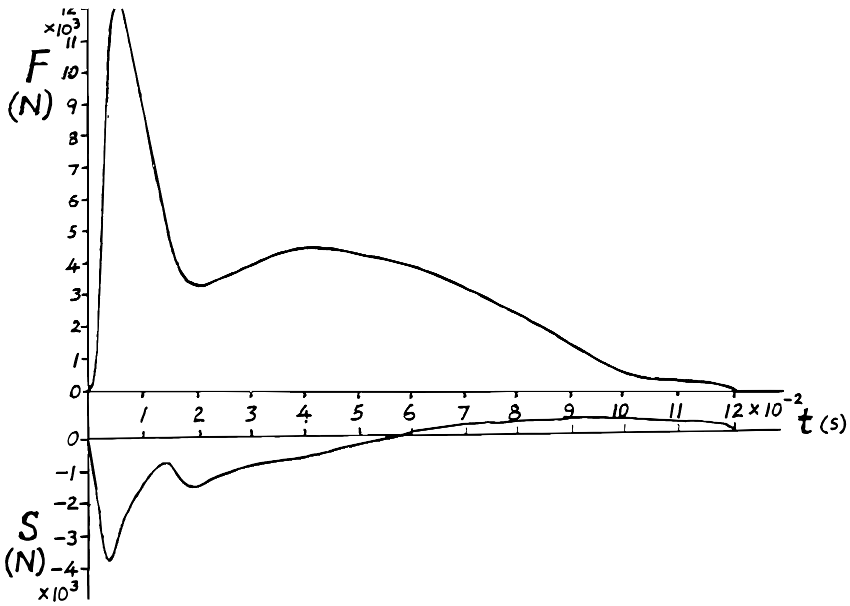

Figure 4: Force components F and S during take-off.

of time were displayed, and these are shown in Figure 4. The horizontal axes for F and S have been separated to avoid overlapping of the graphs.

As expected, the forward component S is negative during the early part of the take-off, and changes to positive about half way through the take-off. The duration of the athlete's contact with the take-off board is 0.12 seconds. Both F and S peak immediately after contact is made with the board because of the nearly impulsive nature of the contact. The maximum value of F is about 12,000 newtons.

The value of the vertical impulse in the right hand side of equation (5) was found by numerical integration, using Simpson's One-Third Rule. It is calculated that

$$\int_0^{0.12} F\,dt = 399\,\mathrm{Ns}.$$

Taking Dombrowski's mass to be $m = 80\,\mathrm{kg}$, equation (5) yields

$$v_0 = \frac{399}{80} - 9.8 \times 0.12 = 3.8\,\mathrm{ms}^{-1}.$$

In the same way it is calculated that

$$\int_0^{0.12} S\,dt = -37\,\mathrm{Ns}.$$

Equation (7) then gives

$$u_0 = 9.0 - \frac{37}{80} = 8.5\,\mathrm{ms}^{-1}.$$

Equations (2) and (3) then show that

$$V = 9.3\,\mathrm{ms}^{-1}, \qquad \alpha = 24^\circ.$$

A reasonable estimate for the difference in the heights of the two positions of G shown in Figure 1 is $h = 0.5\,\mathrm{m}$. Equation (1) then shows the range for the jump to be

$$R = \frac{(9.3)^2}{9.8} \sin 48^\circ + 0.5 \cot 24^\circ = 7.7\,\mathrm{m}.$$

To this can be added the values mentioned in (i) and (ii) of Section 2 of the take-off distance L_1 and the landing distance L_2 shown in Figure 1, giving the length of the jump by Dombrowski as

$$R + L_1 + L_2 = 7.7 + 0.24 + 0.53 = 8.5\,\mathrm{m}\ (27\,\mathrm{ft}\ 10\tfrac{1}{2}\,\mathrm{in}).$$

This calculated value is very close to the actual value of the jump length.

6 Summary and conclusions

The main features and dynamical principles of the long jump have been described. Using data from a film of an Olympic champion athlete, it has been shown that the length of a jump can be calculated from a knowledge of the athlete's speed of approach and the force acting on his take-off foot throughout its contact with the take-off board.

References

[1] M. N. Brearley, "The long jump at Mexico City", *Function Mag.*, **3** (Part 3), Monash University (1979).

[2] K. E. Bullen, *An Introduction to the Theory of Mechanics*, Science Press, Sydney (1971).

[3] Championship Form Series (video), *Biomechanical Analysis of Triple and Long Jump* (V GV1060.179) (1989).

[4] J. G. Hay, *The Biomechanics of Sports Techniques*, third edition, Prentice–Hall Inc. (1985).

TEACHING STATISTICS THROUGH SPORT

K. J. Byun
Department of Statistics
Division of Economic and Financial Studies
Macquarie University
NSW 2109, Australia

Abstract

Sport is a marvellous tool for teaching statistical techniques and can turn an otherwise uninteresting subject for some into an exciting and challenging experience. This paper demonstrates how mathematical models and statistical analysis involving Olympic sporting events can be used to make predictions for the Sydney 2000 games. A simple mathematical model is derived based on previous Olympic data, enabling the prediction of the likely outcome for the upcoming Olympic high jump competitions for both men and women. Also included is a discussion of the asymptotic limits that can be reached in this event provided that the current techniques and rules still apply in the future.

1 Introduction

The use of scientific analysis in sport has enjoyed great momentum, both locally and overseas, with the Sydney 2000 Olympic Games just around the corner. Of great interest to public, athletes and their coaches, are the times, heights and distances that athletes and swimmers have to achieve in order to win a gold medal. This paper uses a simple mathematical model and statistical analysis to determine these values and to predict the winning heights for both the men's and women's high jump at the Sydney 2000 Olympics.

The model developed in this paper also considers the question of whether Olympic performances are likely to improve significantly. Certainly there has been a steady improvement in the past. For example, in the men's high jump, Ellery Clark won the event at the 1896 Olympic Games in Athens (the first modern Olympics) with a jump of 181 cm. Thirty-two years later in Amsterdam in 1928, Robert King jumped 194 cm, and after another 36 years at Tokyo in 1964, Valery Brumel cleared 218 cm. By the time of the Atlanta games in 1996, the winning jump had increased to 239 cm, a 32% improvement over a 100-year period. The Olympic winning heights from 1896 to 1996 are shown in Table 1.

The graph in Figure 1 shows the winning height at each Olympic Games from 1896 to 1996. Note that there are no points for 1916 (World War I), 1940 and 1944 (World War II) since there were no Olympics held in those years.

2 Developing a relationship model

During the past 100 years, the elite athlete has become faster and stronger due to improved diet and training techniques. However, it is reasonable to assume that athletic performances cannot continue to improve forever, since, for example, it is difficult to see a human running 100 metres in three seconds. Under current conditions, it follows that there will be an asymptote for athletic events that will not be surpassed unless there is some significant change in the human body, although this is unlikely to happen during the short term of our forecasts. We will consider a technique available for finding such an asymptote.

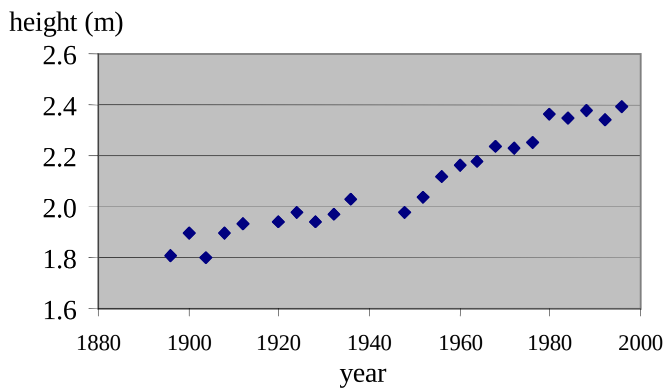

Figure 1: Winning heights for the men's Olympic high jump.

Year	*Height* (m)	*Year*	*Height* (m)
1896	1.81	1948	1.98
1900	1.90	1952	2.04
1904	1.80	1956	2.12
1908	1.905	1960	2.16
1912	1.93	1964	2.18
1916	N/A	1968	2.24
1920	1.935	1972	2.23
1924	1.98	1976	2.25
1928	1.94	1980	2.36
1932	1.97	1984	2.35
1936	2.03	1988	2.38
1940	N/A	1992	2.34
1944	N/A	1996	2.39

Table 1: Winning heights for men's Olympic high jump.

With this in mind, in the medium term of the next twenty years or so at least, it is difficult to imagine an increase of more than, say, 10 cm in the winning Olympic high jump since it has only increased by a total of 5 cm during the previous twenty years. This would place the winning jump in the year 2020 at roughly no more than 2.49 m. But how accurate is this likely to be? This was one of the questions that was open for student discussion.

Information provided to the students included the fact that a Cuban high jumper, Javier Sotomayor, jumped 2.45 m on 27 July 1993 in Salamanca, Spain and set the current world record for the men's high jump. This is already six centimetres higher than the current Olympic record set in 1996. The students were asked to investigate just how likely it is for the Olympic winning height to increase very far beyond this point and to try and estimate a reasonable limiting height that may not be reached, at least for a very long time. In addition, they were asked to make a prediction of both the men's and women's gold medal jump at the Sydney 2000 Olympics.

Bearing in mind that these students were only in their first years of statistical study, one approach they could have taken was the following technique using the Olympic data from 1896 to 1996 (see Searle and Vaile [1], and also [2]).

Although there are many models that could be used in making such a forecast, a least squares technique was used as a demonstration of the sort of analysis that might be undertaken. In this case, both the independent and dependent variables are subject to error. There may well be more efficient methods, and strictly speaking the independent variable should not be subject to random variation, but

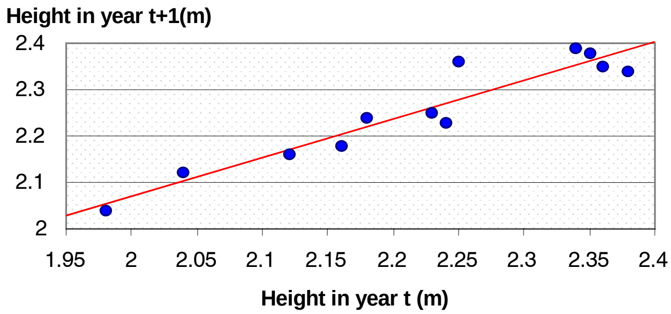

Figure 2: Scatter plot of men's Olympic high jump results of year $t+1$ vs year t since World War II (1948–1996).

according to Weisberg [3] alternative assumptions that allow for errors in the independent variables are much more complicated and there is no agreed correct way to proceed.

In this particular forecasting model, suppose that y_t is the winning performance achieved in the tth Olympics, and that

$$y_{t+1} = a + by_t + z_t, \quad t = 1, 2, 3, \tag{1}$$

where a and b are constants and z_t is a random error term. This means that the *expected* performance (that is, the performance ignoring unpredictable random error) at a given Olympics is linearly related to the performance at the previous Olympics. This relationship can be seen in Figure 2.

In the special case when b is 0, this model just says that the expected performance is equal to some constant a, and does not depend at all on the previous performance.

To estimate the values of the constants a and b, we can make a graph of y_{t+1} versus y_t and fit a straight line to the points in this scatter plot. The fitted line has the equation $y_{t+1} = a + by_t$, so that b is the slope of the fitted line and a is its intercept.

However, as can be seen from Figure 1, since it took some time for performances to recover after World War II, it seems reasonable to ignore all the data before this period. Consequently, only the data after World War II are plotted in Figure 2. This figure shows the scatter plot with the fitted line for the Olympic men's high jump winning heights of $(t+1)$th Olympics vs tth Olympics from 1948 to 1996 where 1948 Olympics is the first point.

The fitted line in Figure 2 has the equation $y_{t+1} = 0.406 + 0.833y_t$, and this model seems to fit the data quite well with $R^2 = 0.8996$. This means that nearly 90% of the variation in the winning height jumped is due to the winning height jumped at the previous Olympics.

3 Developing a mathematical model for prediction

Using the model, we can derive a formula for the expected value of y_t using repeated substitution of earlier data, as follows:

$$\begin{aligned} y_t &= a + by_{t-1} \\ &= a + b(a + by_{t-2}) \\ &= a + b(a + b(a + by_{t-3})) \\ &\ldots \\ &= a + ba + b^2a + b^3a + \ldots b^{t-1}a + b^t y_0 \end{aligned}$$

$$= a(1 + b + b^2 + b^3 + \dots b^{t-1}) + b^t y_0$$
$$= a\frac{1-b^t}{1-b} + b^t y_0.$$

Consider what happens when t is large, that is, a long time in the future. If b is between 0 and 1, b^t approaches 0 as t gets large, so the mean value of y_t approaches $a/(1-b)$. This is the limiting expected performance (i.e. the asymptote) that cannot be surpassed.

For the men's high jump, we obtained estimates of $a = 0.406$ and $b = 0.833$, so our estimate of the limiting value is $a/(1-b) = 2.43$.

The model for the expected performance at the tth Olympics can be written in the slightly different form

$$y_t = \frac{a}{1-b} - \left(\frac{a}{1-b} - y_0\right) b^t, \tag{2}$$

so that the constant term representing the asymptotic limit has been separated.

Since the formula in equation (2) contains y_0 for which we don't have data, its value must be estimated. It may therefore be treated as an unknown constant and estimation made from the data. Given that the expected performances are generally increasing, one method of estimation is to take the *minimum* value of y_t as an estimate of y_0. For the men's high jump, the minimum value over the period 1948–1996 is 1.98 metres, which occurred in 1948.

Substituting the values of a, b and y_0 into equation (2), the expected winning high jump, in metres, for men in the tth Olympics (starting with 1948 as the first) is

$$y_t = 2.43 - 0.451 \times 0.833^t. \tag{3}$$

Figure 3 shows the expected winning heights, using equation (3), from 1948 to 2020 together with the actual winning heights from 1948 to 1996. The predicted value for the Sydney 2000 Olypics ($t = 14$ for this model) is 2.39 metres.

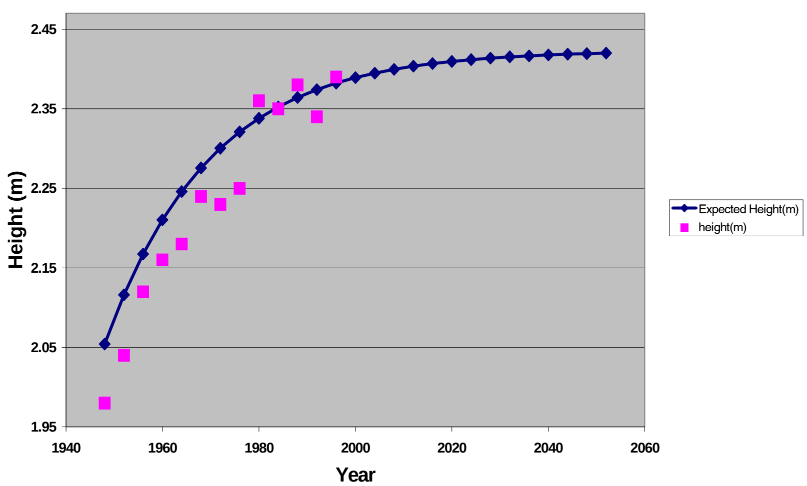

Figure 3: Prediction using equation (3) of men's Olympic high jump ($y_0 = 1.98\,\text{m}$).

Note that most of the observed heights are *below* the curve of the expected heights, indicating that the estimated value of $y_0 = 1.98\,\text{m}$ is too high. To remedy the situation, lower values of y_0 between 1.80 and 1.97 were tried and it was found by observation that the one that provided the best fit was 1.90 m. This resulted in a new predictive equation of

$$y_t = 2.43 - 0.531 \times 0.8337^t. \tag{4}$$

t	*Year*	*Predicted height* (m)	*Actual height* (m)	*Country*	t	*Year*	*Predicted height* (m)
1	1948	1.99	1.98	AUS	14	2000	2.39
2	1952	2.06	2.04	USA	15	2004	2.40
3	1956	2.12	2.12	USA	16	2008	2.40
4	1960	2.18	2.16	URS	17	2012	2.41
5	1964	2.22	2.18	URS	18	2016	2.41
6	1968	2.25	2.24	USA	19	2020	2.41
7	1972	2.28	2.23	URS	20	2024	2.42
8	1976	2.31	2.25	POL			
9	1980	2.33	2.36	GDR			
10	1984	2.35	2.35	FRG			
11	1988	2.36	2.38	URS			
12	1992	2.37	2.34	CUB			
13	1996	2.38	2.39	USA			

Table 2: Winning heights for men's Olympic high jumps.

Figure 4 uses the model in equation (4) and predicts that the winning height for the Sydney 2000 Olympics is 2.38 metres. Note how much better the curve follows the actual data in this new model.

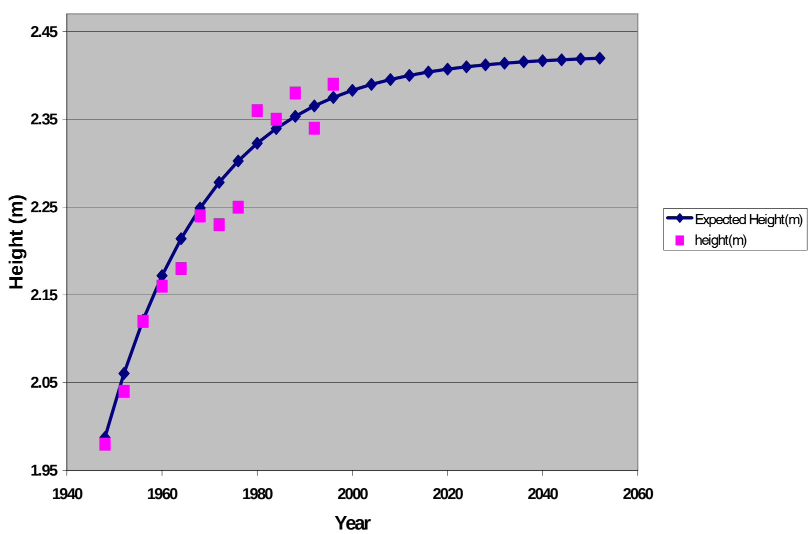

Figure 4: Prediction using equation (4) of men's Olympic high jump ($y_0 = 1.90\,\mathrm{m}$).

Table 2 shows the predicted winning heights from 1948 to 2032, together with the actual winning heights from 1948 to 1996. The mathematical model in equation (4) was used.

From Table 2 it can be seen that there is a prediction of only a small increase of 4 cm during the 36 years from the year 2000 to 2032.

Even though we now have a reasonable model for computing the expected performances, it only provides a *point* estimate of future (and past) winning heights. It does not give any measure of a likely *range* within which these values might lie. To do this we require a forecast interval for each prediction.

4 Forecast intervals

A forecast interval is obtained from the standard deviation of the errors from the prediction. Plots revealed that the residuals of the model in equation (4) had a mean of zero and had an approximate normal distribution. These errors may be estimated from past data, by subtracting the fitted (predicted) winning heights from the observed winning heights to get the residuals, and then computing the standard deviation of these residuals. Using the normal distribution as a model for the errors, we can then get a 68% forecast interval using the fact that 68% of the data under the normal curve lies within one standard deviation of the expected value.

Using the results plotted in Figure 4, the standard deviation of the residuals is 0.027. Figure 5 shows the preceding graph with the corresponding error bars superimposed.

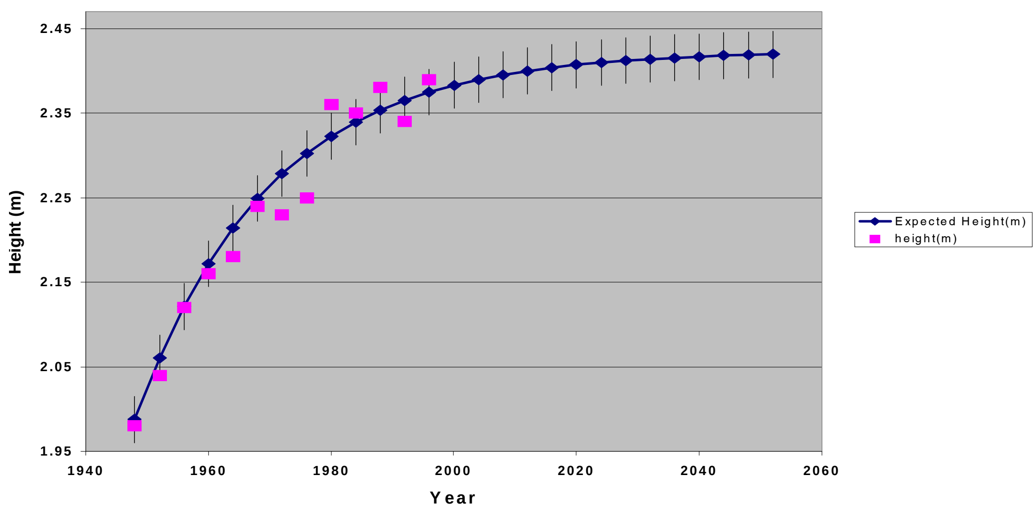

Figure 5: Error bars for the prediction using equation (4) of men's Olympic high jump ($y_0 = 1.90\,\mathrm{m}$).

The model derived based on the data from 1948 to 1996 is shown to give the Olympic men's high jump height limit as 2.43 m and year 2000 prediction as 2.39 m. With the forecast interval however, we can give a likely range of values for those predictions. Hence we can be 68% certain that the ultimate Olympic men's height limit would be somewhere between 2.40 m and 2.46 m, while for the year 2000 performance, the gold medal jump should lie between 2.36 m and 2.42 m.

5 Women's Olympic high jump

The same approach can be applied to the women's Olympic high jump data as was applied to that for men. Table 3 shows the winning height for women for the Olympics between 1928 and 1996.

For women there was an increase of 29% in the winning height jumped between 1928 and 1996. The fitted line to the scatter plot of year $(t+1)$th Olympics to year tth Olympics has the equation $y_{t+1} = 0.308 + 0.853y_t$, and this model fits the data quite well ($R^2 = 0.8362$). The limit then for the women's high jump is shown to be 2.10 m. This is 1 cm higher than the current women's high jump world record held by a Bulgarian high jumper, Stefka Konstantinova (30 August 1987) which has not been broken over a decade. The 68% forecast interval for this limit is (2.06 m, 2.14 m).

Substituting these values into the formula derived earlier, the expected winning high jump, in metres, for women in the tth Olympics (starting with 1948 as the first) is

$$y_t = 2.10 - (2.10 - y_0)0.853^t.$$

Year	*Height* (m)	*Year*	*Height* (m)
1928	1.59	1964	1.90
1932	1.65	1968	1.82
1936	1.60	1972	1.92
1940	N/A	1976	1.93
1944	N/A	1980	1.97
1948	1.68	1984	2.02
1952	1.67	1988	2.03
1956	1.76	1992	2.02
1960	1.85	1996	2.05

Table 3: Winning heights for women's Olympic high jumps.

	Men	*Forecast interval for men*	*Women*	*Forecast interval for women*
Current world record	2.45 m		2.09 m	
Limit to be reached	2.43 m	(2.40 m, 2.46 m)	2.10 m	(2.06 m, 2.14 m)
Year 2000 Olympics gold medal winning height	2.39 m	(2.36 m, 2.42 m)	2.05 m	(2.01 m, 2.09 m)

Table 4: Summary of the predictions for the Olympic high jump.

The prediction for the year 2000 women's high jump gold medal winning height is 2.05 m using y_0 as 1.65 m. The 68% forecast interval for this prediction is (2.01 m, 2.09 m).

6 Remarks

The summary of the predictions for both men and women is shown in Table 4.

It therefore seems that the limit from our model has been reached alreasy as the Cuban high jumper Javier Sotomayor holds the world recored at 2.45 m, although that performance was not at an Olympics. This was also the case with the women's current world record high jump.

Note that Sotomayor will not be competing at the year 2000 Olympics since he has been suspended from international competition as a result of testing positive for cocaine at the 1999 Pan Am Championships.

It is also interesting to note that the difference in performance between the men's and women's winning heights seems to be fairly constant since 1948. One exception was in 1968 when the women's height was 8 cm lower than the previous Olympics and the difference between men's and women's gold medal jump was 0.42 m, the highest difference for any Olympics. For the last five Olympics the differences were 0.34 m (1996), 0.32 m, 0.35 m, 0.39 m. This relationship is shown in Figure 6.

There is a relatively huge gap between 1976 men's winning height (2.25 m) and 1980 men's winning height (2.36 m), despite the fact that the 1980 Olympics at Moscow were subject to the biggest boycott (initiated by USA) in the history of Olympic movement. Perhaps it has something to do with the fact that in the 1980 Olympics, 13 of the 16 finalists of high jump used the *Fosbury flop* as opposed to the *western roll* method used in previous Olympics.

This new revolutionary jumping technique of jumping headfirst, crossing the bar on the back, and landing on the neck and shoulders was named after the 1968 Olympic champion Richard Fosbury who introduced it. On the other hand, the *western roll* or *straddle* method is where the jumper clears the bar face down, virtually rotating the body length-ways around it.

This model is one of many models that seem to fit the Olympic data and naturally there are limitations. It is subject to the value of y_0, the initial height chosen and also the data range the model is based on.

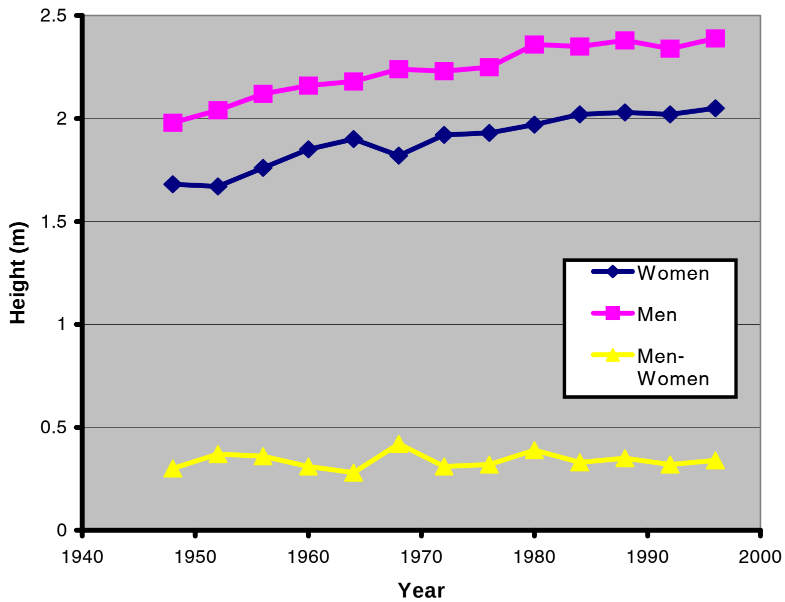

Figure 6: The winning Olympic high jumps for men, women and the difference between heights.

References

[1] C. Searle and B. Vaile, *The IOC Official Olympic Companion 1996*, Random House Australia Pty Ltd (1996).

[2] *Sydney 2000, The Olympic Games*, Dorling Kindersley Limited (1999).

[3] S. Weisberg, *Applied Linear Regression*, second edition, John Wiley and Sons, New York (1985).

VARIATIONS IN RANDOM PAIRING FINALS SYSTEMS

George Christos
School of Mathematics and Statistics
Curtin University of Technology
GPO Box U1987, Perth
Western Australia 6845

Abstract

In previous work we considered two finals systems where the teams are randomly paired to play each other in each round. One option considered was to pair top-half teams with a bottom-half team, and in the other system, the teams are randomly paired to play each other from any ladder position. In this paper we address the question of whether rematches, between the same teams in subsequent rounds, can be avoided in the general random pairing system. We also extend our calculations of "premiership probabilities" for the general random pairing system with no rematches. We also consider the general random pairing system with the added restriction that the top two teams in the final-8 are not allowed to play each other in the first two rounds, so as to allow these teams a separate passage to latter rounds.

1 Introduction

In 1995 the AFL (Australian Football League) adopted the McIntyre final-eight system, which after much public criticism, has since been replaced by a new system which will be used for the first time in the year 2000. One of the main problems (Christos [1, 2]) with the previous McIntyre system was that there was an unfair arrangement of matches in round 2 of the finals. The problem was that a lower ranked team was given an easier assignment than a team ranked above it. Unfortunately the new system (Christos [4]) also suffers from the same problem, which can now occur in both rounds 2 and 3. In previous work (Christos [1, 2, 3]), we have analysed various other systems that are fair (meaning that there are no unfair arrangements of matches in any of the rounds). There are only a handful of possible deterministic systems that are allowed which satisfy this criterion. Another way to get around this problem is to introduce an element of chance, and randomly pair the teams to play each other.

In this paper we investigate some variations on stochastic systems considered previously (Christos [1, 2, 3]). Although these random pairing systems may have "unfair" arrangement of matches this cannot be blamed on the system per se, since the higher the ranking of a team the higher is its probability that it will get an easier match, and a home-final. Higher ranked teams also have a higher probability to proceed to the next round, since only the two lowest ranked losers are eliminated from one round to the next. In previous work we considered two such systems, the so-called "random pairing system", where teams from the top-half are randomly paired to play teams from the bottom-half in each round, and the "general random pairing system", where teams are paired to play each other from any ladder position. In these systems, and in what follows below, we will assume that there are no byes, that is, no teams have a week off during the finals. In both of the final-8 systems that have been used by the AFL, the two highest ranked teams after round one (r1) do not have to play in round two (r2). We believe that byes diminish public interest in the finals, since the public would surely prefer to see the two best teams also play in r2. Incorporating byes also increases the likelihood of a rematch in later rounds.

2 Top half to bottom half system (alias 'random pairing system')

In r1 of this system, each of the top four teams is randomly paired to play a team from the bottom four, such as shown in Figure 1. In what follows, we will label the teams in r1 in order of ranking by A1, A2, ... , A8, the six teams in r2 by B1, B2, ... , B6, and the four teams in r3 by C1, ... , C4.

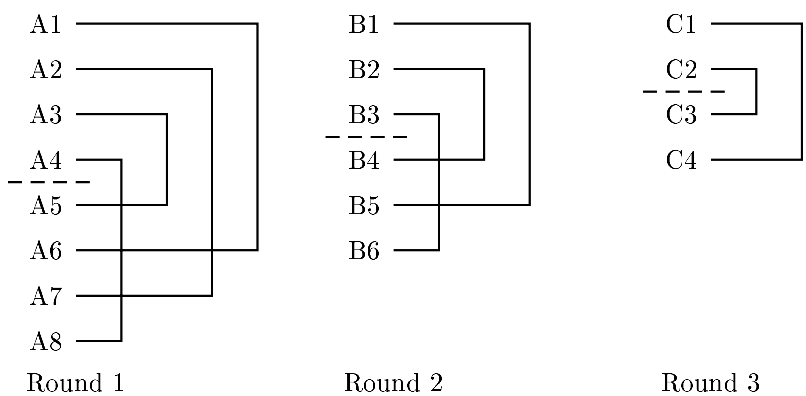

Figure 1: A possible set of matches in the top half to bottom half random pairing system.

In proceeding to the next round the two lowest ranked losing teams are eliminated, which leaves six teams. These teams are reordered with the four winners, preserving their previous relative order in the final-8, followed by the two highest ranked losers, also preserving their previous relative order in the final-8. The six remaining teams are then randomly paired so that each top-3 team plays a bottom-3 team. In proceeding to round 3 (r3) the two lowest ranked losers are eliminated again, which means that the three winners and the highest ranked loser go through. The three winners are ordered with respect to their previous order in r2. In r3 the top two teams play the bottom two teams, in two possible arrangements. We have denoted this system by the notation $R_1R_2R_3$, which symbolises that random pairing is used in each of the three rounds. Each of the finals is played at the home ground of the higher ranked team in each pairing. This means that the top half teams get to play at their home ground in each of the finals. Note also that the top two teams in the final-8 (A1 and A2) cannot be eliminated in r1, because if they lose they will be one of the two highest ranked losers that proceed to r2. The team on top of the ladder, at the start of r2 (that is, the highest ranked winner from r1), namely B1, also cannot be eliminated in r2, whereas the second ranked team B2 can be eliminated, if it loses and B1 also loses. This difference between the top two teams in r2 also effectively separates the premiership probabilities for the two highest ranked teams in the final-8 (A1 and A2), which is a feature which is absent from the systems used by the AFL. Also note that in the $R_1R_2R_3$ system, and in the other systems to be discussed below, the two lowest ranked teams in r1 (that is, A7 and A8) and in r2 (that is, B5 and B6) must win to proceed to the next round.

One can easily calculate the premiership probabilities for each of the final-8 teams in this system, in the so-called equal probability model, where every match is assumed to be an even chance for either team to win. Consider for example the top team in the final-8, A1. If A1 wins (probability $\frac{1}{2}$) it will be on top in r2 (that is, B1) and if it loses (probability $\frac{1}{2}$) it will be in fifth position in r2 (that is, B5). The probability that B1 will proceed to r3 is 1 because B1 cannot be eliminated, whereas B5 proceeds to r3 with probability $\frac{1}{2}$, since it must win its r2 match, or it will be one of the two lowest ranked losers. The premiership probability of A1 is therefore equal to

$$P_{A1} = \left(\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot\frac{1}{2}\right)\cdot\frac{1}{4} = \frac{3}{16} = 0.1875 \qquad [R_1R_2R_3 \text{ system}]$$

The final factor of $\frac{1}{4}$ outside the brackets is the probability that A1 wins its last two matches, in r3 and the Grand Final (GF).

The team A2 in second position in the final-8, will be in second position in r2 if it wins in r1 and A1 also wins (probability $\frac{1}{4}$), in top position in r2 if it wins but A1 loses (probability $\frac{1}{4}$), and in a bottom-2 position (fifth or sixth) if it loses (probability $\frac{1}{2}$). From second position, A2 will proceed to r3 if it wins (probability $\frac{1}{2}$), and if it loses, if A1 wins (probability $\frac{1}{4}$). The probability that B2, second in r2, will proceed to r3 is therefore equal to $\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{3}{4}$, and the premiership probability of A2 is equal to

$$P_{A2}=\left(\frac{1}{4}\cdot 1+\frac{1}{4}\cdot\frac{3}{4}+\frac{1}{2}\cdot\frac{1}{2}\right)\cdot\frac{1}{4}=\frac{11}{64}=0.1719 \qquad [R_1R_2R_3 \text{ system}]$$

The other premiership probabilities for this system (Christos [1]) are given in Table 1.

These premiership probabilities are graded from top position (probability 0.1875) to seventh/eighth position (probability 0.0781). This system has two possible faults. Firstly the teams are not particularly well matched in r1 and r2. If one was to calculate a simple linear difference index sum between the teams that played each other in r1 it would be equal to 16, which is the same as in the McIntyre system, where in r1 A1 v A8, A2 v A7, A3 v A6 and A4 v A5. The second problem relates to possible rematches. Clearly without any constraints the same teams may be chosen to play each other again in r2 that played each other in r1, and the same teams may be chosen to play each other in r3 that played each other in either rounds 1 or 2. We have determined (Christos [1]) that it is possible to avoid rematches (that is, have a redraw) when they occur in almost all cases, except with a probability of $\frac{1}{8}$, that a rematch in r3 from r1 cannot be avoided. In our view however rematches should not be given such a high precedence, particularly when it comes to r1 rematches in r3. Admittedly a rematch in r2 from r1 and a rematch in r3 from r2 may be undesirable. Note that rematches were present in the old VFL (Victorian Football League, which was the precursor to the AFL competition) final-5 system, and still occur in the final-4 system which is currently used in the Western Australian competition.

Final 8	premiership probability
A1	$\frac{3}{16}=\frac{192}{1024}=0.1875$
A2	$\frac{11}{64}=\frac{176}{1024}\doteq 0.1719$
A3	$\frac{37}{256}=\frac{148}{1024}\doteq 0.1445$
A4	$\frac{31}{256}=\frac{124}{1024}\doteq 0.1211$
A5	$\frac{31}{256}=\frac{124}{1024}\doteq 0.1211$
A6	$\frac{25}{256}=\frac{100}{1024}\doteq 0.0977$
A7	$\frac{5}{64}=\frac{80}{1024}\doteq 0.0781$
A8	$\frac{5}{64}=\frac{80}{1024}\doteq 0.0781$

Table 1: The premiership probabilities (equal probability model) in the top-half to bottom-half random pairing system $R_1R_2R_3$.

3 The general random pairing system $G_1G_2G_3$

In this system the teams are randomly paired to play each other from any ladder position. Some possible arrangements of matches in r1 are shown in Figure 2. A similar arrangement of matches would occur in rounds 2 and 3. Some variations on this theme will be considered in subsequent sections. What is interesting about this system compared to the previous system is that the games are generally between better matched teams, and hence more interesting from a spectator point of view. The matches are played at the home ground of the team with the highest ranking for each chosen pair, and so, in this system lower ranked teams (except for A8, B6 and C4) also have a small probability that they may host a home final. This probability is of course dependent on the ladder position of each team, A1 will always host a home final in r1, whereas A2 will host a home-final in r1 with probabilty $\frac{6}{7}$ and A7 will host a home-final in r1 with probability $\frac{1}{7}$. In previous work we have denoted this system by the notation $G_1G_2G_3$.

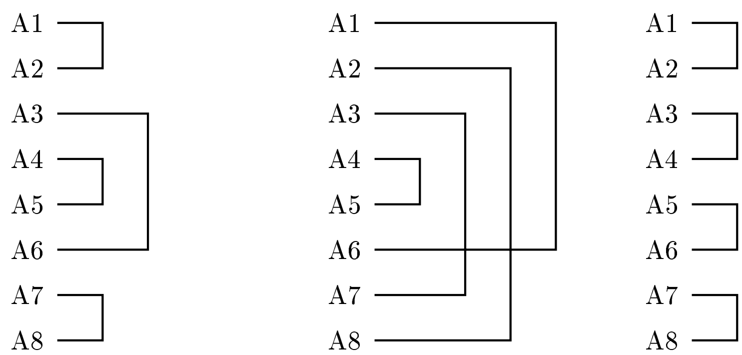

Figure 2: Some possible arrangement of matches in r1 in the general random pairing system $G_1G_2G_3$ where teams are randomly paired to play each other from any ladder position.

The premiership probabilities in this system can also be calculated in the equal probability model. The crucial ingredients in these calculations are the so-called r2 *survival probabilities.* B1 cannot be eliminated in r2 because even if it loses it will be the highest ranked loser, so its r2 survival probability is equal to 1. B2 will be eliminated if it loses, if B1 also loses, because only the highest ranked loser proceeds to r3. The survival probability for B2 is therefore equal to $1 - \frac{1}{2}(\text{B2 loses}) \cdot \frac{1}{2}(\text{B1 loses}) \cdot \frac{4}{5}(\text{B1 does not play B2}) = \frac{4}{5}$. Note that in this calculation we have had to exclude the case where B1 may have played B2 because it would then be impossible for both teams to lose. B3 will survive r2 if it wins (probability $\frac{1}{2}$) and if it loses if B1 and B2 both win. The r2 survival probability for B3 is therefore equal to $\frac{1}{2}(\text{B3 wins}) + \frac{1}{2}(\text{B3 loses}) \cdot \frac{1}{4}(\text{B1 and B2 both win}) \cdot \frac{4}{5}(\text{B1 does not play B2}) = \frac{3}{5}$. B4 will stay in the competition if it loses if both B5 and B6 also lose. Therefore the r2 survival probability for B4 is equal to $\frac{1}{2}(\text{B4 wins}) + \frac{1}{2}(\text{B4 loses}) \cdot \frac{4}{5}(\text{B5 does not play B6}) \cdot \frac{1}{4}(\text{both B5 and B6 lose}) = \frac{3}{5}$. B5 and B6 must win their r2 matches to proceed to r3 so their r2 survival probabilities are both equal to $\frac{1}{2}$. The r2 survival probabilities in the general random pairing system are given in Table 2. (Incidentally, these survival probabilities are different in the $R_1R_2R_3$ system.)

We can now proceed to calculate the premiership probabilities. Consider A1. When A1 wins (probability $\frac{1}{2}$), it will be B1, and when A1 loses it will be B5. Using the r2 survival probabilities (Table 2) the premiership probability for A1 is equal to

$$P_{A1} = \left(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2}\right) \cdot \frac{1}{4} = \frac{3}{16} = 0.1875 \qquad [G_1G_2G_3 \text{ system}]$$

which is the same as in the previous system. The premiership probabilities for A7 and A8 are also easily calculated. A7 and A8 must win in r1 (probability $\frac{1}{2}$). If A8 wins in r1 it will be in fourth position in r2, and if A7 wins it will be in either third or fourth position in r2. These r2 position each have a survival probability of $\frac{3}{5}$, so the premiership probabilities for A7 and A8 are equal to

$$P_{A7} = P_{A8} = \left(\frac{1}{2} \cdot \frac{3}{5}\right) \cdot \frac{1}{4} = \frac{3}{40} = 0.075 \qquad [G_1G_2G_3 \text{ system}]$$

Final-6 (r2)	r2 survival probability
B1	1
B2	$\frac{4}{5}$
B3	$\frac{3}{5}$
B4	$\frac{3}{5}$
B5	$\frac{1}{2}$
B6	$\frac{1}{2}$

Table 2: The r2 survival probabilities for the six remaining teams to proceed from r2 to r3 in the $G_1G_2G_3$ system.

The other premiership probabilities require somewhat more work. The premiership probability for A2 depends on whether A2 plays A1 in r1. The probability that A2 plays A1 in r1 is equal to $\frac{1}{7}$, and the probability that they do not play is equal to $\frac{6}{7}$. The premiership probability for A2 is equal to

$$P_{A_2} = \left(\frac{6}{7}\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{4}{5}+\frac{1}{2}\cdot\frac{1}{2}\cdot 1+\frac{1}{2}\cdot\frac{1}{2}\right)+\frac{1}{7}\left(\frac{1}{2}\cdot 1+\frac{1}{2}\cdot\frac{1}{2}\right)\right)\cdot\frac{1}{4} = \frac{99}{560} \doteq 0.1768.$$

(Note that in Christos [2] this probability is incorrectly given as 0.1679.)

The first three terms in the above expression, multiplied by $\frac{6}{7}$, correspond to when A2 wins (and A1 also wins), A2 wins (and A1 loses), and A2 loses respectively. In these cases A2 will be in second, first and in a bottom-2 position in r2 respectively. The two terms multiplied by $\frac{1}{7}$ (corresponding to A1 v A2 in r1), are for A2 wins and A2 loses respectively. The other premiership probabilities in the $G_1G_2G_3$ system are given in Table 3. Some of these calculations can get quite involved, since the fate of each team depends on who plays whom and on the result of certain matches. For example

$$\begin{aligned} P_{A3} &= \left[\frac{6}{7}\left\{\frac{1}{2}\cdot\frac{1}{4}\cdot\frac{3}{5}+\frac{1}{2}\cdot\frac{1}{4}\cdot 1+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{4}{5}+\frac{1}{2}\cdot\frac{1}{4}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right\}+\frac{1}{7}\left\{\frac{1}{2}\cdot\frac{4}{5}+\frac{1}{2}\cdot\frac{1}{2}\right\}\right]\cdot\frac{1}{4} \\ &= \frac{167}{1120} \doteq 0.1491, \end{aligned}$$

where the terms in this expression correspond to the associated probabilities of [(A1 does not play A2 in r1){(A3 wins r1)(A1 & A2 win r1)(A3 = B3) + (A3 wins r1)(A1 & A2 lose r1)(A3 = B1) + (A3 wins r1)(only one of A1 or A2 wins r1)(A3 = B2) + (A3 loses r1)(A1 & A2 win r1)(A3 = B5) + (A3 loses r1)(one of A1 or A2 loses r1)(A3 = B6)} + (A1 v A2 in r1){(A3 wins r1)(A3 = B2) + (A3 loses)(A3 = B6)}](A3 wins r3 and GF).

The premiership probability for A4 has 14 such terms.

A particularly interesting feature of these premiership probabilities is that they are uniformly graded from top position to 7th/8th position, as is the case with the other random pairing system. What makes this system more interesting than the previous system, other than the fact that the matches are generally between more compatible teams (and hence more interesting) is that there are now many more combinations of matches possible and all rematches can be avoided in later rounds. In r1 there are 105 different possible combinations of matches in the $G_1G_2G_3$ system compared to 24 in the $R_1R_2R_3$ system. For rounds 2 and 3 there are 15 and three possible arrangements of matches in the general system respectively, compared with six and two possible arrangements respectively in the restricted system.

Final 8	premiership probability
A1	$\frac{3}{16} = 0.1875$
A2	$\frac{99}{560} \doteq 0.1768$
A3	$\frac{167}{1120} \doteq 0.1491$
A4	$\frac{93}{700} \doteq 0.1328$
A5	$\frac{77}{700} = 0.1100$
A6	$\frac{105}{1120} \doteq 0.09375$
A7	$\frac{3}{40} = 0.075$
A8	$\frac{3}{40} = 0.075$

Table 3: The premiership probabilities (equal probability model) in the $G_1G_2G_3$ random pairing system.

4 Avoiding rematches in the $G_1G_2G_3$ system

We need to show first that r1 matches can be avoided in r2. Let us label the two teams that played each other in r1 by the same letter, say for example two M's, two N's, two P's and two Q's. In going from r1 to r2, the two lowest ranked losers are eliminated, so it is impossible to eliminate two teams with the same letter, because one of each of these pairs of teams must win in r1. Without any loss of generality we will take the teams that make it to r2 to be

$$(\mathrm{M}, \mathrm{M}, \mathrm{N}, \mathrm{N}, \mathrm{P}, \mathrm{Q}),$$

where the order of presentation of these teams does not reflect their positions on the ladder (or ranking) at this stage. Note that the order on the ladder is not important in the general random pairing system when it comes to pairing teams together to play each other. As we have noted earlier there are 15 different combinations of matches possible in r2. In five of these combinations of matches, a rematch between teams that played in r1 takes place, as shown in Figure 3.

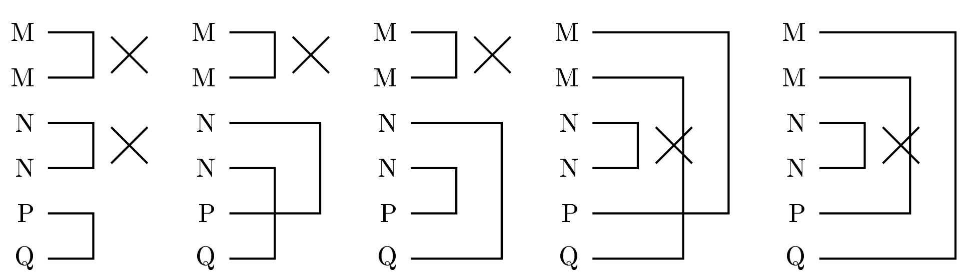

Figure 3: Five combinations of matches in r2 in the $G_1G_2G_3$ system where there is a rematch between two teams that played each other in r1.

In the other ten combinations of matches, shown in Figure 4, no rematch takes place. Clearly it is possible to arrange matters so that there are no r1 rematches in r2.

Similarly, one can also show that r2 rematches in r3 can be avoided. With regard to r1 rematches in r3, there are basically three different scenarios that can arise. The four teams left in r3 may consist of two, one or no pairs of teams that played each other in r1. We can label these combinations by {M1, M2, N1, N2}, {M1, M2, N, P} and {M, N, P, Q}, without any loss in generality, where teams

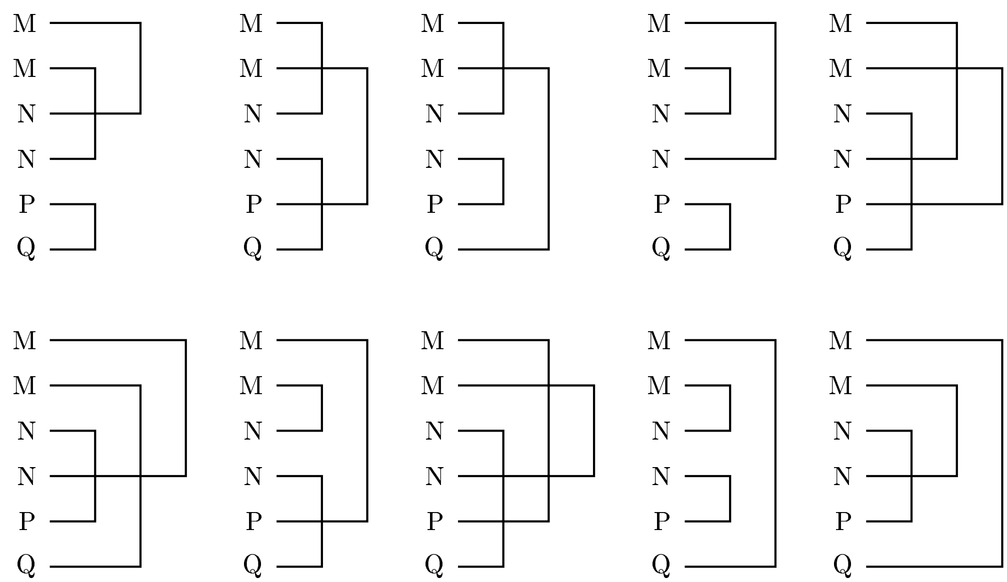

Figure 4: Ten combinations of matches in r2 in the $G_1G_2G_3$ system where there are no rematches between any teams that played each other in r1.

beginning with the same letter played each other in r1. The worst possible scenario arises when there are two r1 pairs in r3. Suppose that M1 played say N1 in r2, then the other combination of matches is always new, since if M1 played N1 then M1 could not have played N2 and N1 could not have played M2. It turns out that it is always possible to find at least one combination of matches in r3, and perhaps two such combinations, where there are no rematches between any of the teams that may have played each other in r1 or r2.

5 Premiership probabilities in $G_1G_2G_3$ system with no rematch option

It is certainly not clear a priori how the premiership probabilities are affected by the added restriction in the $G_1G_2G_3$ system that rematches are to be avoided in rounds 2 and 3. If one looks through the previous calculations given above, and repeats them for this situation one quickly realises that they are unchanged by this added restriction. The premiership probability of A1 is clearly unchanged. Refer to the calculation given previously. The situation is exactly the same for other premiership probabilities. These premiership probabilities are unchanged because the r2 survival probabilities are unchanged under the restriction of no rematches. If one looks back at the calculation of the r2 survival probabilities one can see that they are unaffected by this restriction. The only way that they may be affected is if B1 could not play B2, or B5 could not play B6, because they played each other in round one. This is clearly impossible since the top four teams in r2 correspond to the four winners and they could not have played each other in r1, whereas B5 and B6 are the two highest ranked losers from r1 so they could not have played each other in r1 also. The premiership probabilities in this system where rematches are avoided in subsequent rounds (whether it is r1 matches in r2 and r3, and r2 matches in r3, *or* r1 matches in r2 and r2 matches in r3) are therefore exactly the same as before when the option of no rematches was ignored.

6 Premiership probabilities in general random pairing system with no matches between A1 and A2 in r1 and r2

Another interesting system to consider is where the top two teams A1 and A2 in the final-8 are not allowed to play each other in rounds 1 and 2. This is desirable because one would like to allow these teams separate passage to the later rounds, hoping that they may meet in the Grand Final. We denote this system by the notation $\cancel{G_1 G_2} G_3$, where the strikethrough refers to the fact that a match between A1 and A2 is excluded in rounds 1 and 2. Some of the previously calculated premiership probabilities are affected by this restriction. The main effect is in the r2 survival probabilities, since now B1 may not play B2, as they may correspond to A1 and A2 respectively, and there is no need to exclude this circumstance as in previous calculations. See Section 3. The situation can get quite complicated and so we have only calculated a few of these premiership probabilities. One should refer to the previous calculations in the $G_1G_2G_3$ system to see how the situation changes in this case. Consider A1. In the previous calculation we had two terms, corresponding to when A1 wins in r1 (A1 = B1) and when A1 loses (A1 = B5). A problem may occur in the second scenario if A2 also loses, because then A1 and A2 will correspond to B5 and B6, which is one of the situations that we had to give special consideration to previously. This however does not change the premiership probability of A1 because A1 is in fifth position in r2, and the only r2 survival probabilities that are actually affected by possible matches between B5 and B6 are the survival probabilities for B2, B3 and B4. See the discussion preceding Table 2. As before,

$$P_{A1} = \left(\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot\frac{1}{2}\right)\cdot\frac{1}{4} = \frac{3}{16} = 0.1875 \qquad [\cancel{G_1 G_2} G_3 \text{ system}]$$

The premiership probability for A2 is however affected by the added restriction that A1 and A2 should not play in r1 and r2:

$$P_{A2} = \left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{3}{4} + \frac{1}{2}\cdot\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot\frac{1}{2}\right)\cdot\frac{1}{4} = \frac{11}{64} \doteq 0.172 \qquad [\cancel{G_1 G_2} G_3 \text{ system}]$$

The first term in brackets corresponds to the situation when A1 and A2 both win, but in this case the ladder looks like (in order, where X = some other team)

A1, A2, X, X, X, X

The r2 survival probability of A2 is affected in this case, since now B1 and B2 cannot play each other (refer to previous calculations). A2 will survive r2 so long as A1 does not also lose when A2 loses. The r2 survival probability of A2 is therefore equal to $1 - \frac{1}{2}\cdot\frac{1}{2} = \frac{3}{4} = 0.75$, which should be compared to previous survival probability in the unrestricted model, which was 0.8. The second term in the premiership probability for A2 in the above expression comes from the situation when A2 wins in r1 (probability $\frac{1}{2}$) and A1 loses in r1 (probability $\frac{1}{2}$). In this case the ladder looks like (A2, X, X, X, A1, X), and A2's survival probability is equal to 1. The last term in the premiership probability for A2 corresponds to when A2 loses, in which case A2 will be in either fifth or sixth position, from where it has a survival probability of $\frac{1}{2}$, since it must win.

We have calculated two other premiership probabilities in this general random pairing system without matches between A1 and A2 in rounds 1 and 2. The premiership probability for A8 is given by

$$P_{A8} = \left(\frac{1}{2}\cdot\frac{1}{4}\cdot\frac{5}{8} + \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{3}{5}\right)\cdot\frac{1}{4} = \frac{97}{1280} \doteq 0.0758 \qquad [\cancel{G_1 G_2} G_3 \text{ system}]$$

The first term in the above expression corresponds to when A8 wins in r1 and both A1 and A2 lose, which means that A8 is in fourth position in r2 with A1 and A2 in fifth and sixth position (ladder (X, X, X, A8, A1, A2)). From this position, A8 will survive r2 if it wins, and if it loses if A1 and A2 both lose in r2. Since A1 and A2 cannot play each other in this system, the survival probability for A8 in this case is equal to $\frac{1}{2} + \frac{1}{2} \cdot \frac{1}{4} = \frac{5}{8}$, which is different from the value of $\frac{3}{5}$ in the unrestricted $G_1 G_2 G_3$ system. The second term in the expression for P_{A8} above comes from the situation where A8 wins, and at least one of A1 or A2 also wins (probability $\frac{3}{4}$). In this case A8 is in fourth position in r2 again, but now B5 and B6 may play each other. In this case A8 (= B4) has an r2 survival probability equal to $\frac{3}{5}$.

The fate of A7 seems to depend on whether or not it plays A8 in r1. If A7 plays A8 (probability $\frac{1}{7}$) and it wins, it can finish in fourth position in r2, whereas if it does not play A8 in r1, it can finish in third position (if A8 also wins) or fourth position. Once again one must go through the specifics in this case to see where A1 and A2 finish in r2. The premiership probability for A7 is given by

$$
\begin{aligned}
P_{A7} &= \frac{1}{2}\left(\frac{6}{7} \cdot \left\{\frac{1}{2} \cdot \left(\frac{1}{4} \cdot \frac{5}{8} + \frac{3}{4} \cdot \frac{3}{5}\right) + \frac{1}{2} \cdot \left(\frac{1}{4} \cdot \frac{5}{8} + \frac{3}{4} \cdot \frac{3}{5}\right)\right\} + \frac{1}{7} \cdot \left\{\frac{1}{4} \cdot \frac{5}{8} + \frac{3}{4} \cdot \frac{3}{5}\right\}\right) \cdot \frac{1}{4} \\
&= \frac{97}{1280} \doteq 0.0758 \qquad [\cancel{G_1}\cancel{G_2}G_3 \text{ system}]
\end{aligned}
$$

The factor of $\frac{1}{2}$ in front corresponds to the fact that A7 must win in r1. The first set of terms in curly brackets corresponds to the situation where A7 does not play A8 in r1 (probability $\frac{6}{7}$), and the second term in curly brackets to when A7 plays A8 in r1 (probability $\frac{1}{7}$). The first term inside the first curly brackets correspond to (A8 wins r1)[(A1 & A2 win r1) or (only one of A1 or A2 lose)] and the second term to the same thing except that A8 loses in r1. Note, depending on the situation, A7 has a different r2 survival probability like in the A8 calculation. When A1 and A2 win the ladder looks like (A1, A2, A7, X, X, X), and the survival probability needs to be adjusted as before because B1 = A1 cannot play B2 = A2, whereas when one of A1 or A2 loses the survival probability is as before. The second term in curly brackets corresponds to the situation when A7 plays A8 in r1. In this case, A7 is in fourth position in r2 when it wins. The survival probability needs to be adjusted (first term in curly brackets) if A1 and A2 both lose in r1, as then the ladder looks like (X, X, X, A7, A1, A2) and B5 cannot play B6. The second term in the second curly brackets corresponds to when one of A1 or A2 loses. It is interesting that in the end A7 and A8 have the same premiership probabilities.

The calculated premiership probabilities in this restricted system are listed in Table 4. P_{A1} is the same as in the $G_1 G_2 G_3$ system, P_{A2} is slightly lower compared to the $G_1 G_2 G_3$ system, while P_{A7} and P_{A8} are slightly higher in comparison. We have not calculated the premiership probabilities for A3, A4, A5 and A6, because they are quite involved, and also because we expect them to be approximately the same as in the $G_1 G_2 G_3$ system.

Final 8	premiership probability
A1	$\frac{3}{16} = 0.1875$
A2	$\frac{11}{64} \doteq 0.1719$
A3	NC
A4	NC
A5	NC
A6	NC
A7	$\frac{97}{1280} \doteq 0.0758$
A8	$\frac{97}{1280} \doteq 0.0758$

Table 4: The premiership probabilities in the $\cancel{G_1}\cancel{G_2}G_3$ random pairing system, where teams are randomly paired to play each other in rounds 1, 2 and 3, except that matches between A1 and A2 are not allowed in rounds 1 and 2. NC means that the premiership probability has not been calculated, but these are expected to be similar to the probabilities in the unrestricted $G_1 G_2 G_3$ system.

7 Avoiding rematches in the $\cancel{G_1 G_2} G_3$ system

In this section we wish to address the question of whether rematches can also be avoided in the general random pairing system with no matches between A1 and A2 in rounds 1 and 2. The first thing that one can note is that there are $6 \cdot 5 \cdot 3 = 90$ different possible combinations of matches in r1, and 12 different combinations of matches in r2 in the $\cancel{G_1 G_2} G_3$ system, compared to 105 and 15 different combinations of matches in the unrestricted $G_1 G_2 G_3$ system. This would seem to suggest that this system is not overly constrained by this added restriction.

Referring back to the previous rematch analysis for the unrestricted system, Figure 4 shows the allowed combinations of matches where rematches from r1 are avoided. Some of these matches are however also excluded in this system if we wish there to also be no matches between A1 and A2 in r2. Note that, since A1 and A2 did not play in r1 in this system, they must correspond to different letters. Either A1 and A2 correspond to M and N (or N and M), or to M and P (or M and Q, or N and P, or N and Q), or to P and Q. If A1 and A2 correspond to an M and N labeled team (say the first to an M and the second to an N, without any loss in generality) there are three arrangements of matches shown in Figure 4 that must now be excluded to ensure that A1 and A2 do not play in r2. If A1 = M (or N) and A2 = P (or Q) then two arrangements of matches must be excluded from Figure 4. Finally, if A1 = P and A2 = Q, or vice versa, two arrangements of matches must be excluded from Figure 4 in r2. Therefore one can exclude rematches in r2 from r1 and also matches between A1 and A2 in rounds 1 and 2.

We can also address the question of r1 and r2 rematches in r3. The previous arguments can also be carried over to this system, and there is always at least one combination of matches in r3, where there are no rematches from previous rounds.

8 Conclusion

We have analysed a number of different finals systems where the teams are randomly paired to play each other under certain conditions. We have calculated (most of) the premiership probabilities in these systems in the equal probability model. These calculations can be extended with the aid of a computer to models where the form of the teams is taken into account as outlined by Christos [1, 2] or by Schwertman and Howard [5].

References

[1] G. A. Christos, "More competitive finals systems for the AFL and other team sport competitions", *Curtin University of Technology, Technical Report 22/97* (November 1997).

[2] G. A. Christos, "The AFL finals: it's more than a game", in *Fourth Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar (editors), Bond University, Queensland, Australia (1998), 111–124.

[3] G. A. Christos, "Alternate finals systems for the AFL and other team sport competitions", submitted.

[4] G. A. Christos, "New AFL finals system is unfair!", *Austral. Math. Soc. Gaz.*, **27**(1) (2000), 17–23.

[5] N. C. Schwertman and L. Howard, "Probability models for the Australian Football League finals series", *Austral. Math. Soc. Gaz.*, **17**(4) (1990), 89–94.

HOME ADVANTAGE IN THE OLYMPIC GAMES

Stephen R. Clarke*
School of Mathematical Sciences
Swinburne University
PO Box 218, Hawthorn
Victoria 3122, Australia

Abstract

This paper investigates the extent of home advantage in the Olympic games. The number and type of medals won by each country in the Modern Olympics is analysed. The total number of medals and gold medals won are highly correlated, and many countries show a highly variable performance as measured by the percentage of total medals won. There is a large home advantage or host country effect. The home country wins about three times their away percentage of medals, and about twice their average in the games immediately before and after their home games. There is also evidence that the mix of medals is richer, with home teams winning a proportion of gold medals higher than both their historical average and the proportion available.

1 Introduction

The Olympic games vies with the World Cup of soccer as the World's premier sporting event. The ancient games began at Olympia with a single foot race, and continued for twelve centuries until abolished in 394 AD. The modern Olympics began in 1896, and like its predecessor, the event list has steadily grown. Also like its predecessor, the first modern games did not allow women to compete. However they have steadily increased their participation until they now make up 30% of the competing athletes. These two effects have seen an increasing number of medals awarded. In 1896, 122 medals including 44 gold were awarded. One hundred years later athletes competed for 841 medals including 271 gold.

Success on the sporting field is becoming increasingly important for competing nations. While no official tables are kept, unofficial tallies of the number of gold, silver and bronze medals won by each country are keenly recorded in the media. Final tallies are used by governments to measure the success of sporting policies and allocate funding. How well can Australia expect to perform at the Sydney Olympics? This question could be answered by a detailed analysis of each sport in order to accumulate the total chances of each country winning or placing. For example Dyte and Clarke [2, 5] demonstrate for tennis and soccer a method that could be used to calculate medal chances in most team sports with a knockout or tournament structure. A model is fitted to official rankings to predict a chance of winning any match, and the complete tournament is then simulated to generate probabilities of final finishing positions. In individual events, while several papers have been written predicting winning times and world records [1, 6, 8], there is little in the literature giving an athletes chance of winning or achieving a certain place. Perhaps some of the methods used in horse racing could be adapted. However, clearly an approach that requires detailed investigation in each sport is a very large study. Perhaps the conference as a whole could undertake such a study for the 2004 results with each participant looking at a different sport. This study simply investigates the past global results.

*My thanks to David Johnson, an undergraduate student who collected and computerised data for this paper.

Stefani has carried out several studies [10, 12] into performance at the Olympic games, but from the point of improvement in the various sports. A recent paper [4] uses neural networks and regression to model the number of medals countries won at the Atlanta Olympics, based on various economic variables such as area, population and length of rail road track. Sommers [9] also looks at the Atlanta games, and measures the success of nations per unit of population.

Olympiad	*Year*	*City*	*Country*	*Number of countries to win medals*	*Number of medals*
I	1896	Athens	Greece	10	122
II	1900	Paris	France	18	276
III	1904	St Louis	USA	9	282
–	1906	Athens	Greece	19	226
IV	1908	London	Great Britain	19	323
V	1912	Stockholm	Sweden	18	309
VI	1916	Berlin	Germany	cancelled	
VII	1920	Antwerp	Belgium	22	435
VIII	1924	Paris	France	26	366
IX	1928	Amsterdam	Netherlands	33	327
X	1932	Los Angeles	USA	27	348
XI	1936	Berlin	Germany	32	388
XII	1940	Helsinki	Finland	cancelled	
XIII	1944	London	Great Britain	cancelled	
XIV	1948	London	Great Britain	33	409
XV	1952	Helsinki	Finland	43	459
XVI	1956	Melbourne	Australia	39	470
XVII	1960	Rome	Italy	45	464
XVIII	1964	Tokyo	Japan	42	498
XIX	1968	Mexico City	Mexico	43	525
XX	1972	Munich	Germany	46	596
XXI	1976	Montreal	Canada	41	613
XXII	1980	Moscow	USSR	36	631
XXIII	1984	Los Angeles	USA	46	687
XXIV	1988	Seoul	South Korea	52	739
XXV	1992	Barcelona	Spain	64	815
XXVI	1996	Atlanta	USA	78	841
XXVII	2000	Sydney	Australia		

Table 1: Venues for the modern Olympic games.

Table 1 shows the venues for each of the modern games. Seventeen countries have hosted the games—four countries twice and the USA four times. The modern games had a chequered beginning. The Paris and St Louis games in 1900 and 1904 were overshadowed by the Paris Universal Exhibition and the Louisiana Purchase Exposition and were not a success. (At the St Louis games in 1904, only 12 nations attended and the USA won 84% of the medals.) Interim games were held in 1906 in Athens in an attempt to revive the flagging Olympic movement. While the IOC does not recognise these as official and they are not numbered in sequence, Wallechensky [14] not only includes them but credits them with helping to save the Olympic movement. An Olympiad is a period of four years starting with an Olympics, so the numbering was also upset by the World Wars. Thus while Sydney will be held in the 27th Olympiad it will be the 24th official modern games.

Of the seventeen countries to host the Olympics, fourteen have won their greatest ever percentage of available medals at home. In studying the performance of countries in the Olympics, the effects of home ground advantage arise. The existence of home advantage in sport is well documented, but most study centres around team sports. It is well known that home teams win a majority of the matches at home. The causes of home advantage are usually listed as positive effects for the home side due to ground familiarity and a partisan crowd, and negative effects on the visitors mainly due to travel. Clearly such effects are present during an Olympic games.

While the games are awarded to a host city, the causes of home advantage apply to all athletes from the home country. In Sydney, visiting athletes will suffer from change in season and time zones, while Australian athletes will have a home crowd to spur them on. While most research on home advantage has been within country, Stefani [11, 13] shows that in soccer the effects increase once international travel is involved. While there appears to be little research on the effects of home advantage in individual sports, the performance of French and Spanish tennis players in the French Open is a good example of players performing better on home or near home soil.

There are some other reasons peculiar to the Olympics why the home country can expect to do better than usual. The home country has some choice in the sports that will be offered, and naturally includes sports in which it excels or has a special interest. The host country also fields larger teams and competes in a larger range of events than usual. In addition boycotts have marred several games, reducing the strength of competition. Since no country has boycotted its own games, this has advantaged the home team.

2 Analysis and discussion

The data analysed in this paper consist of the final gold silver and bronze medal tallies of all competing countries. Although unofficial, such tables are of great interest during the conduct of the games, usually published in order of the number of gold medals won. Results can be collected from various print and web sources. For example, Wallechinsky [14] has details for all events up to 1984. The data used here were collected by undergraduate students and consist of the gold, silver and bronze medal tallies for all countries that won medals—a total of 841 observations. There were no data on the countries that competed but won no medals. Table 1 also shows the total number of countries that won medals, and the total number of medals awarded.

There are many factors that could be taken into account when analysing the data. The increasing number of medals awarded means that most strong countries will increase their medal tally throughout the period. For this reason we generally model here the percentage of available medals won by each competing country. Boycotts in particular years will obviously affect the strength of competition. More subtle effects arise with the amalgamation or separation of countries. For example, a field event might now see many strong competitors from countries previously part of the USSR. Such effects would generally strengthen individual events by increasing the number of competitors, but may weaken or strengthen team events, since players are spread more thinly over more teams. To allow for such effects is beyond the scope of this paper. We merely seek to measure in a global way the overall performance of countries and the effect of home advantage.

2.1 Total medals versus gold medals

While many countries measure their success by the number of gold medals won, there is a larger percentage of random element present in this measure than in the total number of medals won, and the latter may be a preferable measure. Figure 1 shows the total number of medals won each year against the number of gold for Australia, with the home performance marked with a +. This illustrates the strong relationship between the two measures that exists for most countries. For Australia, the correlation between the total number of medals and the number of gold medals is 0.87. The correlation between gold and silver is 0.71, and between silver and bronze is 0.69. For the stronger USA, these figures rise to 0.95, 0.92 and 0.90.

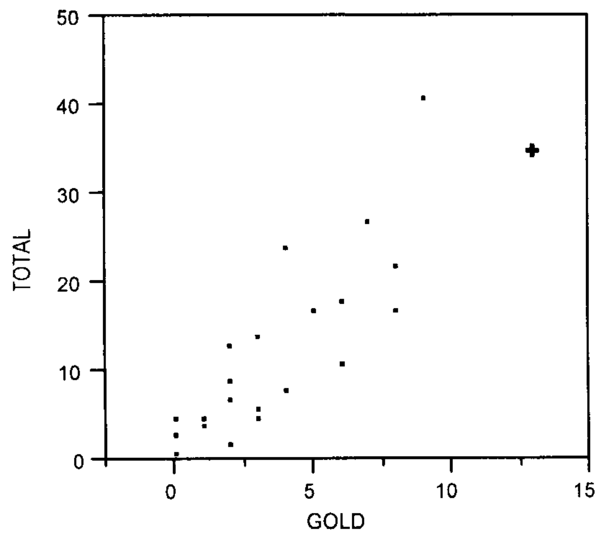

Figure 1: Total number of medals won against number of gold medals won for Australia.

2.2 Performance over time

When the performance of a single country is investigated over time, most countries show some trend in performance. The number of gold medals won by the USA is shown in Figure 2, with home performances marked with a +. This appears to demonstrate a steady increase in success by the USA, with outstanding performances when on home territory.

However when the increasing number of medals awarded is taken into account the story is not so rosy. Figure 3 gives the number of medals won by the USA as a percentage of the medals available. While this is distorted by the performance in St Louis, it shows what now appears to be a slightly decreasing trend. Again, the home performances stand out, with the exception of Atlanta, which now appears as a poor performance.

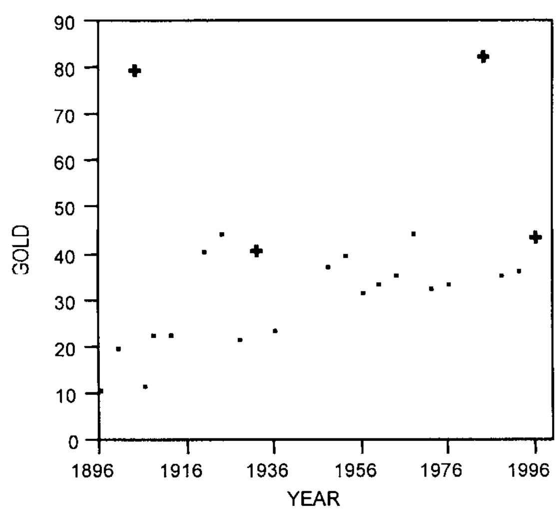

Figure 2: The number of gold medals won by the USA each year.

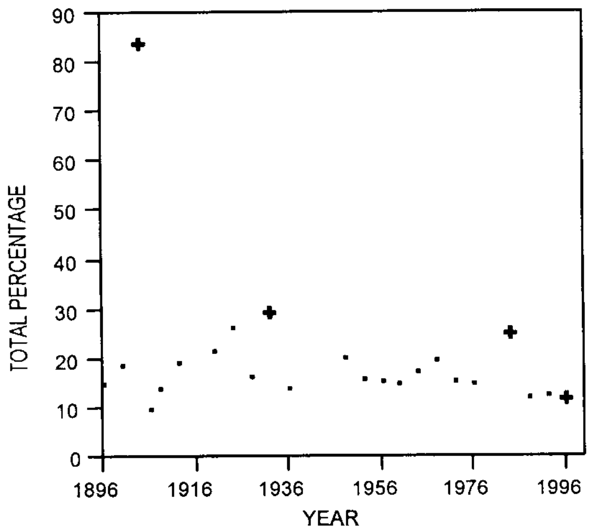

Figure 3: The number of medals won by the USA each year as a percentage of medals available.

2.3 Australia's performance

The performance of Australia in all Olympics is shown in Table 2. The percentage of all medals that Australia has won is given in Figure 4. A smoothing spline highlights the changes in performance. Such rises and falls are often attributed in the media to events such as the establishment of the Institute of Sport. This was proposed following the performance in Canada, where Australia failed to gain a gold medal. While the figures show that Australia's performance has improved over the years, the huge variation in performance to be expected from year to year demonstrates the folly of attempting to predict a tally for an individual country. However home performance is again an obvious outlier, with Australian winning over 50 percent more medals than its next most successful games.

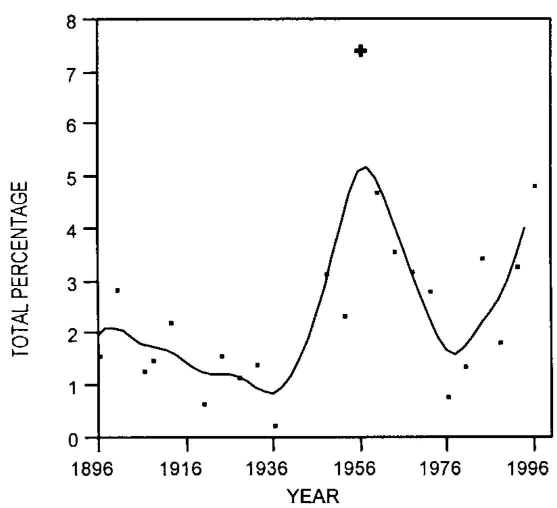

Figure 4: Percentage of available medals Australia won each year.

Year	*Gold*	*Silver*	*Bronze*	*Total*	*Percentage of available medals won*
1896	2	0	0	2	1.64
1900	4	0	4	8	2.90
1906	0	0	3	3	1.33
1908	1	2	2	5	1.55
1912	2	2	3	7	2.27
1920	0	2	1	3	0.69
1924	3	1	2	6	1.64
1928	1	2	1	4	1.22
1932	3	1	1	5	1.44
1936	0	0	1	1	0.26
1948	2	6	5	13	3.18
1952	6	2	3	11	2.40
1956	13	8	14	35	7.45
1960	8	8	6	22	4.74
1964	6	2	10	18	3.61
1968	5	7	5	17	3.24
1972	8	7	2	17	2.85
1976	0	1	4	5	0.82
1980	2	2	5	9	1.43
1984	4	8	12	24	3.49
1988	3	6	5	14	1.89
1992	7	9	11	27	3.31
1996	9	9	23	41	4.88

Table 2: Australia's Olympic performance.

2.4 Comparison of home to away performance

In order to compare the home and away performance we concentrate on the percentage of available medals won for the seventeen countries who have hosted a games. There are several ways of estimating home performance. Table 3 lists the teams that have hosted games, along with their average home performance and away performance. There is clearly a large home advantage. Canada is the only country to do worse at home, and the median ratio is 3.5. Thus 50% of host countries win more than 3.5 times their historical average of medals. Since the original data only includes Olympics where that team won medals, there may be away games where the team won no medals that are not included. If anything, the figures underrate the effect of home advantage.

The high home advantage for Greece in Table 3 demonstrates that the early Olympics, where a few powerful countries were very successful, can distort figures. Various combinations can be tried, but the overall pattern remains. For example, restricting the data to the games after Melbourne, and excluding data from the boycotted games at Moscow and Los Angeles, gives the results shown in Table 4. The USA is now the only team to do worse at home, Germany does the same, but most countries perform much better. Canada is now 40% better at home. Again the median performance is high at 2.5 times better at home. Trends in performance coupled with a bias in the selection process for host cities could distort the above averages. For example, if a rising performance in sport were a factor in gaining selection as host country, home teams would generally perform better at home than they had historically. Comparing the percentage of medals won at their home games, with the average percentage of medals won in the most recent games in which they competed prior to hosting the games and the earliest games in which they competed after they host the games, allows for any increase or decrease in performance over time.

Country	*Percentage of available medals won*		*Ratio*
	Away	Home	Home : Away
Australia	2.3	7.4	3.2
Belgium	1.3	8.1	6.0
Canada	2.1	1.8	0.9
Finland	3.3	4.8	1.4
France	5.0	23.7	4.8
Great Britain	5.3	25.3	4.8
Germany	6.6	14.8	2.2
Greece	0.4	26.8	62.3
The Netherlands	1.7	5.8	3.4
Italy	4.3	7.8	1.8
Japan	3.1	5.8	1.9
Korea	1.2	4.5	3.6
Mexico	0.4	1.7	3.9
Soviet Union	16.9	30.9	1.8
Spain	0.5	2.7	5.1
Sweden	4.5	21.0	4.7
USA	17.0	37.9	2.2

Table 3: Percentage of available medals won by host countries at home and away.

The results are given in Table 5. Again the ratios are generally greater than one with a median of about two. Clearly the home advantage is not due to some ancient performances or the results of boycotts, and is still present in recent performances.

Country	*Percentage of available medals won*		*Ratio*
	Away	Home	Home : Away
Italy	3.1	7.8	2.5
Japan	3.7	5.8	1.6
Mexico	0.2	1.7	8.6
Germany	6.7	6.7	1.0
Canada	1.3	1.8	1.4
Korea	1.3	4.5	3.4
Spain	0.7	2.7	4.2
USA	15.8	12.0	0.8

Table 4: Percentage of available medals won at home and away by host countries at Olympic games 1960–1976, 1988–1996.

2.5 Proportion of gold medals

There is evidence in the literature that in many sports different teams enjoy different levels of home advantage [3, 7, 11]. This would probably be true in the Olympics, where teams may or may not travel across several time zones and even seasons, and to possibly different cultures. The question arises whether home advantage is greater for the better athletes. It seems reasonable that crowd involvement may be greater when a home athlete is a chance for gold than when they are possibly vying for a

Year	*Country*	*Average of performance in games before and after being host*	*Home performance*	*Ratio*
1896	Greece	11.4	38.5	3.4
1900	France	13.4	37.0	2.8
1904	USA	14.7	84.4	5.7
1906	Greece	19.9	15.0	0.8
1908	Great Britain	12.2	44.9	3.7
1912	Sweden	11.1	21.0	1.9
1920	Belgium	1.4	8.1	5.6
1924	France	7.9	10.4	1.3
1928	Netherlands	2.4	5.8	2.4
1932	USA	15.8	29.9	1.9
1936	Germany	5.6	22.9	4.1
1948	Great Britain	3.0	5.6	1.9
1952	Finland	4.0	4.8	1.2
1956	Australia	3.6	7.5	2.1
1960	Italy	5.4	7.8	1.4
1964	Japan	4.3	5.8	1.3
1968	Mexico	0.2	1.7	9.3
1972	Germany	5.7	6.7	1.2
1976	Canada	3.6	1.8	0.5
1980	USSR	19.1	30.9	1.6
1984	USA	14.0	25.3	1.8
1988	Korea	3.6	4.5	1.3
1992	Spain	1.3	2.7	2.1
1996	USA	13.3	12.0	0.9

Table 5: Comparison of percentage of available medals won by host countries at home and in the Olympics before and after home games.

medal. This should produce a greater proportion of gold medals for home teams than normally. A higher proportion of gold medals is also obtained under a model where a country that normally wins the same proportion of gold silver, bronze and fourth places enjoys a home advantage that lifts the same proportion of placegetters up one level. Both these effects would result in a "richer" mixture of winning medals. The actual numbers of medals available alters both in number and proportion at various games. In the early days, the number of gold exceeded the number of silver and bronze, but this is reversed in later games, with more bronze medals than gold being awarded. In total, 7376 gold, 7280 silver and 7642 bronze have been awarded for an overall percentage of gold medals of 33%. Table 6 gives the proportion of gold medals won by the host country at each Olympics, and compares it with both their average away proportion and the actual proportion of gold medals available at their home games. Certainly in the second half of the century, almost all teams have won a richer mixture of medals than both their away games average and the average available at their home games. From the time of the last games in Australia, Canada with no gold is the only country to have a worse winning mixture than was available. The median home performance has a 10% richer mix than is available and than their historical performance. Apart from the early games, most teams win both a greater percentage of gold medals at home, and better than expected at home. Clearly a home country is likely to have a richer mix than their own average in away games and the overall average at their own games.

Country	*Year*	*Gold medals won as percent of total won*	*Percentage of gold medals available*	*Average away percentage of gold medals*
Greece	1896	21.3	36.1	27.8
France	1900	28.4	34.8	31.1
USA	1904	33.6	35.1	43.1
Greece	1906	23.5	34.1	27.8
Great Britain	1908	38.6	34.1	23.5
Sweden	1912	36.9	32.7	29.3
Belgium	1920	40.0	35.9	23.2
France	1924	34.2	33.9	31.1
Netherlands	1928	31.6	33.6	20.8
USA	1932	39.4	33.6	43.1
Germany	1936	37.1	33.5	27.9
Great Britain	1948	13.0	33.7	23.5
Finland	1952	27.3	32.5	35.1
Australia	1956	37.1	32.6	29.9
Italy	1960	36.1	33.0	38.8
Japan	1964	55.2	32.7	26.4
Mexico	1968	33.3	33.1	12.4
Germany	1972	32.5	32.7	27.9
Canada	1976	0.0	32.3	24.7
USSR	1980	41.0	32.3	39.0
USA	1984	47.7	32.9	43.1
Korea	1988	36.4	32.6	11.6
Spain	1992	59.1	31.8	14.7
USA	1996	43.6	32.2	43.1

Table 6: Percentage of gold medals won by host countries.

3 Conclusion

There is a large random element in the performance of countries in the Olympic games. The total numbers of medals and gold medals are highly correlated, and the number of medals awarded has steadily increased over the years. The percentage of all available medals would be a better measure of a country's performance than the number of gold, which is the measure usually trumpeted by the media. Australia's performance has steadily increased over the years, with a more dramatic increase since Montreal. In 2000, Australia has the bonus of what is clearly a large home advantage. Historically the home team wins over three times their usual percentage of medals, but this may be difficult to achieve given the large base level of performance Australia has recently attained. At its last home Olympics, Australia gained 7.6% of available medals, twice the percentage they achieved in the games immediately before and after Melbourne. However the isolation of Melbourne in 1956 resulted in a low number of athletes attending, and the games were also weakened by two boycotts. This time Australia is also coming off a strong performance in Atlanta of nearly 5% so to expect a repeat performance of their last games effort might be optimistic. However there is also strong evidence that the mix of medals is richer for the home teams, so Australia can expect to win a proportion of the gold medals greater than both their long term average of 30% and the actual percentage available at Sydney.

In making predictions there is always the effect of randomness. Most of the above applies equally well to Canada, a country usually about the same level as Australia on the medal tally. Let's hope that after Sydney, Canada remains the only country in the Modern Olympics not to win a gold medal at the games they host.

References

[1] D. C. Blest, "Lower bounds for athletic performance", *The Statist.*, **45**(2) (1996), 243–253.

[2] S. R. Clarke and D. S. Dyte, "Using official ratings to simulate major tennis tournaments", submitted.

[3] S. R. Clarke and J. M. Norman, "Home ground advantage of individual clubs in English soccer", *The Statist.*, **44**(4) (1995), 509–521.

[4] E. M. Condon, B. L. Golden and E. A. Wasil, "Predicting the success of nations at the Summer Olympics using neural networks" *Comput. and Oper. Res.*, **26** (1999), 1243–1265.

[5] D. Dyte and S. R. Clarke, "A ratings based model for World Cup soccer", to appear.

[6] H. J. Grubb, "Models for comparing athletic performances", *The Statist.*, **47**(3) (1998), 509–521.

[7] D. A. Harville and M. H. Smith, "The home-court advantage: How large is it, and does it vary from team to team?", *Amer. Statist.*, **48**(1) (1994), 22–28.

[8] I. T. Heazlewood and G. Lackey, "The use of mathematical models to predict elite athletic performance at the Olympic games", in *Third Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1996), 185–205.

[9] P. M. Sommers, "Meddling with 1996 Olympic results", *Chance*, **9**(4) (1996), 28–30.

[10] R. Stefani, "Predicting outcomes", in *Statistics in Sport*, J. Bennett (editor), Arnold, London (1998), 249–273.

[11] R. T. Stefani and S. R. Clarke, "Predictions and home advantage for Australian Rules football", *J. Appl. Statist.*, **19**(2) (1992), 251–261.

[12] R. T. Stefani, "Applying least squares to team sports and Olympic winning performances", in *First Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1992), 43–59.

[13] R. T. Stefani, "Observed betting tendencies and suggested betting strategies for European football pools", *The Statist.*, **32** (1983), 319–329.

[14] D. Wallechinsky, *The Complete Book of the Olympics*, Penguin, New York (1984).

AN ANNOTATED BIBLIOGRAPHY OF MATHEMATICAL ARTICLES ON SPORT AMENABLE TO ELEMENTARY UNIVERSITY MATHEMATICS TEACHING

G. L. Cohen
Department of Mathematical Sciences
University of Technology, Sydney
NSW 2007, Australia

Abstract

This paper consists of a selection of articles published in mathematics-related journals, and a few books and chapters from books, that are related to sport and suitable for incorporation into elementary university mathematics subjects. Sports are arranged alphabetically and annotations are used to indicate the relevant mathematics.

1 Introduction

This paper consists of a personal and very incomplete selection of articles published in mathematics-related journals, and a few books and chapters from books, that are related to sport and suitable for incorporation into elementary university mathematics subjects. It was compiled originally as part of the development of a new subject called Mathematics in Sport that I have taught at the University of Technology, Sydney, over the last few years, and has since been considerably supplemented.

The format is alphabetical by sports, apart from the first section on articles and books that are of general interest or span a few sports, and the second on draws, tournaments, rankings and ratings. A brief description is given for many of the items listed. The personal nature of the selection must be emphasised—personal in terms of my own interests and background, and personal simply by nature of the fact that an exhaustive search for such articles has not been attempted, and would no doubt be foolhardy to attempt. There must be others with their own lists, perhaps only slightly intersecting this one, and it would be of interest to gather these together.

The articles are not necessarily wholly suitable for use in a first-year mathematics subject, but aspects are and at least the general tenor and the conclusions reached would be of interest. In many cases, articles have lists of further references, so useful project topics abound by simply asking students to follow through the references. The World Wide Web is now also a tremendous resource for information and data in this area, but I have made no attempt here to document the relevant sites. Often, magazines such as *New Scientist* will be secondary sources for interesting items.

The classification for the articles has not always been obvious, and some items of interest may not be in the first place you would look. For example, articles on the rating of teams in the Australian Football League have been placed under Draws, Tournaments, Rankings and Ratings (the second of the following sections), rather than under Football.

There have been four conferences on Mathematics and Computers in Sport held at Bond University, Queensland, and many references are to the proceedings of these conferences. They will be identified below as follows:

Bond 1: *First Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1992).

Bond 2: *Second Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1994).

Bond 3: *Third Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1996).

Bond 4: *Fourth Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar (editors), Bond University, Queensland, Australia (1998).

Copies of these proceedings may be obtained by writing to the School of Information Technology, Bond University, Gold Coast, Queensland 4229, Australia.

2 The bibliography

General interest

1. A. Armenti, Jr (editor), *The Physics of Sports*, American Institute of Physics, New York (1992). A collection of reprinted papers in a huge array of sports.

2. P. Avery, "Mathematics in sport", *Math. Gaz.*, **73** (1989), 1–6. Elementary calculus, probability, combinatorics and statistics applied respectively in rugby, squash, speedway and soccer.

3. J. Bennett (editor), *Statistics in Sport*, Arnold, London (1998).[1] Covers a great many sports and includes a chapter on tournament design.

4. P. J. Brancazio, *Sport Science—Physical Laws and Optimum Performance*, Touchstone, New York (1985). An elementary approach to the physics and mathematics behind many sports.

5. A. Brown, "The probability of breaking sports records", *J. Rec. Math.*, **26**(1) (1994), 42–47. The main application is to baseball.

6. K. D. Buchan and P. M. Sommers, "Judging Olympic judges for political bias", *J. Rec. Math.*, **28**(2) (1996–97), 93–97. Uses paired t-tests to seek evidence of bloc voting in women's gymnastics in the 1992 Olympics.

7. G. L. Cohen, "The subject: Mathematics in Sport", *Bond 4*, 225–235. A full description of the syllabus and assessment methods in this subject.

8. G. L. Cohen and P. Petocz, "Proof by sporting analogy", *Austral. Math. Soc. Gaz.*, **25**(5) (1998), 242–245. The proofs are of some combinatorial identities.

9. M. Cover, "Do longer games favour the stronger player?", *Amer. Statist.*, **43**(4) (1989), 277–278.

10. D. S. F. Crothers, "Mathematics in sport, I", *Int. J. Math. Educ. Sci. Technol.*, **23** (1992), 117–126. Aspects of mechanics, applied to sport, treated with vectors.

11. D. S. F. Crothers, "Mathematics in sport, II", *Int. J. Math. Educ. Sci. Technol.*, **23** (1992), 225–233. An intriguing miscellany, including encryption for passing messages on the rugby field.

12. J. S. Croucher, "Scientific method in sport", in *ASOR '87* (conference proceedings, Australian Society for Operations Research 8th National Conference, Melbourne, 11–14 October 1987), 44–54. Surveys applications of computing software in a number of sports.

13. J. S. Croucher, "Winning with science", *Bond 1*, 1–22.

[1] I am grateful to the referee for suggesting the inclusion of this book.

14. C. B. Daish, *The Physics of Ball Games*, English Univ. Press, London (1972). Includes, for example, the mechanics of where to strike a snooker ball so that it rolls without skidding.

15. N. J. de Mestre, "Mathematics and sport", *Austral. Math. Teacher*, **43**(4) (1987), 2–5. Problems combining mathematics and sport, separated into those suitable at tertiary, secondary and primary levels.

16. N. J. de Mestre, "Mathematics applied to sport", *Bond 1*, 137–148. Survey of a number of areas combining mathematics and sport.

17. N. J. de Mestre, *The Mathematics of Projectiles in Sport*, Cambridge University Press, Cambridge UK (1990).

18. Y. Gerchak, "Operations research in sports", in *Handbooks in Operations Research and Management Science*, (volume 6, *Operations Research and the Public Sector*), S. M. Pollock, M. H. Rothkopf and A. Barnett (editors), North–Holland (1994). Includes a huge list of references.

19. D. Hoffman, "A taxonomy of sporting events applying operations research methodology", *Bond 1*, 101–10. Includes a large bibliography.

20. J. B. Keller, "A characterization of the Poisson distribution and the probability of winning a game", *Amer. Statist.*, **48**(4) (1994), 294–298. Perhaps better suited to a second course in statistics; includes applications to soccer and baseball.

21. S. P. Ladany and R. E. Machol (editors), *Optimal Strategies in Sports*, North–Holland Publishing Company (1977). Includes an extensive annotated bibliography and list of references on pages 206–224.

22. A. G. Mackie, "Mathematics in sport", *Bull. Inst. Math. Appl.*, **16**(1) (1980), 2–6. Includes applications to high jumping, rugby football and ball games.

23. R. D. Mehta, "Aerodynamics of sports balls", *Ann. R. Fluid Mech.*, **17** (1985), 151–189.

24. L. E. Sadovskii and A. L. Sadovskii, *Mathematics and Sports*, American Mathematical Society (1993). The examples are mainly in terms of the techniques of operations research.

25. P. M. Sommers, "Greying of Olympic gold", *J. Rec. Math.*, **28** (1) (1996–97), 18–21. Tests the null hypothesis that the average age of US Olympians in swimming and track and field has remained constant from 1972 to 1988.

26. M. S. Townend, *Mathematics in Sport*, Ellis Horwood, New York (1984).

27. M. S. Townend and D. C. Pountney, *Learning Modelling with DERIVE*, Prentice–Hall, London (1995). Many of the examples are sporting applications.

Draws, Tournaments, Rankings and Ratings

28. G. W. Bassett, Jr, "Robust sport ratings based on least absolute errors", *Amer. Statist.*, **51**(2) (1997), 99–105. A least squares approach to sports ratings, applied to teams in the NFL.

29. F. Budden, "Speedway tournaments in the classroom", *Math. Gaz.*, **61** (1977), 266–272. A combinatorial problem allowing the introduction of some group theory.

30. R. M. Dawes and J. B. Kadane, "Partial round-robin comparisons with perfect rankings", *Amer. Statist.*, **41**(3) (1987), 204–205. An application of probability.

31. N. J. de Mestre, "Speedway problem", *Working Paper 1997-3-019/A* (April 1997), Bond University, Queensland 4229, Australia. Considers a generalisation of the usual draw for 16 riders in the World Speedway Championships.

32. J. H. Dinitz, E. R. Lamken and W. D. Wallis, "Scheduling a tournament", in *The CRC Handbook of Combinatorial Designs*, C. J. Colbourn and J. H. Dinitz (editors), CRC Press (1996), 565–578. Includes such things as softball balanced tournament designs.

33. T. J. Fletcher, "Speedway tournaments", *Math. Gaz.*, **60** (1976), 256–262. The draw is discussed in terms of finite geometries.

34. J. E. Freund, "Round robin mathematics", *Amer. Math. Monthly*, **63** (1956), 112–114. The draw is developed using modular arithmetic.

35. W. A. Glenn, "A comparison of the effectiveness of tournaments", *Biometrika*, **47** (1960), 253–262.

36. D. L. Hoffman, "A team rating system", *ASOR Bull.*, **11**(2) (1992), 6–10; also *Bond 1*, 23–28.

37. R. J. Leake, "A method for ranking teams with an application to college football", *Management Sc. in Sports* (1976), 27–46.

38. S. B. Maurer, "The king chicken theorems", *Math. Mag.*, **53**(2) (1980), 67–80. Aspects of graph theory directed towards finding a winner in a round-robin event.

39. P. A. Rogerson, "Inconsistencies in league standings", *J. Rec. Math.*, **28**(2) (1996–97), 81–84. Demonstrates inconsistencies in different approaches to ranking teams in a league, with application to baseball.

40. N. C. Schwertman and L. Howard, "A probability model for the Victorian Football League final series", *Austral. Math. Teacher*, **7**(10) (1989), 2–3.

41. N. C. Schwertman and L. Howard, "A probability model for the Australian Football League final series", *Austral. Math. Soc. Gaz.*, **17**(4) (1990), 89–94.

42. P. M. Sommers, "Meddling with Olympic results", *J. Rec. Math.*, **28**(1) (1996–97), 49–52. Applies a medal point rate to determine the top countries in the 1992 Olympics.

43. W. D. Wallis, "One-factorizations of graphs: tournament applications", *College Math. J.*, **18**(2) (1987), 116–123. How to construct round-robin tournaments, and why you need to think about this.

44. R. O. Weber and N. J. de Mestre, "Finals draws", *Bond 2*, 105–111. The application is to Australian Rules football.

45. J. J. Wiorkowski, "A curious aspect of knockout tournaments of size $2n$", *Amer. Statist.*, **26** (1972), 28–30. Considers the probability that the second best player should be the runner-up.

Baseball

46. J. Bennett, "Did Shoeless Joe Jackson throw the 1919 World Series?", *Amer. Statist.*, **47**(4) (1993), 241–251. Answer: Probably not. A very detailed statistical analysis of the 1919 World Series.

47. J. D. Gibbons, I. Olkin and M. Sobel, "Baseball competitions—are enough games played?", *Amer. Statist.*, **32**(3) (1978), 89–95. How many games are required in order to state that the probability is at least some specified value that the better team wins the World Series?

48. C. Harman, "'Who's on first!' 'What?' 'What's on second!' and how 'What' got there on an optimal baserunning path", *Bond 4*, 217–223. Using calculus, particularly an analysis of curvature, to determine the optimal path for a baseballer sprinting through first base to second.

49. W. Runquist, *Baseball by the Numbers*, McFarland, North Carolina (1995). An introductory textbook on statistics, with all the standard topics motivated by baseball.

50. P. M. Sommers, "Probably the greatest hitter who ever lived", *J. Rec. Math.*, **26**(1) (1994), 32–35. Uses probability to determine whether Ted Williams was really amongst the best, given that he lost five years due to war service.

51. P. M. Sommers, "Pitcher perfect: all tall talk?", *J. Rec. Math.*, **28**(3) (1996–97), 177–179. A use of the chi-squared distribution to compare left- and right-handers by height.

52. P. M. Sommers, "Home-field advantage in the World Series: myth or reality?", *J. Rec. Math.*, **28**(3) (1996–97), 180–184. The null hypothesis is that winning and home games are independent.

Basketball

53. Y. Gerchak and M. Henig, "The basketball shootout: strategy and winning probabilities", *OR Letters*, **5** (1986), 241–244.

54. N. C. Schwertman, T. A. McCready and L. Howard, "Probability models for the NCAA regional basketball tournaments", *Amer. Statist.*, **45**(1) (1991), 35–38. The models are tested using the chi-squared distribution.

55. P. H. Westfall, "Graphical presentation of a basketball game", *Amer. Statist.*, **44**(4) (1990), 305–307.

Cricket

56. M. N. Brearley, J. C. Burns and N. J. de Mestre, "What is the best way to hit a cricket ball?", *Int. J. Math. Educ. Sci. Technol.*, **21** (1990), 949-961. Some rotational dynamics, backed up by experiment.

57. G. L. Cohen, "One-day cricket: inferences from bowlers' strike rates", *Math. Today (Bull. Inst. Math. Appl.)*, **35** (1999), 45–47. The harmonic mean of individual bowlers' strike rates gives the team bowling strike rate and leads to the probability of bowling the other side out.

58. G. L. Cohen, "One-day cricket: the effect of running out an opposing batsman", *Math. Today (Bull. Inst. Math. Appl.)*, to appear.

59. P. Coutis, "Modelling the projectile motion of a cricket ball", *Int. J. Math. Educ. Sci. Technol.*, **29**(6) (1998), 789–798. Simple use of differential equations and perturbation techniques, resulting from a linear drag model, to determine the trajectories required to hit a six.

60. F. C. Duckworth and A. J. Lewis, "A fair method for resetting the target in interrupted one-day cricket matches", *J. Oper. Res. Soc.*, **49**(3) (1998), 220–227. See also *Bond 3*, 51–68, and Lewis and Duckworth, below.

61. D. Dyte, "Constructing a plausible test cricket simulator using available real world data", *Bond 4*, 153–159.

62. S. Innes and S. Sugden, "One day cricket scoring program", *Bond 4*, 45–49.

63. M. I. Johnstone, S. R. Clarke and D. H. Noble, "Assessing player performance in one day cricket using dynamic programming", *Asia Pacific J. Oper. Res.*, **10** (1993), 45–55. A dynamic programming formulation to develop a method of calculating the contribution of each batsman to the team's score in one-day cricket.

64. C. Kimber and A. R. Hansford, "A statistical analysis of batting in cricket", *J. Roy. Statist. Soc. Ser. A*, **156**(3) (1993), 443–455. In search of a better measure than the batting average.

65. T. Lewis and F. Duckworth, "Developments in the Duckworth–Lewis (D/L) method of target-resetting in one-day cricket matches", *Bond 4*, 131-151. Updates the Duckworth–Lewis article, above. An empirical analysis that has been adopted in international one-day cricket.

66. J. M. Watson, "Howzat hypothesis", *Austral. Math. Teacher*, **45**(1) (1989), 9. Hypothesis testing applied to leg-before-wicket dismissal rates.

67. D. J. Winteridge and D. C. Young, "Mathematics, cricket and common sense", *Math. Gaz.*, **69** (1985), 206–208. On apparent inconsistencies in calculating bowlers' averages.

Cycling

68. W. H. Cogill, "The mathematics of bicycling", *Bond 2*, 37–41. Showing that the potential energy of a bicycle is always larger than its kinetic energy.

69. W. H. Cogill, "The mathematics of bicycling: Part II", *Bond 4*, 245–248. Continuing the theme of the previous note, the consideration here is the overturning moment on a bicycle.

70. J. C. Smith and D. W. Hill, "Mathematical models of the power–time relationship in high intensity cycling", *Med. Sci. Sports Exerc.*, **24** (S74) (1992).

Darts

71. G. L. Cohen and E. Tonkes, "Dartboard arrangements", *Electron. J. Combin.*, to appear. An exercise in combinatorial optimisation, completely solved.

72. H. A. Eiselt and G. Laporte, "A combinatorial optimization problem arising in dartboard design", *J. Oper. Res. Soc.*, **42** (1991), 113–118. Treats the problem as one in linear programming.

73. P. J. Everson and A. P. Bassom, "Optimal arrangements for a dartboard", *Math. Spectrum*, **27** (1994), 32–34.

74. D. Kohler, "Optimal strategies for the game of darts", *J. Oper. Res. Soc.*, **33** (1982), 871–884.

75. M. A. M. Lynch, "Designing a dartboard—an application of graph modelling", *Teaching Math. and Its Appl.*, **16** (1997), 51–54. Useful example of aspects of graph theory.

76. D. Percy, "Winning darts!", *Math. Today (Bull. Inst. Math. Appl.)*, **35** (1999), 54–57. Uses Bayesian decision theory to suggest that non-professionals should simply aim for the bull's-eye.

77. K. Selkirk, "Re-designing the dartboard", *Math. Gaz.*, **60** (1976), 171–178. The problem is to maximise the sum of the absolute differences of adjacent numbers.

Football (of all types)

78. G. Aldis, G. Fulford, R. Weber and N. de Mestre, "Flight of a football", *Bond 1*, 125–136. The model includes lift and drag, and is solved numerically.

79. P. J. Brancazio, “The physics of kicking a football”, *Phys. Teacher*, **23** (1985), 403–407.

80. S. R. Clarke, “Analysing football finals with a spreadsheet”, *Austral. Senior Math. J.*, **7**(1) (1993), 35–44.

81. S. R. Clarke, “Evaluating the fairness and efficiency of the Australian Football League home and away schedule”, *ASOR Bull.*, **17**(3) (1998), 2–11. Uses computer simulations to test the model.

82. A. Hughes, “Conversion attempts in rugby football”, *Math. Gaz.*, **62** (1978), 292–293. Whimsical use of elementary coordinate geometry.

83. D. E. Hughes, “When is a pass forward?”, *Math. Gaz.*, **62** (1978), 44–45. An application of vectors.

84. D. C. Isaksen, “How to kick a field goal”, *College Math. J.*, **27**(4) (1996), 267–271. Uses coordinate geometry, and applies the result to American football.

85. A. Mehrez, J. S. Pliskin and A. Mercer, “A new point system for soccer leagues: Have expectations been realised?”, *Euro. J. Oper. Res.*, **28** (1987), 154–157.

86. R. T. Stefani and S. R. Clarke, “Australian Rules football during the 1980’s”, *ASOR Bull.*, **10**(3) (1991), 11–15.

87. H. Stern, “On the probability of winning a football game”, *Amer. Statist.*, **45**(3) (1991), 179–183.

88. N. Tomecko and J. A. Filar, “Player assignments in Australian Rules football”, *Bond 4*, 171–179. Refers to the assignment problem in operations research.

89. M. S. Townend, “Getting a kick out of numerical differentiation”, *Teaching Math. and its Appl.*, **5**(2) (1986), 57–61.

90. G. Worsnop, “An aid to conversions in rugby”, *Math. Gaz.*, **73** (1989), 225–226. More whimsy with coordinate geometry.

Golf

91. S. R. Clarke and J. M. Rice, “How well do golf courses measure golf ability? An application of test reliability procedures to golf tournament scores”, *ASOR Bull.*, **14**(4) (1995), 2–11. For example, there is often a negative correlation in golfers’ scores from one hole to the next.

92. I. L. Collings and N. J. de Mestre, “A refined aerodynamic model for low trajectory flight”, *Bond 4*, 15–20. Analyses differential equations governing the motion of a golf ball.

93. T. Hale and G. T. Hale, “Lies, damned lies and statistics in golf”, in *Science and Golf*, A. J. Cochran (editor), Chapman and Hall, London (1990), 159–164. Suggests that the scientific value of data collection in golf is questionable.

94. R. Heiny and C. E. Crosswhite, “Best-ball events in golf: an application of the multinomial distribution”, *Amer. Statist.*, **40**(4) (1986), 316–317.

95. P. Moin and J. Kim, “Tackling turbulence with supercomputers”, *Sci. Amer.*, (January 1997), 46–52. Includes a discussion of Reynold’s number and flow past a golf ball.

96. S. M. Pollock, “A model of the USGA handicap system and ‘fairness’ of medal and match play”, in *Optimal Strategies in Sports*, S. P. Ladany and R. E. Machol (editors), North–Holland (1977), 141–150. Uses basic techniques of probability and statistics.

97. L. J. Riccio, “Statistical analysis of the average golfer”, in *Science and Golf*, A. J. Cochran (editor), Chapman and Hall, London (1990), 153–158.

98. F. J. Scheid, "On the normality and independence of golf scores with various applications" in *Science and Golf*, A. J. Cochran (editor), Chapman and Hall, London (1990), 147–152. The scores of individual golfers generally follow roughly a normal distribution.

99. R. C. Stroud, "Mathematical underpinnings of the slope handicap system", in *Science and Golf*, A. J. Cochran (editor), Chapman and Hall, London (1990), 135–140. Applies some very simple coordinate geometry.

Ice Hockey

100. G. Mullet, "Simeon Poisson and the National Hockey League", *Amer. Statist.*, **31**(1) (1977), 8–12. Goals for and goals against, both home and away, tend to follow Poisson distributions.

Jai Alai

101. L. E. Moser, "A mathematical analysis of the game of jai alai", *Amer. Math. Monthly*, **89**(5) (1982), 292–300. Uses game trees and probability for a complete analysis.

Lawn Bowls

102. M. N. Brearley, "A mathematician's view of bowling", *Math. Gaz.*, **80** (1996), 501–510. Discusses the bowl's bias in terms of precession, among other things.

103. M. N. Brearley and B. A. Bolt, "Dynamics of the lawn bowl", *Quart. J. Mech. and Appl. Math.*, **XI**(3) (1958), 351–363.

104. T. Roper, "The mathematics of bowls", *Math. Gaz.*, **80** (1996), 298–307.

105. H. Williams, "The mathematics of flat green bowling", *Math. Gaz.*, **82** (1998), 242–253. Various approaches to determining the bowl's path.

Netball

106. D. H. Noble, "Formation of netball teams for a series of trial matches", *Bond 3*, 97–109. A specific application of the assignment problem.

Rowing

107. M. N. Brearley, "Oar arrangements in rowing eights", in *Optimal Strategies in Sports*, S. P. Ladany and R. E. Machol (editors), North–Holland (1977), 184–185. Calculates moments of the relevant forces.

108. M. N. Brearley and N. J. de Mestre, "Modelling the rowing stroke and increasing its efficiency", *Bond 3*, 35–46.

109. M. N. Brearley and N. J. de Mestre, "The effect of oar flexing on rowing performance", *Bond 4*, 21–32. Practical mechanics, with a useful conclusion.

Squash and Racquetball

110. W. J. Brooks and L. P. Hughston, "A problem in squash strategy", *Math. Gaz.*, **72** (1988), 92–95. The probability of winning a game, given the probability of winning a point.

111. M. Hughes, "Using notational analysis to create a more exciting scoring system for squash", in *Sport, Leisure and Ergonomics*, G. Atkinson and T. Reilly (editors), E. & F. N. Spon, London (1995), 243–247. Data analysis under different scoring systems.

112. J. Marcus, "Probability of winning a game of racquetball", *SIAM Rev.*, **27** (1985), 443–444.

113. D. Watson, "On scoring in games", *Math. Gaz.*, **54** (1970), 110–113. The probability of winning a game, given the probability of winning a point, applied also to table tennis.

Swimming

114. R. E. Machol, "Assigning swimmers to a relay team", in *Optimal Strategies in Sports*, S. P. Ladany and R. E. Machol (editors), North–Holland (1977), 168. A classical application of the assignment problem.

Tennis

115. W. H. Carter, Jr, and S. L. Crews, "An analysis of the game of tennis", *Amer. Statist.*, **28**(4) (1974), 130–134. The probability of winning a game, and the expectation of the length of a game, among other things.

116. J. S. Croucher, "Changing the rules of tennis: who has the advantage?", *Teaching Statist.*, **7**(4) (1985), 82–84.

117. J. S. Croucher, "The effect of the tennis tie-breaker", *Res. Quart. for Exercise and Sport*, **53**(4) (1982), 124–127. In terms of probability and expectation.

118. J. S. Croucher, "The conditional probability of winning games of tennis", *Res. Quart. for Exercise and Sport*, **57**(1) (1986), 23–26.

119. J. S. Croucher, "Using computers and scientific method to determine optimal strategies in tennis", *Bond 4*, 1–14.

120. D. Gale, "Optimal strategies for serving in tennis", *Math. Mag.*, **5** (1971), 197–199.

121. S. L. George, "Optimal strategy in tennis: A simple probabilistic model", *J. Roy. Statist. Soc. Ser. C*, **22**(1) (1973), 97–104.

122. J. R. Magnus and F. J. G. M. Klaasen, "The final set in a tennis match: four years at Wimbledon", *J. Appl. Statist.*, **26**4) (1999), 461–468. Testing hypotheses based on data from Wimbledon.

123. G. H. Pollard, "An analysis of classical and tie-breaker tennis", *Austral. J. Statist.*, **25**(3) (1983), 496–505. In terms of probability and expectation.

Track and Field

124. N. J. de Mestre, M. Hubbard and J. Scott, "Optimizing the shot put", *Bond 4*, 249–257. Uses perturbation techniques on the equations of motion, and tests with real data.

125. D. R. Heffley, "Assigning runners to a relay team", in *Optimal Strategies in Sports*, S. P. Ladany and R. E. Machol (editors), North–Holland (1977), 169–171. An application of the assignment problem, in operations research.

126. K. Prendergast, "What do world running records tell us?", *Modern Athlete and Coach*, **28**(1) (1990), 33–36. A good discussion, but bad mathematics, which makes this worthwhile.

127. R. Tibishirani, "Who is the fastest man in the world?", *Amer. Statist.*, **51**(2) (1997), 106–111. A comparison of world record times for 100 metres and 200 metres; uses principles from mechanics, as well as statistics.

PLAYER RATINGS IN ONE-DAY CRICKET

John S. Croucher
Division of Economic and Financial Studies
Macquarie University
NSW 2109, Australia

Abstract

Perhaps more than any other sport, cricket is brimming with a multitude of statistics and records that provide a feast of data for the enthusiastic statistician to analyse. In this case, consideration is given to the triangular one-day international series between Australia, India and Pakistan played during the 1999/2000 season. A formula is suggested for ranking both the batsmen and bowlers and a simulation made to predict the outcome of the final series of matches.

1 Introduction

Each Australian summer two cricketing nations are invited to Australia to participate in a triangular series of one-day international matches. The rounds of the tournament involve each team playing each other four times so that a total of twelve matches are involved. Teams are awarded two points for a win, one point for a draw and no points for a loss. The two teams with the most points at the end of these initial matches then play in a best of three match final series.

A one-day match consists of each team batting for a maximum of fifty six-ball overs where the team that scores the most runs is declared the winner. The number of wickets lost along the way is not relevant, although the innings is over when all ten wickets are lost.

There is much strategy involved in this type of cricket with issues such as the best batting order, optimal field placings and judicious use of bowlers all given careful consideration since each bowler is allowed a maximum of only ten overs. Decisions such as whether to bat or bowl first if the toss is won can be vital depending on the state of the pitch and the degree to which the batsmen should attack the bowling is of utmost importance. Since the number of balls faced is limited, there has to be a fine balance between scoring runs and losing wickets and not all teams achieved this result successfully. Means of comparing cricket performance of all-rounders can be found in Skinner [13] and Nemeruck *et al.* [12] while improvements to the notion of batting average are given in Kimber and Hansford [10]. One-day cricket, probably since it is easier to analyse because of its nature, is also popular with researchers with scoring programs being developed by Innes and Sugden [8], scoring methods by Christos [1] and de Silva and Swartz [6], scoring policies by Johnson *et al.* [9] and bowlers' strike rates and the effect of run outs by Cohen [2, 3].

Even umpires do not escape attention with Croucher [4] and Crowe and Middledorp [5] setting their sights on the consistency of leg-before-wicket decisions. An interesting paper by Kumar [11] suggests that it is possible in some cases to forecast the run rate per over during the last few overs of one-day matches. But it the work of Duckworth and Lewis [7] that has probably captured the most attention since it is their technique that has been adopted for deciding the required score to win by the batting team in rain interrupted one-day matches.

2 The 1999/2000 series

The series played in Australia during the 1999/2000 summer comprised the nations of Australia, Pakistan and India. It wasn't long before India found themselves in trouble by losing their first four matches. They did not recover and were the team eliminated from the final series. Each team had their strengths and weaknesses in both batting and bowling and team selections were sometimes experimental to give fringe players a chance or to give star players a rest during the rounds. At the end of the rounds, India had won only one of their eight matches, Pakistan had won four and Australia had won seven, all in succession after losing their first match.

The final series therefore consisted of Australia versus Pakistan, with Australia the clear favourites after having won the Test Series against them 3–0 as well as their previous three one-day matches. They had also won the World Cup in 1999. Their strength was reflected in the betting market framed on the morning of the match by *SportsTAB* who offered a $1.35 return for each successful $1 bet on Australia and $2.90 for Pakistan. These represented odds of around 3–1 *on* for Australia and nearly 2–1 *against* for Pakistan.

3 Batting index

Whenever a player comes in to bat the television coverage always displays a number of statistics relevant to their previous batting record and are provided to indicate their strengths or weaknesses. Prominent among these are two crucial figures:

- *The batting average.* This is defined as the total number of runs that the batsman has made divided by the number of times that they were dismissed. This might be considered an indication of an overall "expected" value of their likely score, although it would clearly depend on the quality of the opposition bowling. There could also be considerable variation due to playing conditions and the state of the game during their innings. In practice, the batting average alone is often viewed as a measure of the worth of a batsman, although it can be misleading if a batsman has a relatively high proportion of not-out scores.
- *The batting strike rate.* This is defined as the number of runs the batsman has scored for every 100 balls that they have faced. In other words, this indicates the "speed" at which the batsman has scored. For instance, if a batsman has a strike rate of 73, this means that he has scored 73 runs for every 100 balls faced. Batting strike rates of close to, or exceeding, 100 are only achieved by very few players during their career.

Both of these statistics are very important to the selection of an appropriate player. For example, it is not of much use having a batsman who can average 50 runs if they have a strike rate of only 20 since they would be far too slow and take up too many balls in reaching their score. On the other hand, a batsman with a strike rate of 120 but an average of only 12 would not be of much use as a frontline batsman but might be handy if only a few balls remain in the innings. With this in mind, the ideal batsman in this type of cricket is one who can score a great deal of runs *and* can score them relatively quickly. This naturally leads to the question of how to compare or rate each batsman and there are no shortages of how this might be done. One simple method of achieving this goal is to construct a *Batting Index*, which may be likened to the calculation of a batsman's *momentum*.

The *momentum* of a body is defined in physics as the product of its mass and velocity. In this case, the momentum of a batsman may be considered as the simple product of their *batting average* (analogous to mass) and strike rate (analogous to velocity). In other words:

$$\text{Batting Index} = \text{batting average} \times \text{batting strike rate} \tag{1}$$

Using the definition in (1), a player's batting utility may be ranked according to their Batting Index. With the strike rate calculated as the number of runs scored per 100 balls, values for the Batting Index will almost always be under 5000 and usually around 3000 for a specialist batsman and less than 1000 for a specialist bowler.

Tables 1, 2 and 3 show the values of the Batting Indexes for the players from Australia, Pakistan and India, respectively, at the commencement of the 1999/2000 series. The outstanding player was Michael Bevan whose batting average of 59.01 was 25% higher than any other batsman in the history of one-day international matches. This, coupled with his outstanding strike rate, makes him arguably the most valuable batsman to have ever played the game.

Player	*Innings*	*Strike rate*	*Batting Average*	*Batting Index*
Michael Bevan	111	75.4	59.01	4449
Mark Waugh	197	76.8	38.42	2951
Adam Gilchrist	78	87.4	33.29	2910
Ricky Ponting	95	71.5	40.40	2889
Andrew Symonds	9	82.9	31.50	2611
Steve Waugh	252	82.2	31.66	2602
Damien Martyn	38	78.1	28.20	2202
Shane Lee	22	85.4	18.23	1575
Adam Dale	12	57.4	19.50	1117
Ian Harvey	9	69.2	14.42	997
Glen McGrath	32	49.5	3.46	699
Damien Fleming	23	50.0	7.75	388

Table 1: Career one-day International batting records for Australia.

Player	*Innings*	*Strike rate*	*Batting Average*	*Batting Index*
Saeed Anwar	187	82.8	46.10	3817
Yousuf Youhana	31	72.9	45.28	3301
Inzaman-ul-Haq	192	72.3	38.91	2813
Ijaz Ahmed	221	80.0	32.21	2577
Moin Khan	126	80.6	23.91	1927
Shahid Afridi	88	77.0	23.51	1810
Wasim Akram	224	89.1	15.82	1410
Abdur Razzaq	29	57.4	24.22	1391
Azhar M	58	75.8	18.06	1369
Shoaib Akhtar	11	62.2	18.66	1161
Saqlain Mushtaq	66	52.6	14.24	749
Waqar Younis	89	71.2	10.57	753

Table 2: Career one-day international batting records for Pakistan.

From Tables 1, 2 and 3 it is evident that, in contrast to the Australian and Pakistan teams, the Indian batsmen were going to suffer by comparison since they had only three batsmen with a Batting Index above 2500 compared with six for Australia and four for Pakistan. To make matters worse, their star batsman and captain, Sachin Tendulkar, ultimately turned out to have a poor series (for him) with a batting average of less than 30.

There are, of course, far more complex systems of rating players and in 1998 the global professional services firm PricewaterhouseCoopers introduced ratings for players in limited-overs internationals. Players considered must not have retired and must have appeared in an official match during the previous twelve months. Batsmen were rated on factors including runs scored, the rate at which they scored, the overall rate for the match, the ratings for opposing bowlers and the result of the match. However, the Batting Index is much easier to calculate and can be found by anyone who cares to watch the television broadcasts.

Player	*Innings*	*Strike rate*	*Batting Average*	*Batting Index*
Sachin Tendulkar	225	86.5	42.34	3662
Sourav Ganguly	124	73.0	42.72	3118
Rahul Dravid	104	68.4	37.47	2563
Robin Singh	89	78.3	27.75	2173
H. Kanitkar	23	67.1	17.22	1155
Nikhil Chopra	23	63.3	16.72	1058
Javagal Srinath	93	86.5	11.60	1003
Devang Gandhi	2	45.6	18.00	821
Anil Kumble	90	66.6	10.03	668
Vengatesh Prasad	54	61.5	6.10	375
V.V.S. Laxman	9	36.8	9.37	319
Sameer Dighe	3	19.3	3.66	71

Table 3: Career one-day international batting records for India.

Although the actual PricewaterhouseCoopers formula for the ratings is not revealed, at the beginning of the series in 2000 Australia dominated the top rankings with three players Bevan (first), Ponting (fourth), and Mark Waugh (sixth). Ganguly (third) was the only Indian batsman in the top ten while Anwar (seventh) was the only Pakistan batsman.

These rankings are quite consistent with the simple Batting Index proposed here, although Tendulkar had a current PricewaterhouseCoopers' batting ranking of eleventh (after being first only one year before) and Gilchrist was ranked 13th in the world.

4 Expected number of balls faced

As well as determining the Batting Index, it is easy to calculate the expected number of balls that a batsman might take to achieve their average score. This can be found from:

$$\text{Expected number of balls faced} = \frac{\text{Batting Average}}{\text{Strike rate}} \times 100 \tag{2}$$

For example, using (2), a player with a batting average of 32 and a strike rate of 80 would be expected to score these in $(32/80) \times 100 = 40$ balls. In this way an expected number of balls that might be faced can be calculated for each player. However, when these are added for a team they may well exceed the maximum of 300 balls since there will be variation with some players facing much more than their expected value and other players might not even get a turn at bat.

From the three teams combined, the ten batsmen with the longest expected stay at the crease during their career are shown in Table 4. For comparison, the final column shows their average stay during the 1999/2000 series. The correlation coefficient between the series and career mean balls faced was 0.60 which was only marginally significant. The biggest disappointments for Pakistan included Yousuf Youhana who had a series batting average of only 23 and faced an average of only 36 balls—both about half of what might have been expected. But Inzaman-ul-Haq was even worse with a series batting average of about 12, facing an average of only 23 balls.

Batsman such as Bevan, Ganguly, Dravid and Razzaq were the closest to performing according to their past records. The case of Michael Bevan is particularly interesting since his batting average of nearly sixty was due to the fact that he normally batted at number six and remained not out nearly 40% of the time. However, in the 1999/2000 series he moved up to bat at number four, yet still managed to achieve a series batting average of 55 runs.

Player	*Country*	*No. of innings*	*Career Batting Average*	*Average career balls faced*	*Average series balls faced*
Michael Bevan	Australia	111	59.01	78	77
Yousuf Youhana	Pakistan	31	45.28	62	36
Sourev Ganguly	India	124	42.72	59	64
Ricky Ponting	Australia	95	40.40	57	37
Saeed Anwar	Pakistan	187	46.10	56	39
Rahul Dravid	India	104	37.47	55	58
Inzaman-ul-Haq	Pakistan	192	38.91	54	23
Mark Waugh	Australia	197	38.42	50	28
Sachin Tendulkar	India	225	42.34	49	33
Abdur Razzaq	Pakistan	29	24.22	42	48

Table 4: The ten series batsmen with the longest career average stay at bat per innings and how they performed in the current series.

5 Bowling index

In the same way as the Batting Index can be constructed, it is also possible to create a simple index to rank the utility of bowlers in one-day cricket. This is also based on the notion of momentum where this time the mass is represented by the bowling average (the number of runs conceded per wicket taken) and the velocity is represented by the bowling strike rate (the number of balls bowled per wicket taken). The formula is:

$$\text{Bowling Index} = \text{bowling average} \times \text{bowling strike rate} \qquad (3)$$

Tables 5, 6 and 7 show the career Bowling Indexes for the leading bowlers from Australia, Pakistan and India, respectively at the commencement of the series. In this case the *lower* the index the better it is. A player with a *low* bowling average and *high* bowling strike rate is one who does not concede many runs but does not take wickets quickly. A player with a *high* bowling average and *low* bowling strike rate is one who can take wickets quickly but in doing so will concede many runs. The most valuable bowlers are those who have low values in both aspects and hence have the lowest product or Bowling Index.

Taking into account only those bowlers who had taken at least ten wickets, the best two bowlers based on the Bowling Index were Shoaib Akhtar and Saqlain Mushtaq from Pakistan with a rating of less than 550, while the best for Australia were Fleming and Warne. As was the case with their batsmen, India's bowlers seemed to suffer by comparison with only one bowler, Agarkar, having a Bowling Index less than 1000 compared to Pakistan with six and Australia with five.

PricewaterhouseCoopers also have a system for rating bowlers in one-day internationals, including factors such as runs conceded per over, wickets taken and runs conceded for the match, the overall ratings of the opposing batsmen and the result of the match. At the beginning of the series, the two Pakistan bowlers Saqlain Mustaq and Azhar Mahmood were rated by them to be second and third in the world while for Australia Glen McGrath was rated fourth, Damien Fleming seventh and Shane Warne eighth. The highest ranked Indian bowler was Vekatesh Prasad at 16th. These are again consistent with the Bowling Index with the exception of Azhar Mahmood who has a ranking much higher than his raw figures suggest.

6 The finals series

As mentioned previously, the final series was played between Australia and Pakistan with Australia heavily favoured to win. The first final was played at the Melbourne Cricket Ground as a day/night

Player	*Wickets*	*Bowling strike rate*	*Bowling average*	*Bowling Index*
Andrew Symonds	16	26.2	20.87	547
Damien Fleming	107	33.6	24.05	808
Shane Warne	214	34.7	24.28	842
Glen McGrath	155	36.6	24.30	889
Shane Lee	27	38.1	28.59	1089
Mark Waugh	81	41.5	32.98	1369
Adam Dale	32	49.8	30.59	1523
Steve Waugh	191	45.4	34.41	1562
Michael Bevan	32	53.8	44.15	2375
Ian Harvey	7	59.5	46.42	2762
Damien Martyn	4	77.5	61.00	4728
Ricky Ponting	1	90.0	64.00	5760

Table 5: Career one-day international bowling records for Australia.

Player	*Wickets*	*Bowling strike rate*	*Bowling average*	*Bowling Index*
Inzaman-ul-Haq	2	20.0	26.00	520
Shoaib Akhtar	51	27.1	19.60	531
Saqlain Mushtaq	212	28.0	19.59	548
Ijaz Ahmed	5	126.4	94.80	632
Waqar Younis	290	30.5	23.31	711
Abdur Razzaq	44	33.7	24.13	813
Wasim Akram	397	36.7	23.53	864
Azhar Mahmood	80	43.6	31.83	1388
Saeed Anwar	5	43.6	35.20	1535

Table 6: Career one-day international bowling records for Pakistan.

Player	*Wickets*	*Bowling strike strike rate*	*Bowling average*	*Bowling Index*
A. B. Agarkar	76	30.2	26.50	800
Nikhil Chopra	43	38.7	26.60	1029
Javagal Srinath	247	37.9	27.60	1046
Anil Kumble	246	40.6	28.31	1149
Vengatesh Prasad	164	42.0	32.14	1350
Sourav Ganguly	45	45.2	37.33	1687
Robin Singh	64	49.5	39.42	1951
Sachin Tendulkar	78	58.3	48.06	2802

Table 7: Career one-day international bowling records for India.

match on 2 February 2000 and the Pakistan captain Wasim Akram won the toss and decided to bat.

Such was the strength of the Australian team, they could not find a place for Damien Fleming who was rated the seventh best bowler in the world and second best (behind McGrath) in Australia by PricewaterhouseCoopers. They also omitted the spinner Stuart MacGill who had played in three of the series round matches and was voted man of the match in one of them. Instead, the Australian selectors opted for a newcomer in Brett Lee (the fastest bowler in the country) and the leg-spinner Shane Warne who had missed six of the round matches through injury.

Armed with an endless supply of cricket statistics, it is an interesting exercise to try to determine just how such a cricket match might turn out. With plenty of variables to worry about, however, this is no easy task. To keep matters simple, in this section a basic technique is used to see if the winning team, along with a few statistics, can be predicted. To do this, only the form of the players during the current series was considered (since recent form was considered the most reliable). In particular, only the batting attributes of the players were used to try and predict just how many runs they might score in the final. The relevant series records leading up to the finals, in the correct batting order, for Australia and Pakistan are shown in Tables 8 and 9, respectively.

With Australia having six batsmen with a Batting Index above 2000 while Pakistan had only three, the form suggested that the Pakistan team was going to struggle in the run-making department. The weak link of the specialist batsmen seemed to be Inzamam-ul-Haq, who lived up to expectations with a second ball duck. The strongest in-form batsman on paper was Abdur Razzaq who managed the second top score of 22 in Pakistan's innings in the first finals match.

Player	*Innings*	*Runs scored*	*Balls faced*	*Strike rate*	*Batting Average*	*Batting Index*
A. Gilchrist	8	212	260	81.53	26.50	2161
M. Waugh	8	242	363	66.66	30.25	2016
R. Ponting	8	276	309	89.32	34.50	3082
M. Bevan	8	331	485	68.24	55.16	3765
S. Waugh	7	139	202	68.81	23.16	1594
D. Martyn	7	117	271	69.29	55.66	1420
A. Symonds	6	139	114	121.92	27.80	3390
S. Lee	5	94	69	136.23	31.33	4268
S. Warne	2	25	38	65.78	25.00	822
B. Lee	2	3	6	50.00	1.50	75
G. McGrath	2	0	5	0	0	0

Table 8: Australian batting records during the 1999/2000 series.

Player	*Innings*	*Runs scored*	*Balls faced*	*Strike rate*	*Batting Average*	*Batting Index*
Saeed Anwar	8	214	310	69.03	26.75	1847
Shahid Afridi	5	107	139	76.98	21.40	1647
Ijaz Ahmed	8	263	349	75.36	32.88	2478
Inzamam-ul-Haq	8	97	180	53.88	12.12	654
Yousuf Youhana	8	187	286	65.38	23.37	1529
Abdur Razzaq	7	201	238	85.16	40.20	3425
Anzhar Mahmood	4	97	103	94.17	24.25	2283
Moin Khan	8	125	176	71.02	15.62	1110
Wasim Akram	8	133	219	60.73	19.00	1154
Saqlain Mushtaq	6	80	141	55.94	20.00	1135
Shoaib Akhtar	3	4	19	21.05	–	–

Table 9: Pakistan batting records during the 1999/2000 series.

The Pakistan innings started disastrously since after only 17 balls their top three batsmen were all dismissed for no score—all out to Glen McGrath. Worse was to follow and they soon found themselves

5 for 28 after 14.3 overs and the match was as good as lost. However, as is often the case when there is a dramatic collapse of top order batsmen, the middle and lower orders seem to find an inner reserve. This was the situation here when the team hung on until the 48th over to be all out for 154, a much better score than seemed likely not long after the start.

To determine just how unexpected was a total of 154, Table 10 shows the actual score of each Pakistan batsman along with the score that might be expected if they had scored their series average score. The difference column in Table 10 shows just how far each was above or below this expected score and sadly the predominance of minus signs shows just how badly Pakistan batted. Only Moin Khan managed to score more than his series average, but with no assistance from the rest of his team it was not enough. Pakistan were allowed a simulated value of 18 runs for extras (sundries) since this was their average per 50 overs during the series. Their final total was 54 runs behind their expected value and left Australia with just 155 runs to win the match.

Based on their run rate in the series, the simulation suggested that Pakistan, on average, should have been 8 for 235 at the end of their 50 overs. This left them 81 runs short of what might have been hoped.

Batsman	*Actual score*	*Simulated score*	*Difference*
Saeed Anwar	7	27	−20
Shahid Afridi	0	21	−21
Ijaz Ahmed	0	33	−33
Inzaman-ul-Haq	0	12	−12
Yousuf Youhana	14	23	−9
Abdur Razzaq	24	40	−16
Azhar Mahmood	16	24	−8
Moin Khan	47	16	+31
Wasim Akram	15	16 n.o.	−1
Saqlain Mushtaq	16	5 n.o.	+11
Shoaib Akhtar	3 n.o.	did not bat	+3
Sundries	12	18	−6
Total	154	235	−81

Table 10: The Pakistan innings—actual and simulated scores.

Using equation (2) based on the expected number of balls that each batsman would last, the simulated score at the fall of each wicket and the over in which it would fall can also easily be calculated. It was assumed that batsmen are equally likely to face a ball bowled i.e. if a partnership lasted 30 balls then it is assumed that each batsman would face 15 balls. In this way it was possible to determine the expected *order* in which the batsmen would be dismissed as well as the expected score of the not out batsman.

These simulated values, along with the actual values, are shown in Table 11. Because of the unexpected early batting collapse, the scores at the actual fall of wickets were distorted early on and never really recovered.

With the Pakistan total only a modest 154 runs, this was going to affect the way in which the Australian team batted since there was no real hurry and it was not necessary to take risks. Despite this, they lost their two opening batsmen (Adam Gilchrist and Mark Waugh) to put them at 2 for 27 in the eighth over. However, the innings was steadied by Ricky Ponting with 50 (this being 15 above his series average) and Michael Bevan who scored 54 (only one less than his series average).

The simulation shown in Table 12 is based on Australia having a full 50 overs to bat. This would have been the case had they batted first and the figures suggest that they would have finished with 7 for 259 had they done so. In fact, since they only required 155 runs to win, the simulation predicted that they would do so with the loss of three wickets and in the 33rd over. The actual victory came with the loss of four wickets (Bevan was dismissed with a few runs left to score) in the 43rd over.

Wicket	*Actual (over) (Batsman out)*	*Simulated (over) (Batsman out)*
1st	1 – 1 (0.5) (Shahid Afridi)	1 – 42 (9.2) (Shahid Afridi)
2nd	2 – 4 (2.3) (Ijaz Ahmed)	2 – 60 (13.0) (Saeed Anwar)
3rd	3 – 4 (2.5) (Inzamam-ul-Haq)	3 – 91 (20.2) (Inzamam-ul-Haq)
4th	4 – 12 (7.2) (Saeed Anwar)	4 – 108 (24.0) (Ijaz Ahmed)
5th	5 – 28 (14.3) (Y. Youhana)	5 – 147 (32.2) (Y. Youhana)
6th	6 – 59 (27.1) (A. Mahmood)	6 – 189 (39.4) (Abdur Razzaq)
7th	7 – 78 (29.5) (Abdur Razzaq)	7 – 197 (41.0) (A. Mahmood)
8th	8 – 147 (39.1) (Wasim Akram)	8 – 223 (47.0) (Moin Khan)
9th	9 – 147 (45.3) (Moin Khan)	
10th	10 – 154 (47.2) (S. Mushtaq)	

Table 11: The Pakistan innings—actual and simulated fall of wickets.

Batsman	*Actual score*	*Simulated score*	*Difference*
A. Gilchrist	9	27	−18
M. Waugh	10	30	−20
R. Ponting	50	35	+15
M. Bevan	54	55	−1
S. Waugh	19 n.o.	23	
D. Martyn	4 n.o.	29	
A. Symonds		28	
S. Lee		31	
S. Warne		13 n.o.	
B. Lee		did not bat	
G. McGrath		did not bat	
Sundries	9	17	−8
Total	4 for 155	7 for 259	−32

Table 12: The Australian innings—actual and simulated scores.

Using the same principles as for the Pakistan team in Table 11, the predicted score at the fall of each Australian wicket is shown in Table 13. The order of the first three batsmen dismissed was correct and the fall of the third wicket was out by only 12 runs.

Wicket	*Actual (over) (Batsman out)*	*Simulated (over) (Batsman out)*
1st	1 – 11 (3.6) (A. Gilchrist)	1 – 51 (11.0) (A. Gilchrist)
2nd	2 – 27 (7.4) (M. Waugh)	2 – 72 (15.0) (M. Waugh)
3rd	3 – 104 (29.6) (R. Ponting)	3 – 116 (24.0) (R. Ponting)
4th	4 – 147 (38.4) (M. Bevan)	4 – 164 (35.4) (S. Waugh)
5th		5 – 204 (42.2) (M. Bevan)
6th		6 – 214 (43.2) (A. Symonds)
7th		7 – 259 (50.0) (S. Lee)

Table 13: The Australian Innings—actual and simulated fall of wickets.

7 All-rounder index

Since it has been possible to produce a Batting Index as a measure of batting ability and a Bowling Index as a measure of bowling ability, it seems appropriate that these two measures might be combined in some way to rank players who are proficient at both. These players are known as "all-rounders". The higher the Batting Index and the lower the Bowling Index, the better the player in those aspects in one-day cricket. Since these two measures are independent, a simple All-rounder Index could be defined as in equation (4):

$$\text{All-Rounder Index} = 100 \times \frac{\text{Batting Index}}{\text{Bowling Index}} \tag{4}$$

The ratio of the Batting Index to the Bowling Index is multiplied by 100 to yield easy to rank numbers and is rounded to the nearest integer. To give an indication how this measure applies to the Australian team, Table 14 shows the value of the All-Rounder Indexes for the top nine players.

Player	*Batting Index*	*Bowling Index*	*All-rounder Index*
Andrew Symonds	2611	547	477
Mark Waugh	2951	1369	216
Michael Bevan	4449	2375	187
Steven Waugh	2602	1562	167
Shane Lee	1575	1089	145
Shane Warne	1024	842	122
Adam Dale	1117	1523	73
Damien Fleming	388	808	48
Damien Martyn	2202	4728	47

Table 14: The All-rounder Index values for current Australian players.

From Table 14 it can be easily seen that the best Australian all-rounder is Andrew Symonds, largely because of his very low Bowling Index that in turn was due to taking 16 wickets for relatively few runs. Suppose there were minimums of, say, 1000 runs and 30 wickets before a meaningful value could be obtained. This would leave Mark Waugh clearly on top, followed by Michael Bevan and Steven Waugh.

8 Remarks

This paper would not be complete without mentioning the result of the second finals match played at the Sydney Cricket Ground as a day/night fixture on 4 February 2000. This time Australia won the toss and batted, managing to score a then all-time high one day international score (for Australia) of seven wickets for 337 after their allotted 50 overs. Perversely, the only failure was Michael Bevan who scored only three runs. Pakistan were unable to sustain an initial onslaught of the Australian bowling and were dismissed for 185 in 36.3 overs.

Australia thus won the finals series 2–0. Of all the statistics that came from the international series that summer, one of the more unusual was the batting of the Pakistan fast bowler Shoaib Akhtar who batted five times and was never dismissed. Perhaps he could have been moved up the order from his usual number 11 position.

The wealth of cricket statistics provides a gold mine for data analysts who try to make sense of it all. The idea of the Batting Index suggested here has another analogy with momentum. The Principle of Conservation of Momentum states:

In any system of bodies which act and react on each other, the total momentum remains constant.

When likened to the Batting Index, this means that each batsman has an average score and a strike rate figure that is a reflection of his true ability and that the *product* of these will not change. For example, if the match circumstances are such that it becomes necessary for a batsman to score faster than he otherwise normally would, his increase in risk of being dismissed is in proportion to the decrease in his likely score.

For example, suppose that a batsman has a batting average of 40 and a strike rate of 60 to yield a Batting Index of $40 \times 60 = 2400$. Then, if he is forced to score with a strike rate of 80, his expected score decreases to $2400/80 = 30$ to account for the higher risk he must take.

There are many other ways in which the worth of a bowler and batsman can be measured and the techniques suggested in this paper are really just a starting point for discussion as to how this might be done in an effective but easy to understand manner that is simple to explain. For any such index to have credibility, however, it must be based on a minimum number of innings played or wickets taken since otherwise new players may look far better than they actually are before reality sets in!

References

[1] G. A. Christos, "It's just not cricket", in *Fourth Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar (editors), Bond University, Queensland, Australia (1998), 181–188.

[2] G. L. Cohen, "One-day cricket: inferences from bowlers' strike rates", *Math. Today*, **35** (1999), 45–47.

[3] G. L. Cohen, "One-day cricket: the effect of running out an opposing batsman", *Math. Today*, to appear.

[4] J. S. Croucher, "Anglo-Australian cricket dismissals 1946–1980", *Bull. Appl. Statist.*, **9**(2) (1982), 179–193.

[5] S. M. Crowe and J. Middledorp, "A comparison of leg before wicket rates between Australians and their visiting teams for Test cricket series played in Australia 1977–1994", *The Statist.*, **45** (1996), 255–262.

[6] B. M. de Silva and T. B. Swartz, "Winning the coin toss and the home team advantage in one-day international cricket matches", *New Zealand Statist.*, **32** (1997), 16–22.

[7] F. Duckworth and T. Lewis, "A fair method for resetting the target in interrupted one-day cricket matches", in *Third Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1996), 51–68. (See also *J. Oper. Res. Soc.*, **49**(3) (1998), 220–227.)

[8] S. Innes and S. Sugden, "One day cricket scoring programs", in *Fourth Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar (editors), Bond University, Queensland, Australia (1998), 45–50.

[9] M. Johnson, S. Clarke and D. Noble, "An analysis of scoring policies in one day cricket", in *First Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1992), 71–79.

[10] A. C. Kimber and A. R. Hansford, "A statistical analysis of batting in cricket", *J. Roy. Statist. Soc. Ser. A*, **156**(3) (1993), 443–445.

[11] K. Kumar, "Is cricket really by chance?", in *Third Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1996), 69–76.

[12] M. Nemeruck and K. Meher-Homji, "Those mighty all-rounders", *World of Cricket Monthly* (July 1976), 40.

[13] I. Skinner, "A geometrical argument defining an order relation for all-round performances in cricket", in *Fourth Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar (editors), Bond University, Queensland, Australia (1998), 33–44.

THE FLIGHT OF A CRICKET BALL

Neville de Mestre
School of Information Technology
Bond University
Gold Coast
Queensland 4229, Australia

Graeme Cohen
Department of Mathematical Sciences
University of Technology, Sydney
PO Box 123, Broadway
NSW 2007, Australia

Abstract

This article describes the flight of a cricket ball when it is struck so well by the batsman that it flies through the air without obstruction from the centre of the field over the boundary. That is, the ball is hit for six runs, similar to a home run in baseball. Quadratic drag is assumed and the path is determined analytically to the point where parametric equations involving integrals are obtained; the integrals must be evaluated numerically.

1 Introduction

This article concentrates on the flight of a cricket ball when it is struck so well by the batsman that it flies through the air without obstruction from the centre of the field over the boundary. That is, the ball is hit for six runs, similar to a home run in baseball. By assuming a rope boundary, rather than a fence, we can make use of the mathematical simplification that the trajectory of the ball just passes over the rope at a height equal to the height it had when struck by the bat. A similar mathematical analysis can be used for bigger "sixes" or for the trajectory of any cricket hit to the first bounce.

The essential difference between a hit by a cricket bat and that by a baseball bat is that the baseball bat can impart a great deal of spin to the ball, whereas a full-blooded hit in cricket requires the bowled ball to rebound from the flat middle section of the bat with relatively little rotation imparted to the ball compared with its translational speed.

The dynamics of a hit for six in cricket are therefore governed by the forces of gravity and aerodynamic drag. Coutis [1] considered the drag force to vary linearly with speed, because this was easier to handle mathematically than the more realistic quadratic dependency on speed, yet was more sophisticated than zero drag. Although the quadratic drag law can of course be handled by a numerical Runge–Kutta approach, it is possible to solve the resulting differential equations analytically to a certain extent, and it is the purpose of this paper to present the details of that approach.

In essence, the equations can be re-arranged to produce a Bernoulli-type first-order differential equation. This equation was first formulated by Jacob Bernoulli (1654–1705) and solved by Gottfried Leibniz (1646–1716) for a projectile problem in exterior ballistics.

2 The parabolic path — a first approximation

The simplest model of the flight of a cricket ball after it has been struck by the bat is one of zero drag in which the only force acting is gravity. Then, by Newton's second law of motion, the horizontal and vertical components of the acceleration of the ball satisfy the equations

$$m\ddot{x} = 0 \qquad \text{and} \qquad m\ddot{y} = -mg, \tag{1}$$

respectively, where m is the mass of the ball (taken as 0.16 kg) and g is the acceleration due to gravity (9.81 ms^{-2} in Sydney). The dots denote differentiation with respect to time t, and the origin of the coordinate system is taken to be the point where the bat strikes the ball.

This is a system of differential equations. The usual initial conditions assume that the ball is struck with speed u, at a launch angle θ to the horizontal. Then, when $t = 0$, we have $x = 0$, $\dot{x} = u\cos\theta$, $y = 0$ and $\dot{y} = u\sin\theta$.

From (1) by integration, we obtain $x = At + B$ and $y = -\frac{1}{2}gt^2 + Ct + D$, where A, B, C and D are constants, and then, from the initial conditions, we find that $B = 0$, $A = u\cos\theta$, $D = 0$ and $C = u\sin\theta$. Putting this together,

$$x = ut\cos\theta \qquad \text{and} \qquad y = -\tfrac{1}{2}gt^2 + ut\sin\theta,$$

which are parametric equations for a parabola. The rectangular form can be obtained by eliminating t, yielding

$$y = x\tan\theta - \frac{gx^2\sec^2\theta}{2u^2}.$$

The graph of an instance of this inverted parabolic trajectory is shown later in Figure 2. There are three conclusions that we wish to state, arising from this model of the path of a struck cricket ball.

(i) It is symmetrical in time and distance about the point corresponding to the apex of the trajectory.

(ii) Setting $y = 0$, we find $x = 0$ (where the ball started) or $x = R$, called the range of the stroke (where the ball first lands). We have

$$R = \frac{2u^2\tan\theta}{g\sec^2\theta} = \frac{u^2\sin 2\theta}{g}.$$

(iii) The range is therefore greatest when $\sin 2\theta = 1$, that is, when $\theta = 45°$.

3 The effect of drag

For a cricket ball moving through air however, gravity is not the only force acting, and the aerodynamic effect has to be included. As indicated earlier, a struck cricket ball will probably not have excessive spin, and so lift due to spin will not be included in this analysis. On the other hand, the drag force due to passage through the air must be included, and it is well-known that this is a nonlinear effect. Coutis [1] modelled the drag $\boldsymbol{D}$ linearly because it allowed tractable mathematics and still indicated the general tendencies. We shall adopt the more realistic model with

$$\boldsymbol{D} = -\frac{1}{2}\rho A v^2 C_D \widehat{\boldsymbol{v}},$$

where ρ is the density of the air, A is the area of cross-section of the ball, v is its speed, C_D is the drag coefficient and $\widehat{\boldsymbol{v}}$ is a unit vector in the direction of the ball's velocity at any instant. Although in reality the air density varies with height above the ground, the ball is not exactly spherical and the drag coefficient varies with speed v, a suitable model for a cricket ball in flight is to consider the combination of these in the drag formula as approximately constant. Thus we will model the drag force as

$$\boldsymbol{D} = -Kv^2\widehat{\boldsymbol{v}},$$

maintaining the nonlinear effect through the speed only. For a well-hit cricket ball, the speed of the air particles near the ball will probably start off in the turbulent-drag region of flow where $C_D \approx 0.2$ and will then change very early along the trajectory path to the laminar-drag region of flow where $C_D \approx 0.45$. These produce values of K of 0.0005 and 0.0011, respectively, and so the single value of K

used in this paper will be $0.001\,\mathrm{kg\,m^{-1}}$. Consequently, for a big hit where the initial speed off the bat is $40\,\mathrm{ms^{-1}}$, the ratio of the drag force to the gravitational force (Kv^2/mg) is 1.02, that is, the drag of the air in the early part of the trajectory has as much effect on the ball as its weight.

The relevant equations in rectangular cartesian coordinates are therefore

$$\begin{aligned} m\ddot{x} &= -K\dot{x}\sqrt{\dot{x}^2+\dot{y}^2}, \\ m\ddot{y} &= -mg - K\dot{y}\sqrt{\dot{x}^2+\dot{y}^2}, \end{aligned} \tag{2}$$

with the same initial conditions as before, namely $x = 0$, $\dot{x} = u\cos\theta$, $y = 0$ and $\dot{y} = u\sin\theta$ when $t = 0$. These equations may be solved numerically using a Runge–Kutta approach, but they may also be partially solved analytically, as we now illustrate.

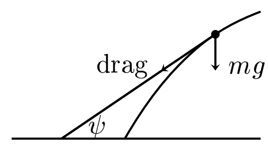

Figure 1: Forces on a cricket ball in flight.

We recast the equations in the tangential and normal directions (see de Mestre [2]), as opposed to the more conventional horizontal and vertical directions. This yields

$$\begin{aligned} m\dot{v} &= -mg\sin\psi - Kv^2, \\ mv\dot{\psi} &= -mg\cos\psi, \end{aligned} \tag{3}$$

where ψ is the angle between the direction of the velocity of the cricket ball (tangent to the trajectory) and the horizontal at any instant (see Figure 1). The initial conditions are $v = u$ and $\psi = \theta$ when $t = 0$, and the relationship to the original coordinate system is given by

$$\dot{x} = v\cos\psi \qquad \text{and} \qquad \dot{y} = v\sin\psi, \tag{4}$$

with $x = 0$ and $y = 0$ when $t = 0$.

Now the time can be eliminated from the system of equations (3) by division to yield

$$\frac{1}{v}\frac{dv}{d\psi} = \tan\psi + kv^2\sec\psi,$$

where $k = K/(mg) \approx 0.00064$. This can be rearranged as

$$\frac{dv}{d\psi} - v\tan\psi = kv^3\sec\psi$$

and is then recognisable as a Bernoulli-type first-order differential equation. This is solved by dividing both sides by v^3, and using the transformation $z = 1/v^2$ to change the equation to a first-order linear differential equation in z. An integrating factor $\sec^2\psi$ produces the following solution, which incorporates the boundary condition $v = u$ when $\psi = \theta$:

$$\frac{\sec^2\psi}{v^2} = \frac{\sec^2\theta}{u^2} + k(f(\theta) - f(\psi)), \tag{5}$$

where

$$f(\xi) = \sec\xi\tan\xi + \ln|\sec\xi + \tan\xi|.$$

This can easily be rearranged to give $v = v(\psi)$, relating the speed v at any point on the trajectory to the intrinsic angle ψ of the tangent to the trajectory at that point. In particular, the speed at the apex of the trajectory is obtained by putting $\psi = 0$.

The time of flight from $t = 0$ to any instant t may be obtained directly from the second of the equations (3) by first recasting it as

$$\frac{dt}{d\psi} = -\frac{v \sec \psi}{g}. \tag{6}$$

Thus

$$t = \frac{1}{g} \int_{\psi}^{\theta} v(\xi) \sec \xi \, d\xi. \tag{7}$$

The corresponding values of x and y are obtained from the equations (4), using (6), as

$$x = \frac{1}{g} \int_{\psi}^{\theta} v^2(\xi) \, d\xi, \tag{8}$$

$$y = \frac{1}{g} \int_{\psi}^{\theta} v^2(\xi) \tan \xi \, d\xi. \tag{9}$$

4 Applications to a cricket ball

Consider a cricket ball struck so that its initial speed off the bat is $u = 40$ ms^{-1} (or 144 kilometres per hour, not much less than the maximum speed attained by any bowler), at a launch angle of $\theta = 45°$. The graph of the trajectory, plotted from the parametric equations (8) and (9) by *Mathematica*, is shown in Figure 2. On the same axes, the no-drag parabolic trajectory is also shown. The diagram is similar to that obtained by Coutis [1]. This is because Coutis chose a linear expression for his drag function, essentially by replacing $\sqrt{\dot{x}^2 + \dot{y}^2}$ in our equations (2) by the constant u (or $1.1u$, in effect). Consequently his model overestimates the effect of drag (as he suggested it might). Coutis predicts that when $u = 40\,\text{ms}^{-1}$ the range will be about 90 m whereas the quadratic-drag model predicts 95.4 m.

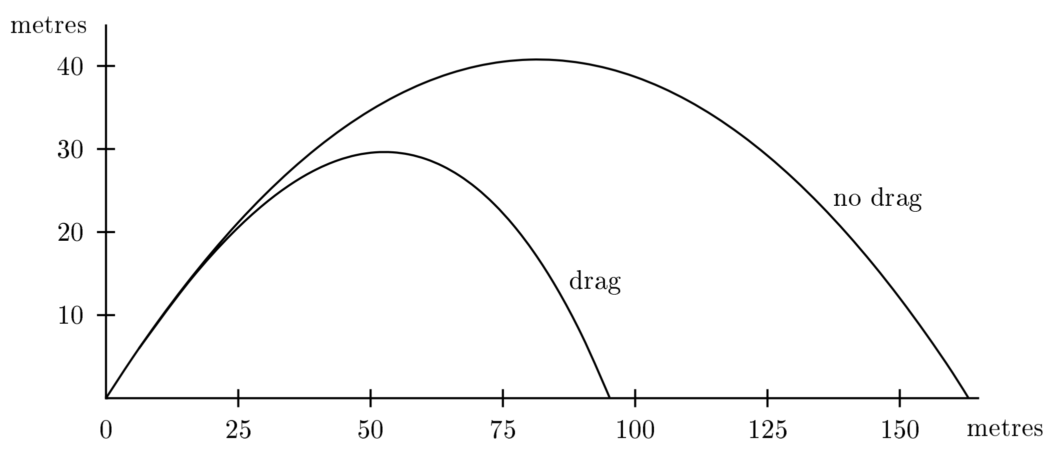

Figure 2: The trajectories for a projectile launched with speed $40\,\text{ms}^{-1}$ at an angle 45° using a no-drag model and a quadratic-drag model.

We turn our attention now to the three items highlighted in Section 2.

(i) When drag is taken into account, the asymmetry of the trajectory is evident. In Figure 2, the cricket ball traverses a longer horizontal distance on the way up (just over 51 m) than it does on the way down (44 m). With no air resistance, the batsman might expect to hit the ball approximately 163 m to the first bounce, about 68 m more than predicted by our model.

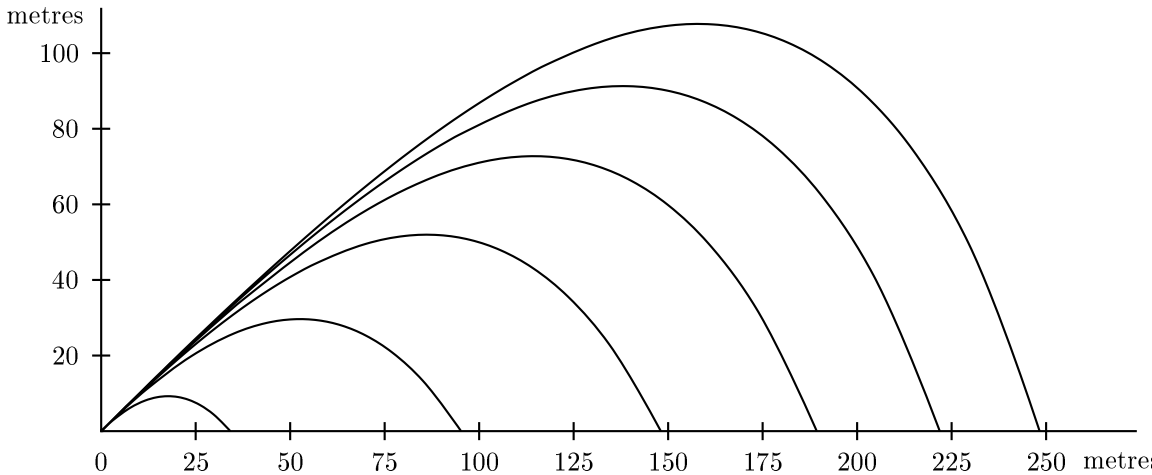

Figure 3: The trajectories for a launch angle of 45° and initial speeds of 20, 40, 60, 80, 100 and $120\,\mathrm{ms}^{-1}$, respectively, from the left.

(ii) Just as for the no-drag model, we may put $y = 0$ in equation (9) to estimate the range R of the batsman's hit for various values of u and θ. That is, we have to solve the equation

$$\int_{\psi}^{\theta} v^2(\xi) \tan \xi \, d\xi = 0$$

for ψ, with $v(\xi)$ given by equation (5). One solution is clearly $\psi = \theta$ (the take-off point), but there is another at a negative value of ψ (the impact angle), which we may evaluate using *Mathematica*, for example. That value of ψ is then inserted into equation (8) and the value of R determined numerically. That has been done, in effect, in Figure 3 which shows the path of the "cricket ball" for a launch angle of 45° and launch speeds of 20, 40, 60, 80, 100 and $120\,\mathrm{ms}^{-1}$. Only the first three speeds are realistic, but the others have been included for comparison with the corresponding figure by Coutis to show that, when quadratic drag is applied to the model of a cricket ball's motion, there is no indication that the range of the stroke is bounded as a function of the speed with which the ball leaves the bat. This contradicts the assertion by Coutis, and we believe that his conclusion is a product of the incorrect use of the linear model for drag. The better quadratic model predicts a range of 249 m when $u = 120\,\mathrm{ms}^{-1}$, and we even calculated a range well over 500 m when $u = 1000\,\mathrm{ms}^{-1}$—but no cricketer can hit that hard. In this latter hypothetical situation, the quadratic-drag model predicts that the ball travels almost along the 45° straight line from impact for about 300 m before it curves sharply and plummets toward the ground. The trajectory is very similar in shape to that of a normal golf ball driven with lift which just counteracts the gravitational pull early in the flight path.

(iii) Figure 4 shows the path for a launch speed of $42\,\mathrm{ms}^{-1}$ and launch angles 10°, 20°, ... , 70°. Again the curves obtained should be compared with those of Coutis [1]. The angle of projection for maximum range when $u = 42\,\mathrm{ms}^{-1}$ appears to be near 40°. When a Runge–Kutta numerical technique is applied to equations (2), the predicted maximum range is 102.1 m at 40.5° with very little variation in this range over the sector 39° to 42°.

We end with an interesting use of the first of the equations (3). From it, we may determine the terminal speed of the projectile. This is reached when $\dot{v} = 0$, and in the limiting case of vertical free-

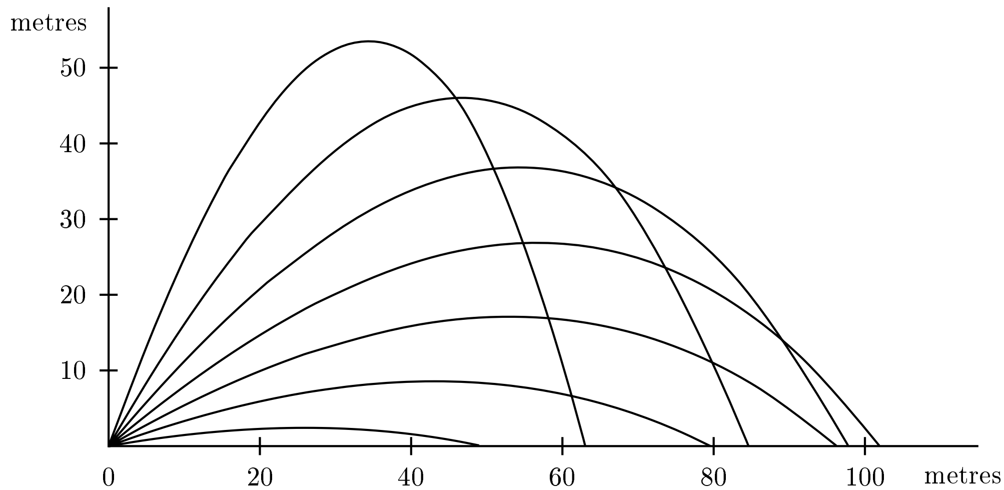

Figure 4: The trajectories for a projectile launched with initial speed $42\,\mathrm{ms}^{-1}$ and initial angles $10°$, $20°$, ... $70°$, respectively.

fall, when $\psi = -\pi/2$. Then, by that equation, the terminal speed will be $\sqrt{mg/K}$, or approximately $37.9\,\mathrm{ms}^{-1}$ for a cricket ball.

References

[1] P. Coutis, "Modelling the projectile motion of a cricket ball", *Int. J. Math. Educ. Sci. Technol.*, **29** (1998), 789–798.

[2] N. de Mestre, *The Mathematics of Projectiles in Sport*, Australian Mathematical Society Lecture Series No. 6, Cambridge University Press (1990).

APPLICATIONS OF THE DUCKWORTH–LEWIS METHOD

Basil M. de Silva
Department of Statistics and Operations Research
RMIT University
GPO Box 2476V, Melbourne
Victoria, Australia 3001

Greg R. Pond
Section of Biostatistics
200 First Street SW
Mayo Clinic
Rochester, MN 55905 USA

Tim B. Swartz
Department of Mathematics and Statistics
Simon Fraser University
Burnaby BC
Canada V5A1S6

Abstract

The Duckworth–Lewis method is steadily becoming the standard approach for resetting targets in interrupted one-day cricket matches. In this note we show that the Duckworth–Lewis resource table can be used to quantify the magnitude of a victory in one-day matches. This simple and direct application is particularly useful in breaking ties in tournament standings and in quantifying team strength.

1 Introduction

There are four possible outcomes in one-day cricket matches:

(i) the team batting first can win in a non-abandoned match,

(ii) the match can end in a tie,

(iii) the match can be abandoned,

(iv) the team batting second can win in a non-abandoned match.

In the first case, the run differential between the two teams is a sensible measure of the magnitude of victory. In the second case, which is rare in practice, the tie itself indicates that there is zero magnitude of victory. In the third case, either the game is declared null or one of the teams is declared the winner. In the latter event, a projected score is determined for the team batting second, and again, magnitude of victory can be assessed by calculating the run differential. However, in the fourth case, the magnitude of victory is unclear because the match terminates as soon as the team batting second scores more runs than the team batting first even though the team batting second may have leftover wickets and overs.

Why should we care about quantifying the magnitude of victory? Without such quantification, statistical analyses are typically based on binary data corresponding solely to wins and losses. For example, de Silva and Swartz [4] estimate the effect of the home team advantage in one-day international cricket matches using logistic regression. It is a generally accepted statistical principle that data is valuable and that one should not "waste" data by needless summarisation. Therefore, by quantifying the magnitude of victory in one-day cricket matches, future statistical modelling can better utilise the information contained in matches. Quantifying the magnitude of victory may aid in assessing team strength, determining betting strategies, breaking ties in tournament standings, etc.

How can we quantify the magnitude of victory? The Duckworth–Lewis method [5, 6, 7] is steadily becoming the standard approach for resetting targets in interrupted one-day cricket matches. At the time of writing, the Duckworth–Lewis method has been adopted for various competitions by the Zimbabwe Cricket Union, the England and Wales Cricket Board, New Zealand Cricket, and most notably, the International Cricket Council. In Section 2, we show that a simple application of the Duckworth–Lewis resource table can be used to quantify the magnitude of victory in one-day cricket matches. This approach and variations of it have also been considered by Allsopp and Clarke [1, 2, 3]. In Section 3, we provide an example whereby a three-way tie could have sensibly been broken using our approach. In Section 4, a second example is given which models the strength of ICC (International Cricket Council) nations in one-day international cricket matches.

2 Quantifying the magnitude of victory

The Duckworth–Lewis resource table (see Table 1) was devised to improve "fairness" in interrupted one-day matches. The resource table is based on the principle that resources are diminished in a shortened match and that targets should be reset according to the resources available. Duckworth and Lewis [5, 6] obtained the entries in the resource table using statistical methods based on historical match data. For a brief introduction to the Duckworth–Lewis method, see the CricInfo website (http://www.cricket.org).

	Wickets lost									
Overs left	0	1	2	3	4	5	6	7	8	9
50	100.0	92.4	83.8	73.8	62.4	49.5	37.6	26.5	16.4	7.6
40	90.3	84.5	77.6	69.4	59.8	48.3	37.3	26.4	16.4	7.6
30	77.1	73.1	68.2	62.3	54.9	45.7	36.2	26.2	16.4	7.6
20	58.9	56.7	54.0	50.6	46.1	40.0	33.2	25.2	16.3	7.6
19	56.8	54.8	52.2	49.0	44.8	39.1	32.7	24.9	16.2	7.6
17	52.3	50.6	48.5	45.8	42.2	37.2	31.5	24.4	16.1	7.6
16	49.9	48.4	46.5	44.0	40.7	36.1	30.8	24.1	16.1	7.6
10	34.1	33.4	32.5	31.4	29.8	27.5	24.6	20.6	14.9	7.5
5	18.4	18.2	17.9	17.6	17.1	16.4	15.5	14.0	11.5	7.0
1	3.9	3.9	3.9	3.9	3.9	3.8	3.8	3.7	3.5	3.1
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 1: Abbreviated version of the Duckworth–Lewis resource table. The table entries indicate the percentage of resources remaining in a match with the specified number of wickets lost and overs available.

Our problem, as stated in Section 1, is to quantify the magnitude of victory when the team batting second wins in a non-abandoned match. Using the Duckworth–Lewis resource table, this is a straightforward task. We determine the resource percentage used U relative to a standard 50-over match. We then solve $UE/100 = R$ for E where E is the number of effective runs and R is the actual number of runs scored. Here, the actual number of runs is simply the proportion of the effective number of runs in accordance with the resources used. The effective number of runs is the same as the *new projected score* as described by Allsopp and Clarke [3]. With effective runs, an effective run differential can then be calculated to quantify the magnitude of victory.

Consider then the most complicated scenario possible where the batting team starts and stops batting n times. When they start batting on the ith occasion they have a_i resources available according to Table 1, and similarly, when they stop batting they have b_i resources available. Therefore, the team has used $U = \sum_{i=1}^{n}(a_i - b_i)$ resources relative to a standard 50-over match. For example, consider a

50-over match where ten overs are played and one wicket is lost. A rain delay occurs and the inning is shortened to 20 remaining overs which are then played out. In this case, $a_1 = 100.0$, $b_1 = 84.5$, $a_2 = 56.7$ and $b_2 = 0.0$.

The simplicity of the approach is part of its appeal. Further, we note that in any shortened match, we may want to scale the number of runs as described above so that they are comparable to a standard 50-over match.

3 An example concerning ties

We refer to the Asia Cup played in Sharjah, UAE, in April 1995. This tournament featured India, Sri Lanka, Pakistan and Bangladesh in 50-over matches with the top two teams advancing to the championship final. After the initial round, Bangladesh had no wins and three losses. The remaining teams each had two wins and one loss. Under the three-way tie, India and Sri Lanka advanced to the finals based on superior run rates.

Now it is widely accepted that run rates can be unfair. Suppose then that we had used the methods of Section 2 to break the three-way tie. The relevant details of the initial round matches involving India, Sri Lanka and Pakistan are given in Table 2. In match 1, Pakistan has a differential of 97 runs over India. In match 2, India had lost only two wickets in 33 overs plus one out of six balls when it exceeded Sri Lanka's run total. Therefore, in a 50-over match, India had 16.83 of its overs left. Interpolating from Table 1, India had 48.17% of its resources remaining. Therefore, we calculate India's effective runs E by solving $(100 - 48.17)E/100 = 206$. This gives India 397 effective runs and an effective run differential of $397 - 202 = 195$ over Sri Lanka. A similar calculation in match 3 gives Sri Lanka an effective run differential of 118 over Pakistan.

Date	*Team 1* (Runs | Wickets | Overs)	*Team 2* (Runs | Wickets | Overs)
7 April	Pakistan (266 | 9 | 50)*	India (169 | 10 | 42.4)
9 April	Sri Lanka (202 | 9 | 50)	India (206 | 2 | 33.1)*
11 April	Pakistan (178 | 9 | 50)	Sri Lanka (180 | 5 | 30.5)*

Table 2: Summary of the relevant initial round matches involving India, Sri Lanka and Pakistan in the 1995 Asia Cup as described in Section 3. We let Team 1 denote the team batting first and Team 2 denote the team batting second. An asterisk denotes the winner of a match.

Putting these results together, India has $195 - 97 = 98$ net runs, Sri Lanka has $118 - 195 = -77$ net runs and Pakistan has $97 - 118 = -21$ net runs. Therefore, our approach would have advanced India and Pakistan to the championship final rather than India and Sri Lanka. We note that the same conclusion is reached when the matches involving Bangladesh are included in the calculation.

4 An example which models team strength

The data used in this analysis are the results of full 50-over one-day international matches involving the nine nations of the ICC. We consider all matches in the 1990s up to and including the final of the 1999 World Cup of Cricket, held in England. There are 623 such matches, the results of which are available from the "Archive" link at the CricInfo website.

As in Section 3, an effective run differential is calculated for every match. For matches in which the team batting first wins, the effective run differential is simply the actual run differential. For matches in which the team batting second wins, the effective run differential is calculated using the actual runs of the team batting first and the effective runs of the team batting second obtained via the Duckworth–Lewis resource table. We let indices $i, j = 1, \ldots, 9$ correspond to the nine ICC nations and let $k = 0, \ldots, 9$ correspond to the site of a match where $k = 0$ refers to a neutral site. We then consider the model

$$d_{ijk} = \tau_i - \tau_j + \gamma_{ijk} + \epsilon_{ijk} \tag{1}$$

where the response variable d_{ijk} is the effective run differential (i.e. team i minus team j) for a match at site k, τ_i is a measure of strength for the ith team such that $\sum_{i=1}^{9} \tau_i = 0$, γ_{ijk} is the home field advantage such that

$$\gamma_{ijk} = \begin{cases} \gamma, & \text{if } k = i, \\ -\gamma, & \text{if } k = j, \\ 0, & \text{otherwise,} \end{cases}$$

and the ϵ_{ijk} are independent and identically distributed Normal$(0, \sigma^2)$ errors. This 10-parameter model is taken over all 623 matches. It is sensible in that the determination of team strength takes into account not only victories and losses, but also the magnitude of the victories and losses, the strength of the opponent and the site of the match. Allsopp and Clarke [3] consider a similar model in the analysis of one-day Australian domestic competition results taken from 1994 through 1999.

To give more emphasis to recent matches, we consider a weighted least squares approach where the weight 1 is assigned to every match in 1999, the weight $\frac{9}{10}$ is assigned to every match in 1998, the weight $\frac{8}{10}$ is assigned to every match in 1997, etc. In S-PLUS, the function glm (Venables and Ripley [9]) is used to estimate the model parameters. The results are given in Table 3. We observe an ordering for the ICC nations where South Africa is the strongest team and Zimbabwe is the weakest. These estimated parameters can be used to forecast the outcomes of matches. For example, should South Africa play Zimbabwe in Johannesburg, we would expect South Africa to win by $27.4-(-37.6)+12.5 = 77.5$ effective runs. We also observe that the home field advantage $\gamma = 12.5$ is meaningful in one-day international cricket matches. Note that the error $\sigma = 52.4$ is somewhat large; this highlights the variability in cricket matches, where unlike sports such as rugby, a considerably weaker team has a realistic chance at upsetting a stronger team.

Parameter	*Wins (matches)*	Est (Std Err)
Australia (τ_1)	114 (184)	17.1 (4.4)
England (τ_2)	31 (74)	−2.1 (6.8)
India (τ_3)	78 (166)	−3.7 (4.6)
New Zealand (τ_4)	54 (142)	−14.3 (5.2)
Pakistan (τ_5)	104 (195)	5.5 (4.3)
South Africa (τ_6)	90 (137)	27.4 (5.0)
Sri Lanka (τ_7)	67 (149)	−4.0 (4.9)
West Indies (τ_8)	57 (116)	11.7 (5.9)
Zimbabwe (τ_9)	18 (83)	−37.6 (6.2)
home field (γ)		12.5 (3.5)
standard deviation (σ)		52.4

Table 3: Summary of the data and the estimated parameters as described in Section 4.

The results in Table 3 are particularly useful for betting purposes. Consider a match between Australia and New Zealand in Auckland. Using model (1) and a continuity correction, the estimated probability of a win by Australia is

$$\begin{aligned} \text{Prob}(d_{144} > 0.5) &\approx \text{Prob}(17.1 - (-14.3) - 12.5 + \epsilon > 0.5) \\ &\approx \text{Prob}(Z > -18.4/52.4) \\ &\approx 0.64 \end{aligned}$$

where Z is a standard normal random variable.

We remark that different weightings have been considered with little effect on the results. Also, we have considered different parametrisations for the home field advantage. For example, some very popular teams (e.g. the West Indies) may actually experience a home team advantage at a neutral site. Again, such changes do not greatly effect the results. Various residual plots indicate that the fitted model is adequate. More extensive discussion of the model and the analysis is found in Pond [8].

References

[1] P. Allsopp and S. R. Clarke, "Measuring the performance of teams in one-day cricket", in *Proceedings of the Mathematics 2000 Festival*, M. Bulmer, B. McCrae and K. Stacey (editors), University of Melbourne, Australian Association of Mathematics Teachers (2000), 1–8.

[2] P. Allsopp and S. R. Clarke, "Objective measures of performance in the World Cup of Cricket", in *Proceedings of the Western Decision Sciences Institute Conference*, Hawaii (2000).

[3] P. Allsopp and S. R. Clarke, "Methods for quantifying performances in one-day cricket", in these *Proceedings* (2000).

[4] B. M. de Silva and T. B. Swartz, "Winning the coin toss and the home team advantage in one-day international cricket matches", *New Zealand Statist.*, **32** (1997), 16–22.

[5] F. C. Duckworth and A. J. Lewis, "A fair method for resetting the target in one-day cricket matches", in *Third Conference on Mathematics and Computers in Sport*, N. de Mestre (editor), Bond University, Queensland, Australia (1996), 51–68.

[6] F. C. Duckworth and A. J. Lewis, "A fair method for resetting the target in one-day cricket matches", *J. Oper. Res. Soc.*, **49** (1998), 220–227.

[7] A. J. Lewis and F. C. Duckworth, "Developments in the Duckworth-Lewis (D/L) method of target resetting in one-day cricket matches", in *Fourth Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar (editors), Bond University, Queensland, Australia (1998), 131–151.

[8] G. R. Pond, *An Analysis of British Columbia Sports Lotteries*, unpublished MSc project, Simon Fraser University, Department of Mathematics and Statistics (1999).

[9] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S-PLUS*, second edition, Springer-Verlag, New York (1994).

CURVATURE EFFECTS IN RUNNING THE 200 AND 400 METRE SPRINTS

Chris Harman
Department of Mathematics and Computing
University of Southern Queensland
Toowoomba
Queensland 4350, Australia

Abstract

A model is presented here for finding the likely effect of curvature on expected running times for the 200 and 400 metre sprints. The model is developed using data from the IAAF World Championships in Athletics held in Seville in 1999. The data chosen for analysis consists of the 50 metre split times for the world champions, Maurice Greene (200 metres) and Michael Johnson (400 metres). Allocation of the eight lanes influences the amount of curvature endured by the runners. This paper extends previous constant speed models by taking into account the speed profile from start to finish.

1 Introduction

There have been many studies of the processes and mechanics of running in a straight line. See for example Fuchs [2], Pritchard [6] and Ward-Smith [7]. Apart from a paper on baseball baserunning by the author (Harman [4]), models for running on a curve only seem to have been successfully devised for circular arcs with constant speed (for example, Greene [3]). Greene's model was supported by experimental evidence and compares well with other models (Behncke [1]). Note that since there are two Greene's involved in this paper, the champion sprinter will always be referred to as Maurice Greene and the biomechanical author will be referred to as just Greene.

Constant speed models have limited applicability. They ignore the important acceleration phase and the considerable variation in speeds in events such as the 400 metre sprint. In this paper, speed against distance sprint profiles are established and modified for curvature effects. The resulting predicted running times are then contrasted between the extreme lanes 1 and 8 for both the 200 and 400 metre events.

2 The IAAF world championship data

At the recent World Championships in Seville, accurate split times were recorded at 50 metre intervals.

In the 200 metres final, the subsequent winner Maurice Greene recorded the following times:

distance (metres)	50	100	150	200
time (seconds)	5.74	10.25	14.94	19.90

and in the 400 metres final, the champion Michael Johnson recorded:

distance (metres)	50	100	150	200	250	300	350	400
time (seconds)	6.14	11.10	16.10	21.22	26.42	31.66	37.18	43.18

It is a pity that more detailed measurements are not taken of the split times, but nevertheless, it will be shown here that it is still possible to interpolate in a way to achieve a speed/distance profile. This then enables the run times to be evaluated for non-uniform speeds and for differing curvature effects depending on lane choice

3 The running track

The standard running track for the 400 metres consists of two 80 metre straights and two semicircular sections with inside track circumference being 120 metres. There are eight lanes. The inside lane runner starts at the beginning of the semicircle (at the end of the straight), runs anti-clockwise, and finishes at the same point after running 400 metres. The lane width is about 1.3 metres and so runners in the outer seven lanes start at varying distances around the semicircle in order to make the total distance 400 metres. For example the runner in lane eight starts from about 57 metres ahead around the semicircle. All runners finish at the end of the second 80 metre straight.

For the 200 metre event, the inside lane starts at the same place at the start of the 120 metre semicircle and finishes at the end of the 80 metre straight. The runner in the outside eighth lane starts about 28.6 metres forward around the semicircle.

Runners in the inside lanes expend more energy in overcoming the greater curvature. There are arguments that favour middle lanes for strategic or psychological reasons, but this paper is only concerned with the mechanics.

4 Michael Johnson's 400 metres sprint speed profile

It has been shown (Pritchard [6]) that an excellent model for a top sprinter's speed-time profile for 100 metres is given by

$$v(t) = P\tau(1 - e^{-t/\tau}),$$

where P is the maximum force per unit mass that the runner can exert and τ is a constant determined by internal resistances in the runner. From this formulation, it can be seen that $P\tau$ is the maximum speed of the runner. Experimental data indicate that τ is usually close to 1. However for sprint runs over a distance of 200 or 400 metres, the speed profile is much more complicated. The approach to be taken in this paper is to use Pritchard's model up to 50 metres and then to use the subsequent 50 metre split times from the 1999 World Championships to build a speed against distance profile for the entire race. In order to contrast the lane curvature effects, the profile will then be moderated accordingly. In order to do this, a variation on Greene's constant-speed curvature adjustment will be used.

The Johnson data are first of all interpolated with a cubic spline. This results in the interpolation function

$$t = f(x), \quad 50 \leq x \leq 400.$$

Then it follows that the velocity v is given by

$$v = \frac{dx}{dt} = \frac{1}{f'(x)}, \quad 50 \leq x \leq 400.$$

For the first 50 metres, it is not possible to infer anything much from the Seville data. So the approach to be taken over this initial distance is to take $\tau = 1$ and to use Pritchard's model for a sprinter with the value of P chosen in some sensible fashion. Now for $0 \leq x \leq 50$, it follows that

$$v = P(1 - e^{-t}).$$

But for $50 \leq x \leq 400$ we have the velocity profile

$$v = \frac{1}{f'(x)}$$

found from the cubic spline interpolation. In fact, the numerical value of the spline for v at $x = 50$ was 10.049 and matching with Pritchard's formulation with the recorded time $t = 6.14$ at $x = 50$, gives the equation $10.049 = P(1 - e^{-6.14})$ and so $P = 10.071$.

Integrating the equation

$$v = P(1 - e^{-t}),$$

gives

$$x = P(t - 1 + e^{-t}),$$

which defines a parametric relationship for v and x for values of x between 0 and 50 metres. A sample of these values of v is then combined with x, v data at the measured x values from Johnson's race and the combined set is then interpolated with a cubic spline F to give the velocity profile $v = F(x)$. Figure 1 shows the result. The circled points up to 50 metres are from the adaptation of Pritchard's sprint model and the remaining points marked with $*$'s are from the interpolated data.

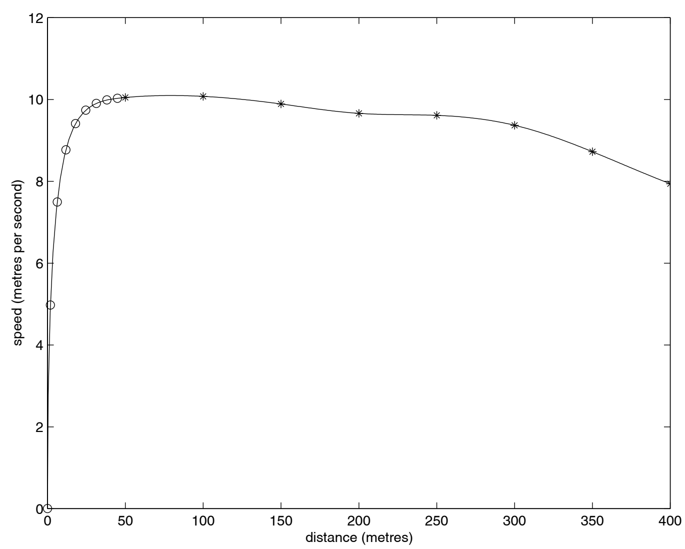

Figure 1: Michael Johnson's velocity profile.

The established velocity profile $v = F(x)$ for Michael Johnson will be used in the following way. First of all the profile will be assumed to represent a typical base running speed profile in the absence of curvature influences. Of course the profile would be slightly different if he were to run in a straight line, but we do not have access to that data. The profile is then used as a basis for velocities on lanes 1 and 8, but with allowances made for the different curvature effects. The running times can then be calculated and compared for significant differences.

In the absence of curvature, the running time T for the 400 metres can be calculated by

$$T = \int_0^{400} \frac{1}{v}\, dx = \int_0^{400} \frac{1}{F(x)}\, dx,$$

where $F(x)$ is cubic spline interpolation of the Michael Johnson velocity profile from Seville, as shown in Figure 1. As a check on the numerical accuracy of the model, this was calculated to give the resulting running time of 43.3 which compares well with the actual measured time of 43.18. This small difference can be expected, since we imposed the Pritchard profile for the first 50 metres. With confidence in the model, we now take the profile $v = F(x)$ as our base and incorporate curvature effects in order to compare running times for lanes 1 and 8.

5 Curvature effects on the velocity profile

Greene [3] modelled the mechanical effects of a sprinter running around a flat circular turn of radius R at constant speed. His model took account of the sprinter's straight line top speed v, foot contact time, ballistic air time, step length, and stride time. A reciprocal Froude number, or dimensionless radius Rg/v^2, enabled him to compare the theory against experiment for a large number of individuals on the same set of axes. The influence of radius of the turn on subsequent velocity was predicted and tested. The agreement between theory and practice was good and was verified for a range of radii between about 4 and 28 yards. More complex models have since been formulated but Behncke [1] considers that "the simplicity of Green's [sic] result and the apparent ease of its derivation make it ... an ideal candidate for the analysis of the track and field situation".

Greene's model requires the solution $V(R, v)$ of the equation

$$V^6 + (Rg)^2V^2 - (Rgv)^2 = 0,$$

where R is the radius of curvature, v is the runner's natural top speed in the absence of curvature, and g is the acceleration due to gravity. Then V is the resultant natural top speed adjusted by the curvature effect. This equation is a cubic in V^2 and so the solution can easily be found to be

$$V = \sqrt{\frac{\left(27R^2g^2v^2 + \sqrt{108R^6g^6 + 729R^4g^4v^4}\right)^{\frac{1}{3}}}{54^{\frac{1}{3}}} - \frac{2^{\frac{1}{3}}R^2g^2}{\left(27R^2g^2v^2 + \sqrt{108R^6g^6 + 729R^4g^4v^4}\right)^{\frac{1}{3}}}}. \qquad (1)$$

For our purposes, at any point x in the run, the value of v can be replaced by the value found from the velocity profile $F(x)$. This then gives a runner's speed V, adjusted if necessary by the curvature influence.

6 The time equation

As detailed above, the 400 metres consists of four sections: curve, straight, curve, straight. Let T_1, T_2, T_3, and T_4 represent the respective running times on the sections and let T denote the race time given by $T = T_1 + T_2 + T_3 + T_4$. For the lane 1 runner, the section times are given by

$$\begin{aligned} T_1 &= \int_0^{120} \frac{1}{V}\,dx = \int_0^{120} \frac{1}{V(v(x))}\,dx, \\ T_2 &= \int_{120}^{200} \frac{1}{v}\,dx = \int_{120}^{200} \frac{1}{F(x)}\,dx, \\ T_3 &= \int_{200}^{320} \frac{1}{V}\,dx = \int_{200}^{320} \frac{1}{V(v(x))}\,dx, \\ T_4 &= \int_{320}^{400} \frac{1}{v}\,dx = \int_{320}^{400} \frac{1}{F(x)}\,dx, \end{aligned}$$

where $v = F(x)$ is Michael Johnson's basic velocity profile shown in Figure 1 and where $V(v(x))$ is the curvature adjusted velocity profile using Greene's method given by (1) above, with radius of curvature of the inside lane given by $R = 120/\pi \doteq 38.2$ metres.

For the lane 8 runner these equations are

$$
\begin{aligned}
T_1 &= \int_0^{91.41} \frac{1}{V}\,dx = \int_0^{91.41} \frac{1}{V(v(x))}\,dx,\\
T_2 &= \int_{91.41}^{171.41} \frac{1}{v}\,dx = \int_{91.41}^{171.41} \frac{1}{F(x)}\,dx,\\
T_3 &= \int_{171.41}^{320} \frac{1}{V}\,dx = \int_{171.41}^{320} \frac{1}{V(v(x))}\,dx,\\
T_4 &= \int_{320}^{400} \frac{1}{v}\,dx = \int_{320}^{400} \frac{1}{F(x)}\,dx,
\end{aligned}
$$

where the radius of curvature of the eighth lane is $R = 120/\pi + 7 \times 1.3 \doteq 47.3$ metres.

7 Results and conclusions

For the 400 metre sprint, using Michael Johnson's velocity profile and calculating the resultant influence on times of curvature as detailed above, the following results were obtained:

Lane 1	Time T (with curvature)	44.01 sec
	Time (ignoring curvature)	43.33 sec
Lane 8	Time T (with curvature)	43.77 sec
	Time (ignoring curvature)	43.32 sec

In the table, the time values (ignoring curvature) are given as an indication of the influence of track curvature on a given lane. The significant result however is the time difference between the lanes 1 and 8 when curvature is taken into account. The time difference of $44.01 - 43.77 = 0.23$ seconds, is a difference at the finish of about 2.5 metres. This could make the difference in an Olympic final between a gold medal and an unplaced result.

The same techniques were applied to Maurice Greene's 200 metre velocity profile from Seville and the following results were obtained:

Lane 1	Time T (with curvature)	20.34 sec
	Time (ignoring curvature)	19.90 sec
Lane 8	Time T (with curvature)	20.20 sec
	Time (ignoring curvature)	19.90 sec

So the time difference for the 200 metres between lanes 1 and 8 amounts to $20.34 - 20.20 = 0.14$ seconds. This would be about 1.4 metres at the finish.

These differences can be contrasted with the results of Jain [5] who extended some empirical results to estimate that the time differences for the 200 and 400 metre events run on an oval track should be 0.07 and 0.14 seconds respectively. Greene [3] used constant speed curvature estimates to predict the time difference for the 200 metres on an Olympic track to be about 0.123 seconds. The results of the present paper indicate that the higher values of the time differences, as given above, are likely.

References

[1] H. Behncke, "Small effects in running", *J. Appl. Mech.*, **10** (1994), 270–290.

[2] P. M. Fuchs, "General mathematical concepts of compensation in sport science with quantitative analysis in the case of sprinting performance", *Math. Methods Appl. Sci.*, **18** (1995), 473–505.

[3] P. R. Greene, "Running on flat turns: Experiments, theory, and applications", *Trans. ASME J. Biomech. Eng.*, **107** (May 1985), 96–103.

[4] C. Harman, "'Who's on first!' 'What?' 'What's on second!' and how 'What' got there on an optimal baserunning path", in *Fourth Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar (editors), Bond University, Queensland, Australia (1998), 217–223.

[5] P. C. Jain, "On a discrepancy in track races", *Res. Quart. for Exercise and Sport*, **51** (1980), 432–436.

[6] W. G. Pritchard, "Mathematical models of running", *SIAM Rev.*, **35**(3) (September 1993), 359–379.

[7] A. J. Ward-Smith, "A mathematical theory of running based on the first law of thermodynamics, and its application to the performance of world class athletes", *J. Biomech.*, **18** (1985), 337–349.

SIMULATION OF FRISBEE FLIGHT

M. Hubbard and S. A. Hummel *
Department of Mechanical and Aeronautical Engineering
University of California, Davis
CA 95616, USA

Abstract

Flight equations of motion of the frisbee are presented. Few reports of aerodynamic force and moment coefficients are available in the literature. Aerodynamic coefficients are estimated using parameter identification by matching predicted and experimental trajectories of markers on the disc. RMS errors as small as 1.4 mm are achieved between predicted and measured marker positions in flights of about 2 m. Longer flights are calculated but performance is very sensitive to the aerodynamic coefficients.

1 Introduction

The frisbee was invented in 1948. Since then it has enjoyed remarkable popularity. Today it is used by millions as a recreational toy and by thousands in the sport of Ultimate Frisbee. Despite its popularity there is relatively little scientific and technical information in the literature documenting the dynamics and aerodynamics of the implement. The book by Johnson [6] is a practitioners' handbook which includes information on the history of the frisbee, frisbee games and throwing techniques. Two brief, somewhat anecdotal, descriptions of the flight dynamics and aerodynamics for the informed layman have been provided by Bloomfield [2] and Schuurmans [12].

A study of a self suspended frisbee-like flare was conducted by Stilley [14]. This work included spinning and non-spinning wind tunnel tests, computer simulations of flight, and experimental flight tests of three disc configurations. A main conclusion was that the effect of spin on the aerodynamic forces is small. Stilley hypothesised, but did not measure, a "Magnus rolling moment", induced by the interaction of spin and velocity. Stability criteria were presented involving relations between the stability derivatives. Mitchell [9] measured the lift and drag forces on three different frisbees over a range of speeds and angles of attack. He confirmed the observation of Stilley [14] that spin affects lift and drag only little, but did not measure pitching moment.

Recently, Potts and Crowther [11] studied the aerodynamics more completely. They accurately measured lift and drag forces and pitching and rolling moment as a function of Reynolds number and spin parameter, the ratio of speed at the edge of the disc due to spin to the speed of the centre. They corroborated the results of Stilley and Carstens [15] at zero spin. Slight differences in lift force and pitching moment were attributed to the different cross-sectional shapes and thickness ratios of the frisbees tested from those of Stilley and Carstens. Potts and Crowther found little effect of Reynolds number on lift and drag except at high angles of attack ($\alpha \geq 20°$). Although they detected virtually no effect of spin parameter on lift and drag, nonzero rolling moments and small but distinct effects of spin parameter on pitching and rolling moments were observed.

In related work, Soong [13] analysed the dynamics of the discus throw, whose shape differs from that of the frisbee mainly in that it has a plane as well as an axis of symmetry. Soong did not include

*This paper was partially researched and prepared while participating in an Ultimate Frisbee tournament in Hawaii. The author is grateful for the hospitality shown by all.

any rolling aerodynamic moments and showed that the main effect of pitching moment is to cause a precession of the spin in roll, thus decreasing lift later in the flight.

Frohlich [4] also investigated the flight dynamics of the discus. With computer simulations that relied on discus lift and drag coefficients determined experimentally by Ganslen [5] and others, he showed that the discus can be thrown farther against the wind than with it. Although this paper included a complete discussion of the effects of aerodynamic moments on the subsequent flight, the analysis remains somewhat hypothetical because of the absence of any experimental data on pitching and rolling aerodynamic moments. Frohlich also noted the remarkable similarity between the discus and the frisbee.

The present paper presents a three-dimensional mathematical model of frisbee flight including the translational and rotational dynamics of the disc driven by the aerodynamic forces and moments caused by the motion. The aerodynamic coefficients used in the model are estimated using experimental flight data. These are also compared to values of similar coefficients from the literature.

2 Flight dynamics model

The frisbee is assumed to be an axially symmetric rigid body with mass m and axial and diametral mass moments of inertia I_a and I_d, respectively. Several reference frames are used in the development of the equations. The frisbee centre of mass (cm) is assumed to be located at coordinates xyz in an inertial reference frame N with the xy plane horizontal and the direction of the x axis chosen (arbitrarily) so that the initial (or release) velocity vector $\boldsymbol{v}$ of the centre of mass lies near the xz plane.

A body-fixed reference frame C (see Figure 1) has its origin at the mass centre and its $\boldsymbol{c_1c_2}$ plane parallel to the plane of the disc. The general orientation of the frisbee (and of the C frame) in the N frame is achieved through three successive Euler angle rotations (a 123 Euler angle set). As shown in Figure 1, first the N frame is rotated about its $\boldsymbol{n_1}$ axis through angle ϕ to reach the intermediate A frame. Next the A frame is rotated about the $\boldsymbol{a_2}$ axis through angle θ to reach a second intermediate frame B. Finally the B frame is rotated about its $\boldsymbol{b_3}$ axis through angle γ to reach the body-fixed C frame. In the derivation of the equations of motion below we often express vectors in the B frame. The three Euler angles ϕ, θ and γ determine the general orientation of the disc. Only ϕ and θ are involved in the transformation matrix T relating the N and B frames,

$$T = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ \sin\theta\sin\phi & \cos\phi & -\cos\theta\sin\phi \\ -\sin\theta\cos\phi & \sin\phi & \cos\theta\cos\phi \end{bmatrix}. \tag{1}$$

The angular velocity of the disc relative to the inertial N frame is given by

$$\boldsymbol{\omega} = \dot{\phi}\cos\theta\,\boldsymbol{b_1} + \dot{\theta}\,\boldsymbol{b_2} + (\dot{\phi}\sin\theta + \dot{\gamma})\,\boldsymbol{b_3}.$$

In the general flight configuration, the velocity of the center of mass is written as

$$\boldsymbol{v} = u\,\boldsymbol{b_1} + v\,\boldsymbol{b_2} + w\,\boldsymbol{b_3}, \tag{2}$$

consistent with the typical procedure followed in analyses of flight vehicles (Etkin and Reid [3]).

The equations of motion, the details of which are summarised in the Appendix, are a Newton–Euler description of the motion of a rigid body. These are expressed in the B frame since, because of symmetry, B is (as well as C) a set of principal axes of inertia for the disc. The equations in vector form are given by

$$\boldsymbol{F} = m\left(\frac{d\boldsymbol{v}}{dt} + \boldsymbol{\omega_b}\times\boldsymbol{v}\right), \tag{3}$$

$$\boldsymbol{M} = I\frac{d\boldsymbol{\omega}}{dt} + \boldsymbol{\omega_b}\times I\boldsymbol{\omega}, \tag{4}$$

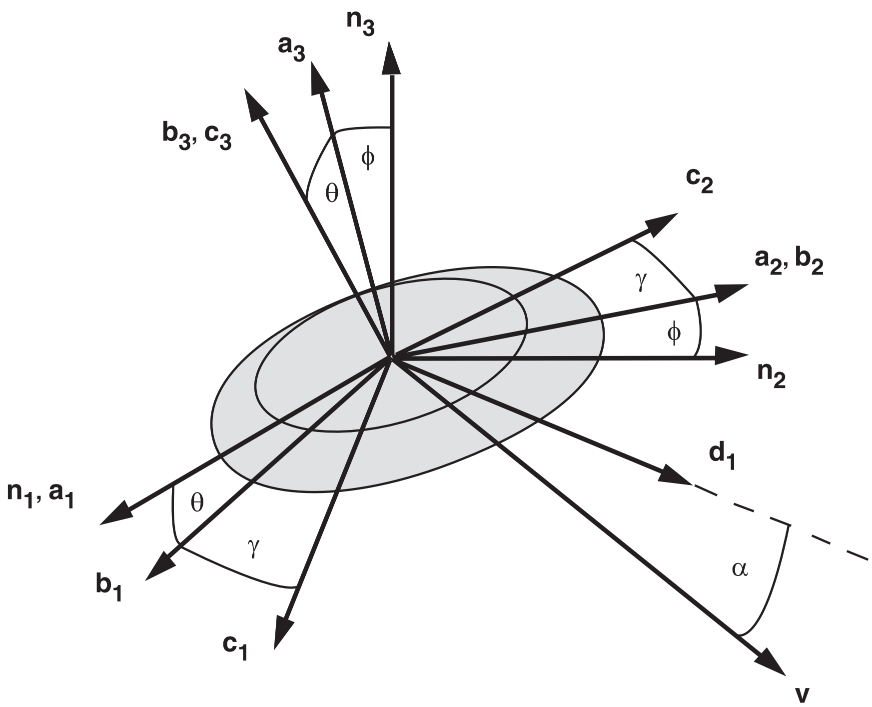

Figure 1: A frisbee body-fixed reference frame is attained with a 123 Euler angle set of rotations.

where $\boldsymbol{F}$ and $\boldsymbol{M}$ are the aerodynamic and gravitational forces and moments applied to the disc, respectively, $\boldsymbol{\omega_b}$is the angular velocity of the B frame relative to N, and I is the moment of inertia matrix

$$I = \begin{bmatrix} I_d & 0 & 0 \\ 0 & I_d & 0 \\ 0 & 0 & I_a \end{bmatrix}.$$

3 Aerodynamic forces and moments

3.1 Forces

The natural directions for decomposition of the aerodynamic forces are determined by the velocity (see Figure 2). Let the unit vector along the projection of $\boldsymbol{v}$ on the $\boldsymbol{b_1 b_2}$ disc plane be denoted by $\boldsymbol{d_1}$. Assuming $\boldsymbol{d_3}$ $(= \boldsymbol{b_3})$ is perpendicular to the plane of the disc, we let $\boldsymbol{d_2} = \boldsymbol{d_3} \times \boldsymbol{d_1}$. The angle of attack α is defined as the angle between $\boldsymbol{v}$ and $\boldsymbol{d_1}$. We assume that the two components of the aerodynamic force along and perpendicular to $\boldsymbol{v}$ in the $\boldsymbol{d_1 d_3}$ plane (called the lift L and drag D, respectively) are each functions of lift and drag coefficients in the standard way:

$$L = C_l A \rho v^2/2, \qquad D = C_d A \rho v^2/2,$$

where here v denotes the magnitude of $\boldsymbol{v}$ (not to be confused with the $\boldsymbol{b_2}$ component of $\boldsymbol{v}$ in equation (2)), ρ is the air density, and A is the projected or planform area of the disc. The lift and drag coefficients C_l and C_d are further assumed (for small angles of attack) to be given by the expressions

$$C_l = C_{l_o} + C_{l_\alpha} \alpha, \qquad C_d = C_{d_o} + C_{d_\alpha} \alpha^2.$$

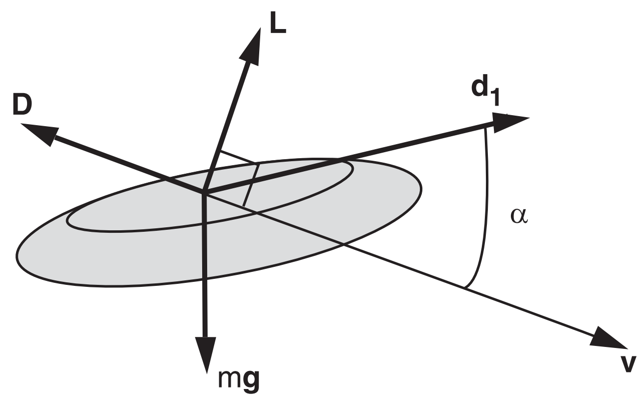

Figure 2: Drag and lift forces act along and perpendicular to the velocity vector $\boldsymbol{v}$ and are functions of α, the angle between $\boldsymbol{v}$ and $\boldsymbol{d_1}$ in the disc plane.

The forms of the relations for L and D above are consistent with the general findings of several researchers that spin does not affect L and D to first order (McShane *et al.* [8], Stilley [14], Mitchell [9]). Note that the lift is linear in α while drag dependence is quadratic. Although this form is exact for a symmetric airfoil, the formulation of Potts and Crowther [11] is probably more accurate in that it predicts the minimum drag to occur at the zero-lift angle of attack, which they measure to be about $-4°$.

It is important to remember that the four coefficients above are chosen (in contrast to using measured forces) for purposes of simplicity and efficiency. This description provides a more practical version of the dependence of the forces on angle of attack, at some cost in accuracy.

Were the lateral projected area of the frisbee (in the $\boldsymbol{d_1 d_3}$ plane) larger compared to its platform area, it would be important to include a third $\boldsymbol{d_2}$ component of the aerodynamic force. For example, in the flight of spherical balls (Alaways and Hubbard [1]) the Robins–Magnus lift force produced by the circulation due to spin about the $\boldsymbol{d_3}$ axis is important, but this force is neglected here. Thus the four constant coefficients C_{l_o}, C_{l_α}, C_{d_o}, C_{d_α} are assumed here to be a simplified but complete parametric description of the force-producing capability of the disc. Decomposing the lift and drag into the D frame gives an expression for the aerodynamic force

$$\boldsymbol{F} = (L \sin\alpha - D\cos\alpha)\,\boldsymbol{d_1} + (L\cos\alpha + D\sin\alpha)\,\boldsymbol{d_3}. \tag{5}$$

3.2 Moments

The net aerodynamic force acts at the centre of pressure (cp) which, in general, does not coincide with the centre of mass. Thus it also exerts a moment about the cm. One of the components is a pitching moment about the $\boldsymbol{d_2}$ axis of the form

$$M_2 = Ad\rho v^2 (C_{M_\alpha}\alpha + C_{M_{\dot\alpha}}\dot\alpha)/2,$$

where the angular stiffness and damping coefficients C_{M_α} and $C_{M_{\dot\alpha}}$, respectively, are constants and d is the diameter of the disc.

For an axisymmetric non-spinning disc, the cp would lie in the $\boldsymbol{d_1 d_3}$ plane. However, the nonzero spin causes the position of the cp to have a small $\boldsymbol{d_2}$ component as well. The coefficient $C_{R_{\dot\gamma}}$ models this effect (called the "rolling Magnus moment" by Stilley [14]) and assumes that a part of the $\boldsymbol{d_1}$ component of the moment is proportional to the spin $\dot\gamma$. A physical explanation of this coefficient is the small effect the spin has on the rotation of the mean flow separation line away from the $\boldsymbol{d_2}$ direction. Its magnitude is typically very small. Potts and Crowther [11] have shown that this coefficient is significantly different from zero only at small angles of attack $\alpha \leq 10°$. The previously introduced damping coefficient $C_{M_{\dot{\alpha}}}$ also relates the $\boldsymbol{d_1}$ component of disc angular velocity to the other part of the moment in the $\boldsymbol{d_1}$ direction yielding

$$M_1 = Ad\rho v^2 (C_{R_{\dot{\gamma}}} \dot{\gamma} + C_{M_{\dot{\alpha}}} \boldsymbol{\omega} \cdot \boldsymbol{d_1})/2.$$

Finally there is a spin-deceleration torque which acts along the $\boldsymbol{d_3}$ axis. This torque differs fundamentally from the $C_{M_{\dot{\alpha}}}$ term in M_2, for example, in that it is due to viscous shear stresses in the fluid rather than from a displacement of the cp from the cm. It is assumed to be proportional to the $\boldsymbol{d_3}$ component of the disc angular velocity $\dot{\phi} \sin\theta + \dot{\gamma}$ through the spin-down coefficient C_{N_r}:

$$M_3 = C_{N_r}(\dot{\phi} \sin\theta + \dot{\gamma}).$$

Justification for a functional dependence of C_{N_r} on α can be found in Stilley [14] (and Nielsen and Synge [10]) which is here neglected. Combining the three components of the aerodynamic moments above yields

$$\boldsymbol{M} = M_1\, \boldsymbol{d_1} + M_2\, \boldsymbol{d_2} + M_3\, \boldsymbol{d_3}. \tag{6}$$

In summary, the aerodynamic forces and moments are parametrised by the eight coefficients shown in the heading of Table 1.

3.3 Evaluation of coefficients

The experimental evidence in the literature for the values of these aerodynamic force and moment coefficients is scanty and incomplete. Stilley [14] contains wind tunnel results for lift, drag and pitching moment as a function of angle of attack for a single frisbee. Mitchell [9] measured lift and drag coefficients in the range of angles of attack $-20° \leq \alpha \leq 20°$ for three different frisbees in a wind tunnel over the range of speeds $2.7 \leq v \leq 28.2\,\mathrm{ms}^{-1}$. Mitchell also found that spin did not affect the lift and drag forces substantially, but he did not measure pitching moment and thus provided no information regarding the moment coefficients C_{M_α}, $C_{M_{\dot{\alpha}}}$, $C_{R_{\dot{\gamma}}}$ and C_{N_r}. The recent, more detailed work of Potts and Crowther [11] does give measurements of pitching and rolling moments. However, wind tunnel tests must be essentially quasistatic, and consequently no experimental evidence is available on the dependence of forces and moments on angular velocities.

In another approach, approximations of the aerodynamic coefficients may be obtained from flight data. Three small circular reflective markers were attached to the top of the frisbee and their three dimensional inertial coordinates tracked during flights using high speed (120 Hz) video. Estimates of the aerodynamic coefficients, together with initial conditions for the short experimental flight, were determined iteratively by modifying the aerodynamic coefficients and initial conditions in a simulation to minimise the differences between predicted and measured marker positions. Coefficients for two flights are shown in Table 1.

Flight	C_{l_o}	C_{l_α}	C_{d_o}	C_{d_α}	C_{M_α}	$C_{M_{\dot{\alpha}}}$ s	$C_{R_{\dot{\gamma}}}$ s	C_{N_r} N m s
fssh3	−0.40	1.89	0.83	0.83	−0.16	0.029	0.012	0.000074
bffl8	1.17	0.28	5.07	0.077	0.22	0.025	−0.0030	−0.00000025

Table 1: Aerodynamic force and moment coefficients for two flights.

Figure 3 shows the projections in the horizontal xy plane of the predicted and measured marker trajectories for flight fssh3. The disc makes almost one complete revolution and the marker trajectories superimpose nearly harmonic motion in x and y on the mean motion of the cm. Figure 4 shows the residuals, the differences between the measured and predicted positions of the three markers, arrayed

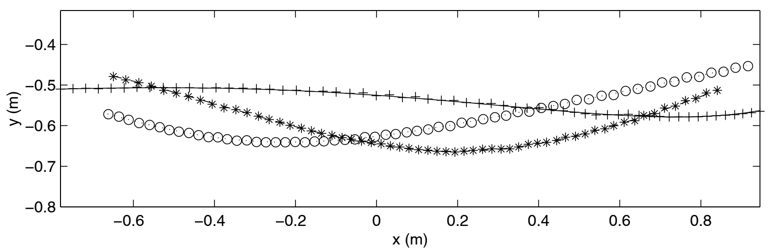

Figure 3: Predicted (lines) and measured (symbols) x and y positions for three markers in flight fssh3.

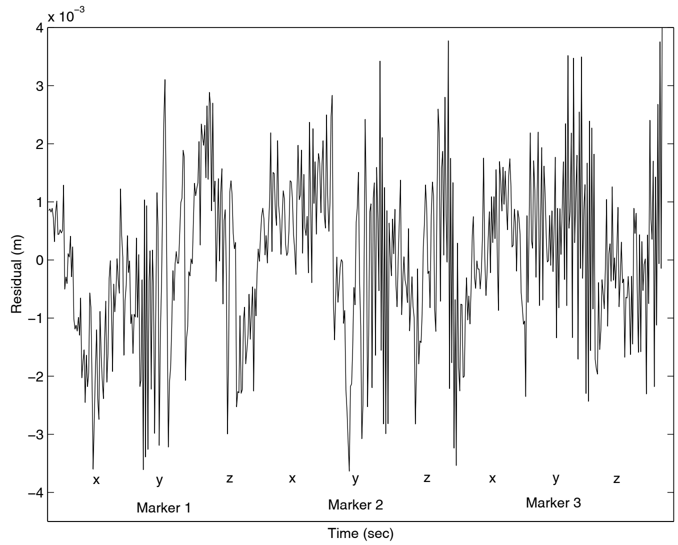

Figure 4: Prediction errors in x, y and z for three markers in flight fssh3 have an rms of only 1.4 mm.

end-to-end. Even though the root mean squared (rms) error in the prediction is only 1.4 mm over the course of the 2 m long flight and the residuals are nearly white, there is some obvious correlation present indicating that a more completely descriptive model of the flight might be possible, including possibly a more accurate drag parametrisation. The aerodynamic coefficients determined from the flight portrayed in Figures 3 and 4 are shown in the first row of Table 1. Coefficients from a second flight are shown in the second row. There exist frisbees of many sizes and cross-sectional shapes, and certainly with correspondingly different aerodynamic coefficients, but a single frisbee was used in all our flight experiments. Even though one would expect the coefficients from two flights of the same disc to be nearly equal, as seen in Table 1 this was not the case. It is curious that the more typical (in the sense discussed below) flight bffl8 yields unlikely and in some cases physically impossible coefficients, while the less typical flight fssh3 produces coefficients that are more similar to those found in the wind tunnel experiments discussed above.

This suggests a modified procedure for the identification of the aerodynamic coefficients in the future. Rather than identifying a different set of coefficients for each flight, a single set of coefficients may be determined which allow a best fit of the marker data from numerous flights, covering a wide range of initial conditions.

It is also worthwhile to note the sensitivity of the identified coefficients to noise in the experimentally measured marker positions. Small errors can require unrepresentative, or in some cases even physically

Parameter	m	I_a	I_d	d	A	ρ	g
	kg	$\mathrm{kg\,m^2}$	$\mathrm{kg\,m^2}$	m	$\mathrm{m^2}$	$\mathrm{kg\,m^{-3}}$	$\mathrm{ms^{-2}}$
	0.175	0.00197	0.00138	0.269	0.057	1.23	9.7935

Table 2: Values of non-aerodynamic simulation parameters.

Flight	x_o	y_o	z_o	u_o	v_o	w_o	ϕ_o	θ_o	γ_o	$\dot{\phi}_o$	$\dot{\theta}_o$	$\dot{\gamma}_o$
	m	m	m	$\mathrm{ms^{-1}}$	$\mathrm{ms^{-1}}$	$\mathrm{ms^{-1}}$	rad	rad	rad	$\mathrm{rad\,s^{-1}}$	$\mathrm{rad\,s^{-1}}$	$\mathrm{rad\,s^{-1}}$
fssh3	−0.71	−0.52	1.13	3.03	−0.45	−2.25	−0.066	−0.34	3.01	−3.03	−0.76	−8.91
bffl8	−1.42	−0.40	0.81	8.40	−0.15	−1.23	−0.081	−0.25	4.40	−6.33	0.37	−37.1

Table 3: Initial conditions for two flights.

impossible, values of coefficients to yield dynamics which will fit the data. This has apparently happened in the second of the flights in Table 1, which yielded a negative, but still extremely small, value for C_{N_r} implying that the spin actually increased very slightly during the flight rather than decreasing. Similar unreasonable estimates of negative drag coefficients have occurred in cases involving errors in measured x position that required the velocity to increase in flight. Thus an essential requirement for this technique of coefficient identification is very accurate marker position data. In addition, flight times are required that are long enough to make the effects of the coefficients apparent through the dynamics. In the simulations of two flights presented below we assume the aerodynamic force and moment coefficients are those given in the two rows of Table 1. Table 2 contains the non-aerodynamic parameters for the test frisbee which are also the parameters for the simulations discussed below. The mass, in particular, is relatively large compared to typical recreational discs, and is the official weight of discs used in Ultimate Frisbee competition.

4 Simulation results

Shown in Figure 5 are time histories of eleven of the twelve state variables ($x\, y\, z\, u\, v\, w\, \phi\, \theta\, \dot{\phi}\, \dot{\theta}\, \dot{\gamma}$) for an atypical right hand, backhand-thrown flight of about 0.5 seconds. The initial conditions for this flight are shown in the first row of Table 3. This is a simulation of a flight which corresponds to that shown in Figures 3 and 4. The disc was thrown rather slowly, with a velocity of $3.77\,\mathrm{ms^{-1}}$, and with a relatively small spin ($8.91\,\mathrm{rad\,s^{-1}}$) and artificially large initial angle of attack. The disc was released 1.13 m from the ground and fell 1.04 m within 0.5 s. The small dynamic pressure (of the order of 10 Pa) makes the aerodynamic forces relatively insignificant compared to the weight and results in a nearly parabolic fall. During the flight it travelled 1.62 m in the x direction, curving slightly right (in y) of the mean direction and then back again to the left. Because of the large initial angle of attack, the component of velocity perpendicular to the disc plane, w, is larger than the vertical component of velocity $\dot{z}$. The reader should recall that the velocities shown in part b) of Figure 5 are components referred to the disc plane and are not the derivatives of x, y and z in part a).

The disc was initially oriented with its far (away from the thrower) edge tilted downward, $\phi = -0.065\,\mathrm{rad}$, and the leading edge tilted upward, $\theta = -0.34\,\mathrm{rad}$. These angles changed significantly over the course of the flight. However, the angular velocity time histories, $\dot{\phi}$ and $\dot{\theta}$, indicated some signs of damping. Because the duration of the flight was so small compared to the characteristic time of spin decay (about 7 s), the frisbee lost only about 10% of its initial spin rate of $-8.91\,\mathrm{rad\,s^{-1}}$.

Shown in Figure 6 are the same eleven state variables for a more typical flight than Figure 5 (i.e. larger initial velocity and spin rate and smaller initial angle of attack). A simulation of 1.0 s was created using the measured coefficients from flight bffl8 (see Table 2). The initial conditions for this flight are listed in Table 3. In this flight the dynamic pressure is more than four times that of the flight shown in Figure 5. As a result the characteristic times of response are correspondingly reduced and the mean aerodynamic

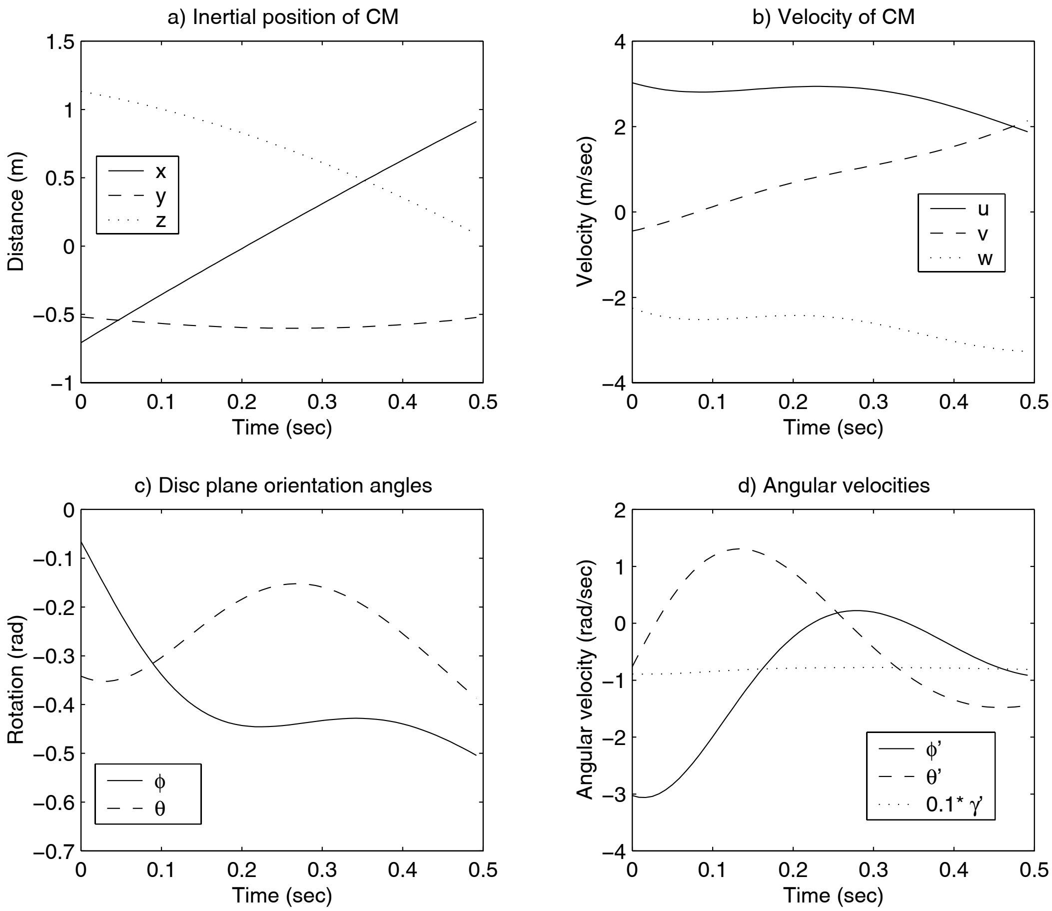

Figure 5: State variables versus time for flight fssh3 of Figs. 3 and 4 and aerodynamic parameters from row 1 of Table 1.

force more nearly equals the weight. The inertial x coordinate increased roughly at the thrown speed of $8.48\,\mathrm{ms}^{-1}$ and the frisbee travelled $6.71\,\mathrm{m}$. The inertial z coordinate increased initially but began to decrease after approximately $0.2\,\mathrm{s}$ and underwent a net drop of $0.84\,\mathrm{m}$. The y component of the disc mass centre remained relatively constant. The disc was initially spinning at $\dot{\gamma} = -37.1\,\mathrm{rad\,s}^{-1}$ but this decreased to $\dot{\gamma} = -35.4\,\mathrm{rad\,s}^{-1}$ over the duration of the flight. Initially the far edge of the disc was tilted down, $\phi = -0.081\,\mathrm{rad}$, and the leading edge was tilted up at $\theta = -0.25$. For about the first $0.3\,\mathrm{s}$ the angular velocities $\dot{\phi}$ and $\dot{\theta}$ underwent damped oscillations. Decreasing attitude oscillations such as these can frequently be observed with the naked eye. After the early oscillations subsided, ϕ then gradually increased through the flight to $\phi = 0.29\,\mathrm{rad}$. The leading edge continued to tilt up as θ decreased to $-0.21\,\mathrm{rad}$.

5 Conclusions

The equations of motion of frisbee flight have been presented. Numerical solutions of these equations can provide quantitative flight information but are sensitive to the coefficients relating body orientation and translational and angular speeds to the aerodynamic forces and moments. Simulations of this type can provide a means for creating sets of initial conditions which produce flights with certain desirable

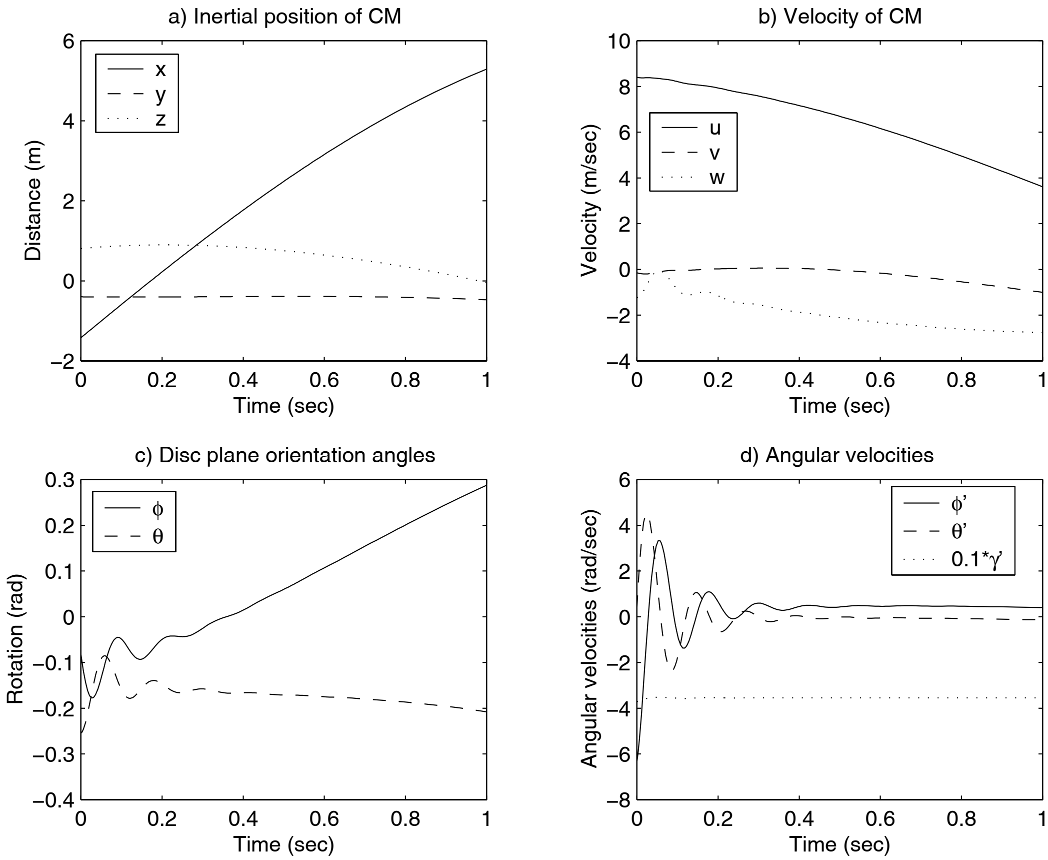

Figure 6: State variables versus time for extended flight bffl8.

characteristics. For example, in game situations one may use throws with sharp curvature in one direction or another since these may be advantageous from a tactical point of view. Alternatively a throw may be desired which hangs or flares over or near a given point on the field. Throws such as these may be investigated and "designed" using the simulation techniques described herein.

Further work is needed in making the estimates of the aerodynamic coefficients from flight data more reliable and meaningful and perhaps in creating more complex parametric descriptions of the aerodynamics. This will require considerably longer experimental flights and more accurate kinematic data acquisition techniques. In addition, the use of data from a wide variety of flight conditions will make the identified coefficients more representative of the dynamic behavior of the frisbee over a similarly broad flight regime.

Appendix Equations of motion

In the derivation of the equations of motion, the aerodynamic forces and moments originally expressed in frame D in equations (5) and (6) are transformed to the B frame by multiplying by the transformation matrix T_2 relating B to D:

$$\mathbf{F_b} = T_2\mathbf{F}, \qquad \mathbf{M_b} = T_2\mathbf{M},$$

where

$$T_2 = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and where β is the angle between the $\boldsymbol{b_1}$ and $\boldsymbol{d_1}$ axes:

$$\beta = \arctan(v/u).$$

Then the differentiations in equations (3) and (4) are carried out and the scalar components are solved for the accelerations yielding

$$\begin{aligned}
\dot{u} &= (F_{b_1} - w\dot{\theta} + v\dot{\phi}\sin\theta)/m, \\
\dot{v} &= (F_{b_2} + w\dot{\phi}\cos\theta - u\dot{\phi}\sin\theta)/m, \\
\dot{w} &= (F_{b_3} + u\dot{\theta} - v\dot{\phi}\cos\theta)/m, \\
\ddot{\phi} &= (M_{b_1} + 2I_d u\dot{\phi}\dot{\theta}\sin\theta - I_a\dot{\theta}(\dot{\phi}\sin\theta + \dot{\gamma})\cos\theta)/I_d\cos\theta, \\
\ddot{\theta} &= (M_{b_2} - I_d\dot{\phi}^2\cos\theta\sin\theta + I_a\dot{\phi}\cos\theta(\dot{\phi}\sin\theta + \dot{\gamma}))/I_d, \\
\ddot{\gamma} &= (M_{b_3} - I_a(\ddot{\phi}\sin\theta + \dot{\phi}\dot{\theta}\cos\theta)\cos\theta)/I_a.
\end{aligned}$$

Finally, the velocity vector $\boldsymbol{v}$ with components u, v and w in the B frame is transformed into the N frame by multiplying by the transformation matrix T from equation (1) to give the rates of change of the inertial coordinates:

$$\begin{aligned}
\dot{x} &= u\cos\theta + w\sin\theta, \\
\dot{y} &= u\sin\theta\sin\phi + v\cos\phi - w\cos\theta\sin\phi, \\
\dot{z} &= -u\sin\theta\cos\phi + v\sin\phi + w\cos\theta\cos\phi.
\end{aligned}$$

References

[1] L. W. Alaways and M. Hubbard, "Experimental determination of baseball spin and lift", submitted.

[2] L. A. Bloomfield, "The flight of the frisbee", *Sci. Amer.* (April 1999), 132.

[3] B. Etkin and L. D. Reid, *Dynamics of Flight: Stability and Control*, John Wiley & Sons (1996).

[4] C. Frohlich, "Aerodynamic effects on discus flight", *Amer. J. Phys.*, **49** (1981), 1125–1132.

[5] R. V. Ganslen, "Aerodynamic and mechanical forces in discus flight", *Athletic J.*, **44** (1964), 50.

[6] S. E. D. Johnson, *Frisbee: A Practitioner's Manual and Definitive Treatise*, Workman Publishing Company (1975).

[7] P. Katz, "The free flight of a rotating disc", *Israel J. Tech.*, **6**(1–2) (1968), 150–155.

[8] E. J. McShane, J. L. Kelly and F. V. Reno, *Exterior Ballistics*, The University of Denver Press (1953).

[9] T. L. Mitchell, *The Aerodynamic Response of Airborne Discs*, MS Thesis, University of Nevada, Las Vegas (1999).

[10] K. L. Nielsen and J. L. Synge, "On the motion of a spinning shell", *Quart. Appl. Math.*, **4** (1946), 201–226.

[11] J. R. Potts and W. J. Crowther, "The flow over a rotating disc-wing", in *RAeS Aerodynamics Research Conference Proceedings*, London, UK (April 2000).

[12] M. Schuurmans, "Flight of the frisbee", *New Scientist*, **127**(1727) (July 28, 1990), 37–40.

[13] T.-C. Soong, "The dynamics of discus throw", *J. Appl. Mech.*, **98** (1976), 531–536.

[14] G. D. Stilley, "Aerodynamic analysis of the self sustained flair", *RDTR no. 199*, Naval Ammunition Depot, Crane, Indiana (1972).

[15] G. D. Stilley and D. L. Carstens, "Adaption of frisbee flight principle to delivery of special ordnance", AIAA paper 72-982 (1972).

AN ANALYSIS OF THE AFL FINAL EIGHT SYSTEM

Jonathan Lowe and Stephen R. Clarke
School of Mathematical Sciences
Swinburne University
PO Box 218
Hawthorn
Victoria 3122, Australia

Abstract

An extensive analysis into the new final eight system employed by the AFL was undertaken using certain criteria as a benchmark. An Excel spreadsheet was set up to fully examine every possible outcome. It was found that the new system failed on a number of important criteria such as the probability of a premiership decreasing for lower ranked teams, and the most likely scenario of the grand final being the top two ranked sides. This makes the new system more unjust than the previous McIntyre Final Eight system.

1 Introduction

Recently, many debates have occurred over the finals system played in Australian Rules football. The Australian Football League (AFL), in response to public pressure, released a new finals system to replace the McIntyre Final Eight system. Despite a general acceptance of the system by the football clubs, a thorough statistical examination of this system is yet to be undertaken. It is the aim of this paper to examine the new system and to compare it to the previous McIntyre Final Eight system.

In 1931, the "Page Final Four" system was put into place for the AFL finals. As the number of teams in the competition grew, so to did the number of finalists. The "McIntyre Final Five" was introduced in 1972, and a system involving six teams was in place in 1991. This was changed to the "McIntyre Final Six" system the next year, and was changed yet again to the "McIntyre Final Eight" in 1994. This system has been used despite much controversy until the year 1999. In that year the Western Bulldogs and Carlton lost the first round of the finals, Carlton played West Coast Eagles in Melbourne and the Western Bulldogs played Brisbane at Brisbane in the second round of the finals. Hence, Carlton, who finished lower on the ladder, played a lower ranked team than the Western Bulldogs. Consequently, the authors received at least ten finals systems from the AFL to analyse. Christos [1] also developed several alternative models to the McIntyre and new system. However, the AFL delivered another different method for the year 2000. This new system uses an interesting combination of single knockout tournament systems and double knockout tournament systems where certain teams are eliminated after one loss or two losses depending on their ranking. However, under certain conditions, a lower ranked side may have a greater probability of winning the premiership than a higher ranked side throughout the tournament if teams are not re-seeded after each round [2, 7]. The McIntyre Final Eight system involved reseeding of teams after Round 1 of the finals, but the new system does not.

Monahan and Berger [11] established some criteria for determining the appropriateness of a fair playoff or finals system. They said that the system has to maximise the probability that the highest ranked team wins the premiership, and maximise the probability that the best two teams play in the grand final. Clarke [3] suggested further criteria for a good system: the probability that a team finishes in any position or higher should be greater than for any lower-ranked team; the expected final position should be in order of original ranking; the probability that a team finishes above a team of lower rank should be greater than 0.5, and should increase as the difference in rank increases; the probability of any two teams playing in the grand final should decrease as the sum of the ranks of those teams decreases. In addition, the organisation of a finals system may have additional requirements such as the number of matches, number of repeat games and closeness of matches.

Using the above criteria, Clarke analysed the McIntyre Final Eight system based on a equal probability of winning. It is assumed that each team has a 50% chance of winning, despite the opposition and the venue of the match. He also used a model based on past results, where victory in each match depended on the match participants and the venue. Other authors have used a variety of different methods to analyse the probability of teams winning matches. The use of paired comparisons in the analysis of round robin and knockout tournaments began with David [5], Kendall [8] and Maurice [9]. McGarry and Schutz [10] constructed a probability matrix of certain teams defeating other teams based on their ranking.

All these methods of analysis assume stationarity and independence within this playing matrix. That is, the probabilities of teams defeating other teams do not change as a function of time (stationarity) and do not change as a function of past events (independence). It is most unlikely that these assumptions are valid, as a team might build confidence following a victory or lose morale following a loss. Although transient probabilities have been considered in paired comparison tournaments [6], this study will assume stationarity and independence within the playing matrix for simplicity reasons. Clarke [4] found that in the years from 1980 to 1995, the average home ground advantage was 58%, and away teams were expected to win 42% of games.

This study will use both the equal probability method and home ground advantage as outlined by the following criteria.

- The probability of a team finishing in any position or higher should be greater than for any lower-ranked team.
- The expected final position should be in order of original ranking.
- The probability of a team finishing above a team of lower rank should be greater than 0.5, and should increase as the difference in rank increases.
- The probability of any two teams playing in the grand final should decrease as the ranks of those teams decrease.
- Systems with no repeat games excluding the grand final are preferable.
- No system should have a fatal flaw. A fatal flaw might include dead matches, where the outcome is inconsequential, or giving teams “unfair” advantages.

2 The new final eight system

In this system, the winners of the top four sides obtain the bye, whilst the losers play at home to the winners of the bottom four teams. The two losers of the bottom four sides are eliminated. After the first round where first, second, fifth and sixth placed sides obtain the home state advantage, only teams who finished the year in the top four get to play at home.

In short, the system is as follows:

Week 1	Game A	Team 1 v Team 4	Winners obtain the bye
	Game B	Team 2 v Team 3	Winners obtain the bye
	Game C	Team 5 v Team 8	Losers are eliminated
	Game D	Team 6 v Team 7	Losers are eliminated
Week 2	Game E	Loser Game A v Winner Game C	
	Game F	Loser Game B v Winner Game D	
Week 3	Game H	Winner Game A v Winner Game F	
	Game I	Winner Game B v Winner Game E	
Week 4	Game J	Winner Game H v Winner Game I	

All teams that are mentioned first receive a home ground advantage (if it exists) with the exception of the grand final (Game J) which is played at the MCG. There is no reseeding in the new model, unlike the McIntyre model that re-seeds after the first round. In the McIntyre system, the outcome of a game could involve one team obtaining a bye, and the other being eliminated; however, this doesn't occur in the new system. The following week's matches in the McIntyre system could not be established until all the matches were complete. The new system is very simple and clear as to what teams are competing in the next round after each match.

3 Performance of the new system using the equal probability model

The following tables were calculated using the equal probability model where it is assumed that each team has a 50% chance of winning the game. Clarke [3] gives similar tables for the McIntyre Final Eight system.

Team	*Current system*	*McIntyre Final Eight*
1	18.75	18.75
2	18.75	18.75
3	18.75	15.62
4	18.75	12.50
5	6.25	12.50
6	6.25	9.37
7	6.25	6.25
8	6.25	6.25

Table 1: Percentage chance of teams winning the premiership given equal probabilities.

Table 1 shows the probability of the premiership is the same for each team in the top four, as well as for teams 5 to 8, although teams 1, 2, 5 and 6 play at home. But what happens if there is no significant home ground advantage, that is, when two teams of the same state play each other? Given this, there may be situations where teams will deliberately lose in the last home and away round so as to play in their home state.

Also note that Table 2 shows that the most likely scenario for the grand final is not the best two teams. It is more likely that Team 2 will play Team 3 or Team 1 will play Team 4, than Team 1 vs Team 2. This means that this system does not give the two best teams throughout the year the greatest chance of meeting in the grand final. This is a major problem with the system. Table 3 shows that Team 1 is more likely to finish above Team 2 and Team 3 than it is Team 4. This is obviously because

Teams	2	3	4	5	6	7	8
1	7.8	7.8	12.5	3.1	1.6	1.6	3.1
2		12.5	7.8	1.6	3.1	3.1	1.6
3			7.8	1.6	3.1	3.1	1.6
4				3.1	1.6	1.6	3.1
5					1.6	1.6	0
6						0	1.6
7							1.6

Table 2: Percentage chance of teams playing other teams in the grand final given equal probabilities.

	Final position								
Team	1	2	3	4	5	6	7	8	EFP
1	18.8	18.8	31.3	6.3	25.0	0	0	0	3.00
2	18.8	18.8	23.4	14.1	18.8	6.3	0	0	3.14
3	18.8	18.8	23.4	14.1	18.8	6.3	0	0	3.14
4	18.8	18.8	15.6	21.9	12.5	12.5	0	0	3.28
5	6.3	6.3	3.1	9.4	12.5	12.5	50.0	0	5.53
6	6.3	6.3	1.6	10.9	6.3	18.8	25.0	25.0	5.86
7	6.3	6.3	1.6	10.9	6.3	18.8	25.0	25.0	5.86
8	6.3	6.3	0	12.5	0	25.0	0	50.0	6.19

Table 3: Percentage chance of teams finishing in certain positions with the Expected Final Position (EFP) using equal probability matches.

	Team j							
Team i	1	2	3	4	5	6	7	8
1	–	57.0	57.0	50.0	82.8	85.2	85.2	82.8
2	43.0	–	50.0	57.0	85.2	82.8	82.8	85.2
3	43.0	50.0	–	57.0	85.2	82.8	82.8	85.2
4	50.0	43.0	43.0	–	82.8	85.2	85.2	82.8
5	17.2	14.8	14.8	17.2	–	66.5	66.5	50.0
6	14.8	17.2	17.2	14.8	33.5	–	50.0	66.5
7	14.8	17.2	17.2	14.8	33.5	50.0	–	66.5
8	17.2	14.8	14.8	17.2	50.0	33.5	33.5	–

Table 4: Percentage chance of team i (row) finishing above team j (column) using equal probabilities.

Team 1 plays Team 4 in the opening round. Likewise, it is more likely to finish above Team 5 than it is Team 8.

Table 3 shows that the expected final position for the top four teams does not differ significantly, but is considerably greater than the expected finals position of the fifth and lower teams. This is also highlighted in the premiership probabilities with fourth being three times more likely to win the grand final than fifth.

The system has removed the possibility of playing repeated games by swapping the semi-finals. If all the favourites were to win leading up to the semi-final, Team 1 would then play Team 3, and Team 2 plays Team 4. This means that Team 2 has an easier game than Team 1. If repeat games were not undesirable, then it is understandable that Team 1 would play Team 4 and Team 2 play Team 3 in the semi-finals. Team 1 should be rewarded for finishing on top of the ladder by playing a less difficult team (Team 4 as opposed to Team 3). But in this case repeat games occur, and the AFL has chosen to "swap" the semi-finals, so that Team 1 plays a harder side than Team 2 so as not to get repeat games. In fact, it is quite possible for Team 2 to play Team 8 in the semi-final, and Team 1 to play Team 3, despite both teams winning the same number of games in the finals. This agrees with Chung [2] and Israel [7] who found that in tournaments that do not re-seed after each round, lower ranked sides could have a greater chance of winning the premiership, which is the case here. This shows an unjustness in the system.

4 Home ground advantage model

If the only difference between teams in the top four and teams in the bottom four is home ground advantage, what happens if there is no significant home ground advantage? No significant home advantage can occur in a number of different scenarios. For example, if Carlton play Geelong, this match would be played at either the MCG or Colonial stadium, giving neither team an advantage. Likewise Richmond can play Melbourne at the MCG, which is a neutral game. Adelaide might meet Brisbane in the grand final at the MCG which is also neutral. Given this, there may be situations where teams will deliberately lose in the last home and away round so as to play in their home state.

As mentioned earlier there is a large problem with home ground advantage. If the first and fourth teams share a home ground, then there is no greater advantage in finishing first than there is fourth. Also as highlighted by an unjust situation that occurred in 1997, Geelong could play away at the MCG versus Melbourne or Richmond, for example, despite finishing higher on the ladder. This problem would not occur in the previous McIntyre system as the top ranked team has a distinct advantage over the fourth ranked team irrespective of any home ground advantage.

In 1999, approximately 58% of matches were won at home. Clarke [4] found similar results to this in the years 1980–1995. Given this distinct advantage for playing a home game, one can analyse the new finals system and compare it to the previous finals system for certain teams. Considering the home grounds of the 16 teams in the AFL, they will be split up into five groups according to their home ground nature.

GROUP 1

Team	*Home Ground*
Brisbane	'Gabba
Sydney	SCG

GROUP 2

Team	*Home Ground*
Carlton	Optus Oval
Geelong	Kardinia Park

GROUP 3

Team	*Home Ground*
Adelaide	Football Park
Port Adelaide	Football Park
West Coast	Subiaco/WACA
Fremantle	Subiaco/WACA

GROUP 4

Team	*Home Ground*
Essendon	Colonial
St Kilda	Colonial
West Bulldogs	Colonial

GROUP 5

Team	*Home Ground*
Collingwood	MCG
Hawthorn	MCG
Kangaroos	MCG
Melbourne	MCG
Richmond	MCG

The reasons these groups have been set up are as follows:

- Group 1 teams do not share their ground with any other team in the league and are guaranteed a home ground advantage if they finish in a position which deems a home ground.
- Group 2 teams do not share their ground with any other team, so will not achieve a home ground advantage in any final. Moreover, they may be at a disadvantage if they play an MCG or Colonial based team.

- Group 3 teams share their ground with one other team, and Group 4 share theirs with two others, and will get a home ground advantage if they finish in a position which deems a home ground, unless they play a co-tenant.
- Group 5 teams share their ground with four other teams and will get a home ground advantage in the finals if they finish in a position which deems a home ground. They are also guaranteed a home grand final.

Quite obviously, Carlton and Geelong are at a disadvantage because they will receive no home ground advantage despite their final position at the end of the year. Whilst teams in Group 5 will have the advantage of playing on their home ground in the grand final if they are to make it, there would be no distinct advantage if they were playing someone else in Group 5. We investigate whether these advantages and disadvantages are greater or less in the new finals system as opposed to the McIntyre system.

Premiership odds for each group given their final position and taking home advantage into account can be calculated. For example, if Adelaide were to finish fifth, then their first game would be at their home ground. However, there is a $\frac{1}{15}$ chance that there will be no home ground advantage if they play Port Adelaide. If Geelong finish second, then it is assumed the match will be played at the MCG. This means that there is a $\frac{1}{3}$ chance that they will play away (a team from Group 5) and a $\frac{2}{3}$ chance that it will be a neutral ground. These premiership probabilities for each of the groups are given in the following table.

Final position	Group 1	Group 2	Group 3	Group 4	Group 5
1	20.9	15.6	20.7	20.5	22.6
2	20.9	15.6	20.7	20.5	22.6
3	18.6	14.9	18.5	18.5	19.7
4	18.6	14.9	18.5	18.5	19.7
5	5.1	4.1	5.2	5.3	6.1
6	5.1	4.1	5.2	5.3	6.1
7	3.8	3.6	3.9	4.1	5.1
8	3.8	3.6	3.9	4.1	5.1

Table 5: Percentage chance of a premiership for each of the five groups of teams given their final position and home ground advantage for the new finals system.

As shown in Table 5, premiership probabilities between Groups 1, 3 and 4 do not differ that significantly. However, Group 2's chances are well below average, whilst Group 5's probabilities are higher than normal despite sharing the MCG with four other tenants. This is largely because Group 5 will always enjoy either a home game or neutral grand final, whilst other groups will either play a neutral or away grand final. This problem is common to any system. The MCG is the largest capacity sporting ground in Australia and well deserves the right to host the grand final.

Of greater importance is that Group 2's (Carlton, Geelong) probabilities are somewhat significantly less than other groups. In fact, they are so low that any other team that finishes fourth has a greater probability of winning the premiership than Carlton or Geelong do if they finish first. Of course, one of the reasons this occurs is that they will never obtain a home ground in any of their matches, but this highlights the problem with the new system in that the top four sides have the same probabilities of winning the premiership given equal probability matches. In the McIntyre system and most other ranking systems, first place has a significant advantage over fourth despite any home ground advantage, but this does not occur in this system.

Carlton and Geelong are 26% less likely to win the premiership given their home ground status than other teams, whereas teams in Group 5 are 11% more likely to win the premiership because of their MCG advantage.

Final position	Group 1	Group 2	Group 3	Group 4	Group 5
1	19.72	15.11	19.46	19.31	21.52
2	19.72	15.11	19.46	19.31	21.52
3	16.29	12.73	16.21	16.13	18.10
4	12.59	12.73	12.54	12.49	14.00
5	10.75	9.62	10.86	10.97	12.76
6	7.27	9.18	7.37	7.47	8.74
7	5.09	4.05	5.10	5.24	6.14
8	5.09	4.05	5.10	5.24	6.14

Table 6: Percentage chance of premiership for each of the five groups of teams given their final position and home ground advantage for the McIntyre Final Eight system

These results can be compared to that of the old McIntyre Final Eight system.

Table 6 shows that the same pattern still occurs; that Group 2's probabilities are below average whilst Group 5's probabilities are greater than average. In fact, the probabilities of each of the groups winning the premiership has barely changed from system to system. However, the new system has given a greater chance of premiership for the top four teams than the bottom four. This is because Team 7 and Team 8 will receive no home games in the new system, but will obtain a home game in the McIntyre system if they win the first game. But probably the greatest difference is that under the McIntyre system, if teams from Group 2 finish first or second, then their probability of a premiership is greater than any other team who has finished fourth. This is not the case in the new system.

For the new system to work properly, every team needs to have their own different and distinct home ground that they will be guaranteed a home game at in the finals if they finish in the required positions. This of course will never happen, with ten teams in Victoria. It is our conclusion that the McIntyre system performs better on the criteria.

The question remains that although the McIntyre Final Eight system takes preference over the new system, is it fair and just? The answer to this is no.

5 Conclusions

These results show that the new finals system is far from perfect. It does not match the criteria that higher ranked sides have a higher probability of winning the premiership, as Team 1 has an equal chance with Team 4, and Team 5 has an equal chance with Team 8. There is also little difference between the expected final position of the top four, and little difference between the bottom four. Fifth place is three times less likely to win than fourth. Also, the best two teams are not the most likely grand final quinella. With the possibility of teams deliberately losing games in the final home and away season round and unfair match-ups in the semi-finals, the new system has several problems. As the new system is so dependent on higher ranked teams having a significant home ground advantage, certain teams may be more likely to win the premiership when finishing fourth, than others finishing first or second. The system would be more just if every team possessed their own separate home ground, but still some problems would occur. Like the McIntyre Final Eight system, the general public will respect the new system until one of its inadequacies is shown up by a particular draw.

References

[1] G. A. Christos, "The AFL Finals: It's more than a game", in *Fourth Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar (editors), Bond University, Queensland, Australia (1998), 111–124.

[2] F. K. R. Chung, "Do stronger players win more knockout tournaments?", *J. Amer. Statist. Assoc.*, **73**(363) (1978), 593–596.

[3] S. R. Clarke, "Calculating premiership odds by computer—an analysis of the AFL final eight playoff system", *Asia Pacific J. Oper. Res.*, **13**(1) (1996), 89–104.

[4] S. R. Clarke, *Performance Modelling in Sport*, PhD thesis, Swinburne University (1997).

[5] H. A. David, "Tournaments and paired comparisons", *Biometrika*, **46** (1959), 139–149.

[6] L. Fahrmeir and G. Tutz, "Dynamic–stochastic models for time dependent ordered paired comparisons", *J. Amer. Statist. Assoc.*, **89**(428) (1994), 1438–1449.

[7] R. B. Israel, "Stronger players need not win more knockout tournaments", *J. Amer. Statist. Assoc.*, **76**(376) (1981), 950–951.

[8] M. G. Kendall, "Further conributions to the theory of paired comparisons", *Biometrics*, **11** (1955), 43–62.

[9] R. J. Maurice, "Selection of the population with the largest mean when comparisons can be made only in pairs", *Biometrika*, **45** (1958), 581–586.

[10] T. McGarry and R. W. Schutz, "Efficacy of traditional sport tournament structures", *J. Oper. Res. Soc.*, **48** (1997), 65–74.

[11] J. P. Monahan and P. D. Berger, "Play-off structures in the National Hockey League", in *Optimal Strategies in Sports*, S. P. Ladany and R. E. Machol (editors), North Holland, Amsterdam (1977), 123–128.

THE USE OF COMPUTERS TO PREDICT THE PERFORMANCE OF GOLF CLUBS

J. S. B. Mather
School of Mechanical, Materials, Manufacturing Engineering and Management
University of Nottingham
Nottingham NG7 2RD, United Kingdom

Abstract

The ever increasing demand of golfers for improvements in the performance of equipment has spawned many research programs in recent years. In the technology of the ball, the work has mainly been successful, but for the club the efforts of the researchers have produced markedly less usable results. The paper discusses why this should be, taking as its basis the swing pattern that needs to be achieved, and considering the effect of club head and shaft design during the swing and impact with the ball. The paper shows that the golf swing and the impact process are very complex and varied and concludes that mathematical modelling and computation will prove to be the only way for future research in this area.

1 Introduction

The performance of golf clubs must be defined by their ability to produce the desired shot. For many golfers this rarely happens and we must question whether this is the unchangeable fault of the golfer or the inability of the manufacturers of the equipment to supply suitable clubs. Part of the answer to this question lies in the history of the technology of the game and part in the complex nature of the swing that is widely advocated by teachers of the art.

Any serious study of golf club design will show that clubs have changed little for over 200 years apart from the use of modern materials. The benefit to the average golfer has been minimal. This is despite quite significant advances made in the understanding of the various mechanisms involved in the game. For confirmation of this and as a background to the topics discussed in the paper, the reader is referred to the many papers and articles in Cochran (1990, 1995), Cochran and Farrally (1994), Farrally and Cochran (1998), and Haake (1996, 1998).

This paper considers the nature and diversity of the very complex swing patterns, and the way they affect the movement of the club, giving examples of and commenting on theories which have been developed and how computational models are used to predict the correct club for each golfer.

The aim of the swing is achieved with the impact of the ball and the face of the club and, again in this area, experiments and theory are developing side by side. Computer models are aiding a more complete understanding of the mechanisms involved in propelling the ball forward and of how high values of backspin are generated.

Once it has left the club, the flight of the ball is governed by the aerodynamics of spinning objects. Most computer models of the flight are based on simple Newtonian mechanics using measured aerodynamic data such as that of Smits and Smith (1994). The computations are mainly used in the choice of the correct values of loft for the clubs and the models are sufficiently accurate for this purpose.

Finally the paper makes a prediction of what now needs to be done and how computers could be used to match the physical abilities of the player to the right equipment.

2 The golf swing

2.1 The unique swing

Early attempts at mathematically modelling the swing came almost before computers were generally available. Figure 1 is a representation of a stick model which is based on a stroboscopic photograph of the swing of Bobby Jones (one of the best ever golfers). This was used directly by Williams (1967) and inferred by Jorgensen (1970). The Lagrangian approach used by the latter and the classical approach of the former both suggest that there is but *one* way of swinging a golf club to utilise the golfer's ability to the maximum.

A distillation of what is proposed for this unique swing is as follows.

- Starting the downswing, the golfer must accelerate the club as a solid body. The arms and wrists hold the club with a constant angle between the arms and the club. This solid body rotation carries on for about 120°, during which the constant angle is maintained by applying an increasing torque through the wrists and hands in opposition to the torque caused by centrifugal acceleration of the clubhead.
- At this point, the hands and wrists cease to resist the torque allowing the club to rotate about a pivot roughly in the middle of the hands until, if the timing of the swing is correct, the head impacts with the ball when it has achieved its maximum velocity and it is travelling along the intended direction of the shot with the face of the club at right angles to that line.

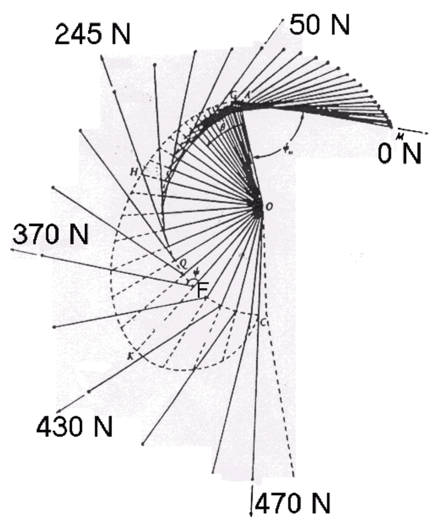

Figure 1: Stick representation of a good swing.

Milne and Davis (1992), with the advantage of computation, solved the complex equations of motion of the swing, taking the analytical model much further. Although the results of his work are extensive and interesting, his main conclusion confirmed the findings of his predecessors.

The geometry and muscular distribution of the golfer must dictate the exact pattern of the swing but the timing of the swing of the top golfers is, with few exceptions, very similar.

The golfer must recognise the importance of these results. To achieve a good swing and utilise his efforts to the full, the golfer should reproduce as closely as possible the timing of the unique swing but at a speed which he can control. Unfortunately, as will be described in the next sections, few golfers understand this requirement and, worse still, very few instructors teach it.

2.2 The real swing

For a very good swing, the forces acting along the line of the shaft are as shown in Figure 1. Provided that the golfer applies a torque sufficient to maintain the club angle (as above) then the forces acting along the arms of the golfer are large only in the last stage of the swing when the club is entering the impact area. The golfer at this point is in an ideal position to withstand the large forces that are transmitted through the legs, which are flexed evenly, through the spine which is untwisted, and into the shoulders and arms which carry the force equally.

If the swing pattern is other than this, and for the greatest majority of golfers it is, then the body is in a poor position to withstand these large forces and hold the club on line. For instance, for the golfer who slices the ball, the wrists do not maintain the angle and the force along the arms is then large in the early part of the downswing. Because the body is in a twisted state, with the weight unevenly distributed on the legs, the golfer inevitably is pulled sideways. By now the club is travelling well outside the correct plane, with the result that the golfer has to pull the left shoulder out of the way to let the club come across and into the ball. This action, as a whole, produces a slice.

This explanation is not given merely as a condemnation of poor golfers. It is done because the market is mainly for these players, and it is essential that *any model of the golfer used for club design replicates this swing*.

2.3 The model of the real swing

The models of Williams, Jorgensen, and Milne are mathematically complex but, even so, have been shown to be limited in realism. If clubs are to be designed for the very wide range of ability in the amateur game, then more sophisticated models are needed. These will be beyond analytical treatment and will rely heavily on the power of the computer to simulate the many elements of the anthropometric system. In this way, the effect of the forces exerted on the body during the swing can be evaluated. Much more work is needed in this area.

As an encouragement to this work, it has already been shown (Mather, Waites and Vardy (1992)) that it is possible to force the golfer into a better swing pattern. In this, the mass distribution of a device representing a club is altered in a series of tests with the individual golfer. The club head speed at impact and the angle and direction of the face of the club are checked. The final distribution is put through the design program to check the shaft flex, lie and loft. The process has been successful with a number of golfers, and is being constantly developed.

2.4 Design and testing of golf clubs using robots

The conclusion of the previous section that the use of the complex design process is essential in design, begs the question of the relevance of robot golfing machines.

Traditionally, these have been used to aid and advance the testing and design processes. Early designs were based on a double pendulum, the two points of rotation representing the middle of the shoulders and the middle of the hands. Three and four link robots have been tried but none reproduce the complex physique and geometry of the real golfer. It is now clear that a multi-link system is required, representing, as a minimum, the trunk, each shoulder, the upper and lower arms, and each wrist. Without this, the robot is singularly unrepresentative and therefore cannot be used to develop golf clubs except, perhaps, for the very good amateur or professional golfer.

Such robots will be extremely expensive and it seems that, if realistic progress is to be made, effort should be put into generating, mathematically, an androidal model of the golfer. Some simplification may still be necessary and this may take the form of assuming, perhaps, that the elements of the body below the waist rotate on a horizontal plane. This concept is the subject of a current study at Nottingham.

2.5 Statistical analysis of the average golfer

Section 2 implied that the average golfer swings the club far too fast. Since this is a fact well-known to all teachers of the game, it is right that the non-golfer should ask why do players continue to do it?

Riccio (1990) shows that accuracy of the shot to the green is by far the most important quantity in playing better golf. For these shots the golfer has simply to produce sufficient head speed for the club he is using to propel the ball the required distance. But the golf industry has cleverly pushed the golfer, by advertising and marketing, into believing that distance from the tee is the essential ingredient to better play. The player is thereby encouraged to swing hard and fast with the most difficult club in the set, the driver, with which, unsurprisingly, he does not remotely approach the correct swing. It is this, more than anything, that gives rise to the widespread incidence of the poor swing pattern referred to in Section 2.

3 Modelling of the swing

The data of Figure 1 show the distribution of high amplitude forces to which the club is subjected during the swing. Since the club is highly elastic, being a heavy weight on a very slender shaft, and the accelerations rise to over 300 g, the shaft must bend and twist significantly throughout the swing. If the golfer applies a different force pattern, the club bends and twists in a different manner.

Many attempts have been made to develop computer programs which model the three-dimensional movement of the club throughout the swing. Starting with the forces and torque introduced to the club by the golfer, they compute the location and attitude of the club. This requires a knowledge of the boundary condition supplied by the hands. In the golf swing, the boundary condition changes markedly during the downswing. At the start, the hands hold the club lightly, just enough to reverse the motion of the club from upswing to downswing and with a force which is normal to the line of the shaft. Once the downward motion has started, the golfer holds the wrists cocked at a fixed angle, which suggests something of a distributed boundary condition. The role of each hand and the forces and torque applied through the fingers of each hand are not the same. Towards the end of the downswing, when the centrifugal forces acting along the shaft rise to over 400 N, the tapered grip of the club jams into the hands, which, with the aid of friction, grasp the club hard enough to prevent slippage. At this stage, the boundary condition could be modelled as a form of elastic clamp.

3.1 A new method of modelling the club

This leads to the conclusion that it would be useful to develop a method of modelling the club movement, based as before on force and acceleration, which avoids the need for the boundary condition. The fundamentals for such a system were reported by Cooper and Mather (1990), who devised a method which measured, for golfers with the complete range of ability, the vector accelerations in the region of the shaft immediately below the grip. In this area the shaft can be assumed to be straight and untwisted. These accelerations are then used to compute the movement of the head and shaft in the swing. The angle of the face and its direction of travel are found throughout the swing and the effects of different shafts are evaluated. This will be a useful tool for the designer.

The results of the calculations are interesting. Figure 2, for example, plots the typical deflection of the head of the club caused by shaft bending in the swing of a good golfer. The maximum deflections both in the plane of the swing (defined locally throughout the swing) are about 20 mm and those perpendicular to this up to 50 mm, depending on the location of the centre of gravity of the head and its mass, the shaft flex and stiffness.

Impact with the ball is at 0.4 seconds, with the shaft bent forwards and downwards (droop) as shown (in exaggerated form) in Figure 3. It is clear that the velocity in addition to the mean head velocity in the swing is almost zero at impact giving a lie to the marketing hype that this final “kick” of the shaft produces extra head velocity. To take advantage of any velocity gain the impact would have to occur, say, at 0.35 seconds and even then the “added” velocity is less than $1\,\mathrm{ms}^{-1}$ or 3% of the mean.

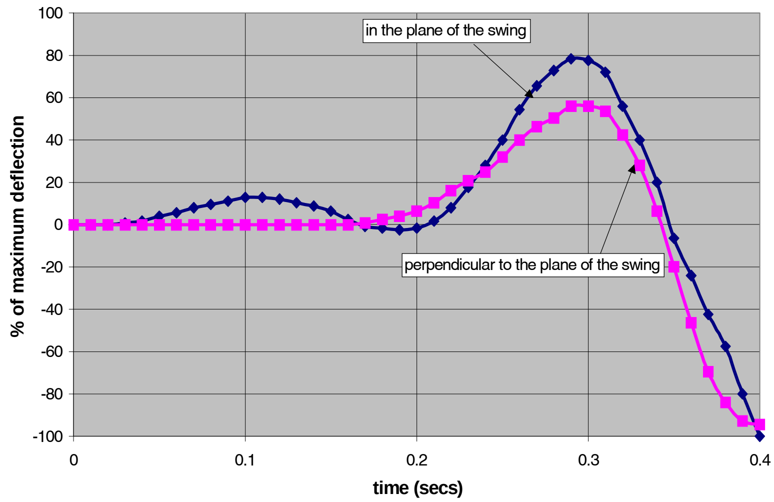

Figure 2: The deflection of the shaft during the swing.

The same computation can be used to evaluate the effect of changing head mass, shaft length and flex—an extremely useful tool for the designer, and one which is being widely used.

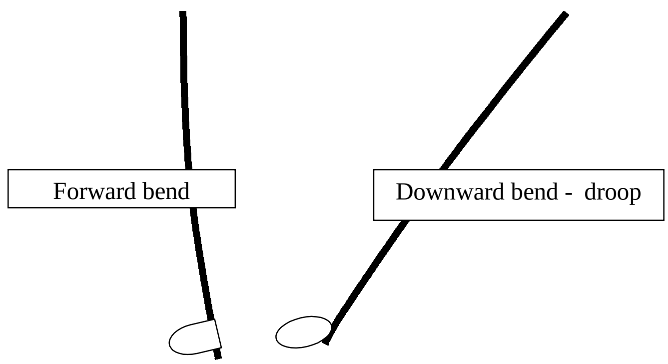

Figure 3: Bend and droop of the club at impact.

3.2 Three-dimensional curvature of the shaft

The overall flex of the shaft is shown in Figure 4. This is one of the figures taken from Mather *et al.* (2000) and shows the curvature of two clubs (each with two swings) immediately prior to the impact of the club and the ball, for the swings of a golfer of good amateur standard.

The radius of curvature in centimetres is plotted against the distance along the shaft. The minimum radius of curvature is generally about 0.5 m and is situated between 35 and 45 cm from the tip of the shaft. The golf industry states that "the location of the kick point governs the gradient of the shaft as it enters the club head and thereby the overall loft of the club; the higher up the shaft the kick point is, the lower the gradient of the club head and the lower the flight of the ball." Apart from the technical non-sequiturs in this concept, the curvature shown in the diagram is found to be mainly in droop and therefore has virtually no effect on the loft of the club head, but a very significant effect on the angle the face presents to the ball. This is vital in the impact process and determines the direction of flight and the amplitude of backspin on the ball.

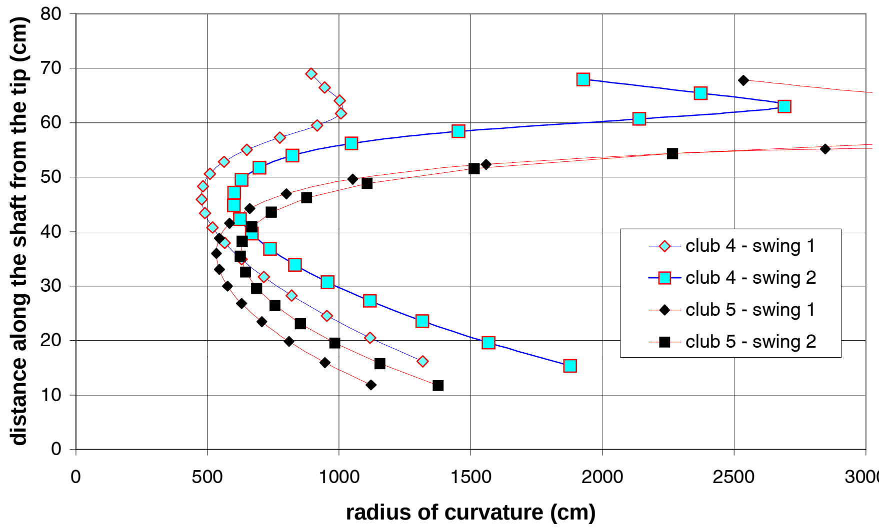

Figure 4: Curvature of shafts before impact.

It is interesting to note, though, the consistency between the performance of club 5 from swing to swing, and the slight inconsistency of the swings with club 4.

As far as is known, such valuable results are only available by using photogrammetric systems such as that described in Smith *et al.* (1998) and the powerful matrix computations needed for this system. With this, results can be obtained for any golfer using any club—again an extremely useful tool. The research in this area is yielding very useful data for the design of clubs specifically for amateur golfers.

4 The impact of the ball and the head

This section deals with the models of the head and ball which allow and assist the designer to predict the outcome of any shot.

4.1 Forces on the face

First assume the swing is correct. The head arrives at the ball travelling along the intended line and with the face at right angles to it. The force of impact is well documented, for example by Lieberman and Johnson (1994), Gobush (1996) and Johnson and Lieberman (1994), and has two components. The first is normal to the face and generates the force necessary to propel the ball forward. The second is perpendicular to this and acts along the face. This generates the spin on the ball. Typical values for these two forces are shown in Figure 5. The negative tangential force towards the end of the impact is probably caused by the effect of a torque created because the normal through the centre of pressure of the contact area does not pass through the centre of gravity of the distorted ball. If correct, this mechanism increases the backspin and probably explains why soft "balata" balls spin more than hard two-piece balls.

The normal force is not insubstantial, rising to over 13,000 N simply because the 43 gm ball is accelerated from rest to 70 ms^{-1} in a time of 500 microseconds. In this time, the collision causes the head to divert from its original path, the shaft to bend and to twist. The ball is still in contact with the head during this movement which, therefore, determines its initial direction, velocity and three dimensional spin, and its subsequent flight.

For the average amateur, the swing is far from correct. The attitude of the head at impact, and the

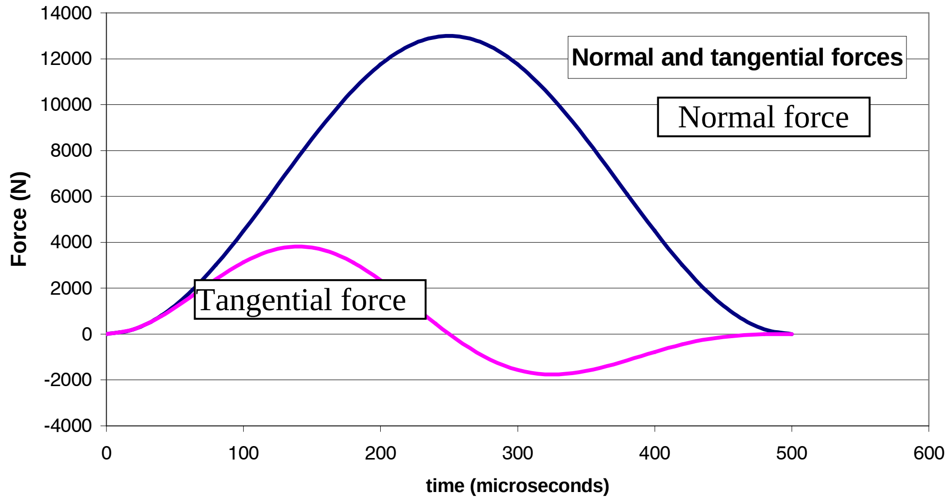

Figure 5: Forces on the club face.

force patterns on the face, particularly in angle, are all different from the data currently available and there is a substantial horizontal component which creates side spin. The movement of the head and the reaction of the ball are bound, therefore, to be different.

4.2 Design principles

The parameters which influence the motion of the head include head mass and mass distribution (and hence the inertia matrix of the clubhead), shaft flex in bending and twist and the attitude of the club at impact, known as prebend.

The aim of club design would, at first sight, seem to be to minimise flight deviation, so encouraging straighter shots. Such clubs, using commercial jargon, would be more forgiving for the golfer. It is often necessary, however, for the good golfer to shape the shot in order to avoid obstacles or to counter natural forces such as the wind, and this could be negated, somewhat, by the forgiveness of the club. The designer needs to consider all sectors of the market and design accordingly.

4.3 Model of the impact phase

Several models have been made of the club at impact. Typical studies are reported in Hocknell *et al.* (1998) and Clafton (1999). Leong (2000) includes the pre-bend of the club at impact as discussed in Section 3 and Daniels (2000) models the time-variant pressure distribution on the face. All of these studies are based on the finite element method but use different base models. In some, only the head is treated, the assumption being made that, in the very short time of impact, the motion is unaffected by the shaft. In others the shaft is modelled as a series of hollow tubes and the head as a point mass concentrated at the centre of gravity together with an inertia matrix about that point. Figure 6 shows such a system with nodes distributed over the face (only a fraction of those used are shown), the centre of gravity displaced away from the face, and the shaft connected to the head model so that the values of the loft and lie angles are maintained throughout the impact, as would be the case for a normal head.

All of the results obtained with a shaft confirm that the boundary condition at the hands has no effect on the impact phase of head and ball.

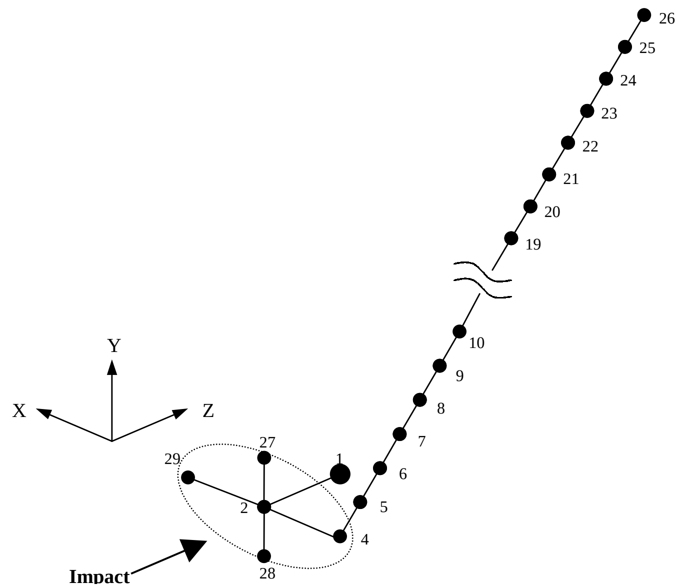

Figure 6: Nodal arrangement for FE analysis.

4.4 Results

The computation proceeds to find the response characteristics of the system. The eigen-frequencies and shapes are derived for a given geometry and the forced response is computed. Measurements of the response of clubs clamped in a manner designed to simulate the hands suggests that only the first five or so modes are required to determine the motion of the head during impact, although this remains to be confirmed by dynamic testing. Figure 7, taken from Leong (2000), shows the motion of the tip section of a club, starting from the original shape with pre-bend and ending with the deflected shape at the end of impact. As expected, the head deflection is about 3 mm.

Figure 8 shows the results of a computation by Clafton (1999) of the response of a head modelled as a semi-ellipsoid. The four results shown are obtained by varying the location of the centre of gravity and therefore the inertia matrix whilst maintaining the overall mass, this being particularly apposite for head design.

The first and second modes (not shown) are cantilevers parallel and perpendicular to an axis through the face of the club. The figure shows the nodal axes of the third, fourth and fifth modes, which pass through the body of the club head at various angles. The ball impact excites these modes and their phased combination determines the head motion.

Experiments in the field with a wide variety of clubs show that heads with different responses (but with the same general parameters such as mass, shaft length and flex) produce different ball flights. Further, a comparison of modal response with ball flight suggests that individual modes or groups of modes can be associated with certain flight characteristics. The details of this are contained in the work of Mather (1996), but generalising the conclusions of that work,

- modes which have nodal axes parallel to the ground control backspin,
- modes which have nodal axes at 90° to the ground control side spin.

Translating these effects to the results in Figure 8(a), mode 5 (nodal axis 5) generally governs the degree of backspin on the ball. Modes 3 and 4 (nodal axes 3 and 4) have more control over the amount of side spin.

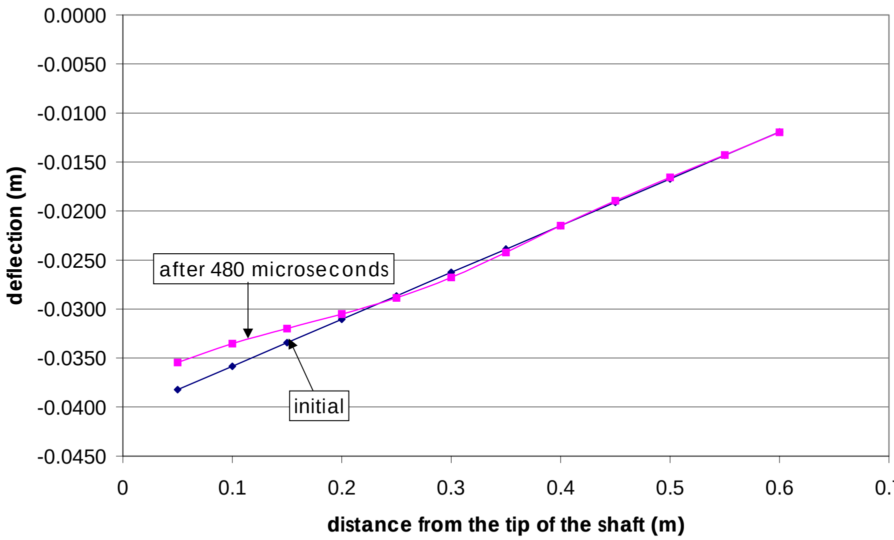

Figure 7: Deflection of the shaft at the end of the impact period.

There has been much said about the "sweet spot" on sports devices and golf is no exception. In Figure 8, the sweet spot might be defined as the impact point or area that induces the minimum head movement. The effect of the cantilever modes is controlled by the flex of the shaft, particularly in the tip region, and the designer can choose the shaft accordingly. For the higher order modes a sweet spot will exist at or near to the junction of the nodal axes. For the results in Figure 8(a) and (d) this would be, say, in the middle of the triangular areas generated by the nodal axes. For Figure 8(b), the axes

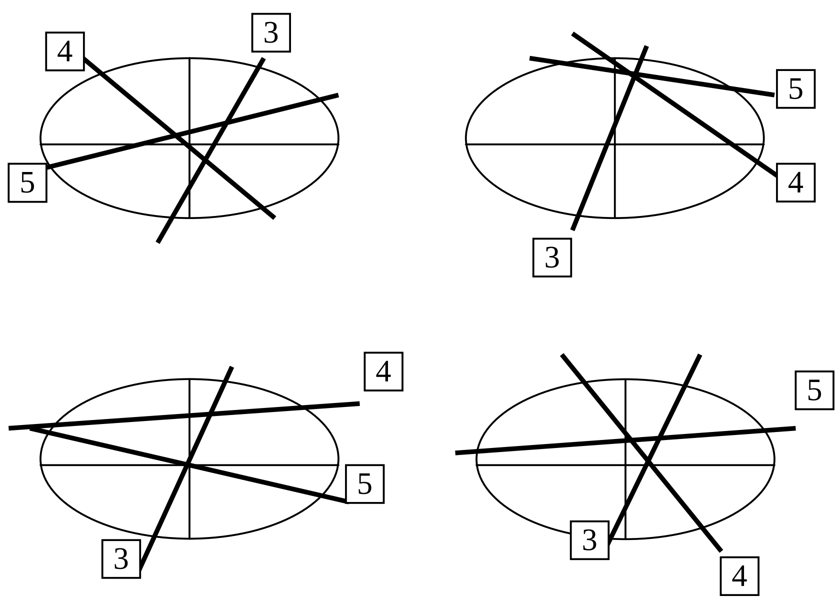

Figure 8: The modal response of an ellipsoidal head.

coincide at a single point but, unfortunately, one which is of no use because of its extreme location. In Figure 8(c), two of the nodal axes cross in the centre of the club, and the third is parallel to one of them. Of the designs shown, this would probably be the most usable club for a good golfer. Shots hit towards the toe or heel will induce side spin bending the ball left (draw) or right (fade) in flight. But the design for a truly forgiving club for the average golfer might be one which counters the ill-effects caused by an incorrect swing. In this respect, the design in Figure 8(d), has a fair degree of backspin, and induces side spin which counters a slice. This is by way of a qualitative argument but one which aids the design process once the experienced designer learns to relate the head response with probable ball flight. The actual values for the movement of the head are always available.

Clearly, the program can be made to design the head to achieve whatever ball reaction the designer wishes. The current design process involves the use of Computer Aided Design to create a usable and marketable shape, from which the location of the centre of gravity and the inertia matrix are found and put into the modal response program. The system then iterates, balancing the aesthetics of the club with the practicality of impact response.

Such a program, or more accurately a suite of programs, is essential for the designer of the next generation of clubs.

5 Modelling the flight

Using simple Newtonian mechanics and the drag and lift data of spinning balls (Smits and Smith (1994)) the flight to impact can be calculated with reasonable accuracy. These calculations require the initial angle of the trajectory, the vector of the velocity and the spin rate. These will vary with:

- the swing pattern of the golfer,
- the type of ball normally used,
- the response of the club,
- the head velocity, and attitude at impact.

As well as tailoring the club to the golfer, the choice of head design might, conceivably, involve the playing location of the golfer. For instance, those who play near to the seaside (links) require shots which penetrate under the wind yet have enough spin to stop on or near to the green. Those who play on courses which might be referred to as having "target golf" with greens and fairways which are well guarded, would prefer high shots which finish with an almost vertical drop to the green. Figure 9 shows examples of these, where loft, head attitude and velocity have been varied.

6 Measurement and analysis of the performance of the golfer

Most of what has been described above requires measurements to be taken of the club during the swing and the impact. Computer modelling is of no use unless it has been confirmed as realistic. Commercially available optical systems often suffer from two deficiencies—the limited number of targets that they can measure and the accuracy that can be achieved bearing in mind that the swing volume is large, being 5 m by 4 m by 2 m deep. High speed stroboscopes and rapid spark guns are now regularly used to determine the location of the shaft (and the golfer) during the swing and impact. In these the number of targets is limited only by analysis time and accuracy by pixel size. Cost is also important and techniques are now being developed using cheap non-metric cameras and transformation techniques based on standard photogrammetric procedures (Smith *et al.* (1998)). In these, the use of large and powerful computers is absolutely vital.

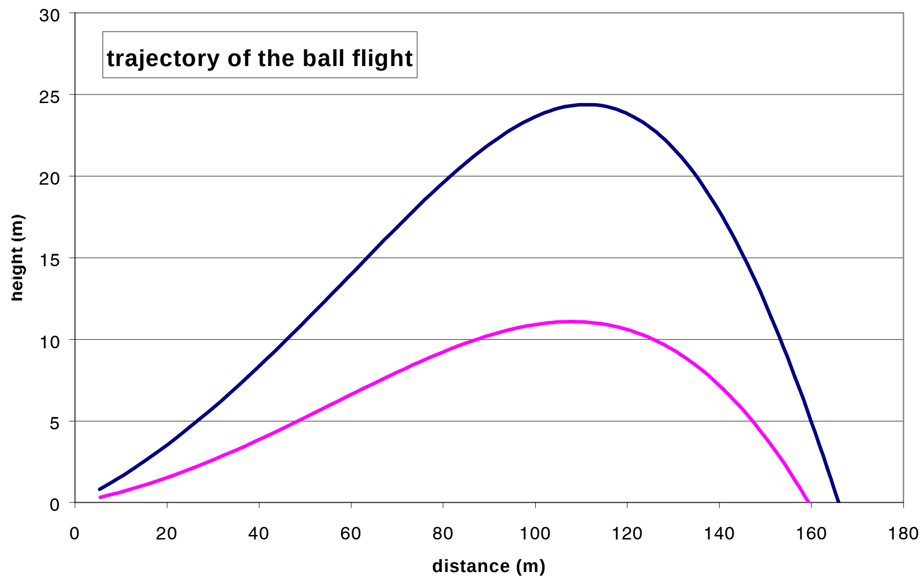

Figure 9: Flight trajectory for the same distance.

7 Summary and conclusions

Ten years ago, the design of golf clubs came from the mind of the marketing man and not the engineer. Competition for the market is more extreme now than it has been for some time and it is essential for companies to take on board the latest technical developments to offer the golfer better products. The problem, inevitably, arises from the fact that this can only happen through a quantum change in culture. The speed with which the technology, referred to above, is developing only adds to this problem. The paper has outlined the use of mathematical modelling and computation in the fields of:

- the swing of the golfer,
- the design of the shaft for the swing,
- the effect of the interaction between the shaft and head,
- the design of the head for impact response,
- the generation of spin on the ball,
- the flight envelope of the ball.

The research is continuing. Certainly, better models for the human form are needed. Testing of club and shaft response during the swing and impact with the ball will yield better designs for different categories of golfer. Finally data on the drag and lift of a spinning ball at low velocities need to be provided.

References

G. Clafton (1999), *The use of finite element analysis for the performance of a golf club*, thesis for the degree of MEng, University of Nottingham, Nottingham, UK (unpublished).

A. J. Cochran (editor) (1990), *Science and Golf* (Proceedings of the First World Scientific Congress of Golf), E. & F. N. Spon, Chapman & Hall, London.

A. J. Cochran (editor) (1995), *Golf the Scientific Way*, Aston Publishing Group, Hemel Hempstead.

A. J. Cochran and M. Farrally (editors) (1994), *Science and Golf II* (Proceedings of the 1994 World Scientific Congress of Golf), E. & F. N. Spon, Chapman & Hall, London.

M. A. J. Cooper and J. S. B. Mather (1994) "Attitude of the shaft during the swing of golfers of different ability", in *Science and Golf II*, A. J. Cochran and M. Farrally (editors) E. & F. N. Spon, Chapman & Hall, London, 271–277.

C. Daniels (2000), *The effect of time based pressure distributions on the response of a golf club*, thesis for the degree of MEng, University of Nottingham, Nottingham, UK (unpublished).

M. Farrally and A. J. Cochran (editors) (1998), *Science and Golf III* (Proceedings of the World Scientific Congress of Golf), E. & F. N. Spon, Chapman & Hall, London.

W. Gobush (1996), "Friction coefficient for golf balls", in *The Engineering of Sport*, S. Haake (editor), A. A. Balkema, Rotterdam, 193–194.

S. Haake (editor) (1996), *The Engineering of Sport: Proceedings of the First International Conference on the Engineering of Sport*, A. A. Balkema, Rotterdam.

S. Haake (editor) (1998), *The Engineering of Sport; Design and Development: Proceedings of the Second International Conference on the Engineering of Sport*, Blackwell Science, Oxford.

A. Hocknell, R. Jones and S. J. Rothberg (1998), "Computational and experimental analysis of a golf club", in *Science and Golf III*, M. Farrally and A. J. Cochran (editors), E. & F. N. Spon, Chapman & Hall, London, 526–534.

S. H. Johnson and B. B. Lieberman (1994), "An analytical model for ball–barrier impact. Part 2, A model for oblique impact", in *Science and Golf II*, A. J. Cochran and M. Farrally (editors), E. & F. N. Spon, Chapman & Hall, London., 315–320.

T. Jorgensen (1970), "On the dynamics of the swing of a golf club", *Am. J. Phys.*, **38** (5), 644–651.

D. Leong (2000), "The effect of pre-bend on the deflection of the club head during impact", *Research report for Decathlon, Lille*, University of Nottingham, Nottingham, UK.

B. B. Lieberman and S. H. Johnson (1994), "An analytical model for ball–barrier impact. Part 1, Models for normal impact", in *Science and Golf II*, A. J. Cochran and M. Farrally (editors), E. & F. N. Spon, Chapman & Hall, London., 309–314.

J. S. B. Mather, B. Waites and D. Vardy (1992), "The design of weighted golf clubs", US Patent No. 5,152,527.

J. S. B. Mather and M. A. J. Cooper (1994), "Categorisation of golf swings", in *Science and Golf II*, A. J. Cochran and M. Farrally (editors), E. & F. N. Spon, Chapman & Hall, London, 65–70.

J. S. B. Mather (1996), "The role of club response in the design of current golf clubs", in *Proc. 14th IMAC*, Dearborn, Michigan USA, 397–403.

J. S. B. Mather, M. J. Smith, S. Jowett, K. A. H. Gibson and D. Moyniha (2000), "Application of the photogrammetric technique to golf club evaluation", *J. Sports Engineering*, to appear.

J. S. B. Mather (2000), "Innovative golf clubs designed for the amateur", in *The Engineering of Sport* (Proceedings of the Third International Conference on the Engineering of Sport), Blackwell Science, Oxford, to appear.

J. S. B. Mather and S. Jowett (2000), “Three-dimensional shape of the golf club during the swing”, in *The Engineering of Sport* (Proceedings of the Third International Conference on the Engineering of Sport), Blackwell Science, Oxford, to appear.

R. D. Milne and J. P. Davis (1992), “The role of the golf club shaft in the swing”, *J. Biomechs.*, **25**, 975–983.

L. Riccio (1990), “Statistical analysis of the average golfer”, in *Science and Golf*, A. J. Cochran (editor), E. & F. N. Spon, Chapman & Hall, London, 153–158.

M. J. Smith, J. S. B. Mather, K. A. H. Gibson and S. Jowett (1998), “Measuring the dynamic response of a golf club during swing and impact”, *Photogrammetric Record.*, **16** (92), 249–257.

A. J. Smits and D. R. Smith (1994), “A new aerodynamic model of a golf ball in flight”, in *Science and Golf II*, A. J. Cochran and M. Farrally (editors), E. & F. N. Spon, Chapman & Hall, London, 340–347.

J. J. Webster (1994), “Response of a golf club during the swing”, unpublished.

D. Williams (1967), “The dynamics of the golf swing”, *Quart. J. Mech. Applied Math.* **XX** (2), 247–264.

MODELLING ENDURANCE TIME AT $VO_{2\,\max}$

R. Hugh Morton*
Institute of Food, Nutrition and Human Health
Massey University
Palmerston North
New Zealand

Veronique Billat
Laboratory in Sport Science
Université Lille 2
2 Avenue Richerand
Paris 75010, France

Abstract

There has been significant recent interest amongst sport scientists and athletes, in the minimal running velocity which elicits $VO_{2\,\max}$. There also exists a maximal velocity, beyond which the subject becomes exhausted before $VO_{2\,\max}$ is reached. Between these limits there must be some velocity which permits maximum endurance at $VO_{2\,\max}$, and this parameter has also been of recent interest. We model the process based on a two component (aerobic and anaerobic) energy system, a two component (fast and slow) oxygen uptake system, and a linear control system for maximum attainable velocity resulting from declining anaerobic reserves as exercise proceeds. The model development produces a skewed smooth curve for endurance time at $VO_{2\,\max}$ with a single maximum. This curve has been successfully fitted to pooled endurance data collected from ten exercising subjects ($R^2 = 0.821$, $p < 0.001$). For this group of subjects the maximal endurance time at $VO_{2\,\max}$ is predicted to be 603 seconds at a running pace corresponding to 88% of the minimal running velocity which elicits $VO_{2\,\max}$ in an incremental running test. This is a longer time than is usually reported in the literature, but not as long as can be achieved if running pace is progressively reduced.

1 Introduction

The relationship between power output and endurance time is a fertile area for the study of human bioenergetics and work performance. For two recent reviews, consult Billat *et al.* [2] and Morton and Hodgson [12]. With very few exceptions, this research has focused on endurance at constant powers, where the critical power (CP) concept (see Hill [5]) has been by far the most commonly adopted model. It has been widely studied, and adapted for swimming, running, rowing, cycling, kayaking and wheelchair exercise. Nevertheless, it is not without its critics (Vandewalle *et al.* [17]).

For some of these exercise modalities, power output can be measured directly on an ergometer. However for running, swimming and wheelchair exercise, velocity and distance take the place of power and work respectively, with corresponding changes to the units of measurement of the parameters of the model. It would be useful if a single forcing variable could be found, one which is independent of exercise modality and which could be used in a totally general setting. Oxygen uptake may be one such candidate.

Oxygen uptake however, is not a simple function of power output or velocity, for it is a function of time as well. Even steady state oxygen uptake is not a linear function of power output beyond a certain level. The slow component of oxygen uptake and increasing oxygen cost of exercise at higher powers complicates the issue (Gaesser and Poole [4]). The slow component has however been successfully modelled, both theoretically (Morton [9]) and empirically (Barstow and Molé [1]); and the energy cost of running can safely be assumed constant (or very nearly so) provided the power or velocity range

*The work of the first author has been made possible through the support of the Centre de Medicin du Sport, CCAS, Paris.

is narrow. These models are not mathematically simple. Perhaps then these difficulties can be largely overcome by considering endurance at a fixed value of oxygen uptake, say at its maximum ($VO_{2\,\max}$).

The power range that will bring on exhaustion in a finite time can be divided into three domains. Power output may be high, that is higher than CP but insufficient to elicit $VO_{2\,\max}$. It may be very high or maximal, sufficient to drive VO_2 to its maximum before exhaustion. Or it may be extreme, such that the subject becomes exhausted before sufficient time has elapsed for VO_2 to reach its maximum. Indeed, the minimum power or velocity just sufficient to elicit $VO_{2\,\max}$ before exhaustion in a subject, and endurance time at $VO_{2\,\max}$, are two phenomena of current interest to exercise physiologists, sports coaches and athletes in training.

The relationship between power output and total endurance has been modeled over the whole power range above CP as referred, but modelling endurance at $VO_{2\,\max}$ is restricted to the narrower mid-range. Given what is already known about the human exercise response, it should be possible to model this latter relationship, perhaps in a similar way to modelling endurance at constant power. To our knowledge, this has never been accomplished, and it is the purpose of this paper to present the mathematics of this modelling process.

2 Methods

2.1 Model background

Quite apart from any philosophical arguments concerning the CP concept itself, there are several practical ones to consider. Several of these are discussed by Vandewalle *et al.* [17]. It is of value in setting the scene for modelling endurance at $VO_{2\,\max}$ to recall the more relevant of these.

It has been clearly shown that when subjects have their endurance at CP (as estimated from the model) tested, they are seldom able to endure for one hour, often much less ([6, 7, 8, 13]). Certainly this falls well short of the "infinitely" long endurance predicted by the model. As a consequence the work-time relationship is not linear but curved downwards and the resulting parameter estimates depend on the selection of powers for the experimental determination of CP (Bishop *et al.* [3]).

At the other extreme, the CP model predicts an infinitely high power as endurance time shrinks to zero. Clearly this cannot be so, as the concept of maximal anaerobic power is well established (Vandewalle [16]). Some finite maximal "instantaneous" power must exist, beyond which no work can be performed and endurance time is zero.

As a consequence one can deduce that some self-preservation control system must be in operation, and that the assumption of the CP model that at exhaustion all of the anaerobic capacity is completely consumed, is erroneous. Indeed Saltin and Karlsson [15] have clearly demonstrated the existence of significant anaerobic reserves at exhaustion at various power outputs.

All of the above difficulties have already been overcome by the adoption of a linear control system for power output based on the extent to which the anaerobic capacity has been consumed. The resulting 3-parameter critical power model is fully discussed by Morton [11]. Nevertheless, two further practical difficulties remain.

Firstly, the adjustment of oxygen delivery to the working muscles as required by the exercise, is not instantaneous as assumed by the CP model. In fact it may take two or three minutes to reach the required level. Wilkie [18] has recognised this problem, though his formulation has other difficulties in common with the CP model. Peronnet and Thibault [14] also recognised this problem, as well as a declining ability to sustain power output at high fractions of $VO_{2\,\max}$. Their model however does contain several arbitrary components, and is not simple. Secondly, and perhaps most problematic of all, the anaerobic reserves are not comprised of a single component, but at least two, accessible through different metabolic pathways. The three component hydraulic model of human bioenergetics proposed by Morton [9] has addressed both these problems. The former is straightforward to overcome, but the latter adds significant complexity to the modelling. Nevertheless, this model has been extended to investigate maximum power and endurance by the introduction of a control system [10].

The approach taken in this paper is therefore to construct a model, building on the previous work of several authors, which represents an energy demand and supply system with the following properties:

- It is based on the two component aerobic and anaerobic critical power concept adapted to running.
- It incorporates a linear control system for power output, dependent on the amount of anaerobic reserve consumed.
- Aerobic energy supply adjusts as a response with single fast exponential kinetics up to the level of critical power.
- It incorporates the slow component (a second exponential) which drives VO_2 beyond the equivalent of CP towards $VO_{2\,\max}$.

In so doing, endurance time both in total and at $VO_{2\,\max}$ can be modelled as a system with six parameters. They are the anaerobic distance capacity (α); critical velocity (CV); maximal "instantaneous" velocity ($V_{\max}$); the minimal velocity sufficient to elicit $VO_{2\,\max}$ (V_{vm}); and two kinetic rate parameters for the fast (r_1) and slow (r_2) components of oxygen uptake.

2.2 Important assumptions

A number of assumptions are inherent in the preceding discussion and in the basis on which the model is constructed as described above. Two of these deserve particular mention.

We assume that at the time VO_2 reaches the equivalent of CV, the primary exponential component is just complete, or very nearly complete, at which point the slow component of VO_2 begins. This assumption allows VO_2 above the equivalent of CV to be treated as a single slow rate exponential process with delay, thus simplifying the mathematics significantly. For the theoretical model of Morton [9] and its empirical verification by Barstow and Molé [1], this seems to be fairly reasonable, as the time of commencement of the slow component occurs part way into the exercise. However, for extremely high exercise levels VO_2 may reach the equivalent of CV quite quickly, say within 30 s or less. In such cases the degree of simplification our model assumes becomes more important, though this is irrelevant if exhaustion occurs before $VO_{2\,\max}$ is reached.

We also assume that once VO_2 reaches the equivalent of CV, the contribution of the aerobic power source to the requirement of exercise stabilises at this level. This derives directly from the usual interpretation of the critical power concept. This is despite the fact that the slow component may drive VO_2 significantly beyond the equivalent of CV. In other words, the contribution which the slow component of VO_2 makes does not enter into energy supply/demand considerations. Indeed Barstow and Molé [1] conjecture whether the slow component " ... represents some energy consuming function that is ancillary to, or even completely separate from, (muscle) contraction". We assume that it does. If it does not, then the critical power concept needs major reinterpretation.

2.3 Model development

For a glossary of all symbols and their definitions, go to the end of the paper.

Linear control system: Morton [10, 11] has conjectured that the maximal velocity that could be developed by a subject at any instant, is controlled by the anaerobic capacity available at that instant; although it is recognised that other causes of local fatigue may be involved. Specifically, this maximum, V_m, declines linearly from $V_{\max}$, the maximal instantaneous velocity when fully rested and nourished, to CV, the critical velocity, as the available anaerobic capacity declines from its replete value α to zero.

That is,

$$V_m = CV + \left(\frac{V_{\max} - CV}{\alpha}\right) AC. \tag{1}$$

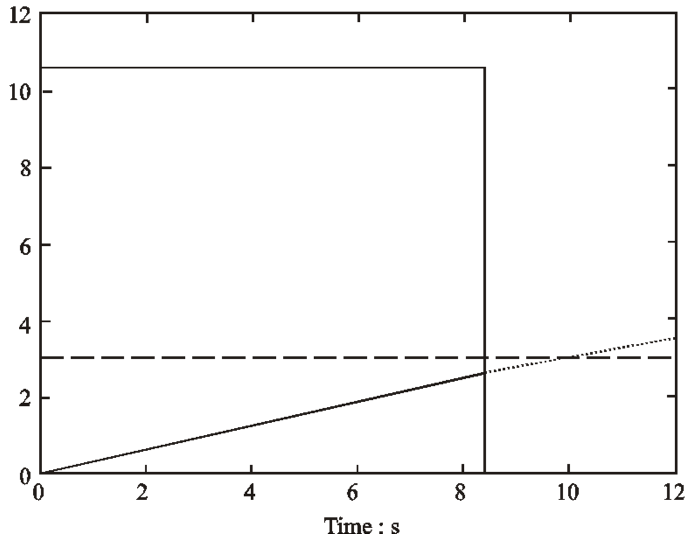

Figure 1: Energy supply/demand at very high velocity. This figure shows the energy demand at a constant velocity of $10.6\,\mathrm{ms}^{-1}$, for which endurance time is $8.42\,\mathrm{s}$. The corresponding rectangular area represents the total work done, expressed in distance units, being the work required to cover that distance. This area is comprised of aerobic work indicated by the area below the rising aerobic supply (VO_2) line, and anaerobic work above this line. Exhaustion occurs prior to $10\,\mathrm{s}$, the time VO_2 would have taken to reach the equivalent of a CV of $3\,\mathrm{ms}^{-1}$ shown by the dashed line.

Exhaustion is precipitated and therefore endurance time determined when AC declines to such a value that V_m just equals the velocity required, V. Thus we must first determine AC as a function of time spent at the velocity V, substitute in equation (1) when $V_m = V$, and solve for t^*, the endurance time at V.

Total endurance time at very high velocity: Upon commencement of exercise at velocity V, the aerobic supply, VO_2, rises mono-exponentially towards the oxygen requirement of V with a rate constant r_1 [1]. That is,

$$VO_2 = V(1 - e^{-r_1 t}).$$

It is conceivable that V may be so high that exhaustion occurs prior to VO_2 reaching the equivalent of CV. That is, for

$$0 < t^* \leq t_c = \frac{-\ln\left(1 - \dfrac{CV}{V}\right)}{r_1}, \tag{2}$$

where t_c is the time required for VO_2 to reach the equivalent of CV.

The energy supply/demand relationship in such a case can be represented by Figure 1.

From this figure we note that AC at t^* is given by

$$\alpha - \left(Vt^* - \int_0^{t^*} V(1 - e^{-r_1 t})\,dt\right) = \alpha - \frac{V}{r_1}(1 - e^{-r_1 t^*}).$$

Thus, using the linear control system of equation (1),

$$V = CV + \left(\frac{V_{\max} - CV}{\alpha}\right)\left(\alpha - \frac{V}{r_1}(1 - e^{-r_1 t^*})\right),$$

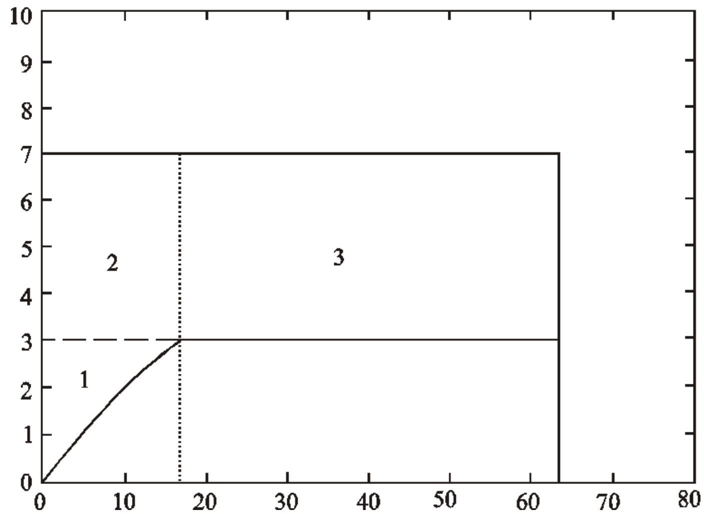

Figure 2: Energy supply/demand at high velocity. This figure shows the energy demand at a constant velocity of $7\,\mathrm{ms}^{-1}$, for which endurance time is 63.7 s. The corresponding rectangular area represents the total work done as in Figure 1. This area is comprised of aerobic work indicated by the area below the aerobic supply line which rises to reach the level equivalent to CV at 16.8 s, as shown by the dashed line at $3\,\mathrm{ms}^{-1}$, and thereafter remains constant at this level until exhaustion; and anaerobic work represented by the areas marked 1, 2 and 3.

which can be solved for t^* to yield

$$t^* = \frac{-\ln\left(1 - \dfrac{r_1\alpha(V_{\max} - V)}{V(V_{\max} - CV)}\right)}{r_1}. \tag{3}$$

This of course only applies for $0 < t^* \leq t_c$ given by equation (2), that is, for a range of velocities V, where

$$V_{\max} - \frac{CV(V_{\max} - CV)}{r_1\alpha} \leq V < V_{\max}.$$

Total endurance time at high velocity: Suppose the velocity required, V, lies in the range

$$CV < V \leq V_{\max} - \frac{CV(V_{\max} - CV)}{r_1\alpha},$$

which ensures VO_2 reaches the equivalent of CV at time t_c prior to exhaustion. The energy supply/demand relationship in this case can be represented by Figure 2.

Area 1, representing a distance in metres, is given by

$$CV \cdot t_c - \int_0^{t_c} V(1 - e^{-r_1 t})\,dt = \frac{CV}{r_1} - t_1(V - CV).$$

More simply, area 2 is given by $(V - CV)t_c$, and we note that areas 1 and 2 sum to a constant, CV/r_1, independent of V.

Area 3 is given by $(V - CV)(t^* - t_c)$.

From Figure 2 we note that AC at t^* is now given by

$$\alpha - \frac{CV}{r_1} - (V - CV)(t^* - t_c).$$

Thus, applying the control system of equation (1),

$$V = CV + \left(\frac{V_{\max} - CV}{\alpha}\right)\left(\alpha - \frac{CV}{r_1} - (V - CV)(t^* - t_c)\right),$$

and substituting for t_c from equation (2) and solving, yields

$$t^* = \frac{-\ln\left(1 - \dfrac{CV}{V}\right)}{r_1} + \frac{\alpha - \dfrac{CV}{r_1}}{V - CV} - \frac{\alpha}{V_{\max} - CV}. \tag{4}$$

We note that if VO_2 was regarded as adjusting instantaneously to CV, then $r_1 \to \infty$ as is the case in the standard critical velocity model, and equation (4) would reduce to

$$t^* = \frac{\alpha}{V - CV} - \frac{\alpha}{V_{\max} - CV},$$

which is exactly the equivalent of the 3-parameter critical power model of Morton [11].

Endurance time at $VO_{2\,\max}$: Above CV, when the first exponential rise in VO_2 is just complete or very nearly complete, the second, or slow component of oxygen uptake enters the model as described by Barstow and Molé [1] and Gaesser and Poole [4]. VO_2 will rise above the equivalent of CV, but may not necessarily rise to reach $VO_{2\,\max}$. In either case, its contribution as the aerobic energy supply is assumed to remain at CV as conjectured by Barstow and Molé [1]. VO_2 may not reach $VO_{2\,\max}$ either because at high velocity the subject becomes exhausted too soon, or because the VO_2 equivalent of the exercise is below V_{vm}. In cases where VO_2 does reach $VO_{2\,\max}$, the time taken in getting there must be subtracted from the total endurance time in order to obtain the time at $VO_{2\,\max}$.

Figure 3 shows the oxygen uptake kinetics, together with energy supply/demand features as were shown in Figure 2.

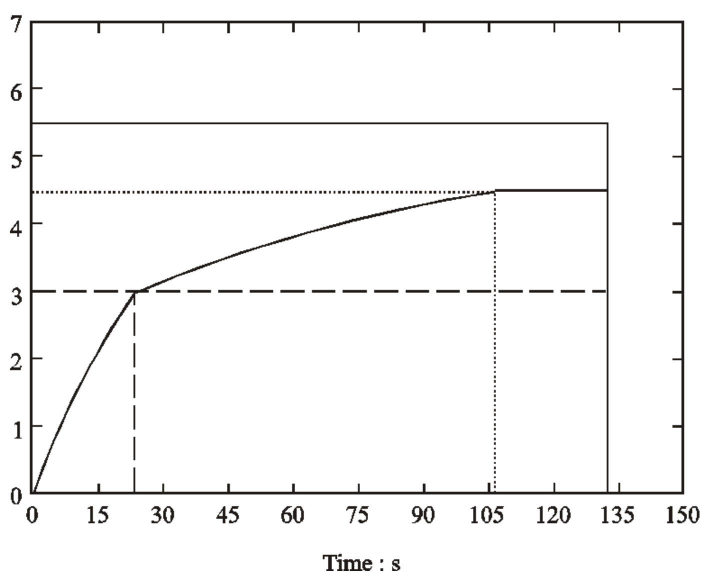

Figure 3: Oxygen uptake and energy supply/demand showing time at $VO_{2\,\max}$. For a constant velocity of $5.5\,\mathrm{ms}^{-1}$, this figure shows oxygen uptake first rising rapidly to the equivalent of a CV of $3\,\mathrm{ms}^{-1}$ by 23.65 s. This is followed by the slow component, driving it further, reaching $VO_{2\,\max}$ at an equivalent of $4.5\,\mathrm{ms}^{-1}$ by 106.1 s. Thereafter, oxygen uptake remains constant at $VO_{2\,\max}$ until exhaustion at 132.1 s. Energy supply/demand features in this illustration are exactly analogous to those of Figure 2.

We have already seen that the time taken for VO_2 to reach the equivalent of CV is given by t_c derived from equation (1). From Figure 3, we note that for the slow component

$$V_{vm} - CV = (V - CV)(1 - e^{-r_2(t_{vm}-t_c)}),$$

which yields

$$t_{vm} = t_c - \frac{\ln\left(1 - \dfrac{V_{vm} - CV}{V - CV}\right)}{r_2}. \tag{5}$$

Hence, since $t^* = t_{vm} + t_a$, equations (4) and (5) yield

$$t_a = \frac{\alpha - \dfrac{CV}{r_1}}{V - CV} - \frac{\alpha}{V_{\max} - CV} + \frac{\ln\left(1 - \dfrac{V_{vm} - CV}{V - CV}\right)}{r_2}. \tag{6}$$

The limits of V between V_m and $V_{\max}$ for which $t_a = 0$ and between which $t_a > 0$, can be found by solving

$$\frac{\alpha - \dfrac{CV}{r_1}}{V - CV} + \frac{\ln\left(1 - \dfrac{V_{vm} - CV}{V - CV}\right)}{r_2} = \frac{\alpha}{V_{\max} - CV}.$$

It will be noted that this admits two solutions, the lower of which is a little greater than V_{vm}, and the upper somewhat less than $V_{\max}$.

Illustration: Suppose $\alpha = 500\,\text{m}$, $CV = 3\,\text{ms}^{-1}$, $V_{\max} = 12\,\text{ms}^{-1}$, $V_{vm} = 4.5\,\text{ms}^{-1}$, $r_1 = \frac{1}{30}\,\text{s}^{-1}$ and $r_2 = \frac{1}{90}\,\text{s}^{-1}$. These values do not represent elite athletes. Equations (3), (4) and (6) are given by

$$\begin{aligned} t^* &= -30\ln\left(1 - 1.852\left(\frac{12 - V}{V}\right)\right), \quad \text{for } 10.38 < V \le 12\,\text{ms}^{-1}, \\ &= -30\ln\left(1 - \frac{3}{V}\right) + \frac{410}{V - 3} - 55.56, \quad \text{for } 3 < V \le 10.38\,\text{ms}^{-1}, \\ t_a &= \frac{410}{V - 3} - 55.56 + 90\ln\left(1 - \frac{1.5}{V - 3}\right), \quad \text{for } 4.73 \le V \le 7.41\,\text{ms}^{-1}. \end{aligned}$$

These curves are depicted in Figure 4.

2.4 Empirical verification

Subjects: Ten physically active male subjects (mean ± SD, age 26.4 ± 4.7 year, weight $79.1 \pm 4.5\,\text{kg}$, $VO_{2\max}$ $59.3 \pm 5.0\,\text{ml}\,\text{kg}^{-1}\,\text{min}^{-1}$) volunteered for this study. Each subject was familiar with the experimental procedures prior to the study, and all gave their written informed consent to participate in accordance with the French National Committee for Clinical Research.

Laboratory Procedures: On five occasions subjects reported to the laboratory when fasted and hydrated. On the first occasion $VO_{2\max}$ and the lowest velocity which first elicited their $VO_{2\max}$, denoted $vVO_{2\max}$, were measured using an incremental test protocol on a treadmill (Gymrol 1800). At the start, speed was set at 12 km/h (0% slope) and was increased by two km/h every three minutes up to 80% of their running speed in a 1.5 km race, and by one km/h every three minutes thereafter, until exhaustion. The criteria used for $VO_{2\max}$ were a plateau in VO_2 despite increasing running speed, a respiratory exchange ratio above 1.1, heart rate above 90% of the age-predicted maximum. The four endurance tests at 90, 100, 120 and 140% of $vVO_{2\max}$ were subsequently performed in random order for each subject at sessions separated by at least one week. At each, following a five minute warm-up

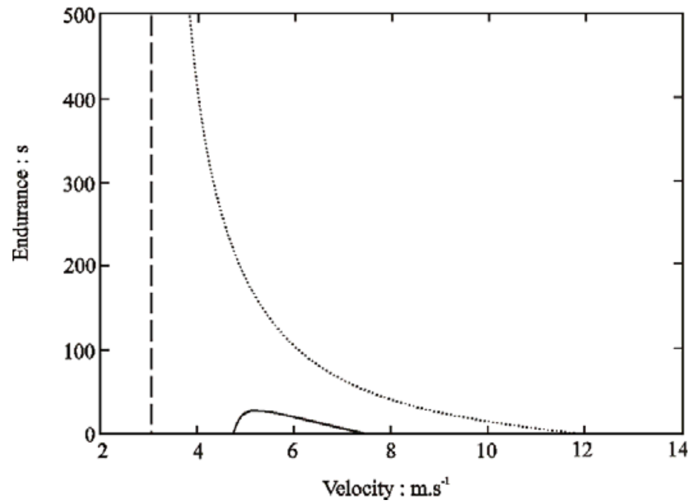

Figure 4: Model illustration. This figure shows graphical traces of the total endurance time (dotted line) and endurance time at $VO_{2\,\max}$ (solid line) as a function of velocity. The equations and parameter values are as given in the text, with $CV = 3\,\mathrm{ms}^{-1}$ shown in the figure by the vertical dashed line. Maximal endurance time at $VO_{2\,\max}$ is 27.8 s, occurring at $V = 5.24\,\mathrm{ms}^{-1}$ with corresponding total endurance time of 153.33 s.

at 60% of their $vVO_{2\,\max}$, speed was quickly increased (over less than 20 s) up to the required velocity. Subjects were verbally encouraged to run to exhaustion. The total endurance time and distance covered together with the endurance time and distance covered at $VO_{2\,\max}$ were recorded at each session for each subject.

Curve fitting: The equations modeling endurance time both in total and at $VO_{2\,\max}$ were fitted to the pooled subject data using the least squares procedure incorporated in SigmaPlot software (Jandel Scientific, San Rafael, CA).

3 Results and discussion

Table 1 lists individual results from all ten subjects.

For six subjects, 90% of their $vVO_{2\,\max}$ as measured on the incremental test was insufficient to elicit their individual $VO_{2\,\max}$. For two of these, and one other, 140% of their $vVO_{2\,\max}$ brought on exhaustion before their $VO_{2\,\max}$ could be attained. Those three subjects who were able to reach their $VO_{2\,\max}$ on all four tests, produced data which appeared skew, with endurance at 90% much longer than at any other percentage.

Equation (6) for endurance time at $VO_{2\,\max}$ contains six parameters, but it can be parametrised more simply as

$$t_a = \frac{a}{V-b} + c\ln\left(1 - \frac{d}{V-b}\right) - e,$$

which contains only five independently estimable parameters (the original formulation contains one redundant parameter). However, since there are at most four non-zero data points for any one subject, this equation cannot be fitted for each subject, and so the data must be pooled over subjects. Furthermore we note that equation (4) for total endurance time contains four parameters (total endurance time provides no information on V_{vm} or on the rate constant for the VO_2 slow component). Thus to overcome the redundancy and provide for more efficient estimation of the four parameters common to

Subject #	$VO_{2\,\max}$ *incremental test* (ml kg^{-1} min^{-1})	*Absolute velocity* (ms^{-1})	*Relative velocity* % $vVO_{2\,\max}$	*Total time limit* (s)	*Time limit at* $VO_{2\,\max}$ (s)	*Total distance limit* (m)	*Distance limit at* $VO_{2\,\max}$ (m)
1.	63	4.00	90	975	0	3900	0
		4.44	100	450	240	2000	1067
		5.33	120	150	90	480	480
		6.22	140	75	30	187	187
2.	55	4.00	90	1200	0	4800	0
		4.44	100	405	210	1800	933
		5.33	120	118	75	629	400
		6.22	140	65	15	404	93
3.	59	4.25	90	800	600	3400	255
		4.72	100	388	300	1832	1417
		5.19	120	138	75	717	390
		6.61	140	89	60	588	397
4.	66	4.44	90	840	0	3733	0
		5.00	100	225	100	1125	600
		6.00	120	75	25	450	180
		7.00	140	45	0	315	0
5.	55	4.25	90	805	525	3421	2231
		4.72	100	337	225	1591	1062
		5.19	120	142	105	737	545
		6.61	140	81	60	535	396
6.	62	4.44	90	495	95	2200	422
		5.00	100	210	135	1050	675
		6.00	120	90	45	540	270
		7.00	140	60	0	420	0
7.	62	4.00	90	770	0	3080	0
		4.44	100	360	150	1600	667
		5.33	120	135	90	720	480
		6.22	140	75	0	466	0
8.	60	4.12	90	1125	870	4640	3588
		4.58	100	524	420	2401	1925
		5.50	120	110	45	605	247
		6.42	140	70	30	449	193
9.	49	4.00	90	420	0	1680	0
		4.44	100	270	150	1200	667
		5.33	120	135	90	720	480
		6.22	140	90	45	560	0
10.	62	4.44	90	885	0	3933	0
		5.00	100	480	345	2400	1725
		6.00	120	150	105	900	630
		7.00	140	90	30	630	210

Table 1: Individual total times and distances run during the all-out runs at 90, 100, 120 and 140% of $vVO_{2\,\max}$, and specific times and distances run at $VO_{2\,\max}$.

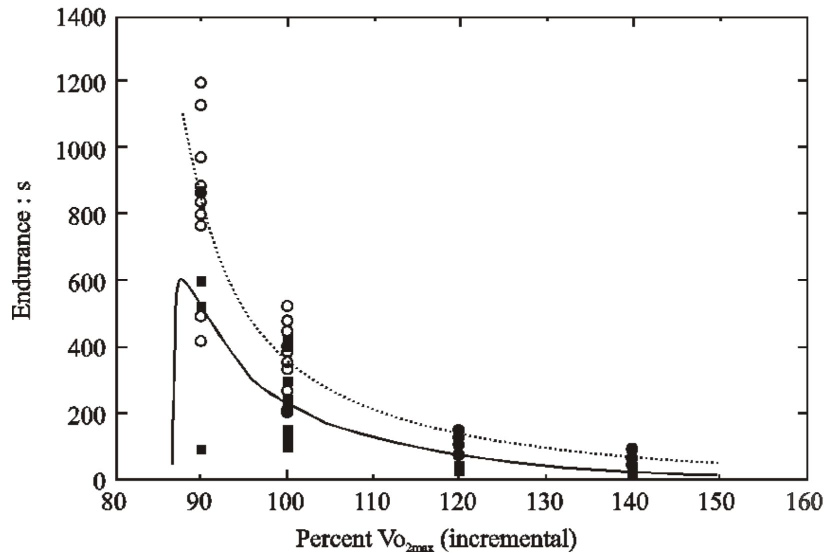

Figure 5: Endurance data and jointly fitted curves for ten subjects. This figure shows individual endurance times (both total and at $VO_{2\,\max}$) versus % $vVO_{2\,\max}$ for all ten subjects; together with fitted model curves for the whole group. Open symbols and dotted fit are for total endurance times, closed symbols and solid fit are for endurance times at $VO_{2\,\max}$. Goodness of fit and estimated parameters are as given in the text.

equations (4) and (6), these two equations are fitted jointly to both sets of endurance data pooled over subjects.

Since all subjects are of differing abilities, their velocity as % $vVO_{2\,\max}$ was used as the ordinate, rather than velocity itself. Doing so alters the units of measurement of several of the parameters, but they can easily be converted back.

Fitting % $vVO_{2\,\max}$ against endurance times yields a good fit, $R^2 = 0.821$ ($p < 0.001$), with a standard error of estimate of 127.6 s. The fitted parameters are $\alpha = 89.01$ s corresponding to 416 m, $CV = 81.4$ % $vVO_{2\,\max}$ corresponding to $3.81\,\mathrm{ms}^{-1}$, $r_1 = 0.0535\,\mathrm{s}^{-1}$ corresponding to a time constant of 18.7 s, $V_{vm} = 86.5$ % $vVO_{2\,\max}$ corresponding to $4.05\,\mathrm{ms}^{-1}$, $r_2 = 0.0034$ corresponding to a time constant of 292.7 s, and $V_{\max} = 203.7\%$ of $vVO_{2\,\max}$ corresponding to $9.53\,\mathrm{ms}^{-1}$. These fitted equations are plotted together with the full data set in Figure 5. The obvious skewness of the curve for endurance time at $VO_{2\,\max}$ is immediately apparent. The maximal endurance time at $VO_{2\,\max}$ is predicted as 603 s for a running velocity of $4.11\,\mathrm{ms}^{-1}$ being 87.9% of $vVO_{2\,\max}$ (incremental).

4 Conclusions

We have agreed that there must exist some running velocity in the range of velocities that elicit $VO_{2\,\max}$, which permits maximal endurance at $VO_{2\,\max}$. Indeed, we have shown that whereas total endurance time plotted against velocity displays a hyperbolic shape, endurance time at $VO_{2\,\max}$ plotted against velocity, displays a maximum. The bioenergetic process which produces such joint data has been modelled, producing a skewed curve for endurance data, both in total, and at $VO_{2\,\max}$ where it has a single maximum. This model has been successfully fitted to endurance data obtained from a group of ten athletes. We find that the minimal velocity to elicit $VO_{2\,\max}$ (V_{vm}) is some 10–13% below that estimated from an incremental test ($vVO_{2\,\max}$).

Glossary

AC Anaerobic capacity, usually expressed in joules, but here expressed as its distance equivalent, in metres, being the work required to cover that distance, independent of velocity.

α The value of AC when the subject is fully rested and nourished, m.

CP Critical power, W.

CV Critical velocity, ms^{-1}.

e The dimensionless exponential constant.

ln The natural logarithm to the base e.

r_1 The rate constant for the primary component of oxygen uptake kinetics, s^{-1}.

r_2 The rate constant for the slow component of oxygen uptake kinetics, s^{-1}.

t The general time variable, s.

t_a The endurance time at $VO_{2\,\max}$, s.

t_c The time taken for the primary component of oxygen uptake to reach the equivalent of CV, s.

t^* Total endurance time at constant velocity, s.

t_{vm} The time taken for oxygen uptake to reach $VO_{2\,\max}$, s.

V The constant velocity required of any given exercise bout, ms^{-1}.

V_m The maximal velocity that could be attained when the anaerobic capacity is less than α, ms^{-1}.

$V_{\max}$ The maximal velocity achievable by the subject when fully rested and nourished, ms^{-1}.

VO_2 Oxygen uptake above rest, usually expressed in $\mathrm{l\,min}^{-1}$ or $\mathrm{ml\,kg}^{-1}\,\mathrm{min}^{-1}$, but here expressed as its velocity equivalent, ms^{-1}.

$VO_{2\,\max}$ Maximal oxygen uptake, $\mathrm{l\,min}^{-1}$ or $\mathrm{ml\,kg}^{-1}\,\mathrm{min}^{-1}$.

V_{vm} The estimated minimal velocity that would drive VO_2 to reach $VO_{2\,\max}$, ms^{-1}.

$vVO_{2\,\max}$ The minimal velocity that would drive VO_2 to reach $VO_{2\,\max}$ as measured in an incremental test, ms^{-1}.

References

[1] T. J. Barstow and P. A. Molé, "Linear and nonlinear characteristics of oxygen uptake kinetics during heavy exercise", *J. Appl. Physiol.*, **71** (1991), 2099–2106.

[2] V. Billat, J. P. Koralsztein and R. H. Morton, "Time in human endurance models: from empirical models to physiological models", *Sports Med.*, **27** (1999), 359–379.

[3] D. Bishop, D. G. Jenkins and A. Howard, "The critical power function is dependent on the duration of the predictive exercise tests chosen", *Int. J. Sports Med.*, **19** (1998), 125–129.

[4] G. A. Gaesser and D. C. Poole, "The slow component of oxygen uptake kinetics in humans", *Exerc. Sports Sci. Rev.*, **24** (1996), 35–70.

[5] D. W. Hill, "The critical power concept, a review", *Sports Med.*, **16** (1993), 237–254.

[6] D. J. Housh, T. J. Housh and S. M. Bauge, "The accuracy of the critical power test for predicting time to exhaustion during cycle ergometry", *Ergonomics*, **32** (1989), 997–1004.

[7] D. G. Jenkins and B. M. Quigley, "Blood lactate in trained cyclists during cycle ergometry at critical power", *Eur. J. Appl. Physiol.*, **61** (1990), 278–283.

[8] T. M. McLellan and K. S. Y. Cheung, "A comparative evaluation of the individual anaerobic threshold and critical power", *Med. Sci. Sports Exerc.*, **24** (1992), 543–550.

[9] R. H. Morton, "A three component model of human bioenergetics", *J. Math. Biol.*, **24** (1986), 451–466.

[10] R. H. Morton, "Modeling human power and endurance", *J. Math. Biol.*, **28** (1990), 49–64.

[11] R. H. Morton, "A 3-parameter critical power model", *Ergonomics*, **39** (1996), 611–619.

[12] R. H. Morton and D. J. Hodgson, "The relationship between power output and endurance: a brief review", *Eur. J. Appl. Physiol.*, **73** (1996), 491–502.

[13] M. L. Pepper, T. J. Housh and G. O. Johnson, "The accuracy of the critical velocity test for predicting time to exhaustion during treadmill running", *Int. J. Sports Med.*, **13** (1992), 121–124.

[14] F. Peronnet and G. Thibault, "Mathematical analysis of running performance and world running records", *J. Appl. Physiol.*, **67** (1989), 453–465.

[15] B. Saltin and J. Karlsson, "Muscle glycogen utilisation during work of different intensities", in *Muscle Metabolism During Exercise*, B. Pernow and B. Saltin (editors), Plenum Press, New York (1971), 289–300, 1971.

[16] H. Vandewalle, G. Peres and H. Monod, "Standard anaerobic exercise tests", *Sports Med.*, **4** (1987), 268–289.

[17] H. Vandewalle, J. F. Vautier, M. Kachouri, J. M. Lechavalier and H. Monod, "Work-exhaustion time relationships and the critical power concept: A critical review", *J. Sports Med. Phys. Fitness*, **37** (1997), 89–102.

[18] D. R. Wilkie, "Equations describing power input by humans as a function of duration of exercise", in *Exercise Bioenergetics and Gas Exchange*, P. Cerretelli and B. Whipp (editors), Elsevier/North Holland Biomedical, Amsterdam (1980), 75–80.

HOME GROUND ADVANTAGE IN THE AUSTRALIAN NETBALL LEAGUE (COMMONWEALTH BANK TROPHY)

Pam Norton*
Department of Mathematics and Statistics
PO Box 28M
Monash University
Victoria 3800, Australia

Stephen R. Clarke
School of Mathematical Sciences
Swinburne University
PO Box 218
Hawthorn
Victoria 3122, Australia

Abstract

This paper investigates home advantage (HA) in Australian netball. Traditional measures of HA for whole competitions, such as the percentage of games won by home teams and alternative measures, such as the average margin of victory (in goals) for the home team, are calculated. Individual HAs for each team are obtained via a linear model, which takes account of teams sharing venues, using least squares methods. It is shown that HA in netball is small in comparison to other sports.

1 Introduction

Netball is the most popular women's sport (in a participation sense) in Australia. There is an estimated total of 1.2 million netball players in Australia currently (including males). Internationally, netball is played in approximately 50 countries, 45 of which are affiliated with the International Federation of Netball Associations.

The game is played between teams of seven players plus reserves on a court of fixed dimensions but varying surface, usually asphalt in lower grades. A match is played for four quarters, each lasting 15 minutes. One is added to the score each time a goal is scored.

The major Australian competition in netball, the Commonwealth Bank Trophy, began in 1997 with 14 rounds between eight clubs, and organised by Netball Australia. Adelaide, Melbourne and Sydney each have two clubs, and Brisbane and Perth one. Matches are played at venues with an indoor international standard court, on a double sprung wooden surface.

The competition is of a fairly standard type, with each team playing each other twice, once at home and once away. Some teams share the same home venue. In 1997, there was a difficulty in obtaining some venues for scheduled matches, and so some teams played others twice at home. Teams receive two points for a win and one for a draw. Ladder position is in order of wins with ties decided upon percentage ($100 \times$ total goals for $\div$ total goals against). The top four teams at the end of the home and away draw play off in a final series to determine the ultimate winner.

Winning teams average about 60 goals per game, while losing teams average about 47 goals per game. In the last three years, there have been only five draws in 168 games. Table 1 gives information about the mean winning and losing scores for the matches played each year, as well as the mean total score per game. These are surprisingly consistent for the three years.

*The authors wish to thank NETBALL AUSTRALIA, particularly Greg Dehn, for providing us with the data and other relevant information.

Year	*1997*	*1998*	*1999*
No. of matches	56	56	56
No. of draws	2	2	1
Mean winning score per match	59.5	60.5	59.9
Standard error of mean winning score per match	1.10	0.93	1.05
Mean losing score per match	46.1	47.8	47.5
Standard error of mean losing score per match	0.93	0.91	1.02
Mean total score per match	105.6	108.3	107.3
Standard error of mean total score per match	1.51	1.42	1.49

Table 1: Basic descriptive statistics for the Commonwealth Bank Trophy, 1997–1999.

Home advantage (HA) is the term used to describe the consistent finding that home teams playing in a balanced home and away sporting competition win over 50% of the games played. It is believed to occur because of circumstances such as familiar surroundings, crowd support influencing both players' behaviour and officials' decisions and travel factors associated with the opposition.

In this paper we investigate home advantage (HA) for the competition overall and for individual clubs, as a first step in explaining the differences in terms of the playing characteristics of the clubs.

When the competition first started in 1997, there were problems obtaining home grounds for three of the matches. These matches were played at the home ground of the "away" team. In addition, home grounds are changing. The AKAI Melbourne Kestrels and the Cenovis Melbourne Phoenix both played at the Waverley Netball Centre in 1997 and 1998, but the Cenovis Melbourne Phoenix had the Melbourne Sports and Aquatic Centre as their home ground for 1999. They both planned to move to a new venue in the Docklands precinct in 2000, but this venue is unlikely to be ready in time. The Decoré Sydney Sandpipers and the TAB Sydney Swifts both played at the Anne Clark Netball Centre at Lidcombe in 1997 and moved to the State Sports Centre in 1998. In 2000, the Decoré Sydney Sandpipers plan to move to the Penrith Sports Stadium.

2 Traditional measures of HA

The phenomenon of HA has been the basis of considerable study since the 1970s. Courneya and Carron [3] give a comprehensive review of this work. In their table surveying the "what" of home advantage, they list many studies which give the home win percentage on the basis of either points or wins/losses (with tied games excluded in the latter case). Two of the studies also give the difference between home and away winning percentages. Other measures of HA, such as winning margins, are not investigated in these studies.

In the first detailed study of HA by Schwartz and Barsky [5], the percentages of matches won by the home team were found to be 53% in major league baseball, 55% in professional (American) football and 59% in college (American) football, 64% in ice hockey and 64% in college basketball. Table 2 gives the percentage of wins by the home team for each year for the round matches in the Commonwealth Bank Trophy, omitting those played by teams sharing a common venue. A draw counts as half a win. These are similar to the percentages for other sports, although the figure in 1998 is surprisingly low.

Table 2 also gives the average margin of victory of the home team. This is quite small in comparison to the total number of goals scored in a game. Since the percentage of home wins depends on the variation in the performance level of the teams as well as their HA, it is difficult to compare HAs between seasons or between different sports. The values of the ratio of the average total goals scored in a match to the average margin of victory by the home team, r, given in Table 2, allows us to make such comparisons. Stefani and Clarke [8] determined the values of r for soccer (three European cups) – 3, hockey (USA) – 10, professional football (USA) – 12, Australian Rules football – 21, and baseball (USA) – 34. A value of r of 71 can be calculated from the data given in Snyder and Purdy [6] for collegiate

Year	*No. of games*	*Percentage of wins by home team*	*Average margin of victory of home team* (HGA)	*Average total goals scored in a match* (GOALS)	*Ratio of* GOALS *to* HGA (r)
1997	50	62	3.78	105.6	27.94
1998	50	50	0.16	108.3	676.88
1999	52	58	2.02	107.3	53.12
All	152	57	2.01	107.1	53.28

Table 2: Traditional measures of home advantage (HA) in Australian women's netball (Commonwealth Bank Trophy).

basketball. The values given for r in Table 2 indicate that netball has a low home advantage compared to other sports, except for basketball. This is probably due to factors such as having a fixed dimension indoor playing surface and a limited number of games during the season to lessen travel factors.

3 Linear model

The percentage of wins by the home team depends as much on the closeness of the competition and the variability of results as on HA. If all teams are of the same strength, then a small HA will result in most home teams winning. However, if teams are wide apart in strength, than a small HA will have little influence on the final result. Hence HAs are best investigated by models that incorporate a strength measure of the individual teams as well as a HA.

There are several linear models that can be used to model results between two teams. Clarke [1] presents three, and the third is chosen here because it assumes that each team has a different home advantage and takes account that some matches may be played between two teams who share a home ground. This model has been used by Stefani and Clarke [7, 8] to investigate individual HAs in Australian rules, by Harville and Smith [4] to investigate HA in basketball, and by Clarke and Norman [2] to find individual HAs for all English soccer teams.

Let w_{ij} be the winning margin (in goals) when the home team i plays away team j (in match k). Let u_i be a rating for team i, which is a measure of team i on a neutral ground. This summarises a team's ability, form or level of performance. Let h_i be the home ground advantage of team i, which includes all that is advantageous for team i playing at home and all that is disadvantageous for any other team playing at team i's home ground. Let c_{ij} be a common home ground factor, which takes the value 1 if team i and team j have a common home ground, and 0 otherwise. Let e_{ij} be a random error, usually assumed to have a mean of zero. Then

$$w_{ij} = u_i - u_j + h_i - c_{ij}h_j + e_{ij}.$$

The term $c_{ij}h_j$ is necessary as some teams share a common home ground.

Since the ratings u_i are relative, we add the constraint that the u_i sum to 800. (The number 800 was chosen so that if all teams were of equal ability, they would each have a rating of 100. Any other number could have been chosen here.) This model, with the additional constraint on the u_i, was fitted to the individual match results for each of the three years with a standard regression package. The values for u and h for each of the teams are given in Table 3. In each case the overall model was significant at the 0.0001 level, with $R^2 = 0.65$ (1997), 0.72 (1998) and 0.75 (1999). The high values of R^2 reflect the low variability in netball. Clarke and Norman [2] obtained a value of R^2 of 0.19 for English soccer, reflecting its high variability, and Clarke [1] obtained a value of R^2 of about 0.40 for Australian Rules football.

The range of the ratings (highest rating − lowest rating) has been increasing. In 1999, when the team with the highest rating played that with the lowest rating, the model predicts that the highest team

	1997		1998		1999	
Team	u	h	u	h	u	h
AAMI Adelaide Thunderbirds	113.36	2.21	113.19	−7.20	116.98	0.56
Firestone Queensland Firebirds	91.26	1.72	97.10	−8.83	96.96	−7.09
Adelaide Wendy's Ravens	95.21	4.48	106.69	−6.80	102.98	3.16
SmokeFree Perth Orioles	91.48	6.74	87.27	−1.17	81.39	11.41
Sydney TAB Swifts	106.75	−5.40	106.32	8.61	111.36	−7.89
AKAI Melbourne Kestrels	92.06	−0.59	101.72	−0.58	99.42	9.24
Decoré Sydney Sandpipers	102.33	2.13	85.49	13.64	98.86	−5.89
Cenovis Melbourne Phoenix	107.55	11.25	102.22	5.33	92.06	10.07
Maximum HA applicable		11.84		13.64		11.41
Range	22.10		27.70		35.59	

Table 3: Individual ratings and HAs for teams in Commonwealth Bank Trophy Women's Netball Competition 1997–1999.

would be ahead by 36 goals before allowances for HA. For 1997 and 1998, the maximum HA applicable was about half the range of the ratings, but in 1999, this was down to about a third.

Some teams are shown with a negative HA in some years. Clarke and Norman [2] show that the apparent HA of any side is affected by the HAs of the others. An end of season ladder for the 1999 Commonwealth Bank Trophy, with the values of u and h included, is given in the appendix. Three of the teams have a negative HA. Sydney TAB Swifts have won more matches away than at home, and also have a larger away goal difference than home goal difference, so their negative HA is understandable. Decoré Sydney Sandpipers have also won more matches away than at home, whereas their home goal difference is not as bad as their away goal difference. The third team with a negative HA, the Firestone Queensland Firebirds, have won as many matches at home as away, and their home goal difference is not as bad as their away goal difference. Certainly the HAs of the other teams have had some effect in determining the HA of the Sandpipers and the Firebirds.

The data from the three years were combined, and the model above fitted to the data (using the REG procedure in SAS), assuming the home advantage for each team remained constant over the three years but allowing the performance of individual teams to vary each year. The p-value for the model was 0.0001, with $R^2 = 0.67$. The HA for each club over the three years is given in Table 4. It is not surprising that the Perth based team has one of the higher HAs, due to travel factors. What is surprising is that one of the Melbourne based teams, Cenovis Melbourne Phoenix, has the highest HA, and appears to have been the most consistent in terms of HA over the three years.

Team	HA, 1997–1999
Cenovis Melbourne Phoenix	9.57
SmokeFree Perth Orioles	5.61
Decoré Sydney Sandpipers	3.21
AKAI Melbourne Kestrels	2.94
Adelaide Wendy's Ravens	0.21
AAMI Adelaide Thunderbirds	−1.59
Sydney TAB Swifts	−1.73
Firestone Queensland Firebirds	−4.88

Table 4: HA of teams, assuming constant from 1997–1999.

4 Further analysis of individual HAs

Do different teams have different HAs or are the above differences due to random variation? For each of the three years, and for the combined three years, an F-test was done under the hypothesis $H_0 : h_1 = h_2 = h_3 = h_4 = h_5 = h_6 = h_7 = h_8$, with p-values given in Table 5.

Year	p-value
1997	0.8111
1998	0.1307
1999	0.0904
1997–1999 combined	0.1145

Table 5: Test results for the h_is being the same.

The results are inconclusive. However, it must be noted that generally for each year the individual HAs have been calculated from only seven (or fewer) observations.

5 Conclusions

The Commonwealth Bank Trophy has only been in operation for three years. The HA for each of the teams over each of the three years and for the three years combined has been calculated. It has varied quite a bit for some teams over the three years. The maximum HA to apply to a game is shown to vary between one-third and one-half of the difference in the ratings of the highest and lowest teams. There is a significant HA in netball, but it appears to be lower than for other sports.

References

[1] S. R. Clarke, *Performance Modelling in Sport*, PhD thesis, Swinburne University (1997).

[2] S. R. Clarke and J. M. Norman, "Home ground advantage of individual clubs in English soccer", *The Statist.*, **44**(4) (1995), 509–521.

[3] K. S. Courneya and A. V. Carron, "The home advantage in sport competitions: A literature review", *J. Sport & Exercise Psychol.*, **14** (1992), 13–27.

[4] D. A. Harville and M. H. Smith, "The home-court advantage: How large is it, and does it vary from team to team?", *Amer. Statist.*, **48**(1) (1994), 22–28.

[5] B. Schwartz and S. E. Barsky, "The home advantage", *Social Forces*, **55** (1977), 641–661.

[6] E. E. Snyder and D. A. Purdy, "The home advantage in collegiate basketball", *Sociology of Sport J.*, **2** (1985), 352–356.

[7] R. T. Stefani and S. R. Clarke, "Australian Rules football during the 1980's", *ASOR Bull.*, **10**(3) (1991), 11–15.

[8] R. T. Stefani and S. R. Clarke, "Predictions and home advantage for Australian Rules football", *J. Appl. Statist.*, **19**(2) (1992), 251–261.

Appendix End of season ladder, 1999 Commonwealth Bank Trophy

Team	HW	HD	HL	Hf	Ha	HGD	AW	AD	AL	Af	Aa	AGD	GD	Points	u	h
AAMI Adelaide Thunderbirds	6	1		432	295	137	5		2	422	299	123	260	23	116.98	0.56
Sydney TAB Swifts	5		2	393	351	42	6		1	420	359	61	103	22	111.36	−7.89
AKAI Melbourne Kestrels	6		1	417	357	60	4		3	353	362	−9	51	20	99.42	9.24
Adelaide Wendy's Ravens	5		2	388	354	34	4	1	2	387	359	28	62	19	102.98	3.16
Cenovis Melbourne Phoenix	4		3	410	403	7	2		5	379	446	−67	−60	12	92.06	10.07
Decoré Sydney Sandpipers	1		6	369	397	−28	2		5	348	397	−49	−77	6	98.86	−5.89
SmokeFree Perth Orioles	3		4	311	380	−69	0		7	274	425	−151	−220	6	81.39	11.41
Firestone Queensland Firebirds	1		6	339	413	−74	1		6	367	412	−45	−119	4	96.96	−7.09

HW – number of home wins
HD – number of home draws
HL – number of home losses
Hf – number of goals for in home matches
Ha – number of goals against in home matches
HGD – goal difference for home matches (= Hf – Ha)
AW – number of away wins
AD – number of away draws
AL – number of away losses
Af – number of goals for in away matches
Aa – number of goals against in away matches
AGD – goal difference for away matches (= Af – Aa)
GD – goal difference for season (= HGD + AGD)
Points – points on ladder at end of season
u – team rating according to model
h – home advantage according to model

OPTIMISING DOWNWIND SAILING

Elliot Tonkes
School of Information Technology
Bond University
Gold Coast
Queensland 4229, Australia

Abstract

High speed sailing craft are capable of superior downwind performance when they utilise downwind tacking instead of following a straight line between buoys. Within each of the downwind tacks, this paper analyses whether a straight line path is superior to an oscillating path.

1 Introduction

In the early days of yacht racing, downwind legs (when the true wind is coming from behind the vessel) were raced by sailing almost directly between the buoys. With the advent of higher speed boats, it has been found to be preferable to zig-zag (tack) downwind. The main reason is associated with the stall of an aerofoil. When sailing dead downwind, the apparent wind is directly from behind and the sail acts as a bluff body. When sailing at an angle to the wind, the apparent wind shifts forward, reducing the angle of attack of wind onto the sail. This allows a smooth flow over the sail and enhanced lift (Tritton [10]). Of course, the compromise is that the force is not directly in line with the motion of the boat and that the boat must travel further. However, these disadvantages are consistently overcome in racing.

It is not difficult, given the velocity profile of a yacht to determine the best angle for piecewise-linear downwind tacking. In this paper, we investigate the dynamic problem of determining whether a straight line is the optimal path.

The assumption of constant wind and no waves is made. Even in this perfect environment, it has been maintained by some sailors that an oscillatory path is preferred over straight line motion. The basic idea is that the yacht turns upwind a little, accelerates and the apparent wind shifts forward. This will then allow the yacht to turn downwind, and while momentum is maintained, the vessel is able to sail deeper downwind. As the yacht slows, she turns upwind to prevent flow separation and the cycle continues.

The brachistochrone is a famous problem in the calculus of variations. Put simply, find the best curve $y(x)$ for a wire, with end points $(0,0)$ and $(a,-b)$ to minimise the travelling time of a bead on the wire. Although the brachistochrone problem appears similar to the sailing problem (see Figure 1), there are fundamental differences which make the yachting problem considerably more difficult. The sailing problem does not reside in the environs of a conservative force field and the forcing function possesses a nonlinear dependence on the speed and direction of the boat at each (x,y). Analytical integration of the differential equations appears difficult, so numerical implementation is required.

An analysis of baseball by Harman [4] finds the best path for a runner to sprint from the home base to the second base subject to the constraint of contacting first base. The sailing problem is more difficult than this analysis because the endpoint of the motion is not fixed. Furthermore, the yacht's velocity must vary in a periodic way. To maintain a sustainable motion, the yacht's exit velocity must correspond with its entry velocity. Otherwise, the initial conditions at the beginning of the cycle cannot be attained and the yacht's motion will be speeding up (initial speed underestimated) or slowing down (initial speed overestimated). At equilibrium, the entry and exit speeds correspond, and we assume that the yacht maintains this periodic motion for the downwind leg.

To date, most work in path optimisation of yachts has been based on steady-state velocities (Philpott *et al.* [9]). For long races the steady-state velocity profiles (and meteorological data) are used to plan courses. The problem in this paper is based on much shorter timescales and hence it is a dynamic problem that must focus on the underlying forces, rather than the steady-state velocities.

The work by Harris [5] analysed the downwind performance of yachts with the influence of waves. The interaction of waves with the yacht produces forces on the hull which in turn affect the speed of the vessel, and consequently influence the relative wind vector. Harris produced a dynamic model which deals with forces, and produced a more accurate velocity profile of the yacht. This is similar to the aim of this paper, except in this case the influence is derived from perturbations in steering the yacht, rather than the effect of waves.

Several articles [2, 3] describe the interaction of relative wind with yacht sails, and the resulting force and drag.

Figure 1 describes the physical arrangement of the problem and introduces some notation.

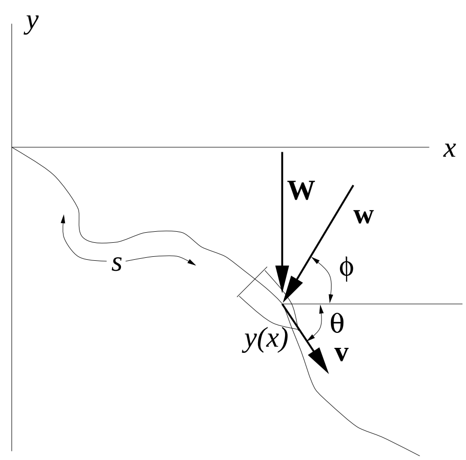

Figure 1: Diagram of nomenclature. The curve is a segment of a long downwind leg.

2 Instantaneous acceleration model

As a preliminary study, suppose that acceleration is instantaneous. A force diagram is then superfluous, and the only required diagram relates speed, boat direction and wind strength. A typical polar velocity diagram is shown in Figure 2.

To optimise performance, one simply needs to maximise the downwind component of velocity, termed the *velocity made good*, $v_{MG} = v(\theta)\sin(\theta)$ (see Figure 1). Any other path is clearly inferior. The velocity profile in Figure 2 for a Tasar establishes the optimal straight-line sailing angle in each particular wind strength.

3 Assumptions

Let the true wind vector be $\mathbf{W} = -W\mathbf{j}$ and the true boat speed vector be $\mathbf{v}$. Suppose the mass of the crew and boat is M. Let the direction of the relative wind be $\mathbf{w} = \mathbf{W} - \mathbf{v}$. Let $(x, y(x))$ describe the path followed by the vessel. Let $s(t)$ represent the arc length of the curve parametrised with time.

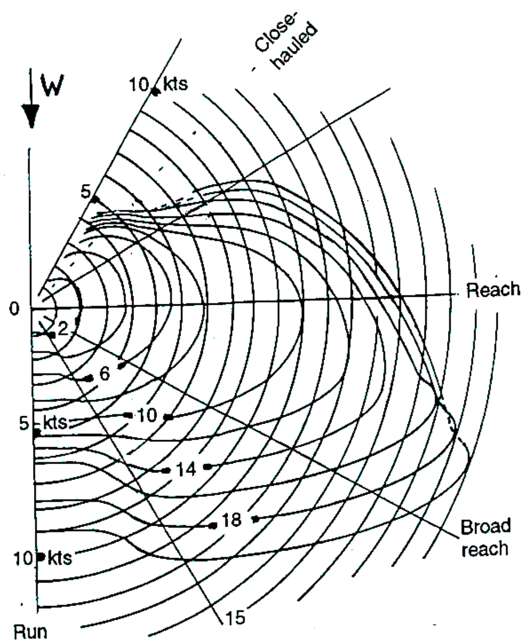

Figure 2: Polar velocity profile for a Tasar dinghy [1]. For a given W, the best v_{MG} is the *lowest* point on the velocity profile.

Assume that the driving force **F** always acts in the direction that the boat is pointing and that it depends on the relative wind vector only. The function **F** will contain a component of drag associated with air drag. Assume that the water drag force **D** also acts in the tangential direction with force dv^2 for some constant d. Quadratic water drag is verified for kayak hulls at low speed (Lazauskas and Winters [8]) and over a larger range of speeds, experimental evidence by Havelock [6] supports the assumption that drag is approximately quadratic with speed. The experimental data by Bethwaite [1] is used to estimate the drag coefficient in this paper. Assume that no energy is lost when turning the vessel (that is, neglect rudder drag). From these assumptions, the motion is parametrised by a single variable, since the path is predetermined by $y(x)$ and velocity $\mathbf{v}$ is completely described by $y(x)$ and $\dot{s}(t)$.

Leeway (slipping sideways in the water) is neglected. Since this problem considers downwind motion, the direction that the boat is pointing closely corresponds with the direction of the instantaneous velocity.

This study considers a continuous dependence of boat speed upon wind speed, and so the possibility of planing is not entertained. A small change in driving force produces a small change in the boat speed.

Differentiation with respect to spatial variables (x or s depending on the context) is denoted ($'$) and differentiating with respect to time is denoted ($\dot{}$).

4 Formulation

Since the forces F and D act in the tangential direction, they may be treated as scalar quantities and by Newton's second law,

$$\ddot{s} = \frac{1}{M}\left(F(\mathbf{w}) - D(\dot{s})\right).$$

The path $y(x)$ is prescribed before the equation of motion is solved. The relationship between $y(x)$ and $s(x)$ is dictated by the usual arc length differential equation:

$$\frac{ds}{dx} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}. \tag{1}$$

The relative wind, in the (x, y) coordinate scheme, is given by $\mathbf{w} = \mathbf{W} - \mathbf{v}$. The velocity of the boat at (x, y) is given by

$$\mathbf{v} = \frac{v}{\sqrt{1+y'^2}}\,\mathbf{i} + \frac{vy'}{\sqrt{1+y'^2}}\,\mathbf{j} = \frac{\dot{s}}{\sqrt{1+y'^2}}\,\mathbf{i} + \frac{\dot{s}y'}{\sqrt{1+y'^2}}\,\mathbf{j},$$

and, consequently, the relative wind velocity is given by:

$$\mathbf{w} = -\frac{\dot{s}}{\sqrt{1+y'^2}}\,\mathbf{i} + \left(-W - \frac{\dot{s}y'}{\sqrt{1+y'^2}}\right)\mathbf{j}.$$

Note that we expect $y' < 0$ on the optimal path. The magnitude of the relative wind speed is

$$w^2 = |\mathbf{w}|^2 = \left(\frac{\dot{s}}{\sqrt{1+y'^2}}\right)^2 + \left(W + \frac{\dot{s}y'}{\sqrt{1+y'^2}}\right)^2. \tag{2}$$

For the *direction* of the relative wind, we really want the direction relative to the boat's bow, not relative to the fixed (x, y) coordinate scheme. Let θ represent the angle of direction of the boat below the x-axis. Let ϕ be the angle of the relative wind above the x-axis. The direction of the relative wind will then be described by $\theta + \phi$.

The angle θ is discerned from $\tan\theta = -dy/dx$. The angle ϕ can be determined from:

$$\tan\phi = \frac{W + \dfrac{\dot{s}y'}{\sqrt{1+y'^2}}}{\dfrac{\dot{s}}{\sqrt{1+y'^2}}} = \frac{W\sqrt{1+y'^2} + \dot{s}y'}{\dot{s}}.$$

For the relative angle $\theta + \phi$,

$$\begin{aligned}\tan(\theta+\phi) &= \frac{\tan\theta + \tan\phi}{1 - \tan\theta\tan\phi} \\ &= \frac{W\sqrt{1+y'^2}}{\dot{s} + y'\left(W\sqrt{1+y'^2} + \dot{s}y'\right)}.\end{aligned} \tag{3}$$

The angle $\theta + \phi$ lies in the range $(0, \pi)$ for a yacht sailing with angle $\theta \in (-\pi/2, \pi/2)$. Thus, the motion is governed by

$$\ddot{s} = \frac{1}{M}\left(F(w, \theta+\phi) - d\dot{s}^2\right) \tag{4}$$

where w and $\theta + \phi$ are prescribed by (2) and (3).

The form of F is not easily established. Typically, the velocity polar diagram provides a description of equilibrium *boat speed* with a given true wind vector, $-W\mathbf{j}$.

The force function $F(w, \theta + \phi)$ will vary between yacht rigs, but it maintains a common shape (Couser [3], Hedges *et al.* [7]). For this paper, an estimate for the function $F(\cdot\,,\cdot)$ was derived from the velocity polar diagram (Figure 2) which provides the steady-state straight-line speed at each true wind strength. Working backwards from any point on the diagram, the relative wind strength and direction can be calculated and using the known quadratic drag function, the force F can be deduced. The resulting data is fitted to a simple function.

Alternatively, a velocity prediction program (VPP) can be used to derive F. In the preparation of this paper, a VPP [11] was trialed. This program is available as a Java Applet on the World Wide Web. From the resulting form of $F(\cdot,\cdot)$, the velocity polars were calculated and compared with known data. However, the fit was inaccurate, and we have relied on the former method to estimate the driving force function.

Suppose the relative wind strength is restricted to a neighbourhood of fifteen knots ($7\,\mathrm{ms}^{-1}$). There is an angle Θ which lies between $\pi/4$ and $\pi/2$ such that for $0 \leq \theta + \phi \leq \Theta$, the driving force is increasing with $\theta + \phi$. Clearly, when a boat sails straight into the wind there can be no forward driving force. As the yacht turns to leeward, the angle of attack increases and more force is imparted. When sailing with relative wind angle Θ, there is considerable forward component of force and the flow is still attached to the sail. However, when the relative wind direction exceeds an angle of Θ, flow separation sets in, and there can be a marked decrease in the lift produced by the sail.

Although flow separation, and the transition from laminar to turbulent flow depends in a complex way on the Reynolds number and the sail shape, if the apparent wind speed does not considerably vary, then the flow remains in the same regime. Consequently, the critical angle Θ does not vary significantly as the wind speed varies and we may assume that $F(w, \theta+\phi) = f(\theta+\phi)g(w)$. The separation of laminar or turbulent flow from the leeward side of an aerofoil (leading to stall) is further discussed in Tritton [10] (Chapter 12). As a consequence of Bernoulli's law for a fixed sail shape, and fixed angle of relative wind, the lift (and hence forward driving force) is proportional to the square of the relative wind speed. This assumption is inherent in the VPP [11]. Thus, we suppose that $F(w, \theta + \phi) = f(\theta + \phi)w^2$.

The drag coefficient from the data in Bethwaite [1] was estimated as $14\,\mathrm{kg\,m}^{-1}$. A good fit for the data was:

$$F(w, \gamma) = 1.2w^2 \frac{(\gamma + 0.5)^{1.4}}{0.2 + (1.4 - \gamma + 0.5)^2}.$$

This force function is plotted in figure 3 and one of its polar velocity profiles is shown. In the wind of $7\,\mathrm{ms}^{-1}$, the velocity polar reveals that $\max_\theta v_{MG} = 2.55\,\mathrm{ms}^{-1}$ when $\theta \approx \pi/4$. We remark that v_{MG} is quite insensitive to θ in this neighbourhood.

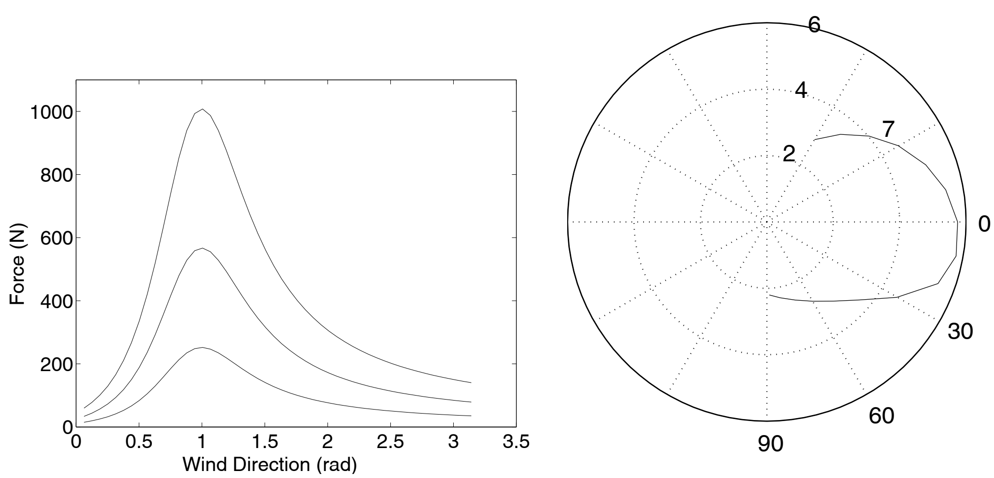

Figure 3: (a) Force function $F(w, \gamma)$ at relative wind speeds of 5, 7.5 and $10\,\mathrm{ms}^{-1}$. (b) A velocity polar derived from (a), with true wind at 7 ms^{-1}.

5 Small perturbations

A simplified version of the equations of motion can be obtained for small perturbations from straight line motion. Suppose that the path of the motion is described by $y(x) = mx + \epsilon\tilde{p}(x)$, where $m < 0$, $|\epsilon|$ is small and $\tilde{p}(\cdot)$ is normalised over the period of motion. For sufficiently small ϵ, the arc length $s(x)$ does not vary significantly from straight line motion. However, the derivative $y'(x)$ will impart significant differences in the velocity of the motion. In order to preserve periodicity and maintain constant end points, constraints are imposed on $\tilde{p}(\cdot)$.

Suppose that the motion traverses $0 \leq x \leq X$, or equivalently, $0 \leq s \leq L$, where $\sqrt{1+m^2}\,x = s$. Define the perturbation in terms of arc length by introducing $p(s) \equiv \tilde{p}(s/\sqrt{1+m^2})$. Thus, $(d/dx)\tilde{p}(x) = ds/dx = \sqrt{1+m^2}\,p'(s)$. The periodicity and endpoint conditions imply:

$$p(0) = p(L) = 0 \quad \text{and} \quad p'(0) = p'(L). \tag{5}$$

These assumptions mean that $F(w, \theta+\phi)$ can be written explicitly in terms of s and $\dot{s}$. For brevity, we shall write $P(s) \equiv \sqrt{1+m^2}\,p'(s)$. Then we can write $F = f(\theta+\phi)w^2$, where

$$w^2 = \left(\frac{\dot{s}}{\sqrt{1+(m+\epsilon P(s))^2}}\right)^2 + \left(W + \frac{\dot{s}(m+\epsilon P(s))}{\sqrt{1+(m+\epsilon P(s))^2}}\right)^2 \tag{6}$$

and

$$\tan(\theta+\phi) = \frac{W\sqrt{1+(m+\epsilon P(s))^2}}{\dot{s} + (m+\epsilon P(s))\left(W\sqrt{1+(m+\epsilon P(s))^2} + \dot{s}(m+\epsilon P(s))\right)}. \tag{7}$$

The tan operator is inverted by taking the branch of $\tan^{-1}$ which assumes the range $(0,\pi)$. Continuity is not contravened because $\theta+\phi$ will never be in a neighbourhood of zero, the location of the discontinuity.

If the mean speed is $\dot{s}_0$, then the extra time to traverse the slightly longer path will approximately be:

$$\delta T(\epsilon) = \frac{1}{\dot{s}_0}\int_0^L \left(\frac{\sqrt{1+(m+\epsilon P(s))^2}}{\sqrt{(1+m^2)}} - 1\right) ds. \tag{8}$$

Linearisation and second order expansion

Make the first order expansion for $s(t)$ in terms of the perturbation parameter ϵ:

$$s(t) = s_0(t) + \epsilon s_1(t), \tag{9}$$

where $s_0(t)$ is the solution to the equations of motion when $\epsilon = 0$, that is, straight line motion.

The steady-state, straight-line, unperturbed motion is solved by $s(t) = \dot{s}_0 t \equiv v_0 t$, with constant relative wind speed w_0 and direction $(\theta+\phi)_0 \equiv \gamma_0$.

The only appearance of the parameter ϵ is in conjunction with $s_1(t)$ or $P(s)$. Consequently, linearisation of the equations of motion in terms of ϵ, yields:

$$M\ddot{s}_1 = A\dot{s}_1 + Bp'(s_0).$$

Since $s_0(t) = v_0 t$, the term $p'(s_0)$ can be written as a function of time. For this paper, we shall consider sinusoidal perturbations, so $p(s)$ assumes the form $\sin(2\pi s/L)$, giving $p'(s) = (2\pi/L)\cos(2\pi s/L)$. Thus $p'(s_0)$ can be replaced with $(2\pi/L)\cos(2\pi v_0 t/L) \equiv (2\pi/L)\cos(\omega t)$.

The initial conditions from the nonlinear system imply that $s_1(0) = 0$. Because the motion is periodic over $s \in [0, L]$, periodic velocity constraints must be enforced: $\dot{s}_1(0) = \dot{s}_1(T)$. Subsequently,

$$s_1(t) = \frac{B\left(M\omega - A\sin(\omega t) - M\omega\cos(\omega t)\right)}{v_0(A^2 + \omega^2 M^2)}.$$

The first order expansion of (8) yields,

$$\delta T(\epsilon) = \frac{1}{\dot{s}_0}\left(p(L) - p(0)\right)\frac{m\epsilon}{\sqrt{1+m^2}} + O(\epsilon^2) = O(\epsilon^2).$$

Clearly $s_1(T) = 0$ and no speed difference is discernible from the first order approximation.

For a second order expansion, suppose that $s(t) = s_0(t) + \epsilon s_1(t) + \epsilon^2 s_2(t)$. Collecting the $O(\epsilon^2)$ terms from the equation of motion yields the governing equation for $s_2(t)$. After simplification,

$$M\ddot{s}_2 = a\dot{s}_2 + b(\dot{s}_1)^2 + c(\dot{s}_1 p') + e(\dot{s}_1 p'') + f(p')^2, \tag{10}$$

where $s_1(t)$ is derived from the first order approximation, $p' = (\omega/v_0)\cos(\omega t)$ and $p'' = -(\omega/v_0)^2\sin(\omega t)$. The constants a, b, c, e and f depend on M, ω, γ_0, W, k, m, v_0 and d.

The solution to (10) with initial condition $s_2(0) = 0$ contains constant terms, linear terms, periodic terms and terms exponential in t. The linear terms do not contain the constant of integration and constitute a growth (or retardation) of $s_2(t)$ over time as the cycles progress. The exponential terms are simply correction terms which will disappear with an appropriate selection of initial conditions $\dot{s}_2(0)$.

From (8), the extra time to travel the convoluted path is of $O(\epsilon^2)$. Thus, the effect of the perturbation $\epsilon p(s)$ on the speed $\dot{s}(t)$ is of order ϵ^2. Unfortunately, the extraction of constants a, b, c, e and f from the expansion of $F(w,\gamma)$ involves prohibitively messy algebra. Although we have established that a second order effect exists, we have not deduced whether the perturbation improves or degrades performance. Instead, our next step involves numerical integration of the differential equations of motion.

6 Large variations

In order to solve the problem for "large" perturbations from straight-line motion, the arc length must be taken into consideration. A large perturbation implies that $s(x)$ is significantly different from $\sqrt{1+m^2}\,x$. As a first step, for a given $y(x)$, equation (1) is solved to yield $s(x)$. A one-to-one relation is thus derived between s and x and the differential equation can be numerically integrated. An iterative scheme is implemented to ensure periodicity in $\dot{s}$. This guarantees that the speed at the beginning of the path is sustainable (equal to the exit speed).

Numerical method

The numerical method follows the procedure outlined below.

Step 1 Guess a path $y(x)$, $0 \le x \le X_{span} \equiv L/\sqrt{1+m^2}$, which satisfies $y'(x) < \infty$ on $[0, X_{span}]$ and $y'(0) = y'(X_{span})$.

Step 2 Solve for $s(x)$ by numerically integrating (1) over $[0, X_{span}]$.

Step 3 Guess initial velocity $v_0 = \dot{s}(t=0)$.

Step 4 The differential equation (4) is numerically integrated using Matlab's ODE45 function:

$$\begin{aligned} \dot{s} &= \zeta, \\ \dot{\zeta} &= \frac{1}{M}\left(F(w, \theta+\phi) - d\zeta^2\right), \\ s(0) &= 0, \quad \zeta(0) = v_0. \end{aligned}$$

The terms w and $\theta+\phi$ are themselves functions of ζ and dy/dx. However, **Step 2** yields a one-to-one mapping between s and x. Consequently, dy/dx can be evaluated for any given s.

Step 5 The exit speed $\dot{s}(x = X_{span})$ is determined by reading off the numerical solution. This information is used to make an improved guess for v_0 by a bisection method: $v_0 \to \frac{1}{2}\left(v_0 + \dot{s}(x = X_{span})\right)$. This value is inserted into **Step 3**, and the process repeated. When a tolerance value is achieved, $|v_0 - \dot{s}(x - X_{span})| < tol$, then the solution is in a steady state.

Step 6 The solution yields the time T_{span} at which $x = X_{span}$ and $v_{MG} = y(X_{span})/T_{span}$ is evaluated.

These steps are repeated for a selection of parameters to obtain the best value of v_{MG}. In our numerical implementation, **Step 1** involves perturbing a straight line path with a sinusoidal oscillation.

7 Computational results

Data for a 14 foot dinghy is used. The drag coefficient $d = 14\,\mathrm{kg\,m^{-1}}$, the wavelength of the perturbation is chosen to be $L = 20\,\mathrm{m}$ and mass $M = 250\,\mathrm{kg}$. Figure 4 shows the effects of small, moderate and large perturbations by plotting the speed as a function of time when $\epsilon = 0.2, 1$ and 4.

For small perturbation parameter ϵ, the linearised solution corresponds closely with the true solution. However, large discrepancies are apparent when the perturbation parameter becomes larger. Figure 4 compares the numerical with the linearised solutions. Linearised solutions are sinusoidal perturbations of the straight-line speed ($3.562\,\mathrm{ms^{-1}}$).

A large portion of the error is attributable to the coarse first order approximation in (8). When $\epsilon = 4$, the arc length is increased by 24%. This is neglected by (8). Discrepancies between the initial speeds of the nonlinear and linear solutions are partially attributable to extra path length. For the nonlinear equations, v_0 is determined so that $v(s = 0) = v(s = L)$ but the linearised solutions are designed to satisfy this condition when the arc length is precisely the straight-line arc length. At $s = 0$, the direction of the yacht is at its maximum difference from straight line motion. Consequently, the deviation of the nonlinear force function from its linearisation is at a maximum at the very beginning, middle and end of the path. As ϵ increases, the discrepancy between initial speeds further increases.

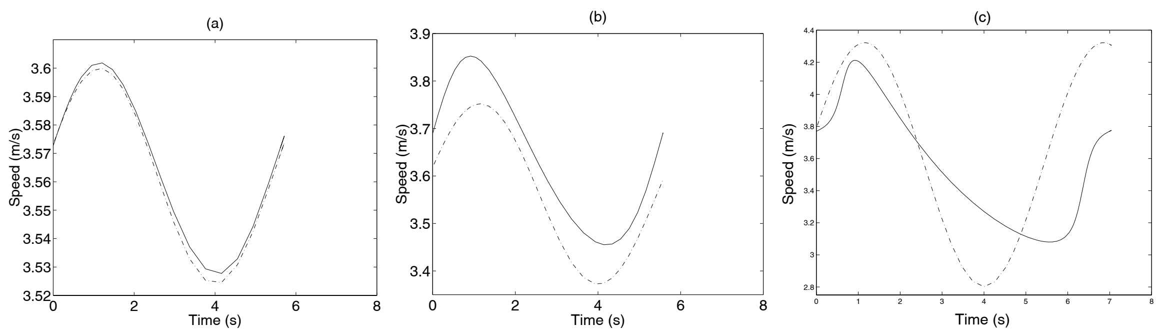

Figure 4: Speed $\dot{s}(t)$ over time for $\epsilon/L =$ (a) 0.2/20, (b) 1/20 and (c) 4/20; $\theta = \pi/4$. Numerical solution is solid curve, linearised solution is broken curve.

Using the full numerical integration, Figure 5 shows the variation of v_{MG} as θ varies. These plots reveal the difficulties associated with the linearised equations. For small ϵ, the improvements in v_{MG} are negligible: the first variation in ϵ is zero. Higher order estimates are imperative for discerning the improvements in downwind speed with ϵ.

The improvements in boat speed are second order effects. To be observed, the perturbations must significantly depart from zero, as revealed in Figure 5. This means that the full nonlinearities must be incorporated creating difficulties for analytical investigation. The simulations reveal that improvements of 3–4% can be made on the downwind leg. Slight improvements to the velocity polar profiles are achieved by an oscillating course, rather than straight-line motion.

At some θ close to $\pi/4$, the graphs in Figure 5 indicate that the philosophy of oscillation becomes beneficial. Empirical evidence indicates the philosophy of an oscillating heading will only improve v_{MG} for yachts with particular characteristics. Although this study has concentrated upon the characteristics of a 14 foot dinghy, it is proposed that yachts with "flatter" downwind velocity profiles will exhibit

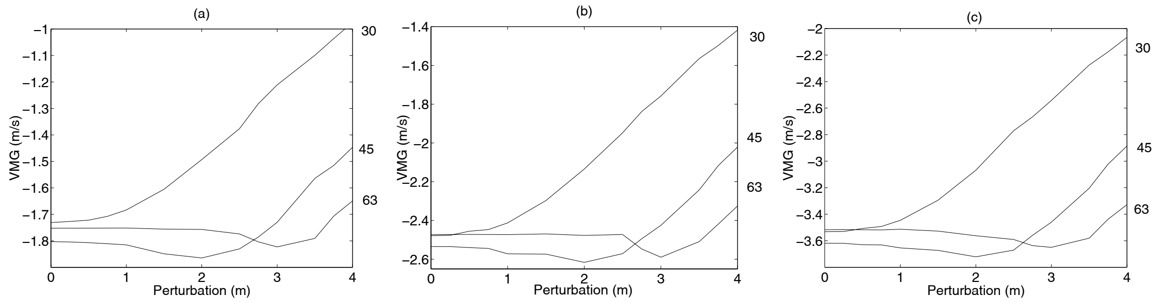

Figure 5: v_{MG} for $\theta = 30, 45$ and 63 degrees. True wind strengths are (a) $W = 5\,\mathrm{ms}^{-1}$, (b) $W = 7\,\mathrm{ms}^{-1}$, and (c) $W = 10\,\mathrm{ms}^{-1}$. Nonlinear integration for large deviations is used.

improvements when the heading oscillates, but a velocity profile which is pointed near the optimal v_{MG} will benefit from a straight heading.

8 Further directions

This same basic method can be applied to windward sailing. However, the no-leeway slip assumption becomes invalid. The force diagram relating F to $\mathbf{w}$ is very sensitive at windward angles and may not provide reliable predictions.

The interaction of the yacht with waves may become important (Harris *et al.* [5]), and this can be incorporated into the analysis in the form of an autonomous forcing term. The interaction of two frequencies (wave frequency and the perturbation frequency) should produce interesting results.

The influence of extra drag, dependent upon the curvature $p''(s)$ should model the rudder drag more accurately. This effect will become more pronounced for larger ϵ and may negate the small benefits that we have demonstrated. Finally, more accurate descriptions of the force F will yield more accurate results for modelling real yachts.

References

[1] F. Bethwaite, *High Performance Sailing*, Waterline, Shrewsbury, England (1996).

[2] T. Charvet, F. Hauville and S. Iluberson, "Numerical simulation of the flow over sails in real sailing conditions", *J. Wind Eng. Ind. Aerodynamics*, **63** (1996), 111–129.

[3] P. Couser, *Computational Methods for Investigating Sail Forces — A Case Study*, Australian Maritime Engineering CRC Online Report, at http://www.curtin.edu.au/curtin/centre/amecrc (1998).

[4] C. Harman, "'Who's on first!', 'What?', 'What's on second!' and how 'What' got there on an optimal baserunning path", in *Fourth Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar (editors), Bond University, Queensland, Australia (1998), 217–223.

[5] D. Harris, G. A. Thomas and M. R. Renilson, "Downwind performance of yachts in waves", in *Proc. Second Australian Sailing Science Conference*, Hobart, Tasmania (1999).

[6] T. H. Havelock, "The wave-making resistance of ships: a theoretical and practical analysis", *Proc. Roy. Soc. Ser. A*, **82** (1909), 276–299.

[7] K. I. Hedges, P. J. Richards and G. D. Mallinson, "Computer modelling of downwind sails", *J. Wind Eng. Ind. Aerodynamics*, **63** (1996), 95–110.

[8] L. Lazauskas and J. Winters, *Hydrodynamic Drag of Sea Kayaks: Prediction and Verification*, preliminary notes (1997).

[9] A. B. Philpott, R. M. Sullivan and P. S. Jackson, "Yacht velocity prediction using mathematical programming", *Eur. J. Oper. Res.*, **67** (1993), 13–24.

[10] D. J. Tritton, *Physical Fluid Dynamics*, Oxford, New York (1988).

[11] WB Sails Homepage: http://www.wb-sails.fi/news/SailPowerCalc/SailPowerCalc.htm, revised 18 July 1998.

A STATISTICAL APPROACH FOR DEVELOPING WEIGHTED NORMS BASED ON DIFFICULTY RATINGS ON SELECTED FITNESS PARAMETERS FOR BOYS IN THE AGE GROUP 13–16 YEARS

J. P. Verma
Lakshmibai National Institute of Physical Education
Deemed University
Gwalior-474002
India

Abstract

The present study is an effort to discuss a method of constructing the norms for any test item based on difficulty ratings. The technique so discussed has been used to develop the norms for the school boys of Gwalior in two age categories *viz.* 13–14 years and 15–16 years on the five test items i.e. sit-ups for abdominal strength, balance, standing broad jump for power, 50 metre performance for speed and 600 metre performance for endurance. The norms so developed were distinct in both the age categories on all the test items. The important finding of the study is that in each of the test items for a given increase in the performance at the lower level a subject is rewarded less but at the higher level the reward is more. This is because the improvement in the performance at the top level has a larger difficulty rating in comparison to that at the lower level.

1 Introduction

Many studies have been conducted by Cohb [5], Front [7] and Ismail *et al.* [8] on developing norms on various motor components. Most researchers use percentile method of constructing norms. Bolcok [2], Box [3], Frenzoni [6], Rosmussen [11], Busch [4] and Zuti *et al.* [12] developed the norms on physical fitness test items in different age categories. Barrow *et al.* [1], Mistkawi [9] and Richardson [10] have developed the norms on the variables other than physical fitness test items. But a score obtained on percentile scale is not the correct indication of one's capability on the test items. Often we require to develop a scale which measures the worth of an individual on the test parameter. The present study is based on the concept that a slight improvement in the performance at the top level should be rewarded more.

This study has two purposes. The first purpose is to discuss a statistical method of developing weighted norms based on difficulty ratings of the performance. The second purpose is to develop the weighted norms on selected fitness parameters for boys in 13 to 16 years age group.

2 Methods

In all, four hundred and eighty boys in the age group of 13–16 years from the central schools of Gwalior were selected randomly as the subjects in the study. The subjects were grouped in the age group of

13–14 years and 15–16 years. When deciding the age group, a boy was classified in the age category of 13–14 years only if his age was exactly above 13 years and below 15 years on the date of testing. Similarly, the subject was classified in the age category 15–16 years if he had completed 15 years but less than 17 years on the date of testing. In each group, two hundred and forty boys were selected. These subjects were tested on the five test items namely, sit-ups, balance, standing broad jump for power, 50 metre performance for speed and 600 metre performance for endurance. The stork balance test was used to test the balance of the subjects. Standard methods were used to measure the test items. To exploit the full potential of the subjects, they were properly motivated by means of intrinsic and extrinsic reinforcement. After obtaining the data on the five test items from the subjects in different age categories, the below mentioned procedure was used to develop the weighted scale.

3 Statistical technique

Here the procedure for developing the weighted scale has been discussed by taking the data on sit-ups for the age group of 13–14 years. For simplicity, the performance has been arranged with a spacing of four counts of sit-ups starting from 8 to 44. However, final results were obtained with the spacing of two counts of situps.

1	Performance	44 A	40 B	36 C	32 D	28 E	24 F	20 G	16 H	12 I	8 J
2	Area covered by the performance $P(X2) - P(X1)$	0.02	0.13	0.12	0.15	0.19	0.09	0.11	0.07	0.06	0.06
3	Area below the performance $P(X1)$	0.98	0.85	0.73	0.58	0.39	0.30	0.19	0.12	0.06	0
4	Lower limit of the sit-ups $X1$	2.05	1.04	0.61	0.20	−0.28	−0.52	−0.88	−1.18	−1.56	$-\infty$
5	Upper limit of the sit-ups $X2$	∞	2.05	1.04	0.61	0.20	−0.28	−0.52	−0.88	−1.18	−1.56
6	Ordinate at the lower limit $O(X1)$	0.049	0.232	0.331	0.391	0.384	0.348	0.271	0.199	0.118	0
7	Ordinate at the upper limit $O(X2)$	0	0.049	0.232	0.331	0.391	0.384	0.348	0.271	0.199	0.118
8	Normative values $w = \dfrac{O(X1) - O(X2)}{P(X2) - P(X1)}$	2.45	1.41	0.83	0.40	−0.04	−0.40	−0.70	−1.03	−1.35	−1.97
9	Linear transformed scores $W = 10w + 50$	74.5	64.1	58.3	54.0	49.6	46	43	39.7	36.5	30.3
10	Cumulative score	496	421.5	357.4	299.1	245.1	195.5	149.5	106.5	66.8	30.3
11	Normative values (NV)	100	84.98	72.06	60.30	49.42	39.42	30.14	21.47	13.47	6.11
12	Normative values (rounded off)	100	85	72	60	49	39	30	21	13	6

Table 1: Computation of weighted norms for the data on sit-ups for the boys in the age group 13–14 years.

The procedure discussed below is based on the concept of normality. It is assumed that in case of large sample human traits behave like a normal distribution. The weight of each grade has been decided by dividing the difference in the ordinates at the boundary points of the grade by the area it covered. To show the procedure, data obtained on sit-up performance in the age group 13–14 years along with its frequency are listed as follows:

Sit-ups	44	40	36	32	28	24	20	16	12	8
Frequency	4	32	29	35	45	21	27	17	15	15

13–14 Years		15–16 Years	
Performance (in number)	Scores	Performance (in number)	Scores
		52	100
		48	93
		46	87
44	100	44	80
42	92	42	75
40	85	40	70
38	78	38	64
36	72	36	58
34	65	34	53
32	59	32	48
30	53	30	44
28	48	28	39
26	43	26	35
24	38	24	30
22	33	22	26
20	28	20	22
18	23	18	18
16	19	16	15
14	15	14	12
12	11	12	8
10	7	10	5
8	3	8	3

Table 2: Weighted norms for the data on sit-ups for the boys in the age group 13–14 and 15–16 years.

The computation of scale values for the various performances on sit-ups is shown in Table 1.

Computation of entries in different rows:

Row 1: Entries in the first row are the performance given in the problem.

Row 2: Convert the frequency corresponding to each performance into its percentage. These entries are denoted by the area in the normal curve covered by the performance $P(X2) - P(X1)$.

Row 3: In the third row area below each performance is computed by subtracting the cumulative entries of second row from 1. For example,

$$\text{Area below A} = 1 - 0.02 = 0.98,$$
$$\text{Area below B} = 1 - (0.02 + 0.13) = 0.85,$$
$$\text{Area below C} = 1 - (0.2 + 0.13 + 0.12) = 0.73.$$

Row 4: In the fourth row, the lower limit of all performances is obtained by using a normal curve area table. Here corresponding to each entry in the third row, the z value is computed.

Row 5: Similarly, in the fifth row, the upper limit of all the performances is obtained by using a normal curve area table. In this row, the first entry would be infinity (∞) and the other entry would be the same as that of the entry in row 4 of the preceding column. For example,

$$\text{Upper limit of B} = 2.05 = \text{same as the entry of the fourth row in the column A},$$
$$\text{Upper limit of D} = 0.61 = \text{same as the entry of the fourth row in the column C}.$$

Row 6: Again using the normal table, the ordinate of each performance level corresponding to the z entries in the fourth row can be obtained.

Row 7: Similarly in row 7, the ordinate corresponding to z entries in row 5 can be obtained by the normal table.
Row 8: In this row, the normative value is obtained by dividing the difference of the sixth and seventh row entries by the corresponding entry in row 2.

Row 9: Normative values obtained in row 8 are converted into the linear derived form by the formula $W = 10w + 50$.

Row 10: Linear normative values are added in a cumulative manner.

Row 11: Each entry of row 10 is divided by the highest cumulative total and multiplied by 100. For example, for the column A, $496 \times 100/496 = 100$ and, for the column D, $299.1 \times 100/496 = 60.30$.

Row 12: Normative values (NV) have been rounded off to the nearest whole number in this row.

4 Results

The weighted norms for the data on sit-ups for the boys in the age categories of 13–14 years and 15–16 years are shown in Table 2. In the 13–14 years age category a boy would get 100 marks if his performance on the sit-up count is 44 whereas in the age category 15–16 years to get 100 marks one would require to perform 52 counts of the sit-ups. On the other hand, in both the categories one would get three marks if their performance is eight counts.

13–14 Years		15–16 Years	
Performance (in seconds)	Scores	Performance (in seconds)	Scores
		245	100
		235	95
		225	89
		215	84
205	100	205	79
195	94	195	74
185	87	185	70
175	81	175	65
165	75	165	60
155	70	155	56
145	64	145	51
135	58	135	47
125	53	125	42
115	47	115	38
105	42	105	34
95	37	95	30
85	32	85	26
75	27	75	22
65	22	65	18
55	18	55	15
45	14	45	11
35	10	35	8
25	6	25	5
15	3	15	2

Table 3: Weighted norms for the data on balance for the boys in the age group 13–14 and 15–16 years.

Another pecularity may be observed in the 15–16 years age category that no score has been shown in front of 50 counts of sit-ups. This is because of the fact that no subject was found to have performance equal to 50. However, it can be interpolated from the listed norms.

Table 3 shows the weighted norms on balance for the boys in both the age categories. In the 13–14 years age category, 100 marks would be given to a boy who performs the balance test for 205 seconds whereas the same marks may be given in the 15–16 years age category for a performance of 245 seconds. On the lower side, for a performance of 15 seconds one would get three marks in the 13–14 years age category and two marks in the 15–16 years age category.

The weighted norms for the data on 600 metre performance in both the age categories are shown in Table 4. In the 13–14 years of age category, 100 marks are given for a performance of 114 seconds whereas to get the same marks in the 15–16 years of age category one has to click 94 seconds on the 600 metre event. On comparing the norms at the lower side, ten marks may be given to a performance of 270 seconds in the 13–14 years of age category whereas only three marks are given for the same performance in the 15–16 years of age category.

13–14 Years		15–16 Years	
Performance (in seconds)	Scores	Performance (in seconds)	Scores
		94	100
		98	94
		102	88
114	100	110	82
118	94	118	76
126	87	126	71
134	82	134	66
142	76	142	61
150	71	150	57
158	66	158	52
166	61	166	48
174	57	174	44
182	52	182	39
190	48	190	35
198	44	198	31
206	39	206	28
214	35	214	24
222	31	222	20
230	28	230	17
238	24	238	14
246	20	246	10
254	17	254	8
262	13	262	5
270	10	270	3
278	8		
286	5		
294	2		

Table 4: Weighted norms for the data on 600 metre boys in the age group 13–14 and 15–16 years.

The norms for the boys in different age categories on the 50 metre performance are shown in Table 5. Anybody in the age group 13–14 years, completing the 50 metre distance in 7.0 seconds would get 100 marks whereas for the same marks one has to click 6.9 seconds in the 15–16 years of age category. On the lower side, three marks would be given to a 10.8 seconds performance in the age category of 13–14 years whereas one has to obtain 10.2 seconds in the age group 15–16 years to get three marks.

Table 6 reveals the norms on broad jump performance for the boys in both age groups. In the age

13–14 Years		15–16 Years	
Performance (in seconds)	Scores	Performance (in seconds)	Scores
		6.9	100
7.0	100	7.0	92
7.1	92	7.1	85
7.2	85	7.2	78
7.4	79	7.4	71
7.6	73	7.6	65
7.8	67	7.8	59
8.0	61	8.0	53
8.2	56	8.2	48
8.4	51	8.4	42
8.6	46	8.6	37
8.8	41	8.8	32
9.0	36	9.0	27
9.2	31	9.2	23
9.4	27	9.4	18
9.6	23	9.6	14
9.8	19	9.8	10
10.0	15	10.0	7
10.2	12	10.2	3
10.4	9		
10.6	6		
10.8	3		

Table 5: Weighted norms for the data on 50 metre boys in the age group 13–14 and 15–16 years.

group of 13–14 years, one would get 100 marks if the performance on broad jump is 2.44 metres whereas to get 100 marks in the 15–16 years age group one has to perform 2.52 metres on the broad jump.

On the other hand in the age group of 13–14 years 2 marks would be given if the performance on the broad jump is 1.52 metre and for getting 3 score in 15–16 years age group, one has to give 1.60 metre performance.

5 Conclusion

In each of the norms you can notice that for a given increase in the performance at the lower level a subject is rewarded less but at the higher level the reward is more. It is because of the fact that the improvement at the top level is more difficult in comparison to the improvement of the same amount at the lower level. In other words difficulty rating is high for improving the performance at the top level in comparison to that of at lower level. The norms were different in both the age groups on all the parameters i.e. to get the same score on any test item a boy in the higher age group will have to perform better in comparison to that of a boy in the lower age group. Further it is suggested that once the norms are developed it should be revised after every year, as the performance of an individual on the test items is very dynamic in nature.

References

[1] H. M. Barrow and R. McGee, *Practical Approach to Measurement in Physical Education*, Lea and Febiger, Philadelphia (1971), 228–229.

13–14 Years		15–16 Years	
Performance (in metres)	Scores	Performance (in metres)	Scores
		2.52	100
		2.48	94
2.44	100	2.44	89
2.40	94	2.40	83
2.36	89	2.36	78
2.32	83	2.32	73
2.28	78	2.28	68
2.24	73	2.24	64
2.20	69	2.20	59
2.16	64	2.16	55
2.12	59	2.12	50
2.08	55	2.08	46
2.04	51	2.04	42
2.00	46	2.00	38
1.96	42	1.96	34
1.92	38	1.92	30
1.88	34	1.88	26
1.84	30	1.84	23
1.80	26	1.80	19
1.76	22	1.76	16
1.72	19	1.72	12
1.68	15	1.68	9
1.64	12	1.64	6
1.60	8	1.60	3
1.56	5		
1.52	2		

Table 6: Weighted norms for the data on broad jump for boys in the age group 13–14 and 15–16 years.

[2] C. F. Bolcok, "Physical fitness of Dalawara boys and girls in grades five through twelve", *Completed Research in Health, Physical Education and Recreation*, **6** (1960), 60–62.

[3] D. L. Box, "Physical ability testing of male students in grades four through twelve", *Completed Research in Health, Physical Education and Recreation*, **9** (1967), 77–79.

[4] J. G. Busch, "A normative study of the AAHPER youth fitness test in grade seventh through tenth in the state of South Dakota", *Completed Research in Health, Physical Education and Recreation*, **12** (1970), 204–207.

[5] R. P. Cohb, "The construction of a motor fitness test battery for girls in lower elementary grades", *Dissertation Abstracts International*, **33** (November 1972), 2146.

[6] M. M. Frenzoni, "A comparison of test evaluation and two teaching programs and the development of fitness and skill norms in elementary physical education", *Dissertation Abstracts International* (January 1994), 2458-A.

[7] D. B. Front, "The determination of the validity of four selected motor fitness items and the construction of norms for Washington State University men", *Completed Research in Health, Physical Education and Recreation*, **5** (1983), 85–87.

[8] A. H. Ismail and C. C. Cowell, "The Purdue motor fitness batteries and development to profile for pre-adolescent boys", *Res. Quart.*, **33** (December 1962), 553–557.

[9] J. J. Mistkawi, "Norms for eight, nine and ten year old boys and Y.M.C.A. athletic achievements test", *Completed Research in Health, Physical Education and Recreation*, **8** (1961), 101–103.

[10] M. J. Richardson, "The development of norms for the Martinek–Zaich–Kowsky self-concept scale for Costa Rican children", *Dissertation Abstracts International* (April 1988), 2572-A.

[11] G. L. Rosmussen, "A normative study of the AAPHER youth physical test for boys in grades seventh through ten in the State of South Dakota", *Completed Research in Health, Physical Education and Recreation*, **9** (1970), 207–209.

[12] W. B. Zuti and C. B. Corbin, "Physical fitness norms for college freshmen", *Res. Quart.*, **48** (May 1977), 499–503.

EXTREME VALUE MODELS IN SPORTING DATA ANALYSIS

Julian Z. Wang
School of Science
University of Western Sydney, Nepean
NSW 2747, Australia

Peter Cooke
Department of Statistics
University of New South Wales
NSW 2052, Australia

Abstract

The theory of extreme values can play a role in sporting data analysis. In sports competitions, the best performances by athletes are of interest. Thus in extreme-value terms, the tail behaviour of the distribution of performance is the focus of our attention. Two extreme-value models are discussed here and a technique based on minimum distance is used to estimate the endpoint of a distribution whose tail belongs to the domain of attraction of the Weibull distribution. Applications are illustrated through examples with real sporting data appropriate to the settings of the two models.

1 Introduction

In sporting events, only the extreme observations leading to records attract people's attention and therefore are readily accessible. The records achieved are usually the extreme performances of the athletes, such as the shortest times, longest distances, highest heights, and so on.

Consider the following problem in sports. A swimming training program is to be evaluated and the objective of the program is to improve swimmers' performances. Assume that ten swimmers are randomly selected for taking part in this program. The usual statistical methods are not applicable in this situation and the sample means and variances of times are inappropriate quantities in making statistical inferences. Instead, the best performances of the swimmers should be considered and analysed. Suppose now the best times before and after the training program by each of the swimmers are recorded. To evaluate the effectiveness of the training program, the ten records after the program are compared with those before the program and the limits, or the endpoints, of the two sets of data have to be estimated. The statistical model in this experiment is one of the extreme-value models, the model of maxima.

This model and the other extreme-value model are described in Section 2. For each model, we assume that the tail of the distribution is in the Weibull domain of attraction. The methods of estimation of the parameters are discussed in Section 3. Section 4 gives some examples.

2 Two models for extreme observations

Let $X_{1n} \geq \cdots \geq X_{nn}$ be the order statistics from a random sample of n independent random variables from a distribution $F(x)$. Suppose that $F(x)$ is in the domain of attraction of a distribution $H(x)$; that is, there are sequences $\{a_n\}$ and $\{b_n\}$ such that

$$P\left(\frac{X_{1n} - a_n}{b_n} \leq x\right) = F^n(a_n + b_n x) \to H(x) \quad \text{as } n \to \infty.$$

It is well known (see, e.g. Galambos (1987)) that, up to a change of location and scale, $H(x)$ must be one of the following distributions:

(1) Gumbel: $\Lambda(x) = \exp\{-\exp(-x)\}, \quad -\infty < x < \infty$,
(2) Frechet: $\Phi_\gamma(x) = \exp\{-x^{-\gamma}\}, \quad 0 < x$,
(3) Weibull: $\Psi_\gamma(x) = \exp\{-(-x)^\gamma\}, \quad x < 0$,

where γ is some positive constant.

Model of Extremes Let $X_{1n} \geq \cdots \geq X_{nn}$ be order statistics as above and, for a fixed integer k, let $Z_{in} = (X_{in} - a_n)/b_n$ for $i = 1, \ldots, k$. Then, as $n \to \infty$, the random vector $\boldsymbol{Z}_{kn} = (Z_{1n}, \ldots, Z_{kn})^T$ converges in distribution to the limit random vector $\boldsymbol{U}_k = (U_1, \ldots, U_k)^T$. The distribution of $\boldsymbol{U}_k$ can be determined if the domain of attraction of $F(x)$ is known; e.g. if $H(x) = \Lambda(x)$, then the joint distribution of the U_i is given by

$$h(u_1, \ldots, u_k) = \exp\left\{-\exp\{-u_k\} - \sum_{i=1}^{k} u_i\right\}, \quad u_1 \geq \cdots \geq u_k.$$

For expectations, covariances of the U_i and estimators $\hat{a}_n$ and $\hat{b}_n$ under $H(x) = \Lambda(x)$ see Weissman (1978).

Model of Maxima Let $X_{1,1}, \ldots, X_{1,n}, X_{2,1}, \ldots, X_{2,n}, \ldots, X_{k,1}, \ldots, X_{k,n}$ be a sample of size $N = kn$ with common distribution $F(x)$. The maxima $M^*_{in} = \max(X_{i,1}, \ldots, X_{i,n})$, for $i = 1, \ldots, k$, are observed, though not necessarily the individual $X_{i,j}$'s. The random variables $X_{i,1}, \ldots, X_{i,n}$ may correspond to data collected from the ith time interval of predetermined length, say, and M^*_{in} is the ith maximum. Let $M_{1n} \geq \cdots \geq M_{kn}$ be the order statistics from $M^*_{1n}, \ldots, M^*_{kn}$ and $Y_{in} = (M_{in} - a_n)/b_n$ for $i = 1, \ldots, k$. Then, as $n \to \infty$, the random vector $\boldsymbol{Y}_{kn} = (Y_{1n}, \ldots, Y_{kn})^T$ converges in distribution to the limit random vector $\boldsymbol{V}_k = (V_1, \ldots, V_k)^T$. If $H(x) = \Lambda(x)$, then the V_i have the joint distribution, e.g.

$$g(v_1, \ldots, v_k) = k! \exp\left\{-\left(\sum_{i=1}^{k} \exp\{-v_i\}\right) - \left(\sum_{i=1}^{k} v_i\right)\right\}, \qquad v_1 \geq \cdots \geq v_k.$$

For expectations, covariances of the V_i and estimators $\hat{a}_n$ and $\hat{b}_n$ under $H(x) = \Lambda(x)$, see Lieblein (1962).

For details of these two models, see Wang, Cooke and Li (1996). Under each model, the commonly used continuous probability distributions with upper tail in some domain of attraction are divided into three types, namely, the Gumbel, Frechet and Weibull. If we are dealing with random variables limited in the upper tail, the Weibull type is appropriate and, if we are dealing with unlimited random variables, Frechet is appropriate.

For physical reasons, the observations in sporting data are limited and so in our statistical model we should assume that F has an endpoint upper or lower, as appropriate. In some highly competitive sporting events, the records could be assumed to be close to their limits.

The most widely applicable extreme value type in many areas of scientific research is arguably the Weibull. The Weibull type contains three parameters, namely, shape, scale and threshold or endpoint. In the following section we will discuss methods of estimating the endpoint.

From here on we ignore the approximation of $\boldsymbol{U}_k$ to $\boldsymbol{Z}_{kn}$ and $\boldsymbol{V}_k$ to $\boldsymbol{Y}_{kn}$ by assuming $n \to \infty$, and denote X_{in} by X_i and M_{in} by M_i for $i = 1, \ldots, k$.

For testing for the domains of attraction and selecting the number k in the model of extremes, see Hasofer and Wang (1992), Wang (1995) and Wang, Cooke and Li (1996).

3 Estimating the endpoint

Assume that we have k ordered observations $T_1 \geq \cdots \geq T_k$ satisfying the linear model $\boldsymbol{T} = \boldsymbol{A\theta} + \mathbf{Error}$, where $\boldsymbol{T} = (T_1, \ldots, T_k)^T$ and $\boldsymbol{\theta} = (\mu, \eta)^T$. The matrix $\boldsymbol{A}$ is a $k \times 2$ matrix $[\mathbf{1}, E(\boldsymbol{U})]$ where $\mathbf{1}$ is the

$k \times 1$ column vector of 1's and $\boldsymbol{U} = (U_1, \ldots, U_k)^T$ is a known random vector with expected value $E(\boldsymbol{U})$ and covariance matrix $\boldsymbol{V}$. The method of least squares or maximum likelihood could be applied to estimate $\boldsymbol{\theta}$. Further assume that the T_i are functions of an unknown parameter ω, and rewrite T_i as $T_i(\omega)$. The estimator of ω is the value of ω which minimises the quadratic form

$$Q(\omega) = [\hat{\boldsymbol{U}}(\omega) - E(\boldsymbol{U})]^T \boldsymbol{V}^{-1} [\hat{\boldsymbol{U}}(\omega) - E(\boldsymbol{U})]$$

where $\hat{U}_i(\omega) = [T_i(\omega) - \hat{\mu}(\omega)]/\hat{\eta}(\omega)$, $i = 1, \ldots, k$. This method of estimating ω was called the minimum distance method in Hall and Wang (1999).

Suppose that the upper tail of the underlying distribution $F(x)$ is in the domain of attraction of $\Psi_\gamma(x)$ (Weibull type) and it has a finite upper endpoint ω_0. The estimation of ω_0 is achieved as follows for the two models.

Model of Extremes Let $T_i = -\log(\omega_0 - X_i)$, $i = 1, \ldots, k$. Then $U_i = (T_i - a_n)/b_n$, $i = 1, \ldots, k$, have the asymptotic density function $h(u_1, \ldots, u_k)$ given above. Thus ω_0 is estimated by the minimiser of $Q(\omega)$. In Hall and Wang (1999) the asymptotic properties of the estimator were derived. It was noticed that minimising $Q(\omega)$ was equivalent to minimising

$$G(\omega) = \frac{\sum_{i=1}^{k-1} i^2 [T_i(\omega) - T_{i+1}(\omega)]^2}{\left[\sum_{i=1}^{k-1} [T_i(\omega) - T_k(\omega)]\right]^2}.$$

When $\omega = \omega_0$, $G(\omega_0)$ is the Greenwood statistic. Thus confidence intervals for the endpoint can be obtained using $G(\omega)$.

Model of maxima Let $T_i = -\log(\omega_0 - M_i)$, $i = 1, \ldots, k$. Then $V_i = (T_i - a_n)/b_n$, $i = 1, \ldots, k$, have the asymptotic density function $g(v_1, \ldots, v_k)$ given above. The endpoint ω_0 is estimated by the minimiser of $Q(\omega)$. Confidence intervals for the endpoint could be constructed by finding the distribution of $Q(\omega_0)$, which could be obtained by Monte Carlo methods. Let q be the upper α-level point of the distribution of $Q(\omega_0)$. Then an α-level confidence interval for ω_0 is defined as the set of values of ω such that $Q(\omega) < q$.

50 m Freestyle (Olympic Games, 25 July 1996)	*Prelims*
A. Popov, RUS	22.22
G. Hall, USA	22.36
F. Scherer, BRA	22.68
C. Jiang, CHN	22.55
B. Dedekind, RSA	22.60
D. Fox, USA	22.64
F. Sanchez, VEN	22.68
R. Busquets, PUR	22.61
… …	…

Table 1: Prelims of men's 50 m freestyle, 1996 Olympic Games. Source: www.swimnews.com

4 Examples

An example for the model of extremes was given in Hall and Wang (1999) where the model was applied to the men's 100 metres records (Track and Field) in the 1988 and 1992 Olympic Games. Here we consider the men's 50 metres freestyle swimming records in the 1996 Olympic Games. Table 1 lists some of the first 40 of the prelims. Since this is a case of minima, we convert it to a case of maxima by taking negatives of the observations. The function $G(\omega)$ reaches its minimum at $\omega = -22.18$ (Figure 1), which is taken as the estimate of the upper limit of the negative times. The upper 10% critical point of the Greenwood statistic for $k = 40$ is 0.0565 (see Stephens (1981)) which gives the 90% confidence interval for the lower limit of the times of (21.20,22.22). The 95% confidence interval is (19.75, 22.22).

Cup	1	2	3	4	5	6	7	8	9	10
Time	24.74	25.69	25.88	25.34	25.86	25.45	24.78	25.17	25.21	25.05

Table 2: FINA World Cups 1–10, women's 50m freestyle. Source: www.swimnews.com

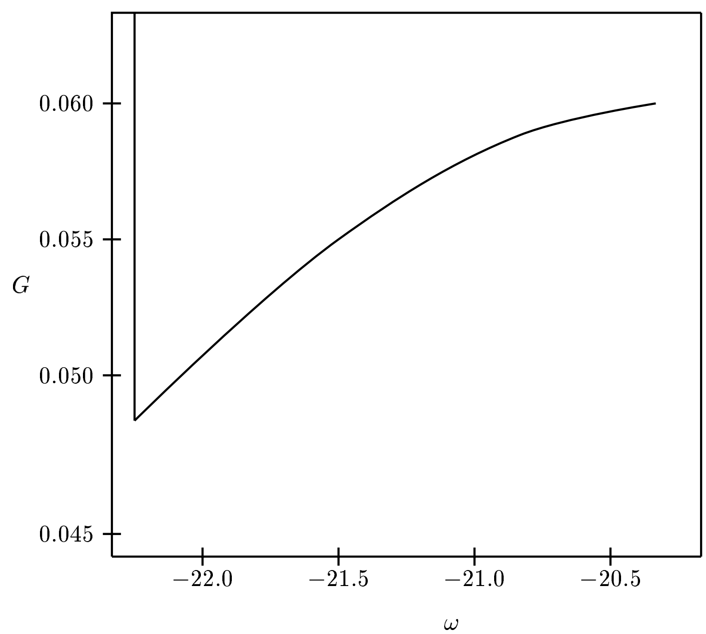

Figure 1: Plot of $G(\omega)$

Our next example uses the model of maxima. Table 2 shows the ten best performances from each of the women's 50 metres freestyle events in the most recent FINA World Cups 1–10 (from November 1999 to February 2000). Again using negatives, Figure 2 shows that the function $Q(\omega)$ attains its minimum at $\omega = -23.70$ and thus the estimate of the limit for the data is 23.70.

References

J. Galambos (1987), *The Asymptotic Theory of Extreme Order Statistics*, second edition, Robert E. Krieger, Malabar, FL.

P. Hall and J. Z. Wang (1999), "Estimating the endpoint of a probability distribution using minimum distance methods", *Bernoulli*, **5**(1), 177–189.

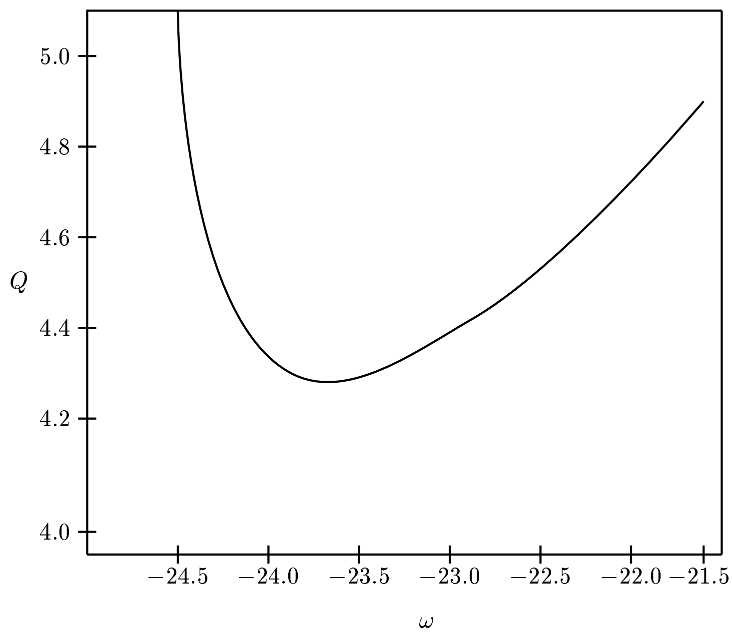

Figure 2: Plot of $Q(\omega)$

A. M. Hasofer and Z. Wang (1992), "A simple test for extreme value domain of attraction", *J. Amer. Statist. Assoc.*, **87**, 171–177.

J. Lieblein (1962), "Extreme-value distribution" in *Contribution to Order Statistics*, A. E. Sarhan and B. G. Greenberg (editors), Wiley, New York.

M. Stephens (1981), "Further percentage points for Greenwood's statistics", *J. Roy. Statist. Soc. Ser. A*, **144**, 364–366.

J. Z. Wang, P. Cooke and S. Li (1996), "Determination of the domains of attraction based on a sequence of maxima", *Austral. J. Statist.*, **38**(2), 173–182.

J. Z. Wang (1995), "Selection of the k largest order statistics for the domain of attraction of the Gumbel distribution", *J. Amer. Statist. Assoc.*, **90**, 1055–1061.

I. Weissman (1978), "Estimation of parameters and large quantiles based on the k largest observations", *J. Amer. Statist. Assoc.*, **73**, 812–815.

CRICKET BOWLING: A TWO-SEGMENT LAGRANGIAN MODEL

R. E. D. Ferdinands
Department of Physics & Electronic Engineering
University of Waikato
Private Bag 3105
Hamilton, New Zealand

K. A. Broughan
Department of Mathematics
University of Waikato
Private Bag 3105
Hamilton, New Zealand

H. Round
Department of Physics & Electronic Engineering
University of Waikato
Private Bag 3105
Hamilton, New Zealand

Abstract

In this study a forward solution of the bowling arm in cricket is made using a rigid body Lagrangian model, coupled with the projectile equations of motion for the free flight of the ball. The model includes two segments, the upper arm and the forearm, and the joints of the shoulder and elbow. The approach is that of an initial value problem. The first step is to determine the initial conditions of the limb angles and velocities, and to choose as system inputs the joint torques and forces. Then the system response is found by numerically solving the equations of motion to determine which combination of ball release speed and arm angle can land a ball at a particular position on the pitch. Derivation of the equations of motion, numerical solving and data processing were performed by the symbolic manipulation software package *Mathematica*, Version 3.0.

An APAS motion analysis system[1](60 Hz) measured the 2D kinematic data of an elite bowler performing two trial types. In the first trial type, the subject was instructed to remain stationary in a side-on delivery position, and, using the bowling arm only, deliver a ball at maximum speed at a target in line with the wickets. No motion of the non-bowling arm was allowed. This effectively isolated the motion of the bowling arm for analysis. In the second trial type, the subject delivered the ball exactly as before, but with the bowling arm now straightened (or locked) earlier. After examining the kinematic data, one typical test bowl was chosen for analysis.

By substituting the 2D kinematic data into the Lagrangian equations of motion the set of time-varying joint torques and forces on the bowling arm during a typical delivery were generated. Running forward solutions with variations in these system inputs allowed us to simulate several realistic trajectories of the bowling arm and test for any corresponding effect on ball release speed. Then the model was validated by comparing the relationship between ball release speed and locking angle in experiment with that predicted by the model. The main conclusions are that ball release speed can be increased by (1) decreasing the locking-angle of the bowling arm, and (2) decelerating the linear motion of the shoulder joint after the bowling arm has rotated above the horizontal.

[1]Courtesy of the Department of Sport and Exercise Science, The Waikato Polytechnic, Hamilton.

A GREYHOUND RACING BETTING STRATEGY

L. H. Liyanage, S. Terry and N. Hopkins
Department of Mathematical Sciences
University of Western Sydney, Macarthur
PO Box 1000
St Mary's
NSW 1790, Australia

Abstract

Betting on greyhound races is one of many popular forms of gambling in Australia and has increased in popularity during the past few years. As with all forms of gambling, the main objective in greyhound racing is to win money. This paper provides a betting strategy, which will significantly improve the chances of winning. Regression techniques are used in selecting the factors which significantly influence the winning time of a greyhound race, out of many influential factors such as box number, best time run at the track, age, etc. Two regression models which best predict the finishing time of an individual greyhound are obtained, one for dogs with a value for the variable "best time" and another for dogs without a value for "best time". Predicted running times obtained from these two models are merged in three different ways to find the overall ranking of the eight runners in a race. This provides three different methods of placing suitable bets. The optimal strategy was obtained and tested by comparing the money won/lost by placing bets according to each method.

SEASONAL VARIATIONS AND CRICKET'S RULE CHANGES

Iain Skinner
Electrical Engineering and Telecommunications
University of New South Wales
NSW 2052, Australia

Abstract

One of the first effects that strikes the numerically literate reader of the history of cricket is how the averages of those who played in the long ago are so much lower than those who played more recently. Simple explanation: conditions were different then. True, but what is the effect of such differences on the players' performances, as typified by averages?

As mathematicians we can quantify these effects, but merely quantifying is not enough. We should be able to model the way the game has changed, even venturing an estimate of the quantitative effects of various specific changes that have occurred.

This presentation will consider the season by season variation of three global measurements of the state of the game—overall average, leading batsmen's average, leading bowlers' average—and from these signals endeavour to determine which changes to the rules and playing conditions of Australian cricket have been the most important in influencing the relative balance between bat and ball. In doing so, some attention necessarily will be paid to the changing standard of cricket that is the first-class game in Australia, and a model developed to relate the three signals to more fundamental measurements: the intrinsic balance (r_0), the batting standard, and the bowling standard.

The signals are obviously very noisy. Simple analysis with Haar wavelets found changes in the intrinsic balance, i.e. discontinuities in r_0, and these jumps were identified with events in the history of the game at the corresponding time. The global variation of both batting and bowling standards show improvement with time. The standard of bowling appears to have improved more (perhaps corresponding to the development of more skills).