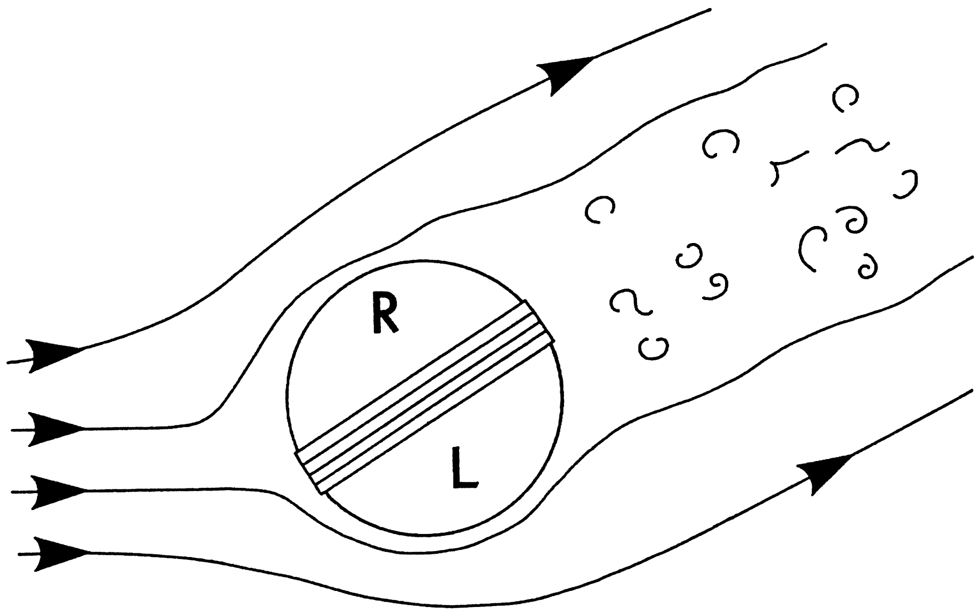
Third Conference on:

MATHEMATICS AND COMPUTERS IN SPORT



Held at Bond University, Queensland, Australia
30th September to 2nd October, 1996

Sponsored jointly by

ANZIAM

and

The Australian Sports Commission

Edited by Neville de Mestre, Associate Professor of Mathematics
School of Information Technology, Bond University.

CONTENTS (Vol 3)

GENERAL

Welcome to the third conference held at Bond University on the use of the mathematics, statistics and computers in sport. Its aim remains to bring together, every two years, researchers in these fields for an exchange of ideas. The conference is now an activity of a special interest group of the Australian and New Zealand Industrial and Applied Mathematicians (ANZIAM). It has been generously sponsored by ANZIAM and the Australian Sports Commission.

There will be many new faces at the conference, but 11 have attended previous conferences including Rao Ayyalarayu, Maurie Brearley and Stephen Clarke who have been to all three.

There are 20 papers to be presented with sporting interests ranging from rowing, cricket, athletics, golf, netball, soccer, BMX to horse-racing.

I would like to thank the two invited speakers — Stewart Townend (Liverpool John Moores University, U.K.) and Hugh Morton (Massey University, New Zealand) for agreeing to participate in the conference and lead discussions. To the other speakers and attendees I say thank you for making the conference successful.

Finally, I must thank my secretary, Jeanette Niehus, for massaging your disks into papers that conform to the final form presented here, and for assisting me in many of the organisational tasks for the conference.

Neville de Mestre
September 1996

Program

Monday 30 September

9.00am Opening by Dennis Green, ex-Olympic kayak paddler

9.10am M.S. Townend (Liverpool John Moores, U.K.) – "Some recent applications of advanced mathematics in sport"

10.10am L. Lazauskas (Adelaide) and E.O. Tuck (Adelaide) – "Low drag rowing shells"

10.40am Morning Tea

11.10am M.N. Brearley (Clifton Springs) and N.J. de Mestre (Bond) – "Modelling the rowing stroke and increasing its efficiency"

11.40am R.H. Morton (Massey, N.Z.) – "An analysis of world records for races run entirely in lanes"

12.10pm Lunch

2.00pm T. Lewis (West of England, U.K.) and F. Duckworth (Royal Stats Soc, U.K.) – "A fair method for resetting the target in interrupted one-day cricket matches"

2.45pm K. Kumar (Bond) – "Is cricket really by chance?"

3.15pm J.E. Baker (Tarata L.S.) – "Ground reaction forces and hub position in the golfswing"

Tuesday 1 October

9.00am R.H. Morton (Massey, N.Z.) – "Mathematics of the Human Engine - Evaluation of power and endurance"

10.00am D.L. Noble (Swinburne) – "Formation of netball teams for a series of trial matches"

10.30am Morning Tea

11.00am S.R. Clarke (Swinb) – "Home advantage in balanced competitions: English Soccer '90-'96"

11.30am P. Norton (Monash) – "The 'Leveller System' tennis tournament"

12.00noon Lunch

2.00pm I. Heazlewood (ACU) and S. Politi (ACU) – "Exercise physiological, biomechanical and kinanthropometric predictors of bicycle motor cross in young adults: a preliminary study"

2.30pm I. Smith (Dept. of Environment, Sport and Territories) – "The National Statistical initiative"

3.00pm A. McNabb (Auckland, NZ) – "Doing mathematics standing on your head"

3.30pm Mathsport meeting and conference photograph

Wednesday 2 October

9.00am G. Lackey (ACU) and I. Heazlewood (ACU.) – "The use of mathematical models to predict elite athletic performance at the Olympic games"

9.30am W. Benter (Happy Valley, HK), G. Miel (Nevada, U.S.) and P.D. Turnburgh (Nevada, U.S.) – "Modelling distance preference in thoroughbred racehorses"

10.00am R. Phatarfod (Monash) – "Betting strategies in horse races"

10.30am Morning Tea

11.00am S. Ganesalingam (Massey, N.Z.), S. Ganesh (Massey N.Z.) and K. Kumar (Bond) – "Statistical analysis of horse-racing data"

11.30am A.J. Hurst (Monash), D.L. Dowe (Monash), G.E. Farr (Monash) and K.L. Lentin (Monash) – "Information-theoretic football tipping"

12.00noon N.J. de Mestre (Bond) and M.N. Brearley (Clifton Springs) – "The car jump"

12.30 pm Close and lunch.

PARTICIPANTS

Dr Rao AYYALARAYU	(Central Queensland)
Dr John BAKER	(Tarata Learning Systems)
Mr William BENTER	(Hong Kong)
Prof Maurie BREARLEY	(Clifton Springs)
Assoc Prof Stephen CLARKE	(Swinburne)
Assoc Prof Neville DE MESTRE	(Bond)
Dr Selvanayagam GANESALINGAM	(Massey)
Assoc Prof Chris HARMAN	(Southern Queensland)
Dr Ian HEAZLEWOOD	(Australian Catholic University)
Mr Neil HOPKINS	(Western Sydney)
Prof John HURST	(Monash)
Dr Kuldeep KUMAR	(Bond)
Mr Gavin LACKEY	(Australian Catholic University)
Dr Leo LAZAUSKAS	(Adelaide)
Mr Tony LEWIS	(West of England, U.K.)
Prof Alex McNabb	(Auckland, N.Z.)
Assoc Prof Hugh MORTON	(Massey, N.Z.)
Dr Kevin NESS	(James Cook)
Mr David NOBLE	(Swinburne)
Dr Pam NORTON	(Monash)
Mr Ted O'KEEFE	(Macquarie)
Dr Ravi PHATARFOD	(Monash)
Assoc Prof Graham POLLARD	(Canberra)
Mr Ian SMITH	(DEST)
Dr Bill SUMMERFIELD	(Newcastle)
Ms Nikoleta TOMECKO	(South Australia)
Prof Stewart TOWNEND	(Liverpool John Moores, U.K.)

SOME RECENT APPLICATIONS OF ADVANCED MATHEMATICS IN SPORT

M. Stewart Townend[1]

Abstract

This paper presents an overview of some novel sports-based applications of established mathematical and computing topics, amongst which are the Fast Fourier Transform (FFT), neural networks (NN) and Computational Fluid Dynamics (CFD).

The applications considered are as wide ranging as the mathematical topics themselves and range from athletics to ocean racing via biomechanics and soccer, including some discussion of behavioural aspects of sport.

Finally, some thoughts as to future policy and direction are presented.

1. THE CONTRIBUTION OF NEURAL NETWORKS TO BIOMECHANICS

A neural network (NN) is a network of processors each of which behaves in the same way as the model of a neuron in the brain.

They may therefore be regarded as simulations of the nerve system and so are able to simulate the decision making of a human expert.

They represent a relatively new method of non-linear multivariate analysis but differ from it in that they can be taught to recognise patterns by examples instead of defining the underlying rules.

They can receive large amounts of data simultaneously (their internal structure preserving any inherent relationships amongst the data) and it is this simultaneity and contextuality which makes them potentially well suited to the automatic recognition of patterns.

For example a NN has been developed which is capable of distinguishing healthy from pathological gait with a success rate of approximately 76% when the gait patterns are represented by the vertical force associated with consecutive foot strikes (Holzreiter and Köhle [1]).

A recent review (Miller et al [2]) described other NNs applied in other areas with similar results while a further recent study has used acceleration data and NNs to successfully recognise incline, speed and distance during unconstrained walking (Aminian et al [3]).

---

[1]Mathematics Education Group, School of Computing and Mathematical Sciences, Liverpool John Moores University, Byrom St, Liverpool L3 3AF, England.

It is apparent that with suitable training, just like developing a human expert, a NN can successfully recognise previously presented patterns from amongst a new set of patterns. Unfortunately the optimal structure of a NN is application dependent, just like the topology of the transputer arrangements in parallel computing (Eberhart et al [4], Zurada et al [5] and Lawrence [6]).

To illustrate the use of neural networks in sport we next describe some recent work into the analysis of gait patterns which is representative of the application of neural networks in a biomechanical context (Barton and Lees [7]). Investigation of a motion as complicated as the human gait cannot be performed in terms of single parameters, rather all the data must be used simultaneously together with any interdependencies – in fact precisely the sort of situation in which neural networks exhibit great potential.

Three conditions of gait were examined (normal, simulated leg-length difference and simulated leg-weight difference). Hip-knee joint angle diagrams were constructed for each of eight subjects under each of the three conditions. The virtue of the hip-knee joint angle is that it represents movement of almost the entire body. Thus it may be a good indicator of a subject's gait and hence provides a basis for gait pattern differentiation. Apart from its potential for fine tuning a runner's action, it has obvious and more worthwhile benefits in assisting in the diagnosis of conditions such as cerebral palsy and spastic diplegia (Gage [8]).

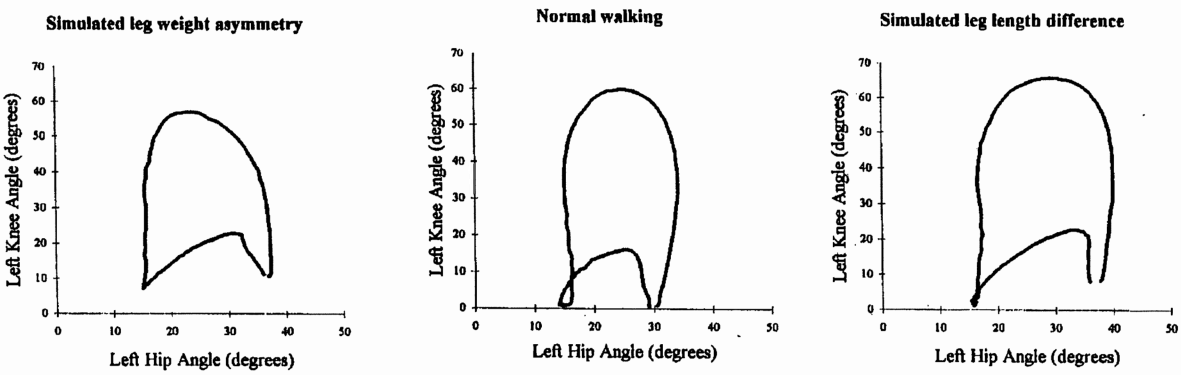


Figure 1: *Hip-knee joint angle diagrams*

Hip-knee joint angle diagrams, see Figure 1, show the changes in the knee-joint angle as a function of the hip-joint angle and thus allow investigation of the relationship between the two angles although the temporal change of each is lost.

Since the temporal dependence has been 'lost', analysis in terms of time seems inappropriate, hence the frequency domain analysis leads to the use of Fourier transforms.

Barton and Lees obtained 128 angle values with constant time interval for each of the hip and knee angles. The fast Fourier transform (FFT) was applied to these resulting in 64 real coefficients and 64 imaginary coefficients for each angle. In fact only the lower frequencies were necessary to define essential characteristics and ultimately 30 pieces of data (2 x 8 real and 2 x 7 imaginary coefficients, the first imaginary coefficient was always zero) were used to represent each of the three conditions for each of the eight subjects, see Figure 2. It is immediately apparent from Figure 2 that the FFT coefficients are very similar across the three gait patterns, with differences of less than one standard deviation between them. This suggests that an isolated coefficient would not be sufficient to separate the conditions and that discrimination requires all the coefficients to be considered together - and this is where the neural network comes in.

The neural network adopted was a four layer, back propagation net with 30 neurons in the input layer (corresponding to the data set) and 3 neurons in the output layer (corresponding to the three conditions). Two hidden layers were included (of five and four neurons respectively), see Figure 3.

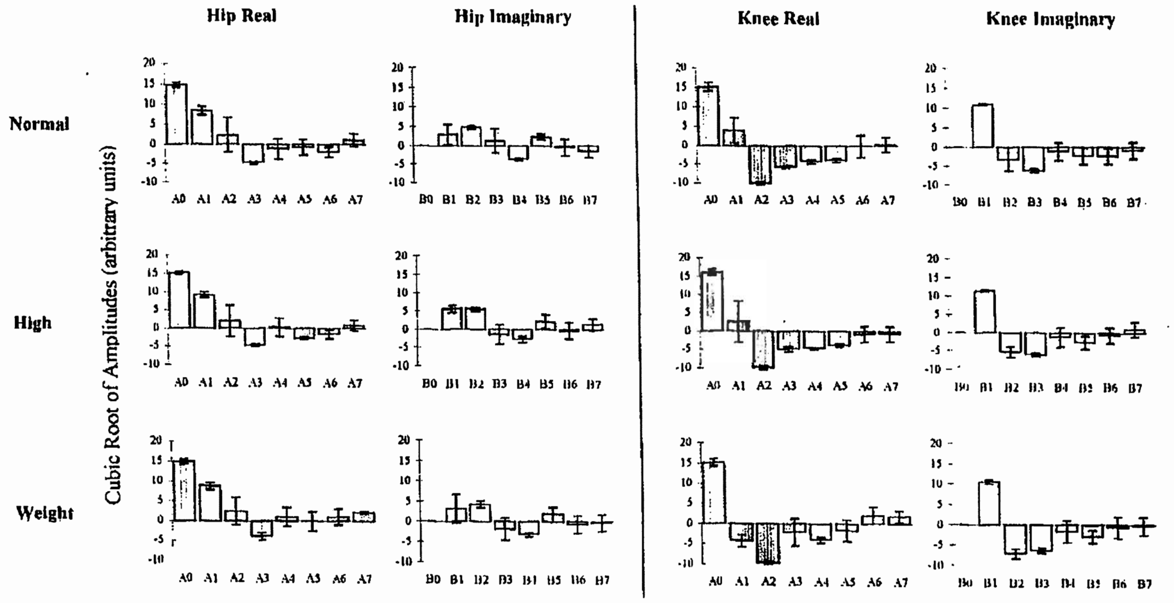


Figure 2: *Fourier coefficients of hip and knee angle data*

The twenty four data patterns (8 subjects x 3 conditions) were split into a training set and a test set. The training set data were presented to the NN one after another and by dynamically changing its internal weights it learned to associate the data patterns with their corresponding gait condition. Once trained, the test patterns were presented and the capability of the NN assessed by the assignment ratio (the ratio of the number of successful recognitions to the total number of test patterns). Four different training and test subsets of the data were used with the test patterns each submitted four times (giving 16 NN in all) and the four best NN were retained. Each of these NN assigned the unknown gait patterns to their correct category in 83.3% of the cases. The results suggest that neural networks can be applied successfully in the automated diagnosis of gait disorders.

Neural networks have also been used for assessing different insole materials used in the construction of running shoes (Barton and Lees [9]). They have proved capable of recognising different foot pressure points associated with different insole conditions. The networks respond accurately even to incomplete data sets (such as would be experienced in the case of a failure of the pressure sole) and, perhaps most important from the experimental perspective, they have proved insensitive to random noise.

As a final example of their potential consider Figure 4 which shows examples of EMG activity of the four large muscle groups of the leg during a number of steps and the vertical component of the ground reaction force obtained from one of these steps.

The EMG values have been used as the input to a neural network which was trained to map the signals to the vertical force component (Barton, Lees and Wit [10]). Once trained, previously unknown EMG data were presented to the NN and the results indicate that the proposed method acceptably predicts the vertical ground reaction from EMG data (high correlation coefficient 0.90±0.05 and low r.m.s. error 0.06±0.02). The implications for biomechanicians are enormous - data can be collected during unrestrained activity (as opposed to a force platform test) over several strides (rather than just one) and then analysed later using the neural network.

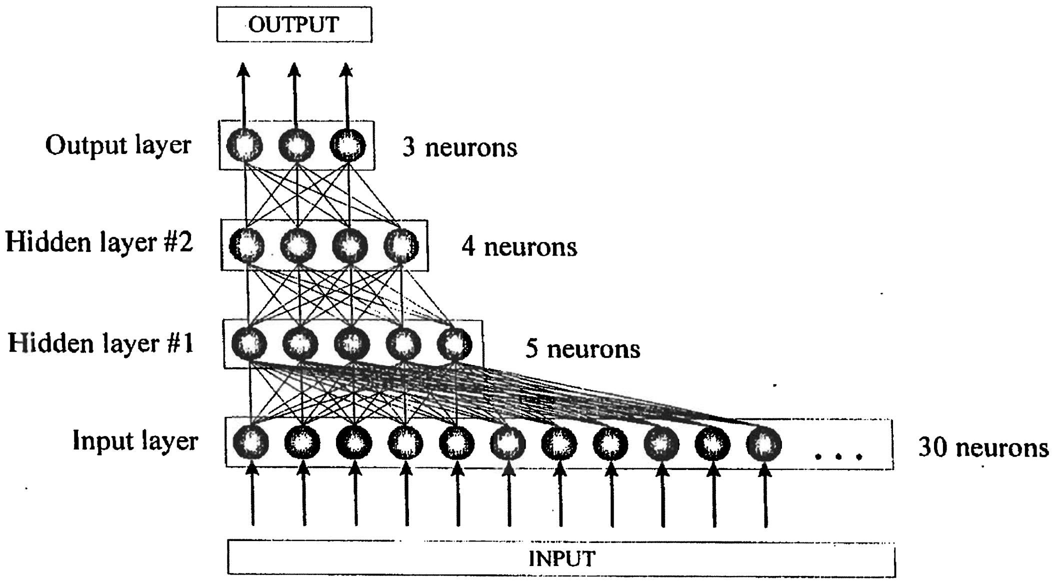


Figure 3: *Four layer neural network with two hidden layers*

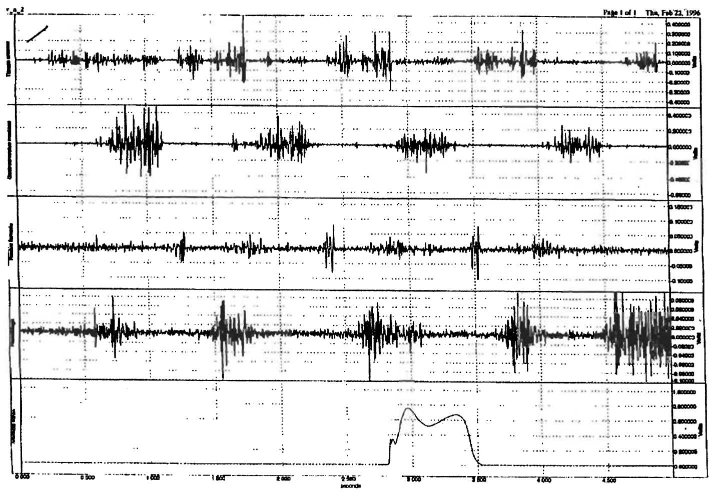


Figure 4: *EMG activity of muscle groups of the leg and vertical component of ground reaction force*

2. A GENDER SPECIFIC MODEL OF THE TORSO

The most significant modern study of modelling of the human body is an analysis conducted on male cadavers and dates from 1955 (Dempster [11]). One of the most popular models still in widespread use today uses Dempster's data and dates from 1964 (Hanavan [12]).

More recently other experimental techniques have been developed including radiation techniques and photogrammetry (Miller and Nelson [13]), immersion techniques (Plagenhoef et al [14]) culminating in 1994 with the use of moment tables and pendulums (Nigg [15]).

In the majority of these studies women have been treated as men with just more of the same fat (Vogel and Friedl [16]). In fact, fat regulation in women is considerably more elaborate with more and different sites for storage and a larger proportion of fat distributed to the extremities and in subcutaneous locations (Vogel and Friedl [16]). There also exist structural differences such as a wider pelvis and narrower and lighter shoulders.

Over the past four decades there have been only four significant studies devoted to the female body (Mori and Yamamoto [17], Fujikawa [18], Plagenhoef et al [14] and Hatze [19]). The first two studies used Oriental subjects and hence their results do not transfer directly to women in other ethnic groups, Plagenhoef's study yielded information about mass distribution while Hatze's model is of extreme complexity and allows for gender, variable density, obesity and pregnancy but requires 242 anthropometric measurements per subject.

Hanavan's model and its derivative (Robertson [20]) are based on a 15 segment model using truncated cones, spheres and elliptic cylinders. Robertson reports that the segments which deviate most from experimental analysis are the upper and lower trunk. These are also of course the most obviously different segments with respect to gender.

Given the equality of opportunity which now exists for both sexes (ranging from female fighter pilots to increased female participation in sport) the time is right to develop a more accurate biomechanical model of the female gender.

Recent unpublished work (Gourley [21]) supervised by this author within the Centre for Sport and Exercise Science at Liverpool John Moores University has been the development of a model for the upper two thirds of the trunk. The thorax and abdomen are modelled using stadium shapes, see Figure 5, a shape which was first proposed in 1990 (Yeadon [22]). Gourley has used the model to determine segmental masses, centre of mass locations and the moments of inertia about different axes, see Figure 6.

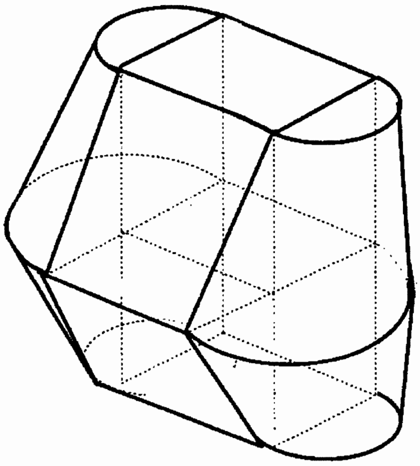


Figure 5: *Proposed model of thorax and abdomen*

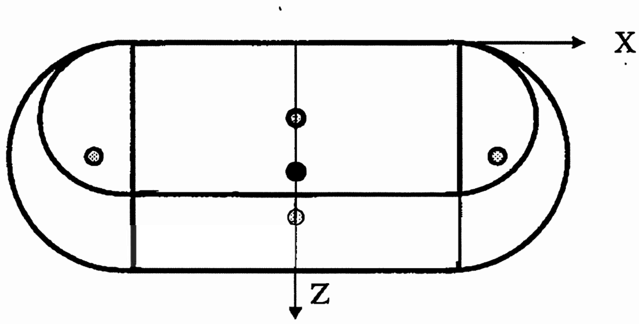


Figure 6: *Axes used for centre of mass location and moment of inertia calculations*

Based on a sample of 20 female subjects and 8 measurements per subject (plus another 26 to permit comparison with Robertson's model) the present study revealed the following results:

Centre of mass	$\bar{y}$ thorax sig. diff. w.r.t. Robertson model
	$\bar{y}$ thorax sig. diff. w.r.t. Plagenhoef model
	$\bar{y}$ abdomen not sig. diff. w.r.t. Plagenhoef model

Moments of Inertia	I_X marginal sig. diff	w.r.t Robertson model
	I_Y sig. diff.	
	I_Z sig. diff.	

Segmental mass sig. diff. w.r.t. Plagenhoef model

The model suggests that in this context the differences which exist with respect to earlier models are sufficiently significant to warrant further development of this gender specific model. Future developments will include first the modelling of the pelvic region followed by refinement to allow for the presence of body cavities.

To bring the story completely up-to-date a recent study has been undertaken in the United States (Ackerman [23]) using CAT scans and MRI images of transverse sections of the body at 1 mm intervals, although the results based on the female data have yet to be published. We may, therefore, be at the threshold of Eiben's prediction that by the millennium the categorisation of the human physique of both genders will be more thorough and objective than had hitherto been thought possible (Eiben [24]).

3. AN APPLICATION OF CHAOS THEORY AND NOTATIONAL ANALYSIS

Analysis of performance is central to the coaching process and has evolved over the past decade into a discipline known as Notational Analysis. Methodologies have been developed for analysing in detail the methods that teams and individuals adopt in their efforts to win competitions and matches in soccer, rugby, karate and horse jumping (Hughes and Reilly [25]). The analysis also examines the tactics used and the contributions and fitness of individual players where appropriate.

As the methodologies have been refined, the information retrieved by them has increased in volume and complexity. For example one notational analysis system developed for the analysis of soccer matches (Hughes [26]) associates twenty four variables with the play. Hughes has designed a 'concept keyboard' which divides the pitch into a matrix having one hundred and twenty eight touch sensitive cells. Around this are the 'keys' of the keyboard representing the team members, symbols for functions and outcomes (e.g. GK for goalkeeper, DRIB for 'dribble' etc.). Matches can be analysed either from video tape or directly if the operator is skilled.

The printout of the match analysis is considerable, some 3 cm thick, and shows a wide range of information such as the passing distribution, how possession was lost, how free kicks were conceded and so on.

Although the computation time of the analysis is quite fast it could potentially be speeded up if the analytical program was made more discerning so that coaches and managers could be selective in the parameters they wished to investigate.

Also if both the analysis and the computation time were speeded up, for example by resorting to parallel computing algorithms, the analysis could be performed as the match was being annotated so that strategies could be modified in the light of the information presented.

There is so much activity on the pitch that no one can mentally analyse everything that is going on. Hughes has evidence that coaches have recalled incorrectly up to 70 per cent of what happened in games when their accounts were compared with videotapes! The match statistics resulting from a notational analysis are seen therefore to possess greater reliability than that of even an 'expert witness'. The technique could thus usefully be applied to warm-up games before a tournament since that is where managers experiment with different playing styles and different combinations of players. Reliable soccer statistics are also used as a source of data for the analysis of injury patterns within the game, leading to changes in the rules to make the game safer.

Prior to the technological explosion of the IT age it was assumed that the efficient solution of non-linear problems only required the gathering and processing of more information.

However, the first conclusions of chaos theory have shown that simple deterministic systems with only a few varying parameters can generate purely random behaviour which cannot be removed by gathering more information. This fundamental randomness has come to be known as chaos. A chaotic system may thus be defined as one which for some condition becomes sensitive to initial conditions.

A simple and accessible example of the phenomenon is provided by a popular snooker trick shot shown in Figure 7.

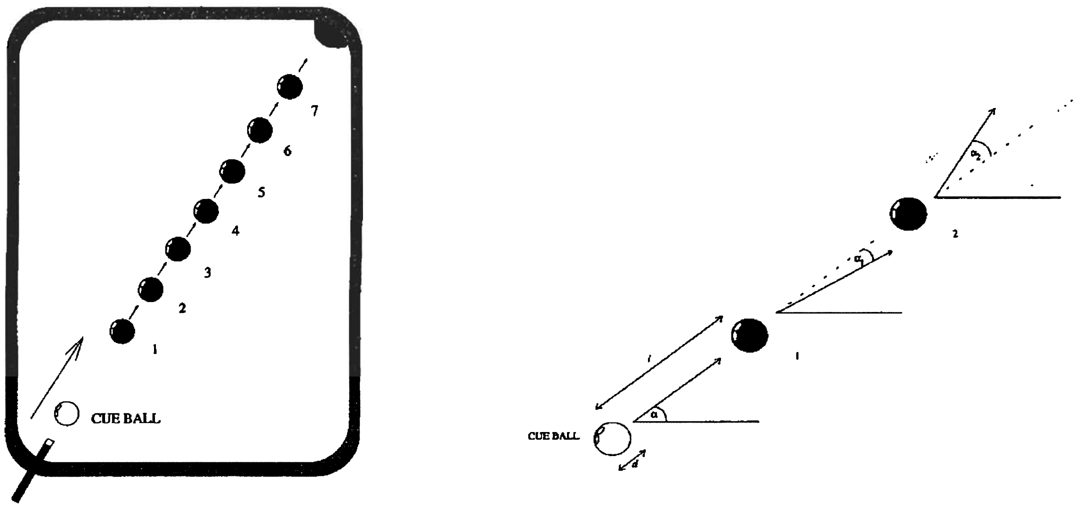


Figure 7: *Snooker trick shot and associated geometry*

The shot involves striking the snooker ball labelled 1 with the cue ball so that a 'cannon shot' occurs (ball 1 then collides with ball 2 which then collides with ball 3 .... until finally ball 7 is potted). This is not a trivial task as the following analysis demonstrates.

Referring to the notation shown in Figure 7 it can be shown that

$$\alpha_1 = \frac{(L - d)}{d}\alpha$$

If we consider a typical case of $L \cong 20$ ins $\cong 508$ mm and $d \cong 50.8$ mm then $\alpha_1 \cong 9\alpha$. This angle then becomes the 'input' angle for the ball 2 - ball 3 collision so that

$$\alpha_2 \cong 9\alpha_1 \cong (9^2)\alpha \, .$$

The final error angle is thus

$$\alpha_7 = (9^7)\alpha \cong 4782969 \; \alpha \; ,$$

an initial error of just $0.00001°$ will cause a final error of $47.83°$!

These results exhibit the sensitivity to initial conditions which is typical of chaotic systems.

Recent work at the University of Wales Institute in Cardiff (Hughes and Lyons, reported in [26]) investigates the potential of applying chaos theory to the coaching of a soccer team. Although at first sight a soccer match may seem to be a largely random affair involving twenty two players who can kick a ball anywhere on a pitch it is in fact a highly structured game in which both teams follow well-established patterns of play. Amidst these patterns there are often four or five occasions on which the game deviates from its rigid structure and their evidence suggests that it is the teams which create and exploit the perturbations which are more likely to be successful.

Their second application of chaos theory has been to use it to examine the evolution of a team and its long term strategies throughout a tournament or a season (or seasons). The team's patterns of play are modelled from the extensive notational database built up by Hughes. With an intimate knowledge of an opponent's patterns of play a team can endeavour to develop tactics to upset (i.e. perturb) these patterns.

Of course the sceptics will point out that these 'magic moments' are evident to all - they certainly don't need a sophisticated mathematical theory to identify them. To a certain extent that is true but there are many critical incidents which are either not remembered or recalled incorrectly by the pundits. A complete notational analysis of a match reveals the orderliness behind the play and makes it easier to identify 'the ripples of player-induced chaos that upset these patterns.'

Spotting a pattern is, of course, only half the story, exploiting it is the important stage. While outstanding players create their own perturbations not every team can afford such players, hence the importance of being able to identify any perturbations and their effect on the result of a game so that they can be incorporated into the team's coaching.

As stated above Hughes and Lyons have also studied team performances over a period of time (for example a tournament or consecutive seasons of play) and found that many teams have periods of equilibrium (successful or otherwise) comprising of a settled team and a comfortable style of play. Introduction of a new player or a change of style 'leads to a second phase - a period of chaos for which the outcome is uncertain.' Ultimately a new equilibrium is reached which may or may not be more successful than the first stage.

Their research aim is to identify how the successful teams emerge even better from the second stage. As Lyons remarks (reference [26]) 'investing in chaos may prove beneficial for teams if they have been relatively unsuccessful over a period of time.

Perhaps we need to introduce a culture where informed risk taking is the norm otherwise systems never move into the second stage of chaos that allows evolution of a new team'.

4. MODELLING THE EFFECTS OF TRAINING

As athletes continually strive to improve their performance there is a very real risk that in their enthusiasm they will overtrain and either injure themselves or adversely affect their performance due to fatigue.

A recent study (Morton [27]) has modelled an athlete's response to training by adapting the established dosage response model used in pharmacological studies. The training input (measured in arbitrary units) is equivalent to the drug dosage and the output (fatigue, performance response, ...) is equivalent to the performance response to the drug. Figures 8 and 9 show respectively the time dependent response of a subject to the administration of a drug and the general pattern of the effect of a daily 'dose' of training on an athlete's fitness (or fatigue). The parallels are self evident.

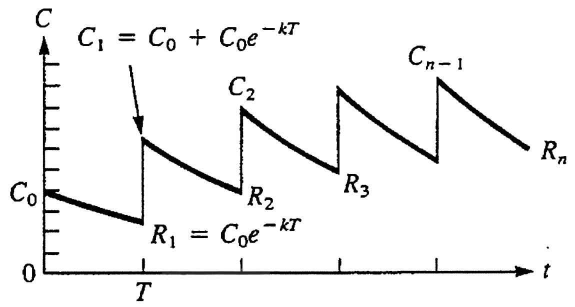


Figure 8: *Increase in drug concentration due to repeated dosage*

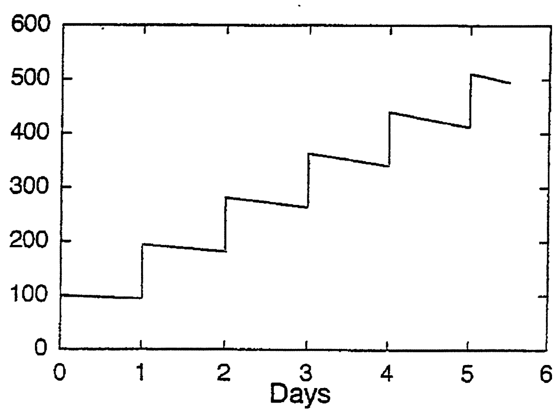


Figure 9: *General pattern of the effects of a daily dose of training on fitness or fatigue (arbitrary vertical scale)*

For the athlete and coach the benefit of such a model is that it offers the potential to maximise future performance while minimising the risks of overtraining. It offers an alternative to models based on optimal control theory as exemplified by Maronski (references [28], [29]).

Morton's model first defines a unit of measurement which provides a quantitative measure for the 'dose' of training administered. The effects of training are to increase both fitness and fatigue, both of which will subsequently decay at different rates if no further training is undertaken. Successive bouts of training will enhance fitness but with a progressively smaller incremental benefit and concurrently fatigue levels will also increase towards a plateau. The performance of the athlete is defined in terms of an index equal to the difference between fitness and fatigue and is shown over a longer period in Figure 10, where there is an initial decrease in performance in response to a new training load. Curtailment, or cessation, of training results in fitness, fatigue and performance patterns as shown in Figure 11 from which the benefits of tapering the training are immediately apparent.

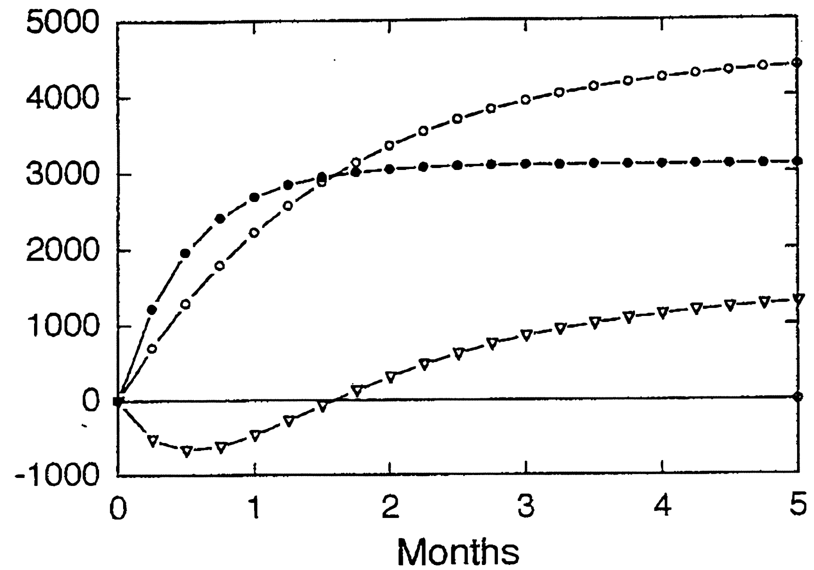


Figure 10: *Fitness (open circles), fatigue (closed circles) and performance (open triangles) as functions of daily dose of training; after Morton*

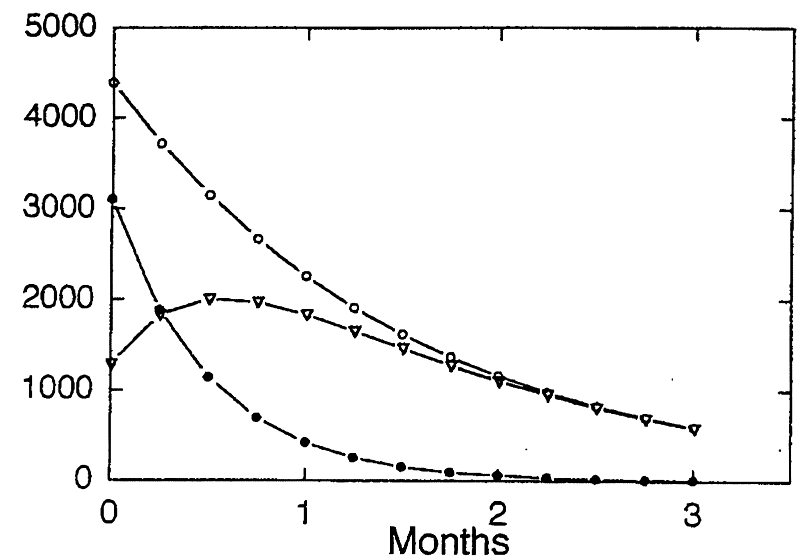


Figure 11: *Fitness (open circles), fatigue (closed circles) and performance (open triangles) after cessation of training; after Morton*

Quantitative measures of fitness, fatigue and so on are provided by various biomechanical markers such as iron status in female long distance runners, serum enzyme activities and glutamine (Banister and Hamilton [30], Banister et al [31] and Rowbottom et al [32]).

These data have enabled Morton to carry out many simulations of his model. He concludes that for optimal performance the heaviest training loads should be applied on alternate days during the twelve to fourteen weeks prior to competition and that a period of about two weeks of very light training should immediately precede the competition. His conclusions reflect the training programmes adopted by many athletes in sports as diverse as track and field, swimming and weightlifting. The model is therefore extensively validated and its development could lead to further increases in performance.

5. COMPUTATIONAL FLUID DYNAMICS (CFD) AND YACHT DESIGN

In 1907 F.W. Lanchester stated that 'the problem of yacht mechanics resolves itself into an aerofoil combination in which the aerofoil acting in the air (the sail) and that acting under the water (the keel) mutually supply each other's reaction'.

This is as true today when the rewards for success in ocean racing can be enormous and hence design techniques which will enhance a yacht's performance are eagerly sought. The methods of computational fluid dynamics (CFD) are now used to optimise the hull design of yachts and catamarans.

The method represents a sophisticated modelling technique which replaces much of the expensive and time consuming tank testing thus reducing development time and improving design. CFD techniques have already been used to optimise the hull design of some competitors in the American cup races and currently are used to assist in the hull design of the ILC40 class of yachts.

The design of the hull(s) of a yacht or catamaran interacts with the rudder design since the hull decelerates and diverts the flow around it.

A CFD analysis represents the hull-rudder assembly by a large number of panels. Solution of the hydrodynamic equations is mathematically challenging due to the size of the problem, the numerical instabilities which exist in the wake and the need to perform numerical differentiation of the potential function in order to obtain the pressure distribution over the hull-rudder surface. CFD allied to a parallel computing methodology represents a practical design tool which permits geometric variations in hull design and the analysis of a range of flow conditions.

Above the surface, the distributed loads on the sails are needed for structural calculations. A method has been developed for predicting the viscous flow past yacht sails with structural deformation (Fiddes [33]). The method synthesises aerodynamic and structural methods, the finite element based structural code having originally been developed to predict the behaviour of membrane structures such as the Olympic Stadium in Munich.

Multihulls present special problems despite their inherent advantage of large transverse stability. The potential for improved design here lies in enhancing their seakeeping qualities. For example the height of the cross structure is important in order to avoid too much slamming in head seas, while the hull separation is important in order to avoid any hull-hull interactions. For such interactions to be avoided then the waves generated at the bow of one hull must not actively interfere with the other, from which it follows that the bow waves must be swept back at an angle less than

$$\arctan (H/L_w)$$

relative to the centre line of the hull, where H is the hull separation and L_w is their waterline length.

Most recently a radical new design of ocean racing yacht has appeared (Aqua Quorum designed by Adrian Thompson) which her skipper, Peter Goss, has likened to a fifteen metre surfboard. Apart from her lightness (the bare hull will float in 0.3m of water and can be lifted by six men) the interest to this conference lies in her compound pendulum keel.

The keel consists of a lead bulb, of mass two thousand five hundred kilograms, suspended at the end of a steel pendulum. Inside the yacht two hydraulic rams can push the keel out to windward to an angle of up to thirty degrees. The designer claims that this keeps the yacht upright (therefore sailing faster) than a fixed keel. The 'surfboard' is achieved by the combination of retractable dagger boards (as found on a sailing dinghy) and the swinging keel.

The acid test for the design will come in November 1996 in the Vendée Globe single handed, non-stop around-the-world race.

Is it not amazing that despite the application of complex techniques such as CFD, yacht design may still be significantly influenced by something as mundane as a compound pendulum?

6. THE FUTURE

The next few years will surely see more widespread sports-oriented applications of the following mathematical techniques discussed.

(i) Parallel Computing – to process the increasingly large data sets which result from ever increasingly sophisticated experiments and the application of advanced numerical methods.

(ii) Computational Fluid Dynamics (CFD) – providing optimisation of hull and sail design of yachts and catamarans. There has already been a noticeable impact in the Americas Cup races. The extremely complicated fluid motions around the swimmer may also be amenable to CFD analysis.

(iii) Chaos and Catastrophe Theory – these theories which currently provide qualitative explanations for sudden changes in the behaviour of a system could be applied to examine and possibly predict the social problems associated with sudden outbreaks of violence both on and off the soccer pitch.

(iv) Neural Networks – their use as a diagnostic tool will surely become more widespread in areas such as biomechanics and exercise physiology as the experimental designs adopted there became more complex.

(v) Optimisation techniques – although these have been illustrated with applications to the training process they are already widely used for tactical and strategic analysis both within sport and elsewhere.

To pass judgement today on the potential impact of any of these mathematical fields is rather like trying to estimate the size of a wave which is still on the horizon in Newton's 'great ocean of truth.' All we can say is that it began deep in the ocean and that it has travelled faster than most. These are precisely the attributes of tidal waves, the ones which sooner or later affect us all.

REFERENCES

[1] Holzreiter S.H. and Köhle M.E., Assessment of gait patterns using neural networks, *J. of Biomechanics,* Vol 26, 645-651 (1993).

[2] Miller A.S., Blott B.H. and Hames T.K., Review of neural network applications in medical imaging and signal processing, *Medical and Biological Engineering and Computing,* Vol 30, 449-464 (1992).

[3] Aminian K., Robert P., Jequier E. and Schultz Y., Level, downhill and uphill walking identification using neural networks, *Electronic Letters* Vol 29, 1563-1565(1993).

[4] Eberhart R.C. and Dobbins R.W., Neural Network P.C. Tools, A Practical Guide, (Academic Press, San Diego, 1990).

[5] Zurada J.M., Introduction to Artificial Neural Systems (West Publishing Company, St. Paul, 1992).

[6] Lawrence J., Introduction to neural networks (California Scientific Software, California, 1993).

[7] Barton J.G. and Lees A., 'An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams', *Gait and Posture,* to appear.

[8] Gage J.R , 'Gait analysis for decision making in cerebral palsy', *Bulletin of the Hospital for Joint Diseases Orthopaedic Institute*, Vol 43, 147-163 (1983).

[9] Barton J.G. and Lees A. The feasibility of a neural network approach for assessing insole materials: a preliminary study, XV Congress of the Intnl.Soc. of Biomechanics, Jyväskylä, Finland, (July 1995).

[10] Barton J.G., Lees, A and Wit, A. 'The use of neural networks to estimate ground reaction forces from electromyographic activity', 2nd Nat. Conf. of the IPEMB and 3rd Biennial Radiotherapy Physics Meeting, Leeds UK, (September 1996).

[11] Dempster W.T., 'Space requirements of the seated operator: geometrical, kinematic and mechanical aspects of the body with special reference to the limbs', *WADC Technical Report*, 55-159 (1955).

[12] Hanavan E.P., 'A mathematical model of the human body', Technical Report AMRL TR64-102, Wright-Patterson A.F.B, Ohio, USA (1964).

[13] Miller D.I. and Nelson R.C., Biomechanics of Sport (Lea and Febiger, Philadelphia, USA, 1973).

[14] Plagenhoef S., Evans F.G. and Abdelnour T., 'Anatomical data for analysing human motion', *Research Quarterly*, Vol 54, 167-178 (1983).

[15] Nigg B.M., 'Inertial properties of the human or animal body' in Biomechanics of the musculo-skeletal system by Nigg B.M. and Hertzog W. (J. Wiley and Sons, Chichester, 1994).

[16] Vogel J.A. and Friedl K.E., 'Body fat assessment in women; special considerations', *Sports Medicine*, Vol 13, Pt4, 245-269 (1992).

[17] Mori and Yamamoto (1959). Cited by Luttgens K., Deutch H. and Hamilton N., Kinesiology-Scientific Basis of Human Motion, 8th edition, (WCB Brown and Benchmark, Oxford, 1992).

[18] Fujikawa (1963). Cited by Luttgens K., Deutch H. and Hamilton N., Kinesiology-Scientific Basis of Human Motion, 8th edition, (WCB Brown and Benchmark, Oxford, 1992).

[19] Hatze H., 'A mathematical model for the computational determination of parameter values of anthropomorphic segments'; *Journal of Biomechanics,* Vol 13, 833-843 (1980).

[20] Robertson K., 'The development of a computerised mathematical model to represent the human body'. Unpublished thesis (Liverpool John Moores University, England, 1985).

[21] Gourley C., 'Design and Validation of a mathematical model of the torso which accounts for gender differences'. Unpublished project dissertation (Liverpool John Moores University, England, 1996).

[22] Yeardon M.R., 'The simulation of aerial movement - II. A mathematical inertia model of the human body', *Journal of Biomechanics*, Vol 23, Pt 1, 67-74 (1990).

[23] Ackerman M.J., 'Accessing the visible human project',
http://www.dlib.org/dlib/october95/10 ackerman.html (1995)

[24] Eiben O.G., 'The physique of women athletes', (The Hungarian Scientific Council for Physical Recreation, Budapest, 1972).

[25] Reilly, T. (editor), Science and Soccer (Chapman and Hall, UK, 1996).

[26] O'Hare M., 'Chaos Pitch,' in New Scientist, pp. 24-28, 8 (June, 1996).

[27] Morton, R.H., 'Modelling Training and Overtraining', Proceedings of International Symposium on Fatigue during competitive sport: nutritional and training strategies to optimise performance, Cape Town (1996), to be published in *J. Sports Sciences.*

[28] Maronski, R., 'Optimal distance from the implement to the axis of rotation in hammer and discus throws', *J. Biomechanics*, Vol. 24, No. 2, pp 999-1006 (1991).

[29] Maronski, R., 'On optimal velocity during cycling', *J. Biomechanics*, Vol. 27, No. 2, pp 205-213 (1994).

[30] Banister, E.W. and Hamilton, C.L., 'Variation in iron status with fatigue modelled from training in female distance runners', *European J. of Applied Physiology*, Vol. 54, pp 16-23, (1985).

[31] Banister, E.W., Morton, R.H. and Fitz-Clarke, J.R., 'Dose response of exercise: elevation of serum enzyme activity'. In: First IOC World Congress on Sports Sciences, Colorado Springs, Colorado, U.S.A., pp 62-63 (1989).

[32] Rowbottom, D.G., Keast, D. and Morton, A.R., 'The emerging role of glutamine as an indicator of exercise stress and overtraining', *Sports Medicine*, Vol. 21, pp 80-97 (1996).

[33] Fiddes S.P., 'An aerostructural model for yacht sails', CFD in Ship and Yacht Design, EPSRC Community Club in CFD, Proc. 18, University of Newcastle, UK, (1994).

LOW-DRAG ROWING SHELLS

E. O. Tuck[1] and L. Lazauskas[2]

Abstract

A displacement vessel of a given loaded weight has a theoretical optimum length which minimises its total (viscous plus wave) calm-water drag. This length is usually somewhat greater than that of conventional merchant or naval ships but is in an appropriate range for competition boats such as rowing shells. Some simple examples are given to illustrate this property. Genetic algorithm techniques are then used to find optimum dimensions for rowing shells over a wide range of speeds and displacements, with a fixed assumption about the waterline, cross-section, and buttock shapes. Michell's integral is used for the wave resistance, the 1957 ITTC line for the skin friction, and a simple empirical formula for the form drag.

1. INTRODUCTION

Consider first a class of monohull ships moving steadily ahead in flat calm deep water. Fix the displacement, draft, speed, and hull shape. The length is then essentially the only variable allowed. At any given length, adjust the beam by uniform scaling of all offsets, so as to achieve the prescribed displacement; longer ships are thinner. Now vary the length until the total (viscous plus wave) drag Rt=Rv+Rw is minimised.

The above simplified ship optimisation problem, with length as the only variable, usually possesses a non-trivial solution, i.e. a finite optimum length, for the following reason. Viscous drag Rv is predominantly skin friction, which is proportional to surface area, and as a body of a given volume gets longer and thinner, its surface area increases. Hence viscous drag increases with length at fixed displacement.

On the other hand, for conventional ships at conventional speeds, wave resistance Rw generally decreases as the shiplength increases. Since we are holding the speed U fixed, as we increase the length L we are decreasing the length-based Froude number F=U/sqrt(gL). At fixed displacement, and at most relatively low Froude numbers, wave resistance is a (rapidly) increasing function of Froude number, and therefore decreases with increasing length. This wave-reducing advantage of long ships is very much part of the naval architectural art.

[1]Department of Applied Mathematics, The University of Adelaide, Adelaide SA 5005

[2]Department of Applied Mathematics, The University of Adelaide, Adelaide SA 5005

Since there are opposite trends with length in the two constituents of the total drag, there must be a minimum for their sum, at some non-trivial intermediate length. That length is generally somewhat larger than for conventional ships, but not necessarily so for rowing shells and other competition boats.

In fact, sometimes there is more than one local minimum in the graph of total drag versus length, and this phenomenon is discussed in more detail in the following section. There is often a delicate interplay between local and global optima, which makes for an optimisation process that is quite difficult to analyse. In order to deal with this problem, we use here a powerful general purpose technique, described later, called "genetic algorithms".

In the present paper, we perform an exhaustive treatment of this optimisation problem for a family of monohull vessels, covering a large range of speeds . We hold the hull shape, displacement and speed fixed, and allow the draft as well as the length to vary until the minimum total drag is achieved. We treat viscous drag as the sum of skin friction (estimated by the 1957 ITTC line) and a generally small but sometimes crucial form drag contribution which is estimated by an empirical formula. We use Michell's integral for the wave resistance, which is only accurate for thin ships. However, this is a more than usually good assumption for the class of extremely fine hulls that arise from this optimisation process.

The main purpose of the present study is to provide a benchmark, from which extended studies can follow. One class of such extensions obviously involves allowing the shape of the hull to vary. We present here results for a very fine type of hull, appropriate for high-speed and sporting-type vessels. When there are length (or other) restrictions, and hence (for shorter-than-optimal ships) a greater contribution of wave resistance to the total drag, multihulls can have less total drag than monohulls of the same length, because of the potential for favourable hull-hull cancellation of wave resistance. Work on both of these extended studies is nearly complete and will be reported elsewhere.

However, perhaps of greater importance is inclusion of further constraints, such as constraints on maximum length or minimum beam, which arise inevitably from commercial, structural, safety, seakeeping, or sporting requirements. When these constraints are imposed, the ship proportions will return to the more conventional range, but at a price in terms of increased total drag. It is of value to know just how much of a price is being paid.

1.1 An illustrative example

In the present section, we first give an example illustrating the character of the results obtained in the present study. Further results are presented in more generality and in nondimensional form later. For this example, we confine attention to a "ship" of one-tonne displacement, representative of a (large) rowing shell, and use dimensional units.

Figure 1.1 shows two typical examples of graphs of total drag versus length (in metres) at a fixed speed, for such a vessel. For the present purpose, it is not essential how the drag is determined or scaled, but we should note that it does include an allowance for form drag, discussed later. The solid curve is at a fixed speed of 5.56 knots and the dashed curve at only a very slightly higher speed of 5.59 knots. In both cases, there are two prominent minima, i.e. two distinct (and remarkably different) lengths are locally favourable, and define "best" and "second-best" boats. At the lower speed, the longer boat (13.2 metres length) is better than the shorter boat (9.8m), whereas at the higher speed, the shorter boat (9.6m) is superior to the longer boat (12.3m). Thus, as we vary the speed and other parameters, there may occur an interchange between two local optima, so that the optimum length may appear to change discontinuously. These changes can occur over a remarkably narrow range of speeds.

This type of discontinuity in the optimum length is shown in Figure 1.2, again taken from the family of one-tonne monohulls. This figure gives the optimum length in metres as a function of the speed in knots. The discontinuities indicated above occur only at relatively low speeds, notably at about 5.6 knots (where the change between the two curves of Figure 1.1 occurs) and 4.3 knots, with smaller discontinuities at even lower speeds.

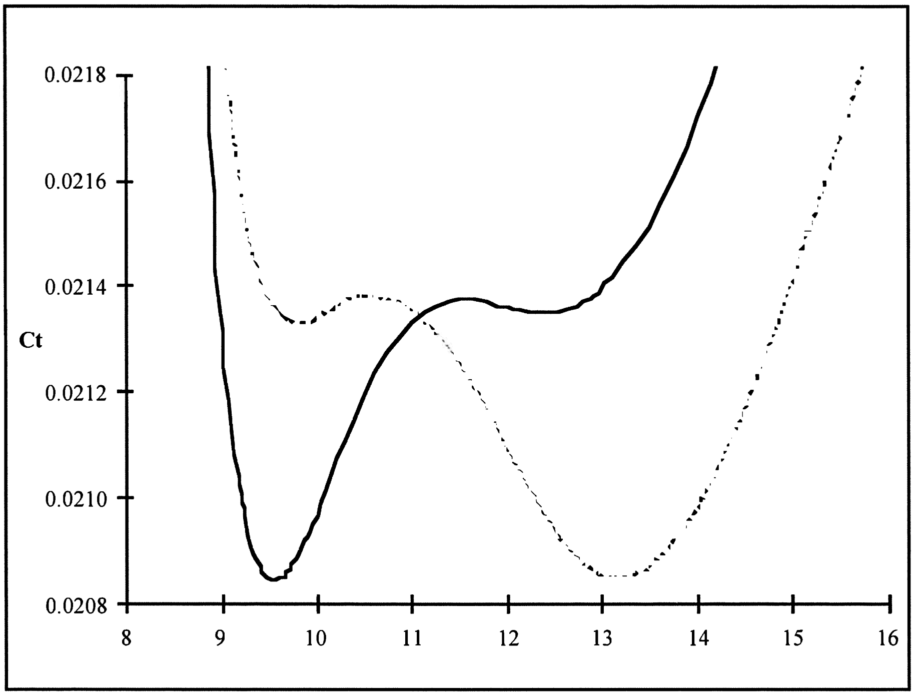


Figure 1.1: *Comparison of total resistance for two one-tonne monohulls.*

At speeds between the discontinuities, the Froude number based on boatlength remains essentially constant, and examination of the variation of wave resistance with Froude number indicates that this constant value corresponds to a local minimum of wave resistance. What is happening as we increase the speed is that, in attempting to design for minimum total drag, we simultaneously increase the boatlength, in order to stay at that local minimum. This continues as long as possible while we increase the speed, and when it is no longer possible, the optimum boat suddenly decreases its length, so that the Froude number suddenly jumps to the next higher local minimum, avoiding the local maximum in between. This process is intuitively like changing gears!

The length variation in the example of Figure 1.2 is continuous for all speeds above 5.6 knots. However, as is discussed later, if form drag is neglected, there can also be an apparent high-speed discontinuity. It is important to note that, as indicated by

Figure 1.1, there is no discontinuity in the actual total drag at these speeds, merely an interchange in the roles of "best" and "second-best" boats. At the speeds where the optimum length changes discontinuously, the residual total drag tends to reach a local maximum, where its rate of change with respect to speed changes discontinuously.

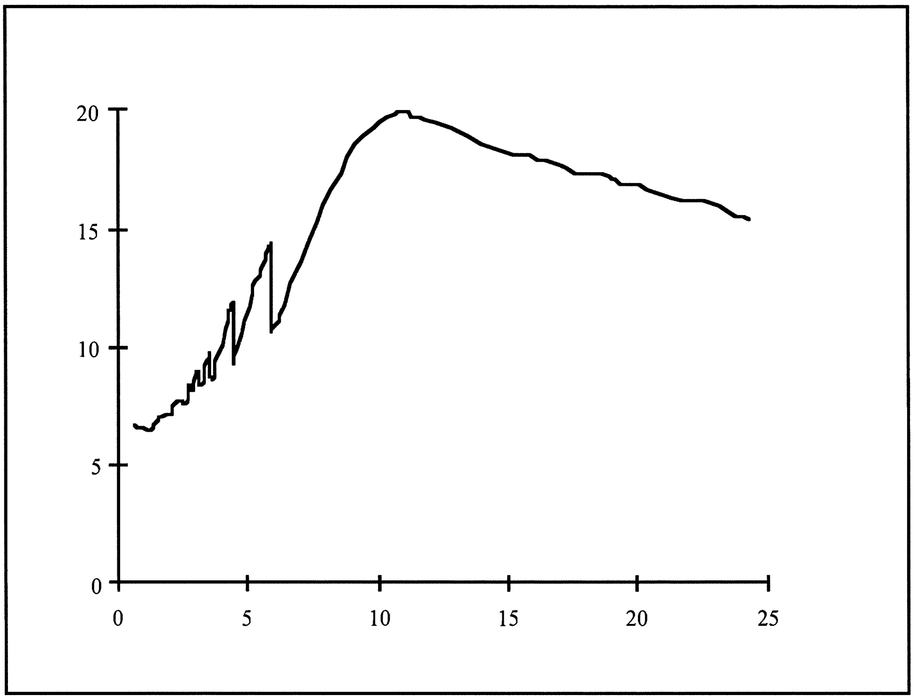


Figure 1.2: *Optimum length for a one-tonne monohull*

Although these discontinuities are of interest in their own right, they are not necessarily the most important feature of Figure 1.2. They depend on the fact that the wave resistance possesses minima, and these minima are to a certain extent magnified by the theoretical procedure (here Michell's integral) used to compute wave resistance. If more empirical means are used to estimate wave resistance, with the effect of smoothing out the humps and hollows in the wave resistance variation, there will be a consequent reduction in the size of the discontinuities. However, so long as there are at least two minima in the wave resistance curve, a discontinuity is inevitable, no matter what method is used to estimate wave resistance.

Above 5.6 knots, the optimum length of a one-tonne vessel varies smoothly, and it is unlikely that the optimum length is sensitive to the procedure for wave resistance computation. In fact, the range of speeds above that where discontinuous length changes occur is the one of greatest interest in practice; for example, it is the competitive speed range for rowing shells. In that range, the results are relatively robust, and show no surprising features.

2. Basic Considerations

2.1 Hull geometry

In this study, we present results for one hullform only – a canoe body defined by parabolic waterlines, elliptical cross-sections, and a parabolic keel line. Although this form is an obvious idealisation, there has recently been an appeal (Insel and Molland 1991) for further work on similar hull shapes.

For this hullform, the block coefficient is Cb = 0.417 and the prismatic coefficient is Cp = 4 Cb/pi = 0.531. Clearly this is a much finer type of hull than that of a typical merchant ship, but is relevant to sporting canoes and hulls of special high-speed vessels. It is particularly appropriate for slender vessels with the high length/beam ratios that we shall find optimal.

2.2 Wave Resistance

We use Michell's integral (Michell 1898; see also Tuck 1989) to estimate the wave resistance Rw of the ship. This requires evaluation of a triple integral, one integral in each of the length-wise and draft-wise co-ordinate directions, and one integral with respect to the angle "theta" of propagation of the ship-generated waves. The numerical method used here for evaluating these integrals is described fully in Tuck (1987). We use up to 81 stations, 81 waterlines, and 320 intervals for the integration with respect to theta. This is an unusually high degree of precision, and is sufficient to eliminate any numerical artefacts in the integration, which is a common source of error in the use of Michell's integral

Michell's integral depends for its validity on the ship being thin, and is sometimes considered (perhaps unfairly) to be insufficiently accurate for use with ships of conventional proportions. However, the hulls produced by the optimisation process in this study are significantly thinner than conventional ships, and there is good evidence that for such slender vessels Michell's integral is satisfactory. For example, Hanhirova et al 1995 (see also Tuck 1989 and Chapman 1972) report that for length-based Froude numbers above 0.35, accuracies relative to measured residuary resistance of better than 10% are achieved by Michell's integral for hulls with length/beam ratios of the order of 10.0. The optimised hulls in the present study are even more slender.

In any case, the hulls resulting from the optimisation process also have the property that their wave resistance is generally only about 10% of the total, so that the absolute accuracy of the wave resistance measure is not critical. This proportion of wave resistance to total drag is lower than what is usually encountered with conventional ships, since the present optimum is in part achieved by increasing the length beyond the conventional, so as to reduce the influence of wave resistance. Even though the wave resistance is then only a small component of the final total drag, it remains a critically important component nevertheless in controlling the optimisation process; after all, if there was no wave resistance at all, short ships of minimum surface area would be preferred.

2.3 Viscous Resistance

The viscous resistance Rv can be written as

$$Rv = \frac{1}{2}\rho U^2 S(Cv)$$

where ρ is the water density and S the wetted surface area of the hull. When skin friction dominates, the drag coefficient Cv approximately equals Cf, where Cf is a skin friction coefficient which can be estimated using the ITTC 1957 ship correlation line (Proc. 8th ITTC).

$$Cf = 0.075/(\log_{10}R-2)^2$$

where $R = UL/\nu$ is the Reynolds number; ν (approximately 0.000001 metres-squared/second) is the kinematic viscosity.

We have used the full length of the waterline for L in the definition of the Reynolds number; however there are other possibilities. Gerritsma et al. (1981) use 0.7L in their study of the resistance of a systematic yacht hull series, reasoning that this defines a kind of average length.

2.4 Form Effects

As a correlation line, the ITTC 1957 line already contains some allowance for three-dimensional effects, and two recent ITTC Committees have recommended that additional corrections not be made in routine resistance predictions of high speed craft (Insel and Molland 1991, p. 16). However, including a form factor specific to the hullform under consideration can often give better estimates of the viscous drag. This factor is difficult to estimate and may vary with speed because of (among other things) changes in trim and sinkage.

In their examination of eight-oared rowing shells, which have a hullform not unlike the canoe body examined here, Scragg and Nelson (1993) found a simple empirical formula for the form factor of these hulls. The viscous resistance coefficient is written as

$$Cv = (1+k)Cf$$

where

$$k = 0.0097(\text{theta\_entry} + \text{theta\_exit})$$

Here, theta_entry and theta_exit are the half-angles (in degrees) of the bow and stern, respectively, at the waterplane.

2.5 Some Effects Neglected

Wave-breaking and spray resistance is neglected. Wave-breaking resistance for our fine, sharp-bowed hulls, would be negligible at relatively low speeds. Spray resistance seems to be one of the reasons form factors are difficult to calculate at high speeds.

We assume that there is no effect of dynamic vertical forces, which at low speeds account for sinkage and trim. At high speeds, dynamic forces are upward and yield a lift rather than sinkage; hence planing, and we neglect that. The present results are for displacement rather than planing conditions, although for completeness we exhibit them even in speed ranges where planing would be expected.

Shallow-water effects can be important in some applications, e.g. see Millward (1992) for catamarans, and Scragg and Nelson (1993) for eight-oared rowing shells. However, we retain the infinite-depth assumption here. We also neglect any lateral flow domain restrictions; see Doctors and Day (1995) and Day and Doctors (1996) for the case of a ship moving in a channel.

For rowing vessels, we do not make any allowance for changes in the centre of gravity or the consequent change in trim due to the continually changing position of the rowers. The present analysis is done on a steady-flow basis, and hence relates to the average conditions during a race, neglecting speed variations due to racing conditions, as well as unsteady variations during an individual rowing stroke.

3. Prediction of Optimal Parameters

Once we have a theory that gives reasonable predictions of the total resistance, it seems natural to search for "sensible" parameter configurations minimising that resistance. Many engineering design problems can be cast into the form of optimisation problems. For example the problem addressed in this paper can be formulated as:

Minimise the real-valued function $f(x_1, x_2, ..., x_n)$, with each real parameter x_i subject to (domain) constraints $a_i \leq x_i \leq b_i$ for some real constants a_i and b_i.

Many techniques exist for solving optimisation problems such as the one described above, but these vary greatly in efficiency and the quality of the final solution for a given number of function evaluations. No single technique is best for all design problems. Gradient-based methods work well with smooth, unimodal functions, but may yield local optima for multimodal functions. Heuristic algorithms can increase search efficiency, but at the expense of guaranteed optimality – they do not always find the global optimum.

3.1 Genetic Algorithms (GAs)

GAs are adaptive search methods that use heuristics inspired by natural population dynamics and the evolution of life. They differ from other search and optimisation schemes in four main respects (Dhingra and Lee 1994):

- Search proceeds from a population of points, not from a single point.
- They use a coding of the parameters, not the parameters themselves.
- Objective function values guide the search process. They do not use gradients or other problem-specific information.
- State transition rules are probabilistic, not deterministic.

In the present study, we use a non-traditional GA similar to Eshelman's (1991) CHC, augmented with, among other features, hill-climbing routines, cataclysmic restarts and incest prevention. The resulting computer program, called "GODZILLA" for Genetic Optimisation and Design of Zoomorphs, is described in Lazauskas (1996 in preparation).

3.2 GODZILLA

GODZILLA's general operation can be described quite succintly: create and evaluate new (candidate) designs until some termination criterion is met. Termination can occur when a certain number of designs have been evaluated, or after a prescribed amount of time has elapsed, or when the algorithm seems to be making no further progress.

GODZILLA begins the optimisation process by creating an initial population of (real-valued) design vectors and calculating the total resistance for each design. Initial designs are randomly generated, although the population can also be "seeded" with previously found good designs.

Genetic operators and hill-climbing operators are used to create candidate designs. Genetic operators create new (offspring) vectors from two parent vectors in the population, using heuristics inspired by the recombination of DNA. There are too many varieties to here discuss individual strengths, deficiencies and peculiarities. GODZILLA's primary genetic operator is one gleaned from fuzzy set theory described in Voigt et al (1995). After evaluating the total resistance of the offspring, GODZILLA replaces the worst individual in the population with the offspring if the offspring's total resistance is lower. This replacement strategy guarantees that the best individual in the population is never replaced by an inferior individual.

The method used to select parent vectors from the population can have a substantial influence on the performance of GAs. GODZILLA uses binary tournament selection. In this method, two individuals are selected without replacement from the population. The individual with the lower total resistance becomes the first parent. A second binary tournament determines the other parent.

One form of hill-climbing operator used by GODZILLA, Stochastic Bit-climbing, creates a candidate vector by adding or subtracting small increments from each of the parameters of the best design vector found so far. This allows the program to explore more closely promising regions of the search space found by the genetic operators. GODZILLA also incorporates another hill-climbing technique called the Simplex Search Method. This method, which is not to be confused with the Simplex Method of linear programming, is described in Reklaitis et al (1983).

The field of evolutionary computation is expanding very quickly, and almost all communication occurs via the electronic Internet. The Usenet group, comp.ai.genetic, is a very useful and important resource.

4. RESULTS

4.1 Method of presentation

Since there is no length restriction, but the displacement D is fixed, the appropriate length parameter for scaling is the cube root L*=D^(1/3) of the displacement. Results are presented in a non-dimensional manner as a function of the (volumetric) Froude number Fnv=U/sqrt(gL*) based on that artificial length. In fact, were it not for scale (Reynolds number) effects, all results would be universal functions of this Froude number, and displacement would be irrelevant. For example, the final minimum total drag Rt=Rv+Rw is expressed in terms of the coefficient

$$Ct=Rt/((1/2)\ \rho\ U^2\ L^2)$$

which would be a function of Fnv alone were it not for the fact that the skin friction coefficient depends on Reynolds number.

In order to exhibit this scale effect of displacement, we carry out the optimisation at fourteen fixed (dimensional) displacements relevant to rowing classes, ranging from 0.075 tonnes up to one tonne. In fact we have also computed results for even larger vessels, up to one million tonnes.

For definiteness, we give most results for the fixed displacement of one tonne. Some such results have already been given in Figures 1.1 and 1.2. It is notable that for this particular displacement, L*=1 metre, so that the non-dimensional length can also be interpreted as the actual length in metres. The volumetric Froude number is also uniquely proportional to the actual speed in metres/second or knots, and Fnv=1 occurs at 6.1 knots for a one-tonne vessel.

It is important to bear in mind that none of the Figures 2 and 3 to follow, where the total drag coefficient Ct is plotted against the volumetric Froude number Fnv, can be interpreted in the usual naval architectural manner as a graph of drag versus speed for a given ship. As Fnv varies, the ship itself changes its shape, and in particular its length, so as to keep the drag as small as possible.

To produce the results in Figures 2(a)-2(e), we performed the optimisations at 83 different speeds. For the case with no form drag, the volumetric Froude numbers corresponding to these speeds are 0.100, 0.125,...,0.675; 0.680, 0.685,...,0.700; 0.725, 0.750...,0.900; 0.905, 0.910...,0.925; 0.950, 0.975; 0.980, 0.985,...,1.0; 1.025, 1.050,...,1.2; and 1.3, 1,4,...,4.0. The very small speed increments at the low end of the range were needed to capture the discontinuities described earlier. Similar ranges and increments were used to produce the plots for the case where form drag has been included.

We used a population of 64 during the optimisation. Each of the design problems was run with at least five different initial populations. A minimum of 5,000 resistance evaluations were performed during each run. To investigate further the previously discussed discontinuities, additional runs were performed with the search domain constrained in such a manner as to disallow one of the alternative solutions. This is an extremely tedious process because we don't know in advance exactly where the discontinuities might occur. Whether we like it or not, human input and understanding is still essential to complex engineering design systems.

4.2 Monohull without form drag

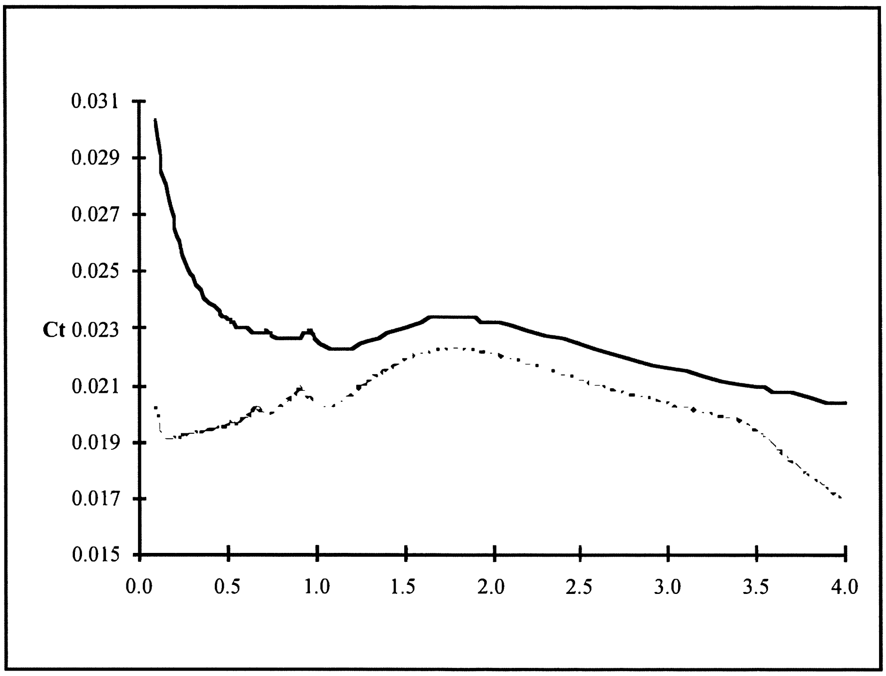


Figure 2(a): *The effect of form factor on the optimum total resistance of a one-tonne monohull.*

The dashed curve of Fig. 2(a) shows Ct as a function of volumetric Froude number Fnv, for a one-tonne boat. This is the residual value of the total drag coefficient, after the boat's dimensions have been optimised to minimise Ct without any allowance being made for form drag. The hull parameters that produce these optimal Ct's are shown as the dashed curves in Figures 2(c)-2(e).

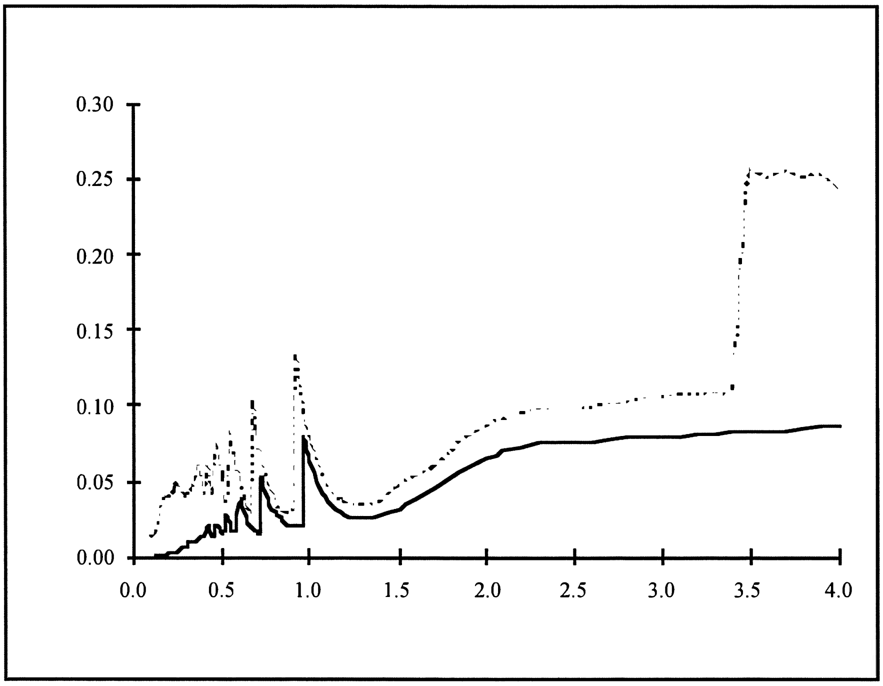


Figure 2(b): *The effect of form factor on the optimal proportion of wave resistance of a one-tonne monohull.*

Fig. 2(b) shows wave resistance as a fraction of the total drag. There is considerable scatter at low Fnv. This could be due to long shallow regions in the "fitness" landscape, where for example, one length is as good as another. Although Ct remains the same, Cw/Ct may vary. GODZILLA searches for the lowest total resistance and if it encounters two or more combinations of parameters with almost the same Ct, it cannot prefer one to the other. In these regions, it could be important to perform the integrations more accurately. In any case, wave resistance is less than 12% of the total for all Fnv<3.25.

The most obvious feature of the dashed curve in Figure 2(b), however, is the sudden increase in the proportion of wave resistance for Fnv>3.25, a rather high speed (of the order of 20 knots for a one-tonne vessel) near the upper end of the range being considered in this study. Figure 2(c) shows that the optimum length also drops sharply to a very low level at this speed. This discontinuity is essentially an interchange in the roles of two local minima, as in Figure 1.1. For Fnv<3.25, the longer boat is best; for Fnv>3.25 the shorter boat is best, and in the present case, the shorter boat is so short as to be quite unrealistic. Indeed, this boat almost eliminates its wave resistance by going to a very high rather than a very low conventional Froude number. Minimum viscous drag dictates minimum surface area, and that inevitably pushes the optimum toward a hemispherical geometry. In the present case, other constraints prevent this hemisphere being achieved exactly, but this class of "optimum" boat does tend to have length comparable to the beam and draft. Clearly this is not a realistic conclusion, and in particular would lead us to question the validity of neglecting form drag.

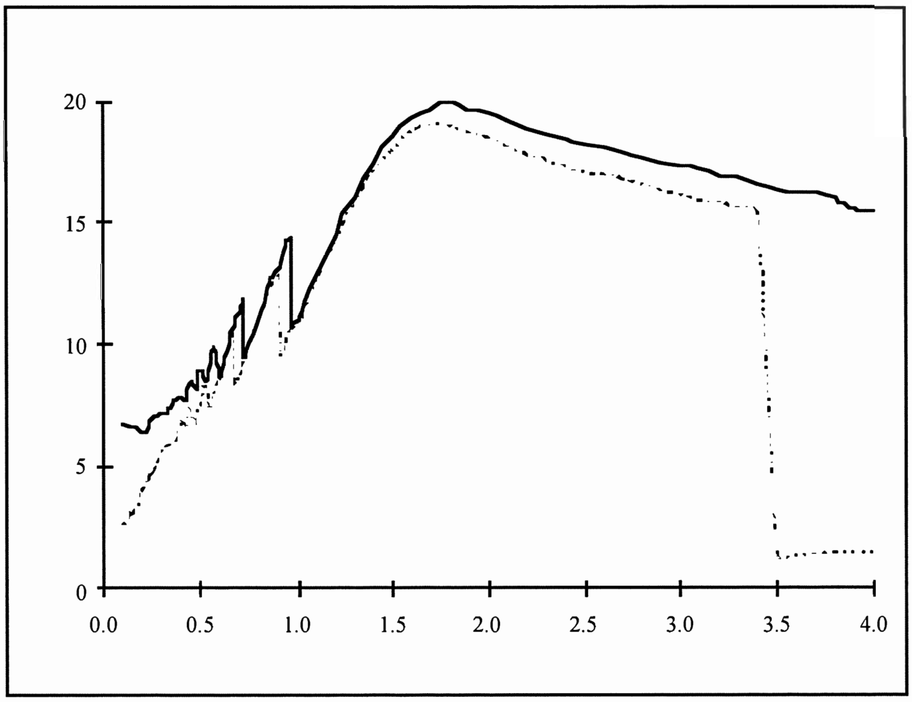


Figure 2(c): *The effect of form factor on the optimal length of a one-tonne monohull.*

Returning to the "realistic" boats produced for lower speeds, with Fnv<3.25, as the "design speed" increases from zero in that range, the optimum length L shown in Figure 2(c) increases to a maximum of about 22 at a Fnv value of about 1.8 before decreasing slowly as the speed increases further. This volumetric Froude number of 1.8 corresponds to a conventional (actual length-based) Froude number of 0.38, or a speed of 11 knots for a one-tonne vessel. At speeds below this value, the usual very dramatic large rise in wave resistance occurs as the length-based Froude number increases. Not surprisingly, longer boats are then preferred as the speed rises.

This trend cannot continue for ever. Eventually, the optimal boatlength reaches a maximum, and further increases in speed can no longer be met by increasing length to keep operating well below the wave resistance main peak. Instead, the length-based Froude number passes (quite rapidly) through the value where wave resistance is maximal, but the proportion of wave resistance is nevertheless kept sufficiently low to achieve an optimal design because of the large boatlength. Eventually as the speed increases further, the optimal boatlength starts to decrease again, since we are now operating at a length-based Froude number above the main wave resistance peak. Then the wave resistance decreases with Froude number, and hence shorter boats have less rather than more wave resistance at any given speed, and are preferred in the optimisation.

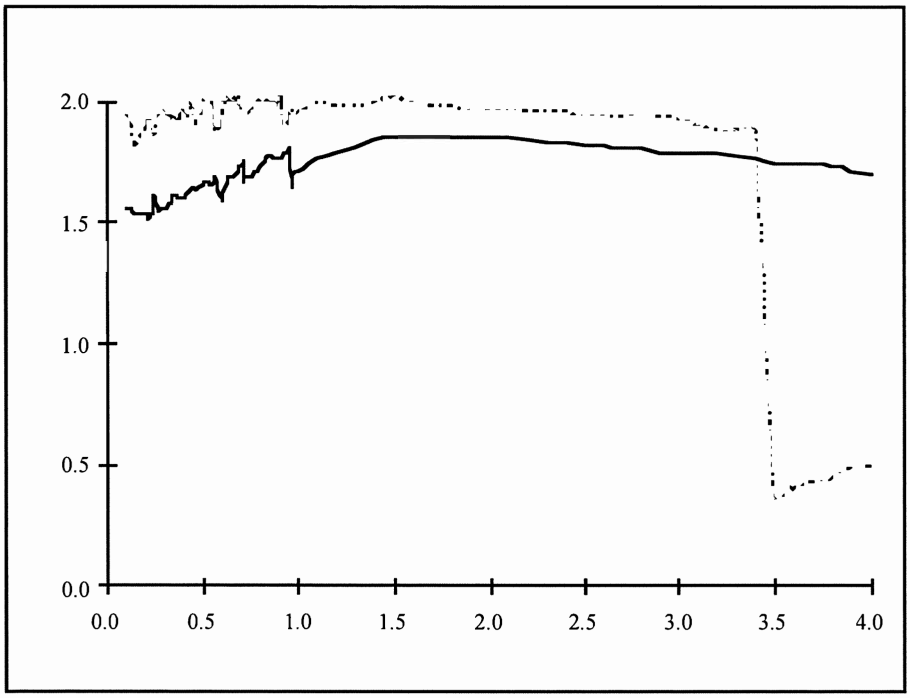


Figure 2(d): *The effect of form factor on the optimal beam-to-draft ratio of a one-tonne monohull.*

When the length is so great, the surface area strongly controls the optimisation, and to minimise the increase in frictional resistance, semi-circular sections tend to be preferred. This is clear in Figure 2(d), where it can be seen that the beam-to-draft ratio B/T stays at a value of roughly 2 for Fnv between 1.0 and 2.5.

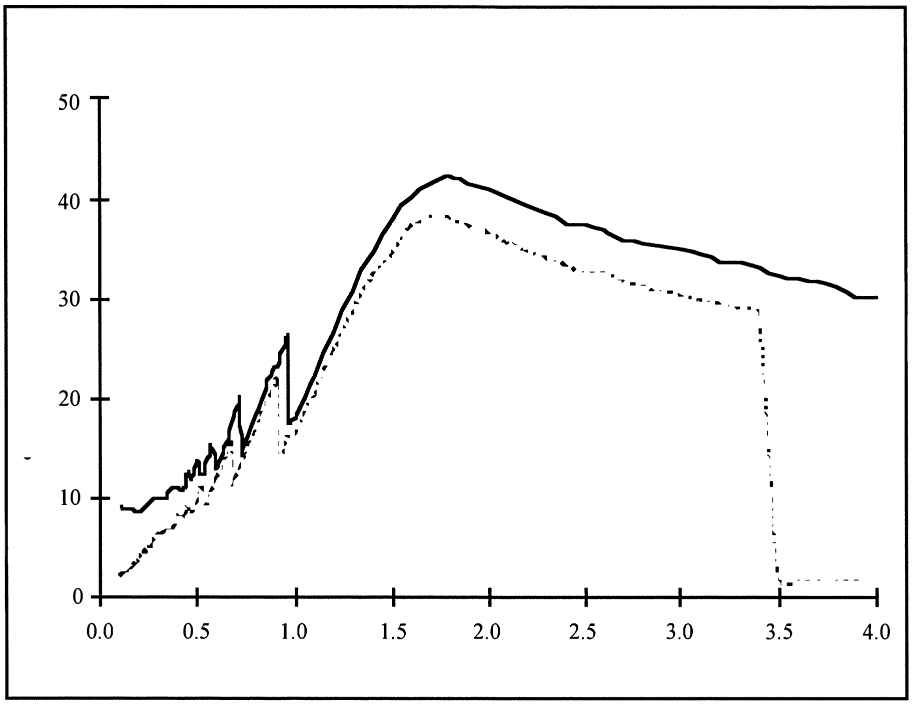


Figure 2(e): *The effect of form factor on the optimal length-to-beam ratio of a one-tonne monohull.*

The optimum boats are very slender. Figure 2(e) shows the length-to-beam ratio, which is very high indeed (reaching a maximum of about 42 at Fnv=1.8) by conventional ship standards, though not entirely unreasonable for rowing shells.

4.3 Monohull with form drag

Figures 2(a) - 2(e) also show (solid curves) the same monohull calculations as in the previous section for a one-tonne boat, but here the total resistance now includes Scragg and Nelson's (1993) form factor.

Figure 2(a) indicates that there is only a quite small increase in the residual total drag Ct for all speeds, consistent with the fact that the form drag is small, especially for the present very fine hulls. The greatest impact of form effects on the optimisation process occurs at very low and very high Fnv. The solid-line Ct curves of Figure 2(a) are smoother at low Fnv than the dashed curves, and the ultimate decrease in Ct at high Fnv is no longer as rapid.

Figure 2(b) shows that with form drag included, the proportion of wave resistance now remains below 10% for all speeds and all displacements. The scatter at low Fnv is not so pronounced as in the optimisations without form effects. Most important of all, however, is that there is no longer a sudden discontinuous increase in the proportion of wave resistance for Fnv>3.25. We have already anticipated this, since the very short boats that were suggested at high speeds by the optimisation without form drag are now heavily penalised by their large entrance and exit angles, and fail in total drag competition with a local minimum corresponding to a longer boat.

Figure 2(c) confirms this point, indicating that the optimum boat stays "long" for all speeds, with no discontinuity at any high-end speed. Indeed, with the inclusion of form effects, there is a tendency towards slightly longer optimum boats. The beam-to-draft ratios shown in Figure 2(d) are generally about 10% smaller with form drag included. For our canoe body, small entrance and exit angles can only be achieved by reducing the beam, so there is a slight tendency toward non-circular cross-sections, with B/T<2.

At the intermediate speeds which are of the greatest practical interest, there is only a small effect of the form factor on all outputs, and the qualitative discussion in the previous section about transition through the speeds where the wave resistance and hence the optimum length is maximal applies equally with or without form factor. Nevertheless, because as we have seen, inclusion of a form factor makes for a smoother and more realistic optimisation process at all speeds, such a factor is included in all of the remaining computations presented here.

4.4 Variation in Displacement

To demonstrate the effects of displacement, we optimised boats with D=0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0 tonnes. The 0.075-tonne boat is of similar displacement to the lightweight women's single scull class; the one-tonne boat would correspond to a coxed eight with a very burly crew. For each of the fourteen displacements, optima were sought at eight volumetric Froude numbers, Fnv = 1.4, 1.6, ..., 2.8.

One commonly-used method to rate rowing performance is the U.S. gold medal standard time over 2000m. For example, standard time over 2,000m. for the lightweight women's single sculls (D=0.075 tonnes) is 7 min. 42 sec., for the open men's coxless fours (D=0.4 tonnes) standard time is 5 min. 55 sec., and for the open men's coxed eights (D=0.9 tonnes) it is 5 min. 29 sec., (Gwadz, M., 1996). Using these standard times to calculate average boatspeeds, the corresponding volumetric Froude numbers of the three classes are approximately equal to 2.0, 2.1 and 2.0, respectively. Thus the range of volumetric Froude numbers we are considering here extends above and below the standard times for each class of rowing, and the central value Fnv=2 seems representative of good speeds for a wide range of displacements.

Since we are not constraining length or draft, the following results also apply to canoes and kayaks. Of course, the narrow hulls that result from the optimisation process have low static stability and are probably less suitable for paddling styles. It is important, however, to know at what cost in total resistance the necessary extra stability can be achieved.

To reduce computation time, 33 waterlines, 33 stations, and 240 intervals of theta were used in calculations. Length was constrained to lie in the range $0.5\text{m} \leq L \leq 30.0\text{m}$; draft was limited to the range $0.01\text{m} \leq T \leq 2.0\text{m}$. For the optimisations, we used a small population size of 32. Each of the 112 design problems (14 displacements, 8 speeds) was run with three different initial populations, and a minimum of 5,000 resistance evaluations were performed during each run. Optimisations were performed on 14 Sun workstations (some IPX models, some Sparc4s, some Sparc 10s) over four nights.

In all cases convergence was quite fast due to the small number of design parameters, and because the wave resistance varies smoothly with length-based Froude number greater than about 0.35. In hindsight, the number of evaluations performed seems quite excessive for this set of problems. In nearly all cases, the eventual optimum was found (to at least the 4th decimal place in Ct) during the first run. Subsequent runs made little difference to either the optimum total resistance, or the parameters producing that optimum resistance. Of course, if we had used smaller search domains, for example small regions surrounding the dimensions of existing shells, we could have used hill-climbing operators alone. However, we are looking for unusual hullforms, ones that we as humans might not normally conceive of. If the search domain is too small, we could miss, for example, the low-speed and high-speed discontinuities discussed previously.

Figures 3(a) and 3(b) show the variation with displacement of the total resistance coefficient and length respectively. Results for the one-tonne boats are the same as in figures 2(a) and 2(c). Note that the proportionality constant relating actual speed to volumetric Froude number varies as the one-sixth power of displacement. Specifically, the actual speed at the central Froude number Fnv=2.0 is 7.9 knots for a 0.075-tonne vessel, 10.8 knots for a 0.5-tonne vessel, and 12.2 knots for a 1.0-tonne vessel. Actual speeds at other Froude numbers are obtained by scaling these central speeds.

We do not present graphs of Cw/Ct, beam-to-draft ratio, or length-to-beam ratio to save space, and also because they are very similar to those of the one-tonne results in figures 2(b), 2(d) and 2(e).

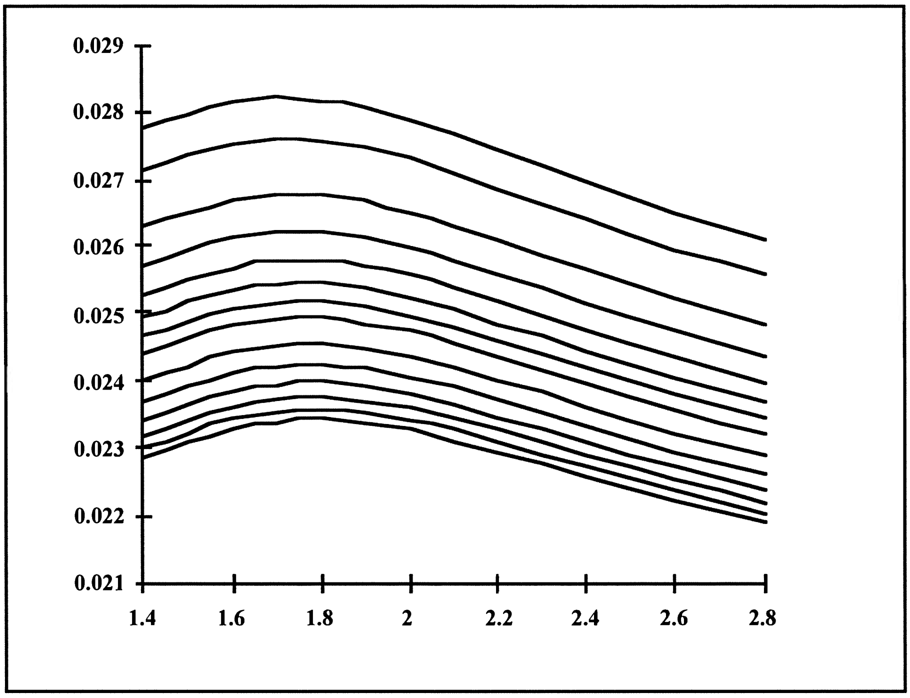


Figure 3(a): *The effect of displacement on the optimal total resistance of a monohull.*

In figure 3(a) the curves at the top of the graph are for 0.075-tonne boats; the curves at the bottom are for the one-tonne shells.

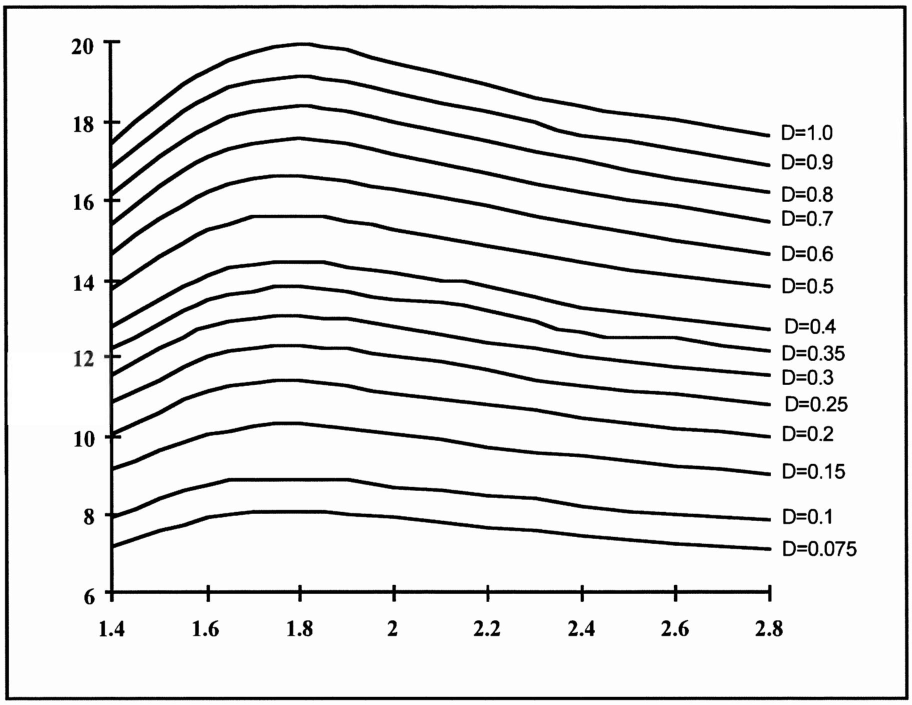


Figure 3(b): *The effect of displacement on the optimal length of a monohull.*

Smaller boats have larger drag coefficient Ct because they are shorter, their Reynolds numbers are smaller, and consequently the skin friction coefficient is larger. Of course the actual total drag Rt is much larger for larger boats, once we multiply Ct by (L*)^2=D^(2/3).

In this range of speeds, the dependence of the results on displacement is quite smooth and predictable by interpolation within the curves presented here.

4.5 Comparison with Existing Eight-Oared Shells

In their study, Scragg and Nelson (1993) compared predictions of total resistance for three existing competitive eight-oared shells – the Vespoli A, Vespoli B, and Janocek models. At a speed of 11.7 knots and a nominal displacement of D=0.871 tonnes, they estimated that the three shells each had a total resistance coefficient of approximately Ct=0.0248. Using figures 3(a) and 3(b) above as simple design charts, we see that at Fnv=2.0, our optimum shell for D=0.871 has Ct approximately equal to 0.0235. This represents a 5% improvement over the hulls in Scragg and Nelson's study. Figure 3(b) shows that the optimum length is about 18.5m. The beam of this optimal shell is 0.45 metres, which is somewhat lower than conventional.

To provide a more realistic comparison, we repeated the optimisation of a 0.871-tonne boat at a speed of 11.7 knots, but now with the beam constrained to B=0.57m., roughly the same as that of the shells considered by Scragg and Nelson. GODZILLA found that this sub-optimum hull had Ct=0.0243, which still represents a 2% improvement over the Ct estimated by Scragg and Nelson for the extant hulls. Our 0.57m-beam shell has L=16.9m. which is about 10% shorter than the above optimum length of 18.5m. On the other hand, this length compares well with the three existing shells: the Janocek hull has L=16.3m; the Vespoli A has L=17.0m; the Vespoli B has L=16.6m. However, the draft for our optimum shell, T=0.216m., is 15.0% greater than the extant hulls which have drafts of about T=0.187m.

Scragg and Nelson's experiments with the Vespoli B and other hulls showed considerable scatter around 11-12 knots and Ct's between 0.0234 and 0.0247 were reported. The scatter was so considerable in fact, that they concluded (page 98):

> "...the empirical approach is less reliable in discerning small differences in performance than the systematic results obtained from numerical hydrodynamics."

The present authors take great heart from this!

5. CONCLUSION

We have found optimum boats for minimum total drag over a large range of speeds and displacements. Results were obtained both with and without form drag corrections. Although the net contribution of form drag is small, it can nevertheless be important in determining the optimum. The optimum boats tend to be longer and have a lower wave resistance relative to viscous resistance than conventional boats. The genetic algorithm tool GODZILLA has proved useful in searching for the global minimum in the presence of two or more local minima, and will be essential in extended work involving multihulled vessels, shape variations and other constraints.

ACKNOWLEDGEMENTS

This work was supported by the Australian Research Council.

REFERENCES

[1] Chapman R.B. "Hydrodynamic design of semisubmerged ships", J. Basic Eng. ASME, pp 874-884, (1972).

[2] Dhingra, A.K., and Lee, B.H., "A Genetic Algorithm Approach to Single and Multiobjective Structural Optimization with Discrete Continuous Variables", Int. J. Num. Meth. Eng., Vol. 37, pp. 4059-4080, (1994).

[3] Eshelman, Larry J., "The CHC Adaptive Search Algorithm: How to Have Safe Search When Engaging in Nontraditional Genetic Recombination ", Foundations of Genetic Algorithms, Rawlins, G. J. E. (ed.), San Mateo, CA, Morgan Kaufmann, pp. 265-283, (1991).

[4] Gerritsma, J., Onnink, R. and Versluis, A., "Geometry, Resistance and Stability of the Delft Yacht Hull Series", Int. Shipb. Prog. Vol 28, pp. 276-297, (1981).

[5] Gwadz, M., Personal communication, (March 1996).

[6] Insel, M. and Molland, A.F., "An Investigation into the Resistance Components of High Speed Displacement Catamarans"", RINA , pp. 1-20 (1991).

[7] Proceedings of the 8th ITTC, Madrid, Spain, published by Canal de Experiencias Hidrodinamicas, El Pardo, Madrid, (1957).

[8] Lazauskas L. "User's Guide for GODZILLA", The University of Adelaide, Applied Mathematics Report L9601, (1996), in preparation.

[9] Lin W.-C., "The force and moment on a twin hull ship in steady potential flow", 10th O.N.R. Symp. Naval Hydro., MIT, pp.493-516, (1974).

[10] Michell J.H., "The wave resistance of a ship", Phil. Mag. (5) vol 45, pp. 106-123, (1898).

[11] Millward, A., "The Effect of Hull Separation and Restricted Water Depth on Catamaran Resistance", RINA,, pp. 341-349 (1992).

[12] Reklaitis, G. V., Ravindran, A. and Ragsdell, K. M., "Engineering Optimization: Methods and Applications", John Wiley and Sons Inc., U.S.A., pp. 684, (1983).

[13] Salvesen N., von Kerczek C., Scragg C.A., Cressy C.P. and Meinhold M.J., "Hydrodynamic design of SWATH ships", Trans SNAME vol 93, pp. 325-348, (1985).

[14] Scragg, Carl A. and Nelson, Bruce D. "The Design of an Eight-Oared Rowing Shell", Marine Technology, Vol. 30, No. 2, pp. 84-99, (April 1993).

[15] Tuck, E. O., "Wave Resistance of Thin Ships and Catamarans", The University of Adelaide, Applied Mathematics Report T8701, (Jan. 1987).

[16] Tuck, E. O., "The Wave Resistance Formula of J.H. Michell (1898) and its Significance to Recent Research in Ship Hydrodynamics", Journal of the Australian Mathematical Society, Series B, Vol. 30, pp. 365-377, (1989).

[17] Voigt, H.-M., Muhlenbein, H. and Cvetkovic, D, "Fuzzy Recombination for the Breeder Genetic Algorithm", 6th Int. Conf. on Genetic Algorithms, pp. 104-111, Morgan Kaufmann, (1995).

MODELLING THE ROWING STROKE AND INCREASING ITS EFFICIENCY

Maurice N. Brearley[1] and Neville J. de Mestre[2]

Abstract

The forces acting during a complete stroke in a racing shell are modelled mathematically. The model is then used to derive familiar aspects of boat behaviour, such as the variation in speed during a stroke. The analysis is conducted for a typical eight, but would apply equally well to fours and pairs, and to single, double or quad sculls.

A method is suggested for increasing the efficiency of the stroke and hence improving race times. It involves storing as elastic potential energy, some of the kinetic energy of the rowers' bodies as they near the end of the recovery phase, instead of requiring the rowers' legs to do all the work needed to annihilate their motion. During the early part of the power stroke the stored energy is used to help in accelerating the rowers' bodies.

1. INTRODUCTION

An investigation is made of the effects of the forces which operate during the rowing of racing shells. The analysis will apply equally well to eights, fours, pairs and double or quad sculls, and even (with obvious verbal changes) to single sculls.

The rowing stroke is divided into two parts: the power stroke, during which the blades of the oars are in the water and the rowers pull on the oar handles and straighten their legs, thus moving their bodies towards the bow on their sliding seats; and the recovery phase, during which the blades are clear of the water and the rowers move stern-wards by bending their legs and leaning forward.

The main resistance to the forward motion of a boat is provided by the drag of the water. Air resistance plays a much smaller part, in general, and is neglected in the analysis. A formula for the drag on the hull of a typical racing eight is obtained from experimental data in a Report of the U.K National Physical Laboratory (Wellicome [1]), and this enables numerical results to be obtained for a typical eight. Boat-flexing, pitching and "fish-tailing" are all neglected, their influence being negligible compared with that of the forces considered in the analysis.

Assumptions are made in later Sections about the variation of the forces on the oar handles and of the rowers' displacements on their slides. To our knowledge only one other author (Millward [2]) has endeavoured to construct a mathematical model of

---

[1]85 Dandarriga Drive, Clifton Springs Vic 3222

[2]School of Information Technology, Bond University, Gold Coast Qld 4229

rowing. His model ignored any movement of the rowers within the boat. The test of any model is how well its predictions correspond to observed results in a practical situation; in Section 5 such things as boat velocity and race duration are predicted by our model for an eight in a 2000 m race. Graphs are obtained of the variation in boat velocity while accelerating from a stationary start and during a full stroke after top speed has been reached.

In Section 6 an idea for improving the performance of a rowing crew will be examined. It involves converting to elastic potential energy some of the kinetic energy possessed by the rowers as they move on their slides towards the catch position. This is achieved by tethering the seats by shock cords (elastic ropes) to anchor points in the boat.

2. NOTATION AND CONVENTIONS

The water on which the boat travels may, of course, be regarded as a fixed reference frame.

A modern racing oar has a mass of only about 2 kg, so the masses of the oars may reasonably be neglected in the analysis. Let

m = mass of boat (including the cox, if present),

M = combined mass of rowers,

t = time from start of power stroke,

v = velocity of boat at time t,

$f = dv/dt$ = acceleration of boat at time t,

τ_1 = duration of power stroke,

τ_2 = duration of recovery phase,

$t' = t - \tau_1$ = time from start of recovery phase,

D = drag of the water on the hull.

In the Appendix it will be shown that for a typical racing-eight hull,

$$D = a + bv + cv^2, \tag{1}$$

where a, b, c, are constants which are calculable from data in Wellicome [1]. It will be assumed that this formula remains valid over the whole range of boat velocities considered in this analysis of rowing.

It is important to distinguish the directions of the forces which operate. The word "forward" will be taken to mean the direction in which the boat is moving, and "backward" to mean the opposite direction.

The word "rowers" will be used instead of the traditional (and sex-discriminatory!) word "oarsmen".

3. THE POWER STROKE

This occurs during the time interval $0 \le t \le \tau_1$.

The feet of the rowers are strapped to footrests fixed to the hull. During the power stroke the rowers straighten their legs and exert a combined backward force, Q say, on the footrests and hence on the boat. By Newton's Third Law, an equal and opposite force is exerted on the rowers by the footrests. These forces are depicted in Figure 1 for a single oar, but the forces shown are the combined values for all rowers in the boat.

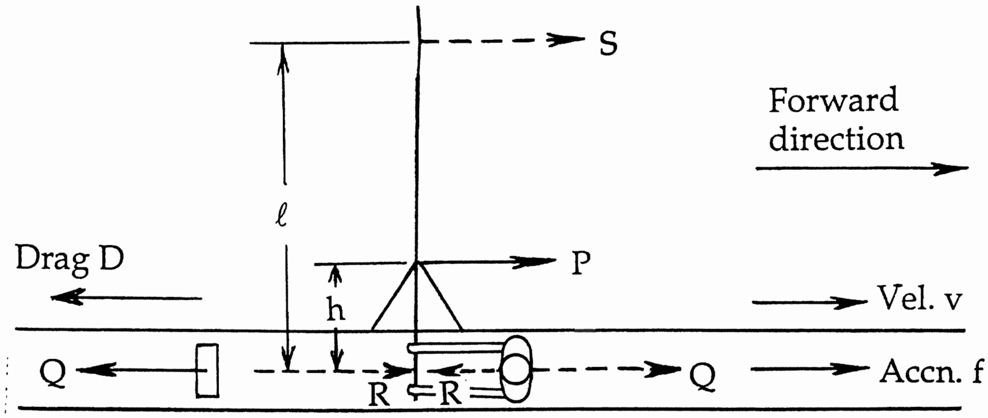


Figure 1. *Plan view of the power-stroke situation.*

In Figure 1, forces acting on the boat itself are drawn with full arrows; others are shown dotted. Since force components perpendicular to the direction of travel will cancel for each pair of oars on opposite sides of the boat, it is sufficient to consider only components parallel to the direction of travel.

The rowers exert a combined forward force R on the oars, and experience themselves an equal and opposite backward force. Not being forces on the boat, both of these forces are shown dotted in Figure 1.

The water exerts forces on the blades of the oars, the combined component in the forward direction being denoted by S. It is shown dotted in the figure since it does not act on the boat. It may be considered as acting at the centre of each blade. It has been observed by us that the blades move very little through the water during the power stroke; in this model they are regarded as fixed fulcrums of the levers formed by the oars. A case can be made for allowing for some small motion of the blades through the water by taking the fulcrum of each oar to be inboard of the blade itself. Such a change would affect some of the numerical work in Section 5 but would not alter the basic form of the mathematical model.

The oars exert forces on the boat at the swivels, their combined components in the forward direction being denoted by P in Figure 1. The mechanical advantage of the oar-lever system ensures that P > Q. It is the combined difference P – Q for all oars that drives the boat forward against the drag D of the water during each power stroke.

As depicted in Figure 1, let

ℓ = oar length from centre of grip to centre of blade,

h = distance from centre of grip to swivel.

Because the mass of the oars is being neglected, taking moments about the centre of the blade gives

$$R\ell = P(\ell - h).$$

It follows that

$$P - R = (h/\ell)P, \tag{2}$$

and this relationship evidently holds throughout the power stroke, irrespective of the angle which the oars make with the boat.

Oar flexing is ignored in this approach, even though it may modify slightly the values of P and ℓ. Its effect will be over-ridden by the necessary assumption of a particular maximum value for P which will be made when a numerical example is considered in Section 5.

Relative to the boat, the rowers begin and end their forward motion with zero velocity, and attain smoothly a maximum velocity at about the centre of their travel. This relative motion is such that it may reasonably be taken as half of a cycle of simple harmonic motion (SHM). This suggests writing the relative forward displacement of the rowers from the central point of their travel during $0 \le t \le \tau_1$ as

$$x_1 = -a_1 \cos n_1 t, \tag{3}$$

where the forward direction of the boat is taken as positive and

a_1 = the amplitude averaged over all rowers of the SHM of the centres of mass of the rowers' bodies,

$n_1 = \pi/\tau_1$ = the circular frequency of the SHM.

The forward acceleration of the rowers relative to the boat is $\ddot{x}_1$, and relative to the water is $\ddot{x}_1 + f$. The equation of motion of the rowers in the forward direction is therefore

$$Q - R = M(\ddot{x}_1 + f) = M(n_1^2 a_1 \cos n_1 t + dv/dt).$$

The equation of motion of the boat is

$$P - Q - D = m\ dv/dt.$$

Adding these two equations and using (2) produces

$$(m + M)\ dv/dt = (h/\ell)P - M n_1^2 a_1 \cos n_1 t - D.$$

The force P begins and ends with small magnitudes in $0 \le t \le \tau_1$, and attains smoothly a maximum near the centre of this interval (Mason et al [3]). The salient features of P will be adequately represented in a mathematically tractable way by taking

$$(h/\ell)P = P_m \sin n_1 t, \tag{4}$$

where $n_1 = \pi/\tau_1$ as before, and P_m is the maximum value of $(h/\ell)P$.

The previous differential equation then becomes, by virtue of (1),

$$(m+M)\,dv/dt = P_m \sin n_1 t - M n_1^2 a_1 \cos n_1 t - a - bv - cv^2.$$

This power stroke equation may also be written as

$$dv/dt = K_1 \sin n_1 t + K_2 \cos n_1 t + A + Bv + Cv^2, \tag{5}$$

where $0 \le t \le \tau_1$, $n_1 = \pi/\tau_1$, and

$$K_1 = P_m/(m+M), \; K_2 = -Mn_1^2 a_1/(m+M), \tag{6a,b}$$

$$A = -a/(m+M), \; B = -b/(m+M), \; C = -c/(m+M). \tag{7a,b,c}$$

4. THE RECOVERY PHASE

This occurs during the time interval $\tau_1 \le t \le \tau_1 + \tau_2$, or $0 \le t' \le \tau_2$ where $t' = t - \tau_1$.

During the recovery the rowers bend their legs to draw themselves on their sliding seats towards the stern of the boat. The combined force F on the rowers which produces their motion is shown dotted in Figure 2, and the equal and opposite force F on the footrests is shown as a full arrow because it is a force on the boat. Just as for Figure 1, it is enough to illustrate the situation for a single oar.

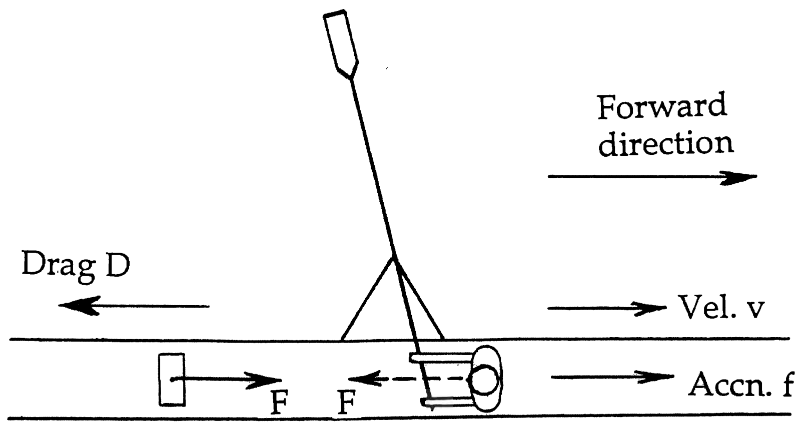


Figure 2. *Plan view of the recovery-phase situation.*

During the first half of the recovery phase the directions of the forces F are as shown in Figure 2, and the boat accelerates in the forward direction. For the second half of the recovery the forces F and the boat's acceleration reverse their directions, and this will shortly be seen analytically.

During the recovery phase the positions occupied by the rowers' bodies are very similar to those during the power stroke, but assumed of course in the opposite direction and over a different time span τ_2. The relative forward displacement of the rowers from the central point of their travel during $0 \le t' \le \tau_2$ may thus be taken as

$$x_2 = a_1 \cos n_2 t', \tag{8}$$

where the amplitude a_1 is the same as in equation (3), and

$$n_2 = \pi/\tau_2 = \text{the circular frequency of the SHM.}$$

The forward acceleration of the rowers relative to the boat is $\ddot{x}_2$, and relative to the water is $\ddot{x}_2 + f$. The forward equation of motion of the rowers is therefore

$$-F = M(\ddot{x}_2 + f) = M\left(-n_2^2 a_1 \cos n_2 t' + dv/dt'\right).$$

The equation of motion of the boat is

$$F - D = m\ dv/dt'.$$

On adding the last two equations and using (1) it is seen that

$$(m + M)\ dv/dt' = M\, n_2^2 a_1 \cos n_2 t' - a - bv - cv^2.$$

Dropping the dash from t' (for convenience only), this recovery phase equation may also be written as

$$dv/dt = K_3 \cos n_2 t + A + Bv + Cv^2, \tag{9}$$

where $0 \le t \le \tau_2$, $n_2 = \pi/\tau_2$, and

$$K_3 = Mn_2^2 a_1/(m + M), \tag{10}$$

and A, B, C are given by (7 a,b,c).

5. A PARTICULAR NUMERICAL EXAMPLE

To apply the foregoing theory to a particular case, a racing eight will be considered. The constants a, b, c in equation (1) are known for such a boat from work done in the Appendix. It is shown there that the drag on the hull (in newtons) is given by

$$D = 24.93 - 11.22v + 13.05v^2,$$

where the boat speed v is in m/s. The values of the constants in (1) in SI units are therefore

$$a = 24.93,\ \ b = -11.22,\ \ c = 13.05.$$

A video of an Australian Olympic eight in action over a 2000-metre course enabled estimates to be made of the power-stroke and recovery durations, and it was decided to take $\tau_1 = 0.7\text{s}$, $\tau_2 = 0.9\text{s}$. Of course in a race the durations would vary, but for purposes of calculation they are taken as constant. The time for a complete stroke is 1.6 s, which corresponds to a stroke rate of 37.5 per minute.

The amplitude a_1 of the motion of the centres of mass of the rowers' bodies is estimated to be 0.36 m.

The mass of the boat plus cox, and the combined masses of the eight rowers are taken to be respectively m = 146 kg, M = 680 kg.

The component in the forward direction of the maximum force exerted by each rower during the power stroke must be guessed, and is assumed to be 100 lb wt = 45.36 kg wt = 444.5 N which lies within the range obtained by Mason et al [3]. For all eight rowers combined, the maximum value of the force R in Figure 1 is then 3556 N. Equations (2) and (4) show that

$$P_m = \left[(\ell/h) - 1\right]^{-1} \times \max R,$$

where ℓ and h are the lengths depicted in Figure 1. If the estimates $\ell = 3.385$ m and h = 1.02 m are used, then

$$P_m = 0.4313 \times 3556 = 1534\ \text{N},$$

and this is the value that is used in (6a) to calculate the value of the constant K_1.

In the power-stroke differential equation

$$dv/dt = K_1 \sin n_1 t + K_2 \cos n_1 t + A + Bv + Cv^2, \tag{5}$$

the relevant domain is $0 \le t \le 0.7$, and $n_1 = \pi/0.7$ rad/s. The values of the other constants in (6 a,b) and (7 a,b,c,) are found to be (in S.I. units)

$$K_1 = 1.8577, \ K_2 = -5.9695,$$

$$A = -0.030182, \ B = 0.013584, \ C = -0.015799.$$

In the recovery phase equation

$$dv/dt = K_3 \cos n_2 t + A + Bv + Cv^2, \tag{9}$$

the domain is $0 \le t \le 0.9$, and $n_2 = \pi/0.9$ rad/s. The values of A, B, C are as listed above, and $K_3 = 3.6112$.

Equations (5) and (9) are not very amenable to analytical solution. A computer was therefore used to solve them numerically, using the Runge-Kutta method. To investigate the acceleration of the boat from a stationary start it was assumed that (5) and (9) applied from the outset, and the following iterative procedure was used.

An initial velocity of $v = v_{01} = 0$ was used, and (5) was solved to find the velocity $v = v_{11}$ at t = 0.7, the end of the first power-stroke. This value v_{11} was used as a starting value with the recovery equation (9), which was then solved numerically to yield the velocity $v = v_{21}$ at $t = 0.9$, the end of the first complete stroke. The whole procedure was then repeated, using in (5) the new initial velocity $v_{02} = v_{21}$ and arriving at a new value for v_{22} at the end of the second complete stroke. The iteration was continued until the value v_{2n} was repeated to sufficient accuracy after successive strokes, showing that the boat had reached a "steady state".

The distance traveled by the boat during each stroke was calculated by numerical integration of the velocity, from which the mean boat velocity $\bar{v}$ during each stroke was found by dividing the distance by the stroke duration of 1.6 seconds. Because the velocity varies greatly during a stroke, $\bar{v}$ is a more suitable quantity to plot as a function of time than the instantaneous velocity. Figure 3 was formed by drawing a smooth curve through the points $(t, \bar{v})$, where t is the time from the start to the middle of the stroke to which the $\bar{v}$ value refers.

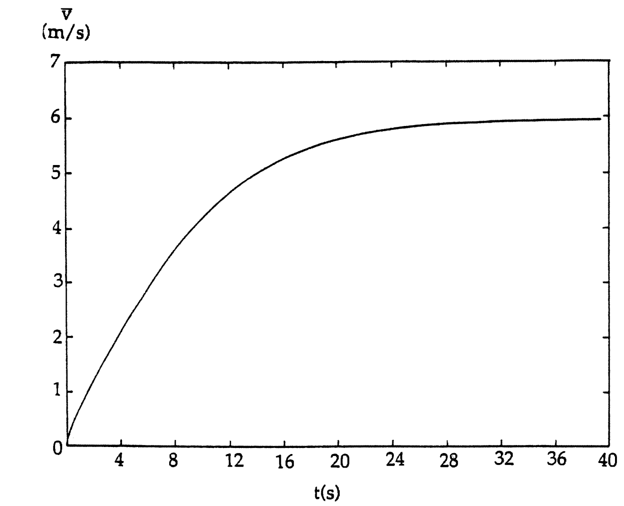


Figure 3. *Mean boat velocity $\bar{v}$ versus time t during acceleration.*

Figure 3 shows that the boat reaches a constant mean speed after about 40 seconds, which corresponds to 25 complete strokes.

The computer solution was also used to find the boat velocity at 0.1 second intervals throughout a complete stroke after the "steady state" had been achieved. Figure 4 was formed by drawing a smooth curve through the resulting points (t, v).

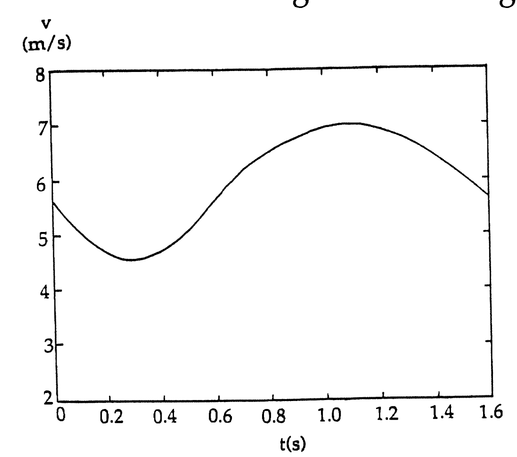


Figure 4. *Boat velocity v during one stroke in the "steady state" situation.*

Figure 4 shows that the boat speed drops to below 4.6 m/s near the middle of the power stroke, and that it reaches nearly 7 m/s near the middle of the recovery phase. This speed variation is mainly the result of the forces exerted by the rowers on their footrests; during the power stroke they are driving the boat back against its predominantly forward motion, and during the recovery phase they are dragging the boat forward and augmenting its velocity. From a dynamical viewpoint one would say that there is an exchange of momentum between the rowers and the boat, with the motion of their combined centre of mass being much more uniform than the motion of either component.

The distance traveled by the boat during one "steady state" stroke was also calculated and found to be 9.488 m, which corresponds to a mean velocity $\overline{v}$ of 5.93 m/s.

The time taken for the eight to row the 2000 m course can now be calculated. During the acceleration phase of 40 seconds duration, the distance travelled by the boat was found from the computer solution to be 200.7 m. The time taken to cover the remaining 1799.3 m at the mean "steady" speed of 5.93 m/s is 303 s, making the total race time 5 m 43 s. This would be a reasonable time for an Olympic eight, and a very good time for a club crew. It suggests that the estimate made for the maximum force exerted by each rower was a reasonable one.

The principles used in the foregoing example of a racing eight would apply equally well to fours and pairs, and even to double, quad and single sculls, provided the obvious modifications were made for the numbers of rowers and oars involved, and for the mass and drag of the hull.

6. An Energy Storing System

An idea for improving the performance of rowers is to have each seat connected to a fixed point nearer the bow of the boat by a length of shock cord (elastic rope). The length of the shock cord is to be such that it comes under tension when the seat is at the midpoint of its motion during the recovery stroke. If

$2d$ = total distance moved by a seat on its slides,

T = maximum tension in the shock cord,

the potential energy stored in the shock cord at the end of the recovery stroke is $\frac{1}{2}Td$ on the assumption that the cord obeys Hooke's Law.

For an eight, the total energy E so stored in all eight cords would be 4Td. Because the shock cords help to bring the rowers' bodies to relative rest in the boat, the rowers' legs are spared an amount E of the work they normally do during the second half of each recovery.

During the first half of the power stroke the tension in the shock cord helps each rowers' legs to straighten. Therefore in an eight the rowers' legs would again be spared an amount E of work. The total saving of work for the legs of all eight rowers in one stroke would be

$$2E = 8Td. \tag{11}$$

It is reasonable to expect that the work saved by such an arrangement of shock cords would be available to increase the effort expended by the rowers in driving the boat forward. On this assumption an estimate may now be made of the expected improvement in the performance of an eight in a race, using the data of the particular example discussed in Section 5.

Experiments were conducted with experienced rowers on a rowing ergometer, the seat being tethered by polyamide shock cord of 3mm diameter. It was found that four strands of this cord produced a tension of 10 kg wt (= 98 N) at the end of the recovery phase, and that this was the greatest load with which the rowers felt comfortable. Measurements showed that a typical value for the seat travel 2d was 0.68m. The total work saved per stroke is shown by (11) to be

$$2E = 8 \times 98 \times 0.34 = 267 \text{ Nm}.$$

The rowing stroke is not 100 per cent efficient for several reasons, including blade slip and the angle of inclination of the oars. It would be difficult to calculate the efficiency accurately, and for the purpose of determining the improvement in performance to be expected from the tethered seat system it will be sufficient to use an estimate. If the efficiency is taken to be 70 per cent, the work W_1 per stroke saved by using tethered seats would be

$$W_1 = 267 \times 0.7 = 187 \text{ Nm}.$$

The useful work per stroke, W_2 say, is that done against the water resistance D during the duration of the stroke. It is given by

$$W_2 = \int_0^{1.6} Dv\, dt \tag{12}$$

where in Section 5 it was found that

$$D = 24.93 - 11.22v + 13.05v^2 \qquad \text{(Newtons)}.$$

The boat velocity v throughout the stroke in the "steady state" situation is as shown in Figure 4.

The value of W_2 may be found from (12) by using Simpson's 1/3 Rule at time increments of 0.1 seconds. It is found that $W_2 = 4220$ Nm.

The ratio of W_1 to W_2 in the "steady state" is

$$W_1/W_2 = 187/4220 = 0.0443.$$

In Section 5 the duration of a 2000 metre race for the particular eight considered was found to be 5m 43s. The improvement in this time that would result from using tethered seats would be of the order of $343 \times 0.0443 = 15.2$s. At the mean boat speed of 5.93 m/s found in Section 5 this would correspond to an improvement of 90.1 m, or about 5 boat lengths.

7. CONCLUSIONS

A mathematical model was set up to represent the rowing stroke in a racing shell. An eight was used for the numerical work, but the principles involved apply also to fours and pairs and to single, double and quad sculls. The validity of the model is verified by its success in predicting quantitatively the familiar variation in boat speed during a stroke. The reason for this speed variation is revealed precisely by the model.

The idea was introduced of tethering each seat by shock cord to a fixed point of the hull behind the rower. The arrangement enabled energy to be stored in the cords, thus saving the legs of the rowers some work during both the recovery and power strokes. Calculations suggest that the system could improve the performance of an eight by as much as five boat lengths in a 2000m race.

Even allowing for the doubt about the value used for the efficiency of the rowing stroke, and for the assumption that any work saved will appear as useful work to propel the boat, it seems likely that a significant improvement in race performance could be expected from the tethered seat arrangement. A practical test under race conditions is obviously indicated.

8. ACKNOWLEDGEMENTS

The authors are grateful to racing shell builder Mr Jeff Sykes, of Jeff Sykes and Associates, Geelong, for information and advice; to Professor B.C. Rennie for advice and for drawing their attention to the NPL Technical Memorandum mentioned in the Reference below, and for providing a copy of it; and to British Maritime Technology Limited for permission to reproduce in the Appendix a graph from that Technical Memorandum.

REFERENCES

[1] J.F. Wellicome, Report on resistance experiments carried out on three racing shells, Nat. Phys. Lab., Ship T.M., 184, (1967).

[2] A. Millward, *A study of the forces exerted by an oarsman and the effect on boat speed*, J. Sports Sci., 5, 93-103, (1987).

[3] B.R. Mason, P Shakespear, P. Doherty, The use of biomechanical analysis in rowing to monitor the effect of training, Excel, 4 (4), 7-11, (1988).

APPENDIX

Wellicome [1] describes resistance measurements made on racing eight hulls in a water tank at the Ship Division of the National Physical Laboratory. The results of one set of experiments are shown in Figure 5, reproduced unchanged from the above Reference.

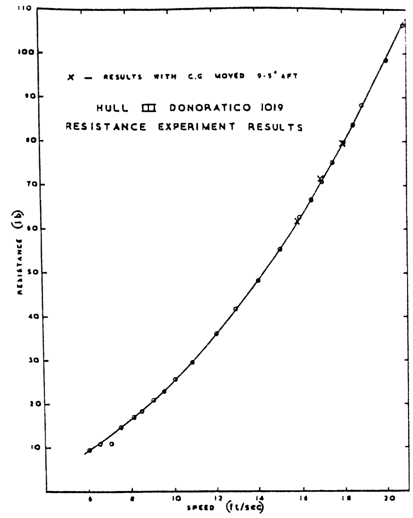


Figure 5. *Hull resistance in lb wt versus speed in ft/sec for a typical racing eight hull (reproduced from Wellicome (1967) by permission of British Maritime Technology Limited).*

The method of least squares was used to fit a quadratic velocity equation to the drag curve of Figure 5. In terms of the units used in that Figure the fitted equation is

$$D = 5.6027 - 0.7685\,V + 0.2725\,V^2 \quad \text{(lb wt)},$$

where V is the boat speed in ft/s. In S.I. units this is

$$D = 24.93 - 11.22\,v + 13.05\,v^2 \quad \text{(N)},$$

where v is in m/s. This is the origin of the equation stated near the beginning of Section 5.

AN ANALYSIS OF WORLD RECORDS FOR RACES RUN ENTIRELY IN LANES

R Hugh Morton[1]

Abstract

A compilation of world records for races run entirely in lanes has recently been published. These are analysed for the effects of track curvature, gender differences and the imposition of hurdles. There is of course strong evidence that males run faster than females and that hurdles slow runners down. However, there is little evidence that the less curved lanes can be traversed faster. Since the central lanes are not necessarily the fastest either, the requirement that the faster runners be seeded to the central lanes, seems unnecessary.

1. INTRODUCTION

The 200m, 400m and 400m hurdles (400H) for both men and women are run entirely in lanes. It is natural to ask what effect, if any, lane allocation may have on the times taken to run these distances. Common sense suggests that the innermost lanes are more affected by track curvature such that slower times would be expected; and vice versa. On the other hand, the innermost lanes enjoy the visual and psychological stimulus of seeing one's competitors, apparently in front due to the staggered start, thereby urging a faster time; and vice versa also.

Naturally one must also ask what gender differences exist, and what effect the imposition of hurdles may have. These effects seem intuitively predictable. Of more interest are interaction effects; such as whether the gender difference is greater or lesser for hurdle versus flat events (particularly knowing that womens hurdles are lower than mens). There is little published scientific information available on these questions, but the recent publication by Brickner [3] of a compilation of the best ever recorded times in each of the (usually eight) lanes in the six events mentioned, provides us with some high quality data to answer them.

2. STUDIES OF RUNNING ON CURVED TRACKS

Jain [5] noted that times for 200m races run on straight track (which used to be officially recognised years ago) are consistently lower than those run on curved tracks. The average difference is about 0.4 sec. By assuming that the difference in times taken to run a given distance on a semi-circular path and on a straight-path is inversely proportional to the radius of the curve, Jain shows that on a standard 200m track the advantage of the outer lane 8 over the inner lane 1 due to lesser curvature, is about 0.07 sec. At world class 200m speeds, this time converts to a distance

[1]Department of Statistics, Massey University, Palmerston North, New Zealand

advantage of a little under one metre. Jain recognises this as a approximation, and argues that for several reasons, it is probably an underestimation.

Alexandrov and Lucht [1] have considered the physics of sprinting. An equation of motion is derived assuming that (a) the runner has available a propulsive force F per unit mass, taken as constant for the duration of the race; (b) there exists a resistive force opposing the direction of motion, taken as proportional to the velocity of motion υ;(c) in order to keep running within the allocated lane, 100m along the arc of a semicircle of radius r, the propulsive force must supply a component giving rise to a centripetal acceleration of magnitude υ^2/r. Using world class 200m speeds on a standard track, Alexandrov and Lucht show that the advantage of lane 8 over lane 1 is about 0.12 sec or a distance of about 1.3 metres.

In a similar theoretical paper, Green [4] shows that the time differences obey an inverse square law and that Jain's estimates are indeed low. An advantage of about 0.123 sec is estimated. Likewise, Behncke [2] estimates an advantage of about 0.106 sec.

These latter estimates are in good agreement. For less capable athletes the time advantage may be larger, but at slower speeds, the distance advantage is probably unchanged. Since most such sprints are won or lost by very close margins, these effects seem quite significant.

3. DATA ON WORLD RECORDS FOR LANES

I have adopted a multiple regression approach to analyse these data on records times compiled by Brickner [2]. Time (T, sec) is taken as the dependent variable. As explanatory variables, both linear and quadratic effects of lane allocation (A) were investigated, and indicator variables utilised to account for distance (D), gender (G), and hurdle (H) effects. Because of collinearity, lane numbers 1 to 8 were scaled to have a mean of zero and range -1 to +1; L = (A - 4.5)/3.5. All first order interactions between these explanatory variables were included in a first fitting, except for DH which is confounded with H.

Using a backward elimination procedure not all these terms were significant, but retaining all terms significant at $p<0.05$ yielded a final fitted equation.

$$T = 21.71 + 26.75D + 4.52H - 1.87G - 2.64DG - 1.48HG + 0.37DL - 0.30HL - 0.18GL + 0.84HL^2$$

with $R^2 = 0.9997$ (adjusted $\bar{R}^2 = 0.9996$). The standard error of estimation was $s = \pm 0.27$sec.

4. DISCUSSION

The curvature effect, as predicted by the mathematicians and physicists where lane 8 should be fastest is noticeably absent! It holds, as a linear progression from lane 1 to lane 8, only for the mens 200m event. The time advantage estimated from these data is about 0.37 sec; three times the size of the 'theoretical' estimate. Clearly this needs further investigation. Curiously, the very opposite lane effect, where lane 1 is found to be the fastest, hold for both the mens and womens 400m flat events. The estimated advantage of lane 1 over lane 8 is about 0.37 sec for men and 0.75 sec for women.

One reason which may partially explain the absence of the expected curvature effect, is the visual psychological stimulus already mentioned. More likely perhaps, may be the seeding rules for lane allocation employed in championship meetings, where runners who have performed best most recently are allocated to the central lanes. In New Zealand for example, seeds 1-4 are randomly allocated lanes 3-6, and seeds 5-8 are randomly allocated to lanes 1, 2, 7 and 8. This practice however would lead us to expect faster times in the central lanes, and record times to show a U-shape when plotted against lane number. While there is an overall tendency towards this on average, it is statistically evident only in the 400H events for both men and women. Lanes 4 and 5 are the fastest and lanes 1 and 8 the slowest, with an estimated difference of about 0.82 sec for both men and women. While the proponents of lane allocation based on seeding may be quick to place importance on this latter result, it is worth noting that most of the 400H lane records for men have been established by the same individual, Edwin Moses.

The remaining race, the womens 200m shows no evidence of lane effects of any sort. The estimated time for any lane is the same, 21.77 sec. To this basic central lane time for women, we can add the following effects estimated from the record data: if extending the distance to 400m, add 26.75 sec; if imposing hurdles at 400m add a further 4.52 sec. For men over 200m, subtract 1.87 sec; over 400m then add 24.11 sec; while if imposing hurdles at 400m add a further 3.04 sec. The following table gives these fitted values:

Table 1: Fitted times for central lane world records

	200m	**400m**	**400H**
Women	21.71	48.46	52.98
Men	19.84	43.95	46.99

Interactions are clearly present here. For example, the gender difference at 400m (men faster by 4.51 sec, refer Table 1) is more than twice the effect at 200m (1.87 sec). Perhaps the higher physical strength of men applicable during the acceleration phase which is absent during the second half of a 400m race may be one explanation. Also, even though hurdles for women are lower in height than for men, their imposition at 400 metres penalises women both absolutely more (by 4.52 sec) than it does men (3.04 sec), and relatively more (8.53% versus 6.47%).

5. Conclusion

Apart from the obvious and expected effects of gender, distance and the imposition of hurdles, the world record data for the six lane affected races show no consistent pattern. Thus, despite the gross averages suggesting that lane allocation on the basis of seeding may produce faster times in the central lanes, when one takes into account gender, distance and hurdle effects, the evidence is at best rather weak. Also, if one takes each race in isolation, there is a danger of getting the wrong impression by losing sight of the effects of all information contained in the data. Whether athletic authorities should consider changing the conventional method of lane allocation is of course quite a different proposition, for it has tradition and anecdotal evidence on its side.

REFERENCES

[1] Alexandrov I. and Lucht P. Physics of sprinting, *Am. J. Phys.*, 49, 254-257 (1981).

[2] Behncke H. Small effects in running, *J. Appl. Biomech.*, 10, 270-290 (1994).

[3] Brickner J. World records for lanes, *Assoc. Track and Field Stat. Bull.*, 1/95, April, 7-11 (1995).

[4] Green P. R. Running on flat turns: Experiments, theory and applications, *J. Biomech. Eng.*, 107, 96-103 (1985).

[5] Jain P. C. On a discrepancy in track races, *S.N.I.P.E.S. J.*, 5, 38-43 (1982).

A FAIR METHOD FOR RESETTING THE TARGET IN INTERRUPTED ONE-DAY CRICKET MATCHES

Frank Duckworth[1]

Tony Lewis[2]

Abstract

One-day cricket has a history of injustices to one of the teams when, due to a sustained interruption in play, the target of the team batting second is reset. All methods that have been used so far fail to produce fair targets in all circumstances especially when interruptions occur after the start of the second innings. Methods that are currently in use, or that have been tried and abandoned, are described, together with their limitations.

This paper describes a method that produces fair targets in all known situations including single and multiple interruptions to either one or both innings. Unlike other methods it makes due allowance for the stage of the innings when the overs are lost and for the number of wickets that have fallen at the time. With this method an interruption to play does not award an advantage to either team.

We present a model for the average runs that are obtained from the two resources of overs and wickets in combination, and modify this to give the proportion of the run-scoring resources of an innings remaining at any stage of the innings. The values for the parameters of the model are obtained using relevant data from one-day internationals. We show how the relationship obtained enables the target score in an interrupted match to be reduced. This is done by using the proportion that is lost of the run-scoring resources of the innings, a correction which is equally fair to both sides.

Through the use of several examples we demonstrate how the method produces fair targets in practical and hypothetical situations in contrast to existing methods which are shown to yield unfair targets in many circumstances.

This method has been presented to the chief executives of member countries of the International Cricket Council. We believe that the method will be tested soon in one-day international series.

1. INTRODUCTION

One-day cricket began in the United Kingdom in the 1960s. It has been played in many formats with 40, 55, and 65, later reduced to 60, overs per side. The game has since spread around the world, so that now there are many one-day competitions

[1]Moonrakers, Taits Hill, Stinchcombe, Gloucestershire, GL11 6PS, UK.

[2]Faculty of Computer Studies and Mathematics, Frenchay Campus, University of the West of England, Bristol, BS16 1QY, UK.

played not only within cricket playing countries but also between them. Every four years, the major countries and some minor ones, compete in a World Cup.

These games are clearly intended to be completed within one calendar day. Sometimes the rules of the competition allow for a second or even a third day in which to complete the game if bad weather should interrupt play for any substantial length of time.

Some competitions, however, such as the English Sunday League (40 overs per innings) and some games in the World Cups, do not allow for extensions in the time allocated. If playing time has to be reduced, then either one or both teams receives fewer than the maximum number of overs allowed per team.

From the beginning of one-day cricket, a mechanism has been needed by which to adjust the target score of the team batting second so that a result can be achieved in one day when overs have been lost. The length of innings for one-day internationals (ODIs) has been standardised (at 50 overs per side, Wisden [1]) but the mechanism by which the target score is reset has not. Various methods have been tried, with differing philosophies and levels of complexity.

2. Review of Existing Methods

There follows a description of methods that have been used so far in one-day cricket matches. None of the methods takes account of the stage of the innings at which the overs have been lost nor is any allowance made for the number of wickets fallen. Both of these factors have an important bearing on setting a fair target.

Average run-rate (ARR)

This method is currently used in the English Sunday League. The winner is the team with the higher average run-rate per over for the number of overs each has separately received. Although it is a method with a simple calculation, its major problem is that it tends (but not always) to favour the team batting second. For example, Team 1 scored 250 runs in 50 overs, an average run-rate of 5.0 per over. If Team 2 receives only 25 overs then, at 5.0 per over, the target to beat is 125, that is 126 to win. With all ten wickets available in half the overs batsmen can take more risks and, consequently, should achieve a higher scoring rate. Therefore a target of 126 in 25 overs, half that of Team 1, is easier to achieve than the 251 in 50 overs.

Most productive overs (MPO)

This method was used in the 1992 World Cup competition in Australia. The target to beat is determined from the same number of the highest scoring overs of Team 1. For example, Team 1 scored 250 runs in 50 overs but, perhaps due to some very economical bowling at one stage of the innings, ten of these overs produced a total of only five runs. Consequently, the 40 most productive overs produced 245 runs.

If the innings of Team 2 is reduced by ten overs the winning target is 246 in 40 overs which is a tall order even before the innings has begun. This method tends to favour strongly the team batting first. It also produces the illogical situation that the economical bowling that saved runs is the reason that the target is reduced by only a small amount. In other words, when batting, Team 2 is being penalised because there was some good bowling when the team was fielding. A further disadvantage of this method is the book-work that needs to be undertaken to identify the most productive overs and their total runs.

We do not agree that the way in which Team 1 achieved its total should be a criterion in deciding the target of Team 2 in an interrupted game. Just as in an uninterrupted game, it is only the total that matters and not how it was achieved nor how many wickets were lost in the process. In the example, the winning target for Team 2 could be as low as 201 in 40 overs if exactly five runs per over were scored from every over. It could be as high as 251 in 40 overs if no runs were scored from ten overs. Team 1 was not constrained in the method by which the total of 250 was obtained. Therefore Team 2's revised target should not in any way be dictated by Team 1's scoring pattern.

Discounted most productive overs (DMPO)

This method is now used in the Benson & Hedges World Series Cup in Australia. The total from the most productive overs is discounted by 0.5% for each over lost. Thus, using the previous example, if ten overs were lost the discounted score to beat is $245\ (1 - 10 \times 0.005) = 232.75$. The score to win is 233. As with the MPO method there is the disadvantage of the additional book-work in identifying the most productive overs and their total runs.

This is still a difficult target for this particular game and it still suffers from the undesirable feature of using Team 1's scoring pattern. The range of possible winning targets is from 191 to 238, a spread of 47 runs.

The parabola method (PARAB)

This method is based upon an idea of a young South African, and tries to account for the deficiencies of the ARR method (do Rego[2]). He produced a table of 'norms' (y), (reproduced in Table 1) for every number of overs (x), from 25 to 50 rounded to the nearest integer. They were obtained using the parabola $y = 7.46x - 0.059x^2$ (sometimes called the 'parabola formula') as a model of the 'diminishing returns' nature of the average total score as overs in the innings increase. The method finds the winning target by increasing or decreasing the norm pro rata according to Team 1's performance relative to the norm for its innings. For example, in Table 1, the norms for 40 and 50 overs are respectively 204 and 226. If Team 1 scored 250 in 50 overs and Team 2 is to receive only 40 overs then the score to beat is $250 \times 204/226 = 225.7$, that is, 226 to win.

If Team 2 goes out to bat knowing that it is to receive only 40 overs then the target is reasonable. But, as with all the methods tried so far, the timing of the lost overs is not accounted for, nor is the number of wickets lost at the interruption. If the ten overs are lost late in the innings, or several wickets have fallen, then the effect can be extremely unjust to one team or the other.

Another major problem, from a mathematical modelling viewpoint, is that a parabola can not be regarded as an appropriate model since it has a turning point, here at around 63 overs, and so the estimated average total score would start to decline as overs available increased beyond 63. Further, the data do Rego used to estimate the parameters of the parabola are limited to the average total scores achieved in English county one-day games (40, 55 and 60 overs per side), and only one season of these. The model has been applied to international matches without any statistical justification.

The method was agreed, but not called upon, for the South Africa – England one-day tournament in Jan/Feb 1996, but it was invoked in the Singer Cup in Singapore, April 1996. (See the India/Pakistan match, later)

Table 1: Norms and percentage factors for the PARAB and WC96 methods

Overs					25	26	27	28	29	30
PARAB norm					154	158	163	167	171	175
WC96 % factor					66.7	68.4	70.2	72.4	74.2	76.0
Overs	31	32	33	34	35	36	37	38	39	40
PARAB norm	175	178	182	185	189	192	195	198	201	204
WC96 % factor	77.8	79.1	80.9	82.2	84.0	85.3	86.7	88.0	89.3	90.7
Overs	41	42	43	44	45	46	47	48	49	50
PARAB norm	207	209	212	214	216	218	220	222	224	226
WC96 % factor	92.0	92.9	94.2	95.1	96.0	96.9	97.8	98.7	99.6	100

World Cup 1996 method (WC96)

This method is an adaptation of the PARAB method. Each of the norms from 25 to 50 overs was converted into a percentage of the norm for 50 overs, which was taken as 225 rather than 226. Table 1 summarises these percentage factors alongside the corresponding norms.

For the example above, the score to beat would be $0.907 \times 250 = 226.75$, which is 227 to win in 40 overs. The major problems are, again, no allowance for the timing of the lost overs and no allowance at any stage for the number of wickets lost.

The Duckworth/Lewis Method (D/L)

The aims of this approach to the problem are to produce a method which:-

- is independent of Team 1's scoring pattern
- is fair to both sides – leaves the game as balanced or unbalanced as it was before the interruption
- is easy to understand and apply
- gives a fair target in all circumstances including past controversial situations.

Basis of the method

(1) The batting side has two resources at its disposal from which to make its total score. These are *overs* and *wickets.*

(2) The number of runs that may be scored from any position depends on both resources *in combination*. Clearly, a team with 20 overs to bat with all ten wickets in hand has a greater run-scoring potential than a team that has lost, say, eight wickets. The former team has more run-scoring resources at its disposal; that is, it has more of its innings left than the latter team although both have the same number of overs left to bat.

(3) The target score is corrected by the proportion of the run-scoring resources of the innings the second team has been deprived of by the interruption.

(4) Using historical one-day international data a relationship is obtained which yields a table giving the runs remaining to be scored as a proportion of the total for all combinations of overs and wickets as the innings progresses.

The model, stage 1

The first stage of the analysis is to derive a model that adequately describes the average total score Z(u) which is obtained in u overs. An exponential equation is used to model the diminishing-returns nature of the relationship.

$$Z(u) = Z_o[1 - \exp(-bu)] \tag{1}$$

where Z_o is the asymptotic average total score in unlimited overs under one-day rules, and b is the exponential decay constant.

Our initial research used entirely English county games over seasons 1986 to 1994 to estimate Z_o and b. By using such a wide spread of length of games, including many shortened games, useful estimates for these parameters were obtained. Using data collected mainly from Reference [1], estimates were obtained of b=0.0355 and Z_o= 265.6. The estimating process of b and Z_o was undertaken by finding the total least squares weighted deviations of actual average score from expected average score. Microsoft Excel Solver 5.0 was used to perform the calculations. Appendix I summarises the data that were used.

During our initial discussions with representatives of the International Cricket Council (ICC) and the English Test and County Cricket Board (TCCB) it was suggested that to gain international acceptance for use in future one-day internationals (ODIs) the database from which our estimates of the parameters are made should consist entirely of international games.

From Reference [1] and from the Cricinfo database on the Internet, data from over 250 international games from all parts of the world have enabled us to estimate the parameters as b = 0.0315 and Z_o = 283.69. Most of the innings were of the 50-over variety but some shortened games were included in our analysis as were some earlier 60 over innings. Some variation in the length of innings would enable estimates of the asymptote Z_o and the decay constant, b, to be obtained. Appendix 1 summarises the data which were used. Only the data from the average total scores of the team batting first were used to avoid problems of bias.

It will be noted that our two pairs of estimates differ a little. The asymptote Z_0 is higher and the decay constant b is lower for international games. This might be expected when taking into account the greater expertise of international players compared with that of English county players. We are satisfied, therefore, that our estimates for the parameters b and Z_0 are reasonably reliable even though their estimation, so far, is based on a fairly narrow spread of length of innings. As more data become available we will be able to refine our estimates of the parameters.

The model, stage 2

Stage 2 of the modelling process is to introduce a revision of equation (1) for when w wickets have already been lost but u overs are still left to be received. Clearly the asymptote will be lower and the decay constant will be higher and both will be functions of w. The revised model is of the form

$$Z(u,w) = Z_0(w)[1 - \exp\{-b(w)u\}] \qquad (2)$$

Much research has gone into finding sensible yet simple forms for $Z_0(w)$ and $b(w)$. A function $F(w)$ is defined which, in effect, represents the proportion of the asymptotic average total score obtained with w wickets already lost. Clearly $F(0) = 1$ and $F(w)$ is a monotonic decreasing step function.

The function $F(w)$ is used in both $Z_0(w)$ and $b(w)$ to define

$$Z_0(w) = Z_0\,F(w), \text{ and}$$
$$b(w) = b/F(w)$$

Thus, only the single function $F(w)$ is necessary to introduce the effect of wickets into both the average further score and the exponential decay factor.

The full form of $Z(u,w)$ is thus:

$$Z(u,w) = Z_0\,F(w)\,[1 - \exp\{-bu/F(w)\}] \qquad (3)$$

In order to estimate the values of $F(w)$, w = 1,..,9 it was necessary to use data on the score at various stages of the innings over a wide range of games. Reference [1] does not, in the main, provide this level of detail and yet it is usually recorded at televised matches by the television company's scorer.

Obtaining this data has proved extremely difficult despite some assistance from the ICC. At the time of writing insufficient data have been obtained to find sensible and reliable estimates. Consequently, for current purposes, the values of $F(w)$ being used are:

w	0	1	2	3	4	5	6	7	8	9
$F(w)$	1	0.88	0.76	0.64	0.52	0.40	0.30	0.21	0.13	0.06
$\Delta F(w)$	0.12	0.12	0.12	0.12	0.12	0.10	0.09	0.08	0.07	0.06

These values for $F(w)$ summarise the concept that, on average, each of the partnerships for the first five wickets accounts for 12% of the asymptotic average total score. The partnerships for the last five wickets would, on average, contribute respectively, 10%, 9%, 8%, 7% and 6% *with unlimited overs to bat.*

Figure 1 shows the nature of the family of curves described by equation (3)

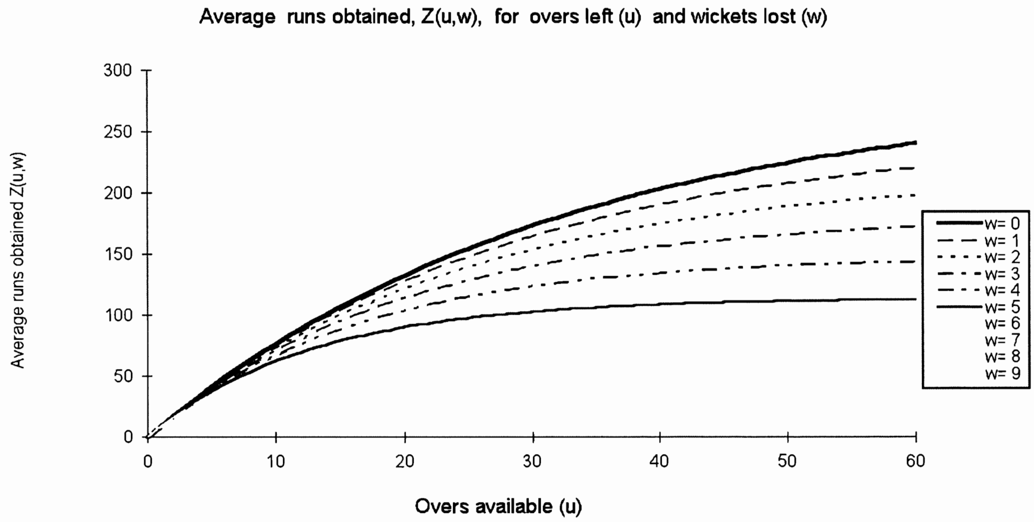


Figure 1: *Graph of the average runs obtained in u overs with w wickets lost.*

From equation (3) we obtain $\partial Z/\partial u = b\, Z_o \exp\{-bu/F(w)\}$.

The function $Z(u,w)$ has the properties that:

(i) $\lim Z(u,w)$ as $u \rightarrow \infty$ = $Z_o F(w)$, proportional to $F(w)$

(ii) $\lim \partial Z/\partial u$ as $u \rightarrow \infty$ = 0

(iii) $\partial Z/\partial u$ at $u = 0$ = $b\, Z_o$, independent of w
$\approx$ 1.5 runs per ball

These properties are consistent with what is expected in one-day cricket and consequently give confidence that our model is a sensible representation of the average further runs that may be scored.

Conversion to proportions

The function $Z(u,w)$ gives the average number of further runs that are scored when u overs are left and w wickets have been lost. Calculating the average number of runs lost by an interruption and adjusting the target score can be undertaken using this function. Recognising that there are some advantages in the simplicity of the WC96 method, we convert the average runs obtained, $Z(u,w)$, into proportions of the 50-over norm.

The D/L norm for 50 overs is Z(50,0) and so a proportion P(u,w) is defined by

$$P(u,w) = Z(u,w)/Z(50,0) \tag{4}$$

The function can be interpreted as representing the *proportion of the combined resources of the innings remaining when u overs are left and w wickets have been lost*. Appendix 2 tabulates this function as a percentage, for u = 50,...,0 and w = 0,...,9.

The advantage of simplicity for 50-over games, however, is at the expense of the complication of adjusting these proportions if the *first* innings lasts less than 50 overs. A multiplying factor is then necessary for each alternative innings length below 50 overs to rebase the starting point at 100%. If the method were to be used in the English Sunday League, a separate table for 40 overs would be produced.

Resetting the target

Team 1 has scored T runs in its allocation of 50 overs. The number of wickets that has been lost is irrelevant. Team 2 begins its reply but a stoppage occurs with u_1 overs left and w wickets lost. Play is resumed with u_2 overs left ($u_2 < u_1$) but still with w wickets lost.

Team 2 has been deprived of u_1-u_2 of its overs resource and so the target score to beat should be adjusted to account for this loss *at the stage of the innings it occurred*.

The proportion of the run-scoring resources of the innings lost in those $u_1 - u_2$ overs is $P(u_1,w) - P(u_2,w)$ and so Team 2 has had available to it the proportion of the run-scoring resources of its innings $1 - P(u_1,w) + P(u_2,w)$. The number of runs that Team 2 needs to beat is this proportion of Team 1's total score, that is, $T[1 - P(u_1,w) + P(u_2,w)]$. The winning target is the next highest integer.

Par score

Formula (4) can also be used to gauge the progress of Team 2 as it attempts to beat the total T of Team 1. The 'par' score is defined as the total that Team 2 should have acquired with u overs remaining and w wickets lost so as to be on course for a total T.

For a 50-over innings P(u,w) is the proportion of the combined run-scoring resources of the innings remaining with u overs left and w wickets fallen. The proportion of resources used at this point is 1 – P(u,w) and so the par score is T [1– P(u,w)].

3. APPLICATIONS OF THE DUCKWORTH/LEWIS METHOD

Some hypothetical and actual examples are provided below that illustrate how the Duckworth/Lewis method produces sensible revised targets under all known circumstances. They also show how targets set by other methods, although sometimes producing realistic targets, more often than not give targets which are not sensible, being either too easy or too difficult.

Hypothetical examples

For ease of understanding of the application of the Duckworth/Lewis method we shall use the same total score for Team 1 and the same number of overs lost, although the method will, of course, apply to any total score and to any number of overs lost. Throughout all these hypothetical examples we shall assume that Team 1 has completed its 50-over innings and scored 250 runs, and that interruptions to Team 2's innings are of 20 overs in length. The interruptions occur at different stages of Team 2's innings and with different numbers of wickets lost. Table 2 summarises all of the situations and the calculations to obtain the D/L revised target score to win.

Example I covers the situation where the interruption occurs before Team 2 commences its innings and so it is known *in advance* that only 30 overs will be received. In this situation ARR falls down, yielding the relatively easy target of 151. With all ten wickets available, but fewer overs to bat, a more challenging target is fair. The Duckworth/Lewis method, which sets 193 runs to win in 30 overs, gives a target as challenging as the original 251 to win in the full 50-overs. Note that in this circumstance, the PARAB and WC96 methods, at 190 & 191 respectively, produce comparable targets to the D/L method. The MPO target at 201 is slightly harder whereas the DMPO target is slightly easier. All the methods, except ARR, produce a target reasonably in keeping with the situation of knowing in advance that the second innings has been shortened.

Example II shows the effect of an interruption part way through Team 2's innings. Team 2 has made a solid start and, at 75 for no wicket, is in a strong position from which to accelerate and score the remaining 176 runs to win in 30 overs. This target is less demanding than in Example I reflecting the fact that Team 2 is in a strong position at the interruption and being well ahead of the D/L par of 57.5. It has used only 22.9% of its run-scoring resources even though 40% of the overs have been bowled.

The D/L target reflects the strength of Team 2's position by providing a revised target of 143 which is a further 68 runs in ten overs. This should normally be achievable with all ten wickets in hand. Of the other methods, only ARR at 151 can be regarded as reasonably fair in this circumstance..

The PARAB, WC96, MPO and DMPO methods yield targets in the range 181 to 201. These are grossly unfair to Team 2, which then requires between 106 and 126 further runs to win in the ten overs. The balance of the game is badly upset because all these methods do not take account of the stage of the innings at which the overs are lost.

Table 2. Calculations of revised target score in hypothetical examples

Hypothetical example no.	I	II	III	IV	V	VI
Team 2 score, chasing 250 (=T) in 50 overs	0	75	120	75	191	180
Wickets lost, w	0	0	0	2	9	4
Overs left at the interruption, u_1	50	30	20	30	20	20
Overs left at the resumption, u_2	30	10	0	10	0	0
Propn. of innings left at interruption $P(u_1,w)$	1	.771	.589	.682	.076	.461
Propn. of innings left at resumption $P(u_2,w)$	.771	.341	0	.325	0	0
Propn. lost in (u_1-u_2) overs=$P(u_1,w) - P(u_2,w)$	.229	.430	.589	.357	.076	.461
Propn. available $1 - [P(u_1,w) - P(u_2,w)]$	.771	.570	.411	.643	.924	.539
Revised score to beat:						
$T\,[1 - P(u_1,w) + P(u_2,w)]$	192.8	142.5	102.8	160.8	231.0	134.8
Revised target score						
D/L	**193**	**143**	**103**	**161**	**232**	**135**
ARR	151	151	151	151	151	151
PARAB	190	190	190	190	190	190
WC96	191	191	191	191	191	191
MPO*	201	201	201	201	201	201
DMPO*	181	181	181	181	181	181
Par score (D/L method)						
$T\,[1 - P(u_1,w)]$	0	57.3	102.8	79.5	231.0	134.8

* The targets by the MPO and DMPO methods cannot be evaluated properly without the actual score cards to find the total of the 30 most productive overs. To obtain some comparative figures we have assumed that the 20 *least* productive overs yielded 50 runs, which is half the average run-rate. Therefore, the 30 most productive overs yielded 200 runs.

Example III takes the situation of the loss of the 20 overs at the end of the innings. That is, Team 2's innings has been prematurely terminated and the game abandoned. Team 2, at 120 for no wicket, is in a very strong position. Although 60% of the overs have been bowled, only 41.1% of the run-scoring resources of its innings have been used. The D/L par for this point in its innings is 102.8 with no wickets lost, and so, at 120, it can be regarded as being well on course for a score of over 250. The team requires only a further 131 runs to win in 20 overs which it would expect to achieve comfortably having all ten wickets still in hand.

A decision on the winner is required. Only our method, the D/L method, would justly declare Team 2 the winner. The other methods would require Team 2 to have scored between 151 and 201 by the end of the 30th over to be declared the winner. All these other methods would clearly produce unfair decisions on the winner in this circumstance. They all take no account of the stage of the innings that the overs have been lost.

The D/L method, with its concept of par score, enables a judgement to be made, at any stage of Team 2's innings, on which team is winning. If Team 2's score is ahead of par it is winning, if it is below par Team 2 is losing. In an abandoned match, the D/L method awards victory to the team that is winning when play is stopped.

Example IV introduces the effect of wickets into the assessment of the target. Compared with the situation in Example II, Team 2 has lost two wickets in scoring the 75 runs and so it is in a weaker position. Our par score is 79.5 and so Team 2 is losing at this point of its innings. The revised target should reflect this. The D/L method does so by requiring a further 86 runs to win in the ten overs with eight wickets left. The method maintains the balance of the game at the stage of the interruption.

The ARR method requires only 76 more runs to win, the same as in Example II. This is rather easier than Team 2 deserves. It has not taken account of the loss of wickets at the stage of the interruption. The other methods still require unfair tasks of scoring between 106 and 126 further runs to win.

All except the D/L method have provided targets which upset the balance of the game as it was at the interruption.

Example V is perhaps a rather extreme case, but it does further emphasise the importance of considering the number of wickets that have been lost. The scenario might be that Team 2 has been batting extremely well at over six runs per over but then collapses to be nearly all out. A further 60 runs are needed to win in 20 overs but, with the last man at the crease, the odds are strongly in favour of Team 1 winning. When the remaining overs are lost, however, and the game is abandoned a decision on the winner is required.

Only the D/L method would fairly make Team 1 the winner; Team 2 being well behind the par of 231 for that stage of its innings. All the other methods would, unfairly, make Team 2 the winner. The weakness of Team 2's position would not have been taken into account.

Example VI takes a not-dissimilar situation to Example V, but now Team 2 has lost only four wickets and has scored 180, 11 runs fewer. Team 2 needs only 71 runs to win in 20 overs and, with six wickets in hand, it is in a very strong position and would be expected to go on to win the game. The D/L requirement, the par, is 135 after 30 overs with four wickets lost. Having already scored 180 runs, our method would fairly make Team 2 the winner.

ARR would also be fair in this circumstance in making Team 2 the winner. The PARAB, WC96, MPO and DMPO methods, however, would unfairly make *Team 1* the winner! At that stage of the innings, the lost overs represent a deprivation of a substantial proportion of the combined run-scoring resources still available in Team 2's innings.

All these hypothetical examples, in various ways, emphasise strongly that, when resetting the target score, there is a need to consider both the stage of the innings when the overs are lost and the number of wickets that have fallen at that point. The Duckworth/Lewis method is the only one of these methods that yields sensible targets in all these situations.

All the other methods yield sensible targets in only a few of these situations. It will be noted from Table 2 that all methods except the D/L method set the same target score for all the examples. There is no variation with the stage of the innings that the overs are lost or how many wickets have fallen.

Actual examples

The examples considered so far have taken hypothetical situations. We now include several applications of the D/L method to actual international games, some of which would have produced results different from those which actually occurred.

Because no game in the 1996 World Cup needed a revised target score, we have taken most of our examples from the 1992 World Cup in Australia. Several of these games were affected by rain, some leading to very controversial situations. In this competition the MPO method was used whereby the same number of highest scoring overs of the team batting first was used to set the target for the second team. Table 3 summarises the situations for three of the games.

In the RSA/ENG game England made a very positive start at 63 for no wicket, and, according to the D/L par score of 28.3, were well on course to achieve the 237 runs needed to win. They were then deprived of nine overs by the weather but only 11 runs came off the target by the MPO method. This was regarded as very unfair, forcing England into a desperate scramble they didn't deserve given the strong position they were in. Although they did make the 226 and win the game the D/L target of 207 to win would have been easier to achieve and much fairer to England.

The other methods produce either a too-easy target of 194 by ARR (unfair to South Africa) or more difficult targets from 215 to 226 (unfair to England). The D/L target is the only one which maintains the balance of the game as it was at the beginning of the interruption.

In the RSA/PAK game, Pakistan got off to a reasonable start at 74 for 2 wickets in 21 overs chasing a below average score. They were ahead of the D/L par of 69.6 for that stage of their innings. The loss of 14 overs in the middle of the innings impinged very badly on their prospects. Using MPO only 19 runs were deducted from their target. The required total of 193 (a further 119 in 15 overs) proved beyond them; they scored only 173 and lost. The D/L method would have given the easier and much-fairer target of 164 (a further 90 in 15 overs) and they would have won. ARR would produce a too-easy target (78 in 15 overs) which would have been unfair to South Africa. The other methods would have produced more difficult targets, unfair to Pakistan.

Table 3. Calculations of the revised target score in actual matches, World Cup 1992

Match (Team 1/Team 2)	**RSA/ENG**	**RSA/PAK**	**ENG/RSA**
Team 1 score(T)	236	211	252
Overs in the innings	50	50	45
Team 2 score	63	74	231
Wickets lost, w	0	2	6
Overs left at the interruption, u_1	38	29	2.1
Overs left at the resumption, u_2	29	15	0.1
Propn. of innings at interruption $P(u_1,w)$	.880	.670	.072*
Propn. of innings at resumption $P(u_2,w)$	.755	.444	0*
Propn. lost in (u_1-u_2) overs $P(u_1,w) - P(u_2,w)$	.125	.226	.075+
Propn. available $1 - [P(u_1,w) - P(u_2,w)]$	.875	.774	.925
Revised score to beat: $T [1 - P(u_1,w) + P(u_2,w)]$	206.5	163.3	233.1
Revised target score			
D/L method	**207**	**164**	**234**
ARR	194	152	241
PARAB	217	180	248+
WC96	218	180	248+
MPO(actual method in use)	**226**	**193**	**252**
DMPO(current Aust. method)	215	179	249
Par score (D/L method) $T [1 - P(u_1,w)]$	28.3	69.6	231.7

where RSA=Republic of South Africa, PAK = Pakistan and ENG = England.

* The proportion for the last two complete overs has been used as an approximation for the proportion of the resources of the innings lost.

+The first innings lasted only 45 overs. The proportion of the run-scoring resources of the innings lost has been rebased so that P(45,0) = 1.00; similarly for the PARAB and WC96 targets.

The **ENG/RSA** game, a semi-final of the World Cup, exposed the problem of the MPO method and led subsequently to more flexible playing conditions when games are interrupted. South Africa's innings in reply was interrupted with 2.1 overs (13 balls) remaining and resumed with 0.1 overs (1 ball) left! Much debate has since centred on why the full 13 balls couldn't have been bowled; there was certainly sufficient time. The umpires applied the rules, however, and the MPO method took England's 43 most productive overs which totalled 251 and so South Africa's target was reduced from 253 to 252 to win. From a position of requiring 22 runs from 13 balls, where they had a fighting chance, they were placed into the impossible position of scoring 21 runs from the one ball! The MPO method had failed. The D/L method would have given a revised target of 234, which is three runs in one ball. South Africa's fighting chance would have been retained.

To obtain this target it has been assumed that fractional overs are handled in one of several ways. These include taking the nearest complete overs, or the score at the end of the previous over, or by linear interpolation. It is quite possible also to extend Appendix 2 to handle every single ball. The ICC has indicated that, whilst recognising the need for a rule for assessing the target when stoppages occur part way through an over, a table on a ball-by-ball basis is not desirable.

For this example we have taken, from Appendix 2, the proportions for the nearest complete overs (that is 2 overs and 0 overs left respectively with six wickets fallen). Using complete overs, the difference in proportions, from Appendix 2 with six wickets lost, is 7.2% but based upon a 50-over innings. For a 45-over innings all these proportions need to be rebased. This is done by dividing all proportions by $Z(45,0) = 0.955$. Therefore, the lost proportion of the innings for South Africa is 7.5%.

4. INTERRUPTIONS TO TEAM 1'S INNINGS

Sometimes it happens that Team 1's innings is interrupted and either prematurely terminated or resumed later to complete a shorter innings. Either circumstance can be unfair to Team 1 and, unless an appropriate correction is made in the calculation of Team 2's target, some injustices can occur. The Duckworth/Lewis method can provide a fair target for Team 2 in both of these circumstances.

India v Pakistan, Singer Cup, Singapore, April 1996 - Premature termination of the first innings

India had scored 226 for 8 wickets in 47.1 out of 50 overs when rain interrupted play. Their innings was terminated and Pakistan were given a revised target of 186 in 33 overs based on the PARAB method. Pakistan won with overs to spare. The unfairness in this target is that India were unexpectedly deprived of 2.5 overs right at the end of their innings whereas Pakistan knew in advance that only 33 overs would be received. The D/L method would provide a fairer target in the following way.

India were deprived of 2.5 overs which represents 8.1% of their innings resources (using linear interpolation on the proportion of the innings for the 48th over with eight wickets lost). Thus, India's 226 was a score obtained from 91.9% of their resources and so a reasonable estimate of their final score would be $226/0.919 = 245.9$. Pakistan's reduced target should therefore be based on 245.9 in 50 overs. With 33 overs to bat the revised target score would be $0.815 \times 245.9 = 200.7$, which is 201 to win and a much fairer target for Pakistan to chase..

England v New Zealand, World Series Cup, Perth, Australia, 1983 - Resumption of the first innings

England, needing to beat New Zealand to qualify for the final stages of the competition, had scored 45 runs for 3 wickets in 17.3 of an expected 50 overs when a heavy rain storm and a long delay led to the deduction of 27 overs from each innings. England thus resumed their innings for a further 5.3 overs and scrambled 43 more runs to reach a score of 88 in the 23 overs.

New Zealand's target in 23 overs was 89 using the ARR method. New Zealand won the game easily. It was clearly an unfair target because of the unexpected and drastic reduction in the number of overs England were expecting to receive, whereas New Zealand knew from the start of their innings that they were to receive only 23 overs and could bat accordingly.

To apply the D/L method to this situation it is necessary to assess what proportion of the run-scoring resources of their innings England lost because of the interruption. Using linear interpolation in Appendix 2 for three wickets lost:-

Propn. left with 32.3 overs to play	0.644
Propn. left with 5.3 overs to play	0.191
Propn. lost in 27 overs	0.453
Propn. of innings available to England	0.547

The score of 88 thus represents 54.7% of their projected total and so their final total score in 50 overs is projected to be 88/0.547 = 160.9 runs. New Zealand, in 23 out of 50 overs, would be required to score 65.0% of this, a target of 104.5, which is 105 to win. While this is still not a very demanding target, nevertheless it gives England compensation for not knowing that the interruption would occur and yet rewards New Zealand for playing England into a fairly weak position at the interruption. The D/L target would have been fair to both teams.

5. MULTIPLE INTERRUPTIONS

The D/L method enables the effect of any stoppage to be assessed including multiple stoppages to either innings. The cumulative proportion of the innings that has been lost is recalculated after each stoppage. The calculations of the target score for interruptions to either the first or the second innings are carried out just as for single interruptions.

6. ACCEPTANCE OF THE METHOD

The method has been presented to the chief executives of the full-member countries of the ICC at their meeting in London in July 1996. There is every indication that member countries of the ICC will conduct tests of the method leading, hopefully, to its ultimate adoption internationally.

7. CONCLUSION

In this paper we have explained the mechanisms of existing methods used for resetting target scores in interrupted one-day cricket matches. Each of these methods yields a fair target in some situations. None has proved satisfactory in deriving a fair target under all circumstances.

We have presented our proposed method of resetting the target score, the Duckworth/Lewis method. Through the examples given, both hypothetical and real, we have shown that our method gives sensible and fair targets in all these situations. They include the circumstances where overs are lost at the start of the innings, part way through, or at the end of the innings when the game is abandoned and a winner has to be decided. The examples have shown the importance of taking into account the wickets that have been lost at the time of the interruption and the stage of the innings at which the overs were lost.

Our method handles situations where there are interruptions, not only to the second innings but also to the first innings.

The basis of the Duckworth/Lewis method is a table of proportions derived from an exponential relationship for the average number of runs which are scored from the remaining resources of overs and wickets in combination.

The method is easy to understand and simple to apply, requiring nothing more than a single- page table of percentages and a pocket calculator.

The Duckworth Lewis method has been presented to the cricketing authorities who are showing interest in using it for resetting target scores in interrupted one-day cricket matches.

REFERENCES

[1] Wisden, The Cricketers Almanack, (Wisden, Guildford, UK, 1987-96)

[2] do Rego, Wayne's System, *Wisden Cricket Monthly*, (November 1995)

COPYRIGHT

APPENDIX 1

Data on English county one-day matches and one-day internationals used to estimate b and Z_o

English County one-day matches, 1987-94

b	0.03347
Zo	265.2

Competition	Overs	Count	Act Mean	Expt Score
Short	10	4	92.3	75.4
Sunday	19	3	126.3	124.8
League	20	5	143.2	129.4
matches	23	4	123.3	142.4
	25	5	136.6	150.3
	28	3	167.0	161.3
	30	3	158.5	168.0
	31	5	171.6	171.2
	34	4	161.8	180.2
	35	4	191.0	183.0
	36	12	181.4	185.7
	37	8	194.9	188.3
	38	13	192.8	190.8
S League	40	212	197.8	195.6
B & H	55	373	219.0	223.1
NatWest	60	253	234.2	229.6
Total		911		

One-day internationals, 1987-96

b	0.03150
Zo	283.69

Overs	Count	Act Mean	Expt Score
60	42	238.5	238.5
50	217	226.5	225.3
47	2	259.0	220.3
30	2	189.5	178.8
25	2	111.5	160.8
Total	265		

Appendix 2: Table of percentages of resources of the innings remaining

Percentage of innings remaining

Overs gone	Overs left	Wickets lost 0	1	2	3	4	5	6	7	8	9
0	50	100	92.4	83.8	73.8	62.4	49.5	37.6	26.5	16.4	7.6
1	49	99.2	91.8	83.3	73.5	62.2	49.4	37.6	26.5	16.4	7.6
2	48	98.3	91.1	82.7	73.1	62.0	49.3	37.6	26.5	16.4	7.6
3	47	97.4	90.3	82.2	72.7	61.8	49.2	37.6	26.5	16.4	7.6
4	46	96.5	89.6	81.6	72.3	61.5	49.1	37.5	26.5	16.4	7.6
5	45	95.5	88.8	81.0	71.9	61.3	49.0	37.5	26.4	16.4	7.6
6	44	94.6	88.0	80.4	71.5	61.0	48.9	37.5	26.4	16.4	7.6
7	43	93.6	87.2	79.7	71.0	60.7	48.7	37.4	26.4	16.4	7.6
8	42	92.5	86.3	79.0	70.5	60.4	48.6	37.4	26.4	16.4	7.6
9	41	91.4	85.4	78.3	70.0	60.1	48.4	37.3	26.4	16.4	7.6
10	40	90.3	84.5	77.6	69.4	59.8	48.3	37.3	26.4	16.4	7.6
11	39	89.2	83.5	76.8	68.9	59.4	48.1	37.2	26.4	16.4	7.6
12	38	88.0	82.5	76.0	68.3	59.0	47.9	37.1	26.4	16.4	7.6
13	37	86.8	81.5	75.2	67.6	58.6	47.7	37.1	26.4	16.4	7.6
14	36	85.5	80.4	74.3	67.0	58.2	47.5	37.0	26.4	16.4	7.6
15	35	84.2	79.3	73.4	66.3	57.7	47.2	36.9	26.3	16.4	7.6
16	34	82.9	78.1	72.4	65.6	57.2	47.0	36.8	26.3	16.4	7.6
17	33	81.5	76.9	71.4	64.8	56.7	46.7	36.6	26.3	16.4	7.6
18	32	80.1	75.7	70.4	64.0	56.1	46.4	36.5	26.3	16.4	7.6
19	31	78.6	74.4	69.3	63.2	55.5	46.0	36.4	26.2	16.4	7.6
20	30	77.1	73.1	68.2	62.3	54.9	45.7	36.2	26.2	16.4	7.6
21	29	75.5	71.7	67.0	61.3	54.3	45.3	36.0	26.1	16.4	7.6
22	28	73.9	70.2	65.8	60.4	53.5	44.9	35.8	26.1	16.4	7.6
23	27	72.2	68.8	64.5	59.3	52.8	44.4	35.6	26.0	16.4	7.6
24	26	70.5	67.2	63.2	58.3	52.0	43.9	35.4	25.9	16.4	7.6
25	25	68.7	65.6	61.8	57.1	51.2	43.4	35.1	25.9	16.4	7.6
26	24	66.9	64.0	60.4	55.9	50.3	42.8	34.8	25.8	16.3	7.6
27	23	65.0	62.3	58.9	54.7	49.3	42.2	34.4	25.6	16.3	7.6
28	22	63.0	60.5	57.3	53.4	48.3	41.5	34.1	25.5	16.3	7.6
29	21	61.0	58.6	55.7	52.0	47.2	40.8	33.7	25.3	16.3	7.6
30	20	58.9	56.7	54.0	50.6	46.1	40.0	33.2	25.2	16.3	7.6
31	19	56.8	54.8	52.2	49.0	44.8	39.1	32.7	24.9	16.2	7.6
32	18	54.6	52.7	50.4	47.4	43.5	38.2	32.1	24.7	16.2	7.6
33	17	52.3	50.6	48.5	45.8	42.2	37.2	31.5	24.4	16.1	7.6
34	16	49.9	48.4	46.5	44.0	40.7	36.1	30.8	24.1	16.1	7.6
35	15	47.5	46.1	44.4	42.1	39.1	35.0	30.0	23.7	16.0	7.6
36	14	45.0	43.7	42.2	40.2	37.5	33.7	29.1	23.2	15.8	7.6
37	13	42.4	41.3	39.9	38.1	35.7	32.3	28.2	22.7	15.7	7.6
38	12	39.7	38.8	37.6	36.0	33.9	30.8	27.1	22.1	15.5	7.6
39	11	36.9	36.1	35.1	33.7	31.9	29.2	25.9	21.4	15.3	7.5
40	10	34.1	33.4	32.5	31.4	29.8	27.5	24.6	20.6	14.9	7.5
41	9	31.1	30.6	29.8	28.9	27.6	25.6	23.1	19.6	14.5	7.5
42	8	28.1	27.6	27.0	26.3	25.2	23.6	21.5	18.5	14.0	7.5
43	7	25.0	24.6	24.1	23.5	22.7	21.4	19.7	17.2	13.4	7.4
44	6	21.7	21.4	21.1	20.6	20.0	19.0	17.7	15.7	12.6	7.2
45	5	18.4	18.2	17.9	17.6	17.1	16.4	15.5	14.0	11.5	7.0
46	4	14.9	14.8	14.6	14.4	14.1	13.6	13.0	11.9	10.2	6.6
47	3	11.4	11.3	11.2	11.1	10.9	10.6	10.2	9.6	8.5	6.0
48	2	7.7	7.7	7.6	7.6	7.5	7.4	7.2	6.9	6.3	4.9
49	1	3.9	3.9	3.9	3.9	3.9	3.8	3.8	3.7	3.5	3.1
50	0	0	0	0	0	0	0	0	0	0	0

IS CRICKET REALLY BY CHANCE?

Kuldeep Kumar[1]

Abstract

In this paper we have tested the hypothesis that run rate per over or runs made in each over in a one day cricket match is random. We have also done a time series analysis of the data. It is observed that in some cases it is possible to forecast the run rate per over.

1. INTRODUCTION

The origin of cricket dates back to the 13th Century, and the first set of rules was written in 1744. The Marylebone Cricket Club, which is the world governing body of sports, was formed in 1787. During England's colonial history, cricket was exposed to countries around the world. Australia, New Zealand, India, England, Pakistan, West Indies etc. are the primary countries that participate regularly in international matches. Some newcomers are South Africa, Sri Lanka, Kenya, Zimbabwe, Holland and UAE.

Cricket matches are divided into innings. A one day match, which is becoming more popular these days as compared to five day test matches, consists of one innings of 50 overs for each team. Each side has one round at bat, and the innings is usually not over until 10 batsmen are dismissed or 50 overs are finished. Each over consists of six balls. (For details about the rules etc. see Encyclopedia Britannica; 15th edition, Vol. 28 Macropaedia, pp 115 – 120.)

2. WHAT IS CHANCE?

Chance always plays a major role in each and every activity of our lives. The fact that the nervous system possesses a random structure and that the genetic transmissions are subject to chance mechanism could be of profound significance. Not only one's time of birth, sex, growth, development, decay, time and mode of death are subject to chance but one's entire life of diverse activities can be interpreted, technically speaking, as a sample path of a multi-dimensional stochastic process. It is a fact that the God of Probability, or the Goddess of Chance, may be ruling our lives in this universe.

One need not, therefore, wonder or speculate why the existing deterministic models used in sports and various other fields are being replaced by stochastic models. For instance, classical mechanics paves the way to statistical mechanics, communication

---

[1]School of Information Technology, Bond University, Gold Coast Qld 4229

theory to statistical communication theory, mathematical genetics to statistical genetics, differential equations to stochastic differential equations etc.

The deterministic models of currently fashionable bio mathematics must, sooner or later, pave the way to more realistic stochastic models, if one has to take into account the chance fluctuations in Nature. Put briefly, this is the era of probability and statistical modelling, which has made inroads to physics, engineering, electronics and communication, biomedical sciences, psychology, economics, industry, natural planning, defence, sports and a host of other human activities.

3. TIME SERIES ANALYSIS

A time series is a collection of observations generated sequentially through time. The special features of a time series are that the data are ordered with respect to time, and that successive observations are usually expected to be dependent. The order of an observation is denoted by a subscript t. Therefore, we denote by z_t the t^{th} observation of a time series. The preceding observation is denoted by $z_{t-1}$and the next observation as z_{t+1}.

Usually in time series analysis measurements or readings are made at predetermined and equally (or almost equally) spaced time intervals to generate hourly, daily, monthly, or quarterly data. However, in some cases data may not be collected corresponding to time as mentioned above. For example, the engineering data may be a series of consecutive yields from a batch chemical process (Jenkins and Watts, 1968) in which yield is measured for each batch.

The purposes of time series analysis are generally two fold: to understand or model the stochastic mechanism that gives rise to an observed series and to predict or forecast future values of a series based on the history of that series.

4. DATA IN THE STUDY

Data in this study has been collected by the author during the last Wills World Cup cricket matches (Feb – March, 1996) played in India/Pakistan/Sri Lanka. Data consists of runs made in each over sequentially in each inning for the following matches.

Match	Countries	Team	Venue	Date	Comments
M1	Aust. vs SL	Aust.	Lahore	17.3.96	Final
M2	Aust. vs SL	SL	Lahore	17.3.96	Final
M3	India vs SL	SL	Calcutta	13.3.96	SF
M4	India vs SL	India	Calcutta	13.3.96	SF
M5	Aust. vs WI	Aust.	Chandigarh	14.3.96	SF
M6	Aust. vs WI	WI	Chandigarh	14.3.96	SF
M7	Aust. vs NZ	NZ	Madras	11.3.96	QF
M8	Aust. vs NZ	Aust.	Madras	11.3.96	QF

SL: Sri Lanka, WI: West Indies, NZ: New Zealand, SF: Semi Final, QF: Quarter Final

In this paper we have considered runs made in each over sequentially throughout the whole inning of 50 overs (or less in case all the players are out). Since the data is collected sequentially over a period of time the runs made in each over or run rate could be considered as a time series. This is a discrete time series which is obtained by accumulating a variable (run) over a period of time (over). This is similar to time series of the yield from a batch process which is accumulated over the batch time. Similar examples are noted in quality control where the number of defectives are counted in each batch over a period of time.

In this paper we have tested the hypothesis that the run rate per over is random against the alternative that it follows some trend. Using the times series analysis we have also tried to forecast the run rate per over.

5. CLASSICAL TIME SERIES ANALYSIS OF THE DATA

The classical approach to time series analysis begins with the premise that a typical time series has the four components – trend, seasonal, cyclical, and irregular to grow or decrease fairly steadily over quite long periods of time, and this pattern is identified as trend. The trend can be described for the time series z_t as

$$z_t = \beta_o + \beta_1 t + \varepsilon_t$$

In case there is no trend then $z_t = \beta_0$ which implies there is no long run growth or decline in the time series over time.

Seasonal variations refer to variations of periodic nature. It is not limited to periodic variation associated with the seasons of the year. Usually in time series analysis the unit of time referred to in discussing seasonal variation is less than a year. Cyclical variation refers to those up-and-down fluctuations that are observable over an extended period of time. These wavelike fluctuations, called business cycles, are different from seasonal fluctuations in that they cover longer periods of time, have different causes and are less predictable. Irregular or random variation is considered to be due to a host of unpredictable influences and is not accounted for by trend, seasonal or cyclical factors. With cricket data the seasonal and cyclical factors can be ruled out and we can assume that there is a trend and irregular components in the data only.

In this section we have tried to check if there is any trend in the data of runs made in each over. We have fitted the model

$$z_t = \beta_0 + \beta_1 t + \varepsilon_t$$

In case β_1 is insignificant it will imply that there is no trend in data. The result for the various innings are given below in Table 1 for runs made in each over.

Table 1

Match	β_0	β_1	t-value	p-value	R-Square
M1	5.04	-0.0086	-0.3447	0.7317	0.002
M2	3.98	0.0485	0.3631	0.1794	0.039
M3	5.35	-0.0123	-0.4007	0.6904	0.003
M4	5.09	-0.0892	-2.4028	0.0221*	0.165
M5	1.21	0.1147	4.7543	1.79E-05**	0.320
M6	3.77	0.0104	0.3802	0.7054	0.003
M7	6.39	-0.0266	-0.8118	0.4208	0.013
M8	3.43	0.1056	2.9991	0.0043**	0.163

Out of 8 innings β_1 is found significant in two innings* at a 1% level of significance and in one innings** at a 5% level of significance. If we look closely at these innings it can be observed that at least one partnership (or pair) survived for more than 15 overs. However, the coefficient of determination R^2 was found to be quite small for almost all the matches, the maximum being 32.0%.

However, if we consider the run rate per over and regress it over time the results are improved as shown in Table 2.

Table 2

Match	β_0	β_1	p-value	R-Square
M1	5.38	-0.0145	0*	0.248
M2	5.23	-0.0116	0.133	0.049
M3	5.47	-0.0971	0.201	0.034
M4	4.79	0.0316	0*	0.433
M5	1.63	0.0443	0*	0.748
M6	3.41	0.0162	0.002*	0.179
M7	7.05	-0.0329	0.012*	0.125
M8	2.47	0.842	0*	0.806

In this case out of 8 innings β_1 is significant in 6 innings*. The coefficient of determination R^2 has also improved and the maximum value is now 80.6%. From Table 1 and Table 2 we can conclude that there is a trend in run rate per over.

6. Time Series Models

Since the appearance of the book by Box and Jenkins (1976) the use of auto regressive moving average (ARMA) models has become widespread in many areas of forecasting. It includes a special case and many other methods including the various forms of exponential smoothing. The whole Box-Jenkins' approach revolves around three basic models – Auto regressive (AR), Moving Average (MA) and mixed auto regressive moving average (ARMA) models. The auto regressive model of order p written as AR(p) is defined as

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots \phi_p z_{t-p} + a_t$$

where a_t is the sequence of random or white noise and it is assumed that it follows a normal distribution.

The moving average model of order q denoted as MA(q) is defined as

$$z_t = a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} \ldots \phi_q a_{t-q}$$

The mixed auto regressive model of order (p, q) denoted as ARMA(p, q) is defined as

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + a_t - \phi_1 a_{t-1} \ldots \phi_2 a_{t-q}$$

A stationary series has a constant mean and variance and a covariance structure which depends only on the difference between two time points. However, there are quite a few time series which are non-stationary. It has been found that if the series is non-stationary and the series is differenced one then it becomes stationary. If a series has to be differenced once to obtain stationarity, then the model corresponding to original series is called an integrated ARMA model of order p, 1, q or an ARIMA (p, 1, q). If differencing has to be performed at times to obtain stationarity the model is called an ARIMA (p, d, q) model.

The Box-Jenkins iterative approach for constructing linear time series models consists of four steps: identification of the model, estimation of the parameters of the model, diagnostic checking of model adequacy and finally forecasting future realisations.

The most crucial step in the Box-Jenkins approach to time series analysis is the specification or identification of the correct model. The rest of the steps are automatic in nature and any standard statistical package, eg. MINITAB or SPSS can do it. The two tools which are commonly used to specify ARMA models are the auto correlation function (ACF) and the partial auto correlation function (PACF). For the observed series $z_1, z_2, \ldots, z_n$ the ACF of order or lag is defined as

$$r_k = \frac{\Sigma(z_t - \bar{z})(z_{t+k} - \bar{z})}{\Sigma(z_t - \bar{z})}, k = 0, 1, 2, \ldots$$

It can be shown that for a moving average process of order q (MA (q)) the ACF is zero for orders greater than q. Hence the sample ACF is a good indicator of the order of the MA process. However, the ACF of auto regressive model do not remain zero after certain lags. The order of an AR model can be determined by using the partial auto correlation function (PACF). PACF at lag k is defined as the correlation between z_t and z_{t+k} after removing the effect of the intervening variables $z_{t-1}, z_{t-2}, \ldots, z_{t-k+1}$ and is denoted by ϕkk. It can be shown theoretically that PACF of AR (p) model is zero beyond lag p.

Briefly, whereas the ACF of AR (p) tails off, its PACF has a cut off after lag p. Conversely, the ACF of a MA (q) process of order q has a cut off after lag q, while its PACF tails off. If both the ACF and PACF tail off a mixed ARMA process is suggested. However, it is difficult to specify a mixed ARMA process by just looking at the ACF and PACF. Kumar (1986, 92) has developed a very simple but powerful method based on the theory of Padé approximation for the identification of mixed ARMA (p, q) model.

Looking at the ACF and PACF of the data the following models were selected as shown in Table 3 for different innings using run rate per over. It was observed that ACF and PACF of runs made in each over do not specify any model in most of the cases.

Table 3

Match	Model (for run rate per over)	Mean Square Error
M1	ARIMA (3,0,0)	0.02468
M2	ARIMA (1,0,1)	0.1708
M3	ARIMA (1,0,0)	0.3528
M4	ARIMA (2,0,0)	0.1349
M5	ARIMA (1,1,0)	0.05184
M6	ARIMA (0,0,2)	0.07786
M7	ARIMA (0,0,1)	0.2323
M8	ARIMA (1,1,0)	0.0936

Based on the above model we try to forecast the runs made in each over and run rate per over for the last 5 overs in each inning. The results are given in Table 4. It is observed that it is easier to get a good forecast of run rate per over as compared to runs made in each over. Also the Box-Jenkins' model works better than the simple regression.

Table 4 Forecast of run rate per over

Last 5	MI		M2		M3		M4		M5		M6		M7		M8	
Overs	A	F	A	F	A	F	A	F	A	F	A	F	A	F	A	F
1	4.63	4.65	4.78	4.78	4.89	4.85	3.76	3.92	3.81	3.80	3.95	3.84	5.63	6.15	6.65	5.51
2	4.61	4.66	4.93	4.81	5.02	4.87	3.70	3.84	3.80	3.81	3.89					
3	4.64	4.65	5.11	4.95	4.93	4.98	3.62	3.78	3.84	3.80	4.04					
4	4.75	4.70	5.15	5.13	5.0	4.91	3.57	3.72	3.89	3.84	4.04					
Last over	4.82	4.82	5.23	5.18	5.04	4.96	3.52	3.66	3.93	3.93	4.04					

A: Actual
F: Forecast

7. THE RUN TEST

Using the run test we can test the null hypothesis that the time series data is random against the alternative that there is a trend or that data exhibit a typical pattern of non-randomness. In case the null hypothesis is accepted it will imply that the runs made in each over is random or in other words the runs made in each over is independent of that in any other.

The use of the runs' test is not limited to testing the null hypothesis that a sample is random. The test can be applied to any sequence for randomness, no matter how the sequence is generated. The runs' test is frequently used to determine whether the residuals observed as part of a regression are likely to have come from a population in which the assumption of independent error term is violated. It is a non-parametric test and no assumption is made about the distribution from which the observations are drawn.

Suppose that we have a time series of n observations . A sequence of signs, with + denoting a value above the median and – a true value below, is formed from these data. Let R denote the number of runs in the sequence. The null hypothesis to be tested is of randomness in the time series. For a time series of more than 20 observations, the distribution of the number of runs under the null hypothesis can be approximated by the normal distribution. The test statistic

$$z = \frac{R - \frac{n}{2} - 1}{\sqrt{\frac{n^2 - 2n}{4(n-1)}}}$$

has a standard normal distribution.

The results for different innings are given below in Table 6.

Table 6

Match	k	p-value
M1	5.0086	0.000*
M2	4.9567	0.000*
M3	5.1985	0.000*
M4	4.1940	0.0073*
M5	2.7629	0.000*
M6	3.8219	0.000*
M7	6.2061	0.000*
M8	4.5338	0.000*

It can be observed that in all the cases the hypothesis is that the data is rejected. It confirms that data exhibits a trend and justifies the time series analysis done earlier in section 5 and 6.

8. CONCLUSION

The above data confirms that cricket is not by chance. Data does not show a random pattern but follows some kind of trend. It is possible to forecast the run rate per over for the last few overs using a Box-Jenkins' approach.

However, more appropriate analyses can be attempted by using sophisticated models like intervention time series models. This is in view of the fact that after a wicket falls there is some intervention otherwise the trend may continue.

REFERENCES

[1] Box G.E.P., and Jenkins G.M., Time Series Analysis: Forecasting and Control, (Holden-Day, Inc., San Francisco, 1976)

[2] Jenkins G.M. and Watts D.G., Spectral Analysis and its Applications, (Holden-Day, Inc., San Francisco, 1968)

[3] Kumar K., Some Topics in the identification of time series models. Unpublished Ph.D. Thesis University of Kent, U.K. (1986)

[4] Kumar K., Padé Approximation and its Application in Time Series Analysis, *Applied Mathematics and Computation*, Volume 50(1), pp 51-58 (1992).

ACKNOWLEDGMENTS

The author is grateful to Prof. R.K. Singh for suggesting this analysis of data.

GROUND REACTION FORCES AND HUB POSITION IN THE GOLFSWING

J. E. Baker[1]

Abstract

Earlier studies of the ground reaction forces that occur during a golfswing have gone a long way to providing a description of what happens during this complex physical action. However, many questions remain concerning why such forces can be observed and what are the implications for helping players to improve their performance. A case study approach was adopted to provide a qualitative description of the golfswing; others having already provided detailed quantitative analyses of swing data obtained through repeated trials. This qualitative approach should be seen as a complement to the quantitative results of others such as Carlsoo [3], Cooper [5], Richard [10], Fishman [7] and Barrentine [1].

1. Introduction

The aim of this paper is to find an explanation of the pattern of ground reaction forces that were observed when a player swings a golf club, and to indicate how this explanation might be of importance to the golf coach. The focus is on the vertical component of ground reaction force, the VGR. VGR will be used in this article to refer to the Vertical Ground Reaction of the player during the swing. VGR is not measured directly by the Kissler Force Plate, but can be derived as the Total Vertical Force experienced by the plate less the player's bodyweight. It is that part of the total vertical force that is caused by swinging the club, rather than just standing on the plate. There have been a number of studies of the VGR of golf players, the purpose of which has been to enable a better understanding of what occurs during a golfswing. One such study, Fishman [7], reported on weight transfer with results that prompted further investigation. His findings indicated that while there is a shift in weight from one foot to the other during the golfswing, the total VGR remains constant. Cooper [5], on the other hand showed substantial variation in VGR during the swing. Two surprising features of Cooper's findings are that:

- The VGR patterns are distinctively different for the 7-iron, 3-iron and driver, suggesting that the golfer swings differently for different clubs.
- The maximum VGR for the driver occurs very early in the downswing, approximately 0.1 seconds after the start of a process that lasts 0.25 seconds. At this stage of the swing, the player is pulling the club and their arms downwards which ought to show as a less-than-body-weight VGR.

[1]Tarata Learning Systems (Technology Consultants), Tarata Rd, Buanaba, Queensland.

The comprehensive study of Barrentine, Fleisig, Johnson and Woolley [1] gives an indirect picture of total VGR, since their data was collected by using separate force plates for left and right feet. In a situation where only one force plate is available to measure the forces, the results of their study are difficult to interpret.

In the first published discussion of ground reaction forces, Carlsoo [3] suggested that "the marked but brief diminution of the load on the left foot after impact is quite certainly a result of the force of support", a force that he attributes to the contact of the clubhead with the ground. If this were the case, the golfer should report experiencing a push through the hands as a result of contact with the ground; but no records of this have been reported.

As will be shown later, a feature of VGR force can be explained in terms of the rotation of the player about the hub of the swing. If the golfswing is considered as a two-arm pendulum, as first proposed by Williams [12], the hub of the swing is defined as the point about which the first pendulum arm rotates. The second arm of the pendulum is the club itself. The motion of the hub during the swing has been a feature of many studies of the golfswing, most notably Sanders and Owens [11] who investigated the swings of novice and elite players to determine where the hub is located. Unfortunately, they had hold of the 'wrong end of the stick', as their paper traces the position of the instantaneous centre of rotation of the clubhead rather than the grip. The hub is the point about which the player's hands, and hence the club grip, rotate rather than the clubhead. During the downswing, the player's body makes a forward movement as indicated by the forward shift of COVP. COVP is the 'centre of vertical pressure'. Movements of the COVP during the downswing correspond to movement of the player's centre of gravity, or weight transfer, as the swing proceeds. The position of the hub appears to shift in a similar way. The shift was noted by Jørgensen [8] and then Cochran and Stobbs [4, page 21] posed the problem of hub position by saying:

> The reasons for this shift are not yet entirely clear, or at least not yet scientifically substantiated.

One aim of this paper is to suggest why the shift is made, and whereabouts in the body the hub is located.

2. METHOD

The golfswings that formed the basis of the analysis were of three young golfers (aged 17) who represented Australia in the World School Championships, 1995, and won convincingly. The subjects were chosen because, from the case study point of view ,their swings showed different characteristics which demanded explanation. Subject 1 was the longest hitting player, Subject 2 was noted for a very accurate short game and steady performance off the tee, while Subject 3 was the only left-hander in the group. Ground reaction force data were collected during three separate occasions using a Kissler force plate. The experimental data output by the Kissler device were analysed as follows:

- The vertical VGR and COVP data were smoothed using a 10-point moving average. Because of the high sample rate, this process removed none of the characteristics of the data, but made the resulting experimental data curves more representative of the trends in force and position.

- The time component was converted to a % of total swing time figure, to facilitate comparisons between swings.

- The theoretical model parameters, described below, were determined by minimising the root-mean-square difference between expected and observed data. The minimisation process was performed by a gradient descent method.

3. PARAMETERS OF THE THEORETICAL VGR

From the point of view of Newtonian mechanics, Figure 1 shows a basic pattern of VGR that theoretical considerations suggest occurs during the golfswing.

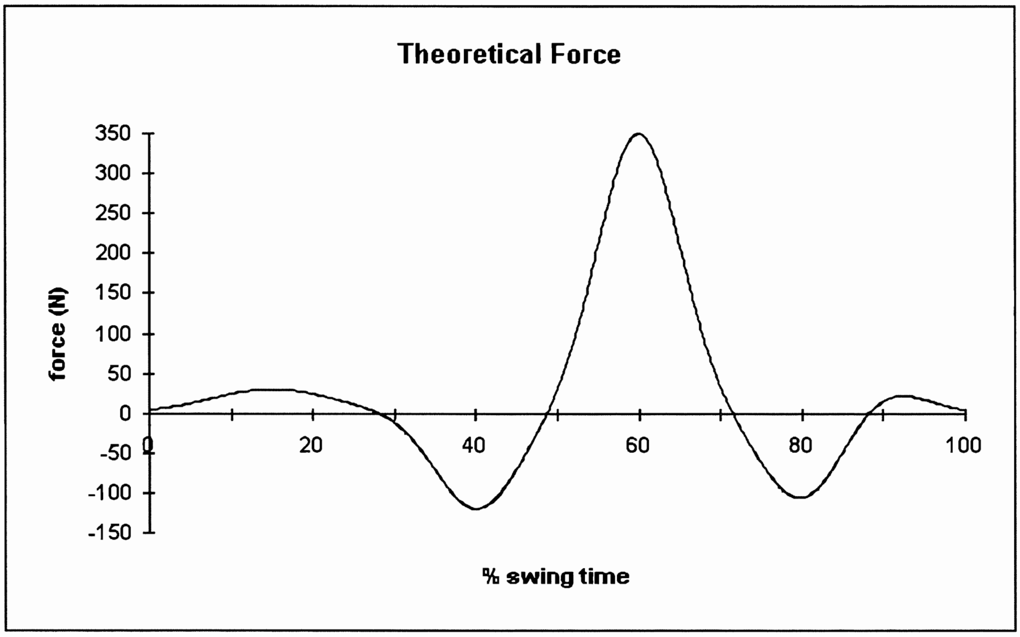


Figure 1: *Theoretical force pattern*

The graph of Figure 1 was based on elementary Newtonian principles and was formed by combining 'normal' functions of the form:

$$y = f \exp\left(\frac{-(t-\mu)^2}{2\sigma^2}\right) \tag{1}$$

where the parameter, f, is a measure of the maximum force occurring during the current time interval and is given in Newtons and the mean, μ, standard deviation, σ, and swing time, t, are given in terms of the percent of the total time taken for the swing. With functions of this kind, their influence is restricted to a region within 2σ of μ, and the influence decays rapidly to zero outside that region. Thus the formula for the theoretical VGR is given as:

$$vgr(t) = \sum_{i=1}^{5} f_i \exp\left(\frac{-(t-\mu_i)^2}{2\sigma_i^2}\right) \tag{2}$$

The purpose of this formulation is to enable comparison between trials and players to be made in reasonably precise terms. The values of the force, mean and standard deviation parameters are given in Table 1.

This basic pattern was present in all the swings of the sample, but there is a further feature whose presence can be noted in the sample data, and which is seen as a dip in the VGR vs % time graph that occurs at about impact.

To model this feature, we introduce a sixth term to Equation 2, formed by multiplying the function of Equation 1 by

$$\sin\left(\frac{\pi(t-\mu)}{2\sigma}\right)$$

which gives the negative and positive parts of the dip, but confined to a region of 2σ about μ. The theoretical VGR function now becomes:

$$vgr(t) = \sum_{i=1}^{5} f_i \exp\left(\frac{-(t-\mu_i)^2}{2\sigma_i^2}\right) + f_6 \sin\left(\frac{\pi(t-\mu_6)}{2\sigma_6}\right)\exp\left(\frac{-(t-\mu_6)^2}{2\sigma_6^2}\right) \tag{3}$$

Table 1: Parameters of a theoretical VGR force

Phase	Force	Mean	StDev	Comment
0 – 30	50.0	15.0	7.5	The first part of backswing stage of the swing involves raising the arms and club above shoulder level. Raising a mass in this way causes an increase in VGR above body mass.
30 – 50	-120.0	40.0	5.0	In this part of the swing, the backswing slows down, and effort is put into pulling the club and arms downwards. The top of the backswing typically occurs at the 35% mark. The slowing down and pulling down processes both show as a less-than-body-mass VGR.
50 – 70	350.0	60.0	5.0	The maximum forces of the swing occur slightly before impact, when the player plants the left foot and makes a large effort to pull the club upwards, against the centrifugal forces generated by the circular motion of the arms and club. Impact occurs at about the 65% mark.

				To counteract the centrifugal forces of the arms and club, the player has to pull upwards, hence the sharp increase in VGR above body mass.
70 – 85	-110.0	80.0	5.0	During the follow through, the momentum of the arms and club is now upwards and the player has to slow the motion down. To slow the motion down, the player has to pull downwards, which shows as a decrease in VGR.
85 – 100	30.0	95.0	2.5	There is a final raising of the body and slowing of the club which results in this increase in vertical force. This shows as a small increase in VGR above body mass.

Figure 2 (see next page) shows the theoretical and actual curves superimposed, and gives a visual indication of the level of agreement.

Of most interest here is the extent to which the dip phenomenon, as measured by the additional sixth term, is present in the subjects' swings. Thus the force, mean and standard deviation values for the impact phase and dip in the theoretical graph of Figure 2 are given in Table 2. Parameters for five other sets of experimental data are also included.

4 EXPERIMENTAL DATA

Experimental data for three subjects were collected for swings with a driver and a 5-iron. Subjects 2 and 3 show similar features to those exhibited by Subject 1. Figures 3 and 4 give the comparative charts for the actual and theoretical force patterns for Subject 2, while Figures 7 and 8 are the comparative charts for Subject 3. Table 2 summarises the theoretical parameters for all subjects using the driver and 5-iron.

Table 2: Parameters of the theoretical curves to fit data from Subjects 1, 2 and 3 Columns 2 – 4 for Driver, Columns 5 – 7 for 5-iron.

	Driver			**5-iron**		
	Subject 1	Subject 2	Subject 3	Subject 1	Subject 2	Subject 3
f_3	426.1	331.3	582.5	360.8	265.5	466.8
μ_3	58.2	58.2	55.4	57.8	50.0	60.5
σ_3	2.7	2.9	3.8	3.0	3.3	3.5
f_6	173.9	546.7	841.5	127.1	289.2	523.4
μ_6	68.2	67.4	61.3	63.1	56.9	64.2
σ_6	2.4	3.2	3.8	4.7	3.0	2.9

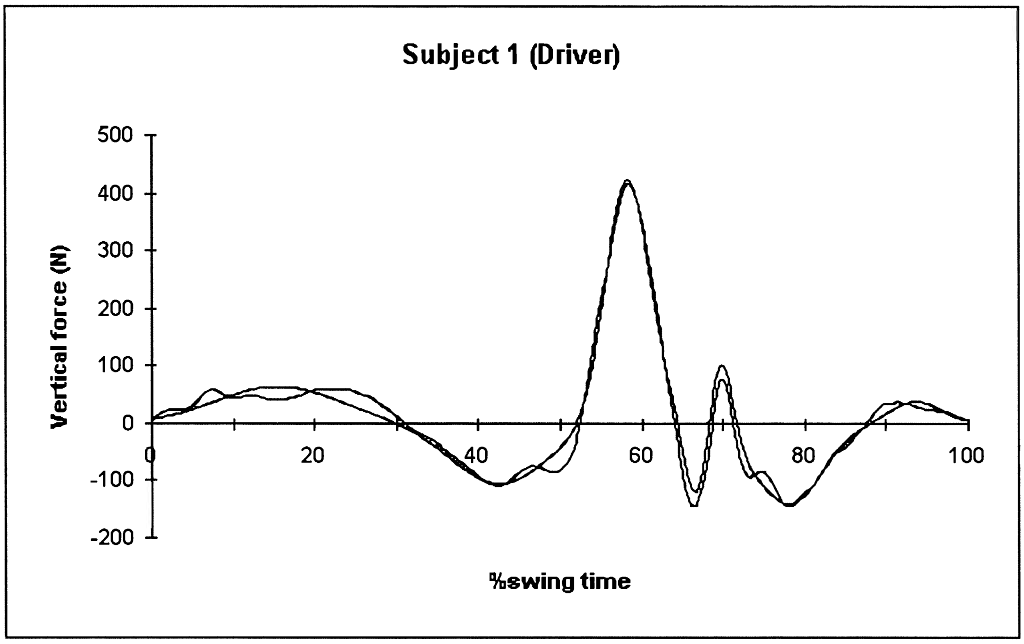


Figure 2: *Actual and theoretical VGR forces for Subject 1 with driver*

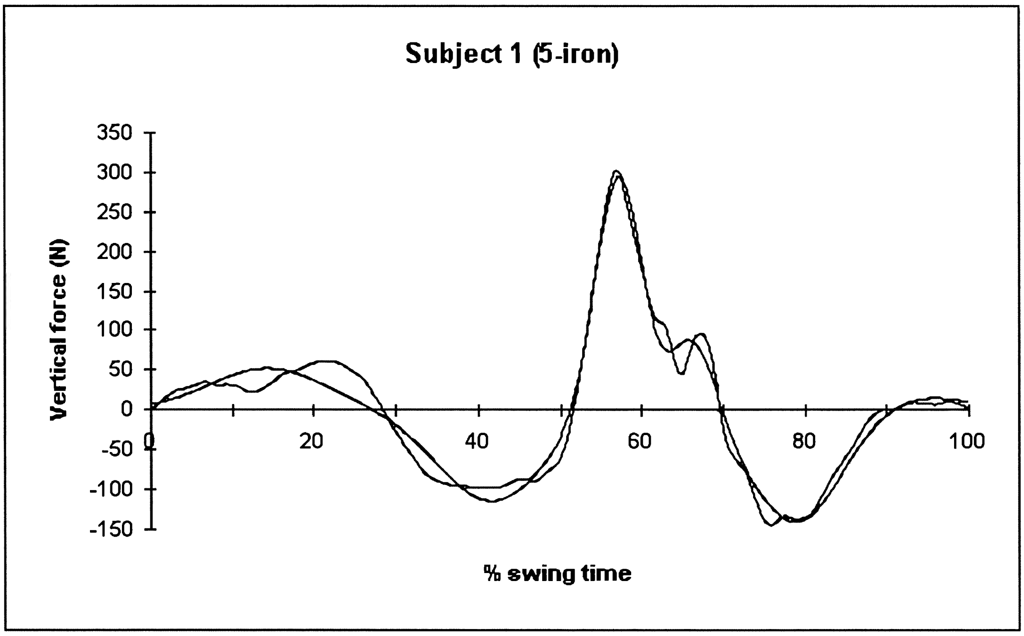


Figure 3: *Actual and theoretical VGR forces for Subject 1 with 5-iron*

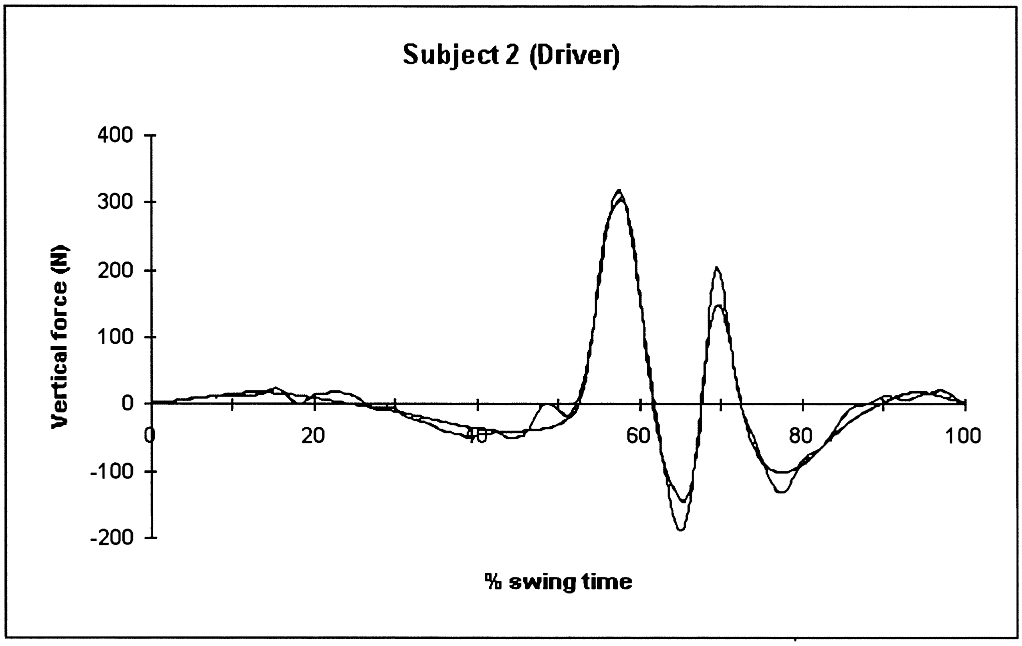


Figure 4: *Actual and theoretical VGR forces for Subject 2 with driver*

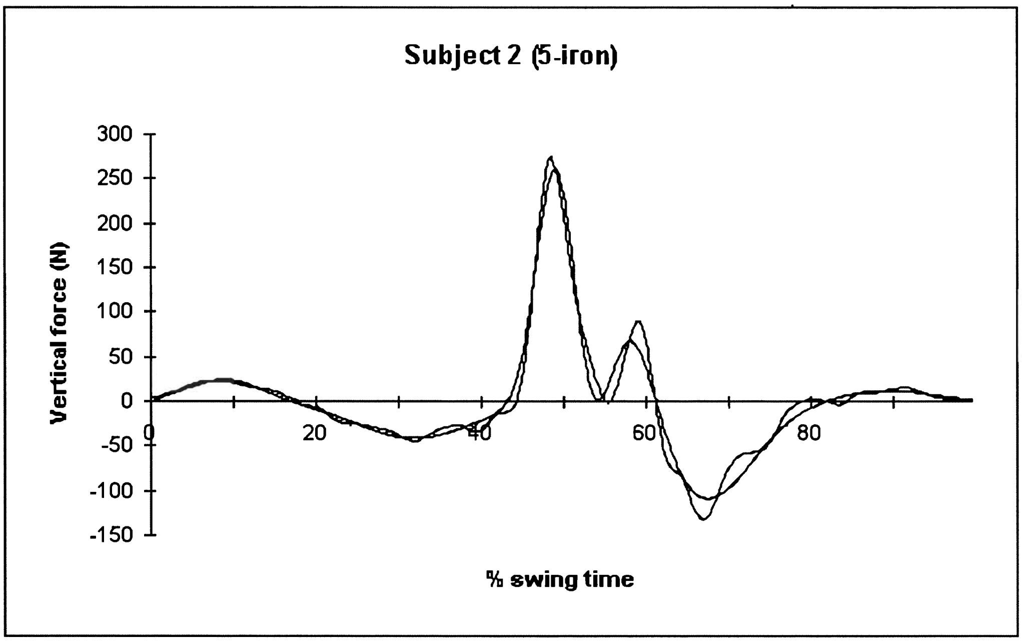


Figure 5: *Actual and theoretical VGR forces for Subject 2 with 5-iron*

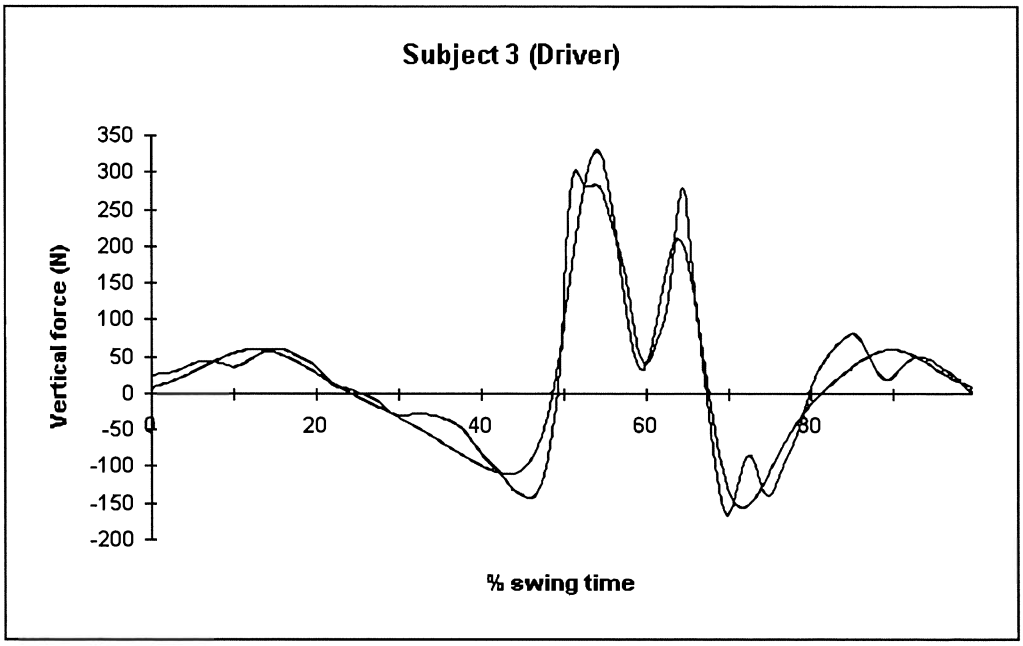


Figure 6: *Actual and theoretical VGR forces for Subject 3 with driver*

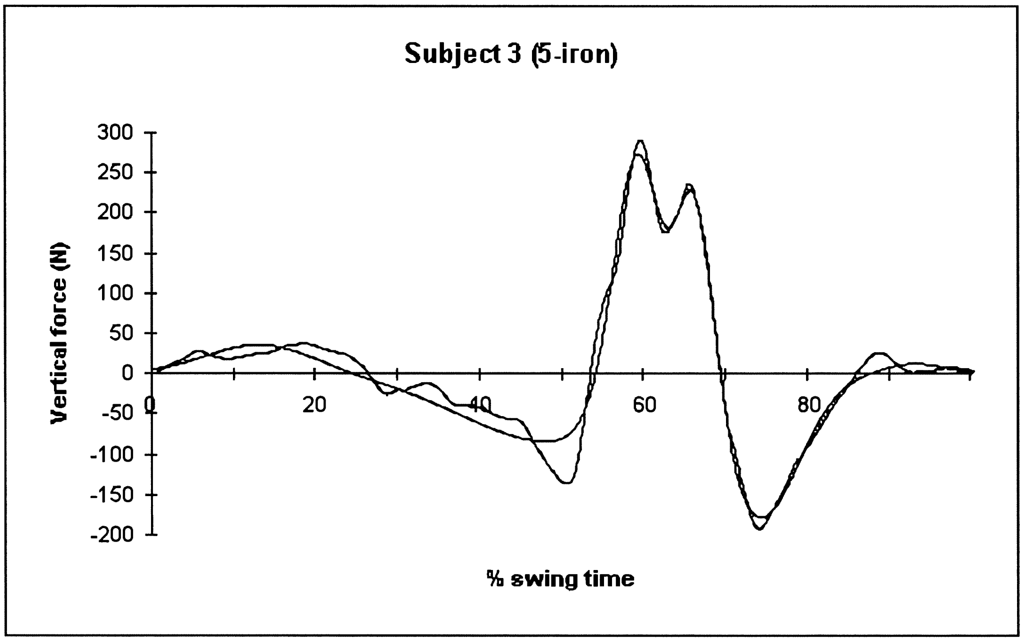


Figure 7: *Actual and theoretical VGR forces for Subject 3 with 5-iron*

What becomes clear by inspecting the data and graphs, is that:

- The theoretical curve of Figure 1, without the dip component, is evident in all three subjects, and independent of the club used. The main difference between players is not so much in the timing of their swings, as shown by variation in μ_i

values, as in the variation in f_i values, with the higher values being recorded for the heavier subjects.

- The dip component is different for each subject, with the dip for the driver being more pronounced than the dip with 5-iron. This trend is supported by the theoretical values for h_6, which show an average drop of 37% from driver to 5-iron values.

- An explanation of the dip phenomenon cannot be found by looking at the club-ground reaction, as suggested by Carlsoo. Since there is such a noticeable difference between the driver and 5-iron, and, of the two, the 5-iron shot tends to make more contact with the ground than does the driver, one should expect a lesser dip with the driver than the 5-iron.

In each of the trials analysed, it was noticeable that the rapid increase in VGR coincided with a movement of COVP in the medial direction towards the target. For example, Figure 8 clearly shows this process, which corresponds to weight transfer from rear to front foot during the downswing.

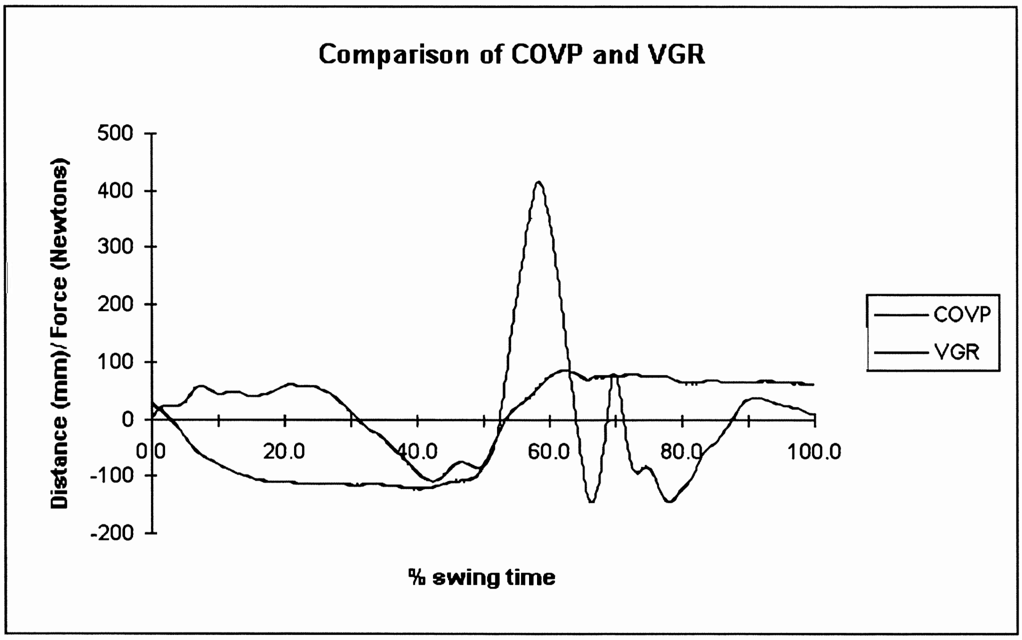


Figure 8: *Comparison of COVP position and VGR force for Subject 1 (Driver)*

It was also noted that weight transfer continued after the maximum VGR and that a slight reversal of the transfer coincided with the local minimum VGR in the dip.

5. CONCLUSIONS

Studies that focus on the mechanics of the golfswing, such as Williams [12] and Daish [6], make it quite clear that to maintain the motion of the clubhead just before and through impact considerable vertical force needs to be applied, which would manifest in a larger-than-bodyweight VGR. Williams [12] calculates this force to be 107 lb wt (477 Newtons) towards the hub of the swing. And yet, for the subjects of this investigation, the VGR dips to below or close to bodyweight at this crucial moment in the swing.

The dip is clearly evident in swings with a driver, and is present, but less pronounced, in swings with the 5-iron. To explain this phenomenon, we need to consider the basis of the mechanical models presented in the literature. The authors cited above consider the swing in terms of a compound 2-arm pendulum in which the hub is considered as a fixed point (physically located between the shoulders on the spine), the first pendulum arm is an imaginary line from hub to grip (not to be confused with the player's arm) and the second pendulum arm is the club itself. Jørgensen [8] adds a forward motion of the hub to the above, and this noticeably improves the match between observed and theoretical data. Suppose that the hub of the swing is positioned not on the spine, but at a point closer to the front shoulder than the rear shoulder, as shown in Figure 9.

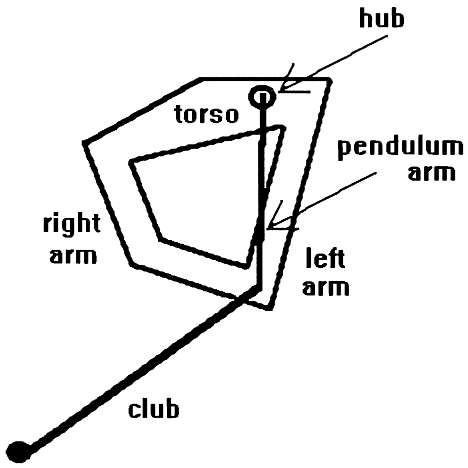


Figure 9: *Location of Hub between spine and left shoulder*

In this case, the motion, in which the player's torso rotates underneath the axis of the hub, would be such as to cause a substantial reduction in VGR, and the near body-weight VGR at impact would be explained. Also, since the player exerts most force by a pulling action along the left arm, the closer the hub is to the left shoulder, the less the amount of torque that the player would have to generate about the hub in order to exert a large pull along the left arm.

The extent of the reduction in VGR or dip, as measured by the f_6 factor in the theoretical model, is affected by the club being used. The dip is most pronounced for the longest club, the driver, and reduces with club range. The experimental data of the study, as modelled by the function of Equation 3, showed an average reduction of 37% between driver and 5-iron, with swings with the seven iron showing a further reduction. This phenomenon can be explained by suggesting that the hub position for the downswing changes with club. The shorter the club, the closer to the spine is the hub position. Swings with a pitching wedge should be made with the hub very close to the spine, while putting is an action that is best carried out with the spine as hub held stationary.

Leadbetter [9] refers to two axes of rotation. For the right-handed golfer, he describes the axes as follows:

Imagine a line drawn down the inside of your right shoulder, through your right hip joint and past the inner part of your right thigh into the ground. Now imagine the same line travelling down the left side of your body. These are the two axis points around which every athletic swinger rotates, back and through.

The experimental data of this study certainly supports Leadbetter's assertion, and goes on to suggest that the position of the hub most certainly varies from club to club, with the hub of the low irons being closer to the spine than the hub of the woods. The golf coach, therefore, should expect a player to exhibit a certain degree of reduction in VGR at the moment of impact, and may want to encourage the player to increase/reduce this effect depending on the level of reduction exhibited and the club being used.

REFERENCES

[1] Barrentine S.W., Fleisig G.S., Johnson H. and Woolley T.W., *Ground reaction forces and torques of professional and amateur golfers*, in Science and Golf II: Proceedings of the 1994 World Scientific Congress of Golf, Ed: Cochran, A. J. and Farrally, M. R., E & F Spon, London (1994).

[2] Budney D.R. and Bellow D.G. , *A dynamic model for a golfswing*, in Science in sports: open papers from the International Congress of Sports Sciences, ed: Terauds, J., Academic Publishers, Edmonton, Canada(1979).

[3] Carlsoo S *A kinetic analysis of the golfswing*, Journal of Sports Medicine and Fitness, 7, 76–82 (1967).

[4] Cochran A. and Stobbs J., *The search for the perfect swing*, Heinemann, London (1968).

[5] Cooper J.M., Bates B.T., Bedi J. and Scheuchenzuber J., *Kinematic and kinetic analysis of the golfswing*, Biomechanics IV: Proceedings of the fourth International Seminar on Biomechanics, University Park Press, Baltimore (1974).

[6] Daish C.B., *The physics of ball games*, English Universities Press, England (1972).

[7] Fishman L., *Weight shift*, Golf Digest, 40.4, 132–141 (1989).

[8] Jørgensen T.P. ,*The physics of golf*, American Institute of Physics, New York (1994).

[9] Leadbetter D., *The Golfswing*, Penguin, New York (1990).

[10] Richard J., Farrell M., Kent J. and Kraft R., *Weight transfer patterns during the golfswing*, Research Quarterly for Exercise and Sport, 56.4, 361–365 (1985).

[11] Sanders R. H. and Owens P. C. *Hub movement during the swing of elite and novice golfers*, International Journal of Sport Biomechanics, 8, 320 - 330 (1992).

[12] Williams D., *The dynamics of the golfswing*, Quarterly Journal of Mechanics and Applied Mathematics, 20.2, 247–264 (1967).

MATHEMATICS OF THE HUMAN ENGINE - EVALUATION OF POWER AND ENDURANCE

R Hugh Morton[1]

Abstract

The human body is often used as a source of mechanical power, mainly in sport and recreation. In the relentless pursuit of the ultimate and in bona fide scientific research, it is natural to ask about the capacity, power and endurance of this engine. Mathematical modelling of the system, together with experimental data collection and analysis provides us with many insights. This paper examines some models of human bioenergetics, describes some typical experiments undertaken in an exercise laboratory, and discusses some applications where the human engine is used.

1. INTRODUCTION

As a mathematician and statistician interested in sport, I am often asked what on earth mathematics or statistics has to do with sport. Through regular practice I am now able to successfully defend my position against all but the most sceptical enquirer. Today however I am preaching to the converted, and the fact that this meeting is the third in what has become a regular and successful series on Mathematics and Computers in Sport is testimony to my views. I wish to share with you some of the recent work on the modelling of the human bioenergetic system to which I have contributed. I shall of course not confine myself to mathematics, but will include brief details of some of the experiments exercise physiologists typically undertake in their laboratories as a means of studying these properties of this powerhouse of energy, and make mention of some of the interesting applications of this most fascinating of engines.

In particular, and with a focus on human performance, I am interested in what power the engine can develop, what its' endurance capability is at various tasks, and how we can modify the engine for increased performance. In a mathematical approach to these questions, one comes face to face with problems regarding the fuels available to the engine, the methods of "burning" them, the capacities of the fuel tanks (not to be confused with the capacity of its stomach, or volume of its cylinders!), its control system, the self preservation of the organism, measurement of the parameters of the system, etc. I shall not avoid these matters either. However, my time is limited, and I must be both selective and succinct yet allow some time for questions and discussion.

---

[1]Department of Statistics, Massey University, Palmerston North, New Zealand

2. MODELLING THE SYSTEM

The basic chemical fuel for muscle contraction is adenosine tri-phosphate (ATP) which splits into adenosine di-phosphate (ADP) and P, releasing energy. It is available *in situ* in limited quantities, and so for continued contractions the ADP and P must be recombined into ATP. This is achieved anaerobically by consumption of either creatine phosphate (CP) or glycogen (the stored form of carbohydrate). CP is immediately available *in situ* also, and is a "high octane" fuel. It exists in limited amounts and resynthesis occurs rapidly but only after exercise ceases. Significant amounts of glycogen are available, though less promptly, both *in situ* and from other storage locations. It is also a "high octane" fuel, but anaerobic glycolysis is chemically inefficient and produces lactic acid as an undesirable by-product. The resynthesis of glycogen is a rather slow process, taking several hours, or even days if depletion has been severe. The resynthesis of ATP can also be achieved aerobically either by the combustion of glycogen, which is chemically much more efficient and does not produce lactic acid, or by the combustion of fat. Aerobic glycolysis is a less prompt mechanism, taking about three minutes or more to stabilise. The oxidation of fat is considerably slower still, and is really only significant in exercise durations of two or more hours. Figure 1 shows a schema of basic muscle bioenergetics.

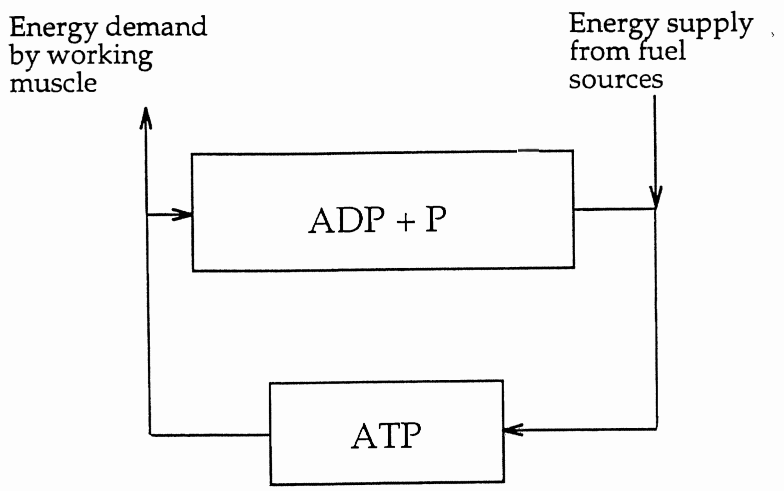


Figure 1: *The cyclic breakdown and resynthesis of ATP*

Mathematical models of the human bioenergetic system can therefore be based either according to the type of fuel utilised, or the mechanism by which it is consumed, or by a combination of both. In any model, certain other functional assumptions also need to be made, as we shall see. The remainder of this section investigates three of the rather few attempts that have been made to model the whole system as a single entity.

The first model was proposed as a result of observing a close linear relationship between the total work, W, performed at constant power by a cyclist exercising to exhaustion, and the time, t, taken to achieve that work (Monod and Scherrer [3]).

$$W = \alpha + \gamma t \tag{1a}$$

Since at a constant power P, W = P.t, this equation has various mathematically equivalent alternatives

$$P = \alpha/t + \gamma \tag{1b}$$

$$t = \alpha/(P - \gamma) \tag{1c}$$

$$W = \alpha P/ (P - \gamma) \tag{1d}$$

Given the errors of measurement in P, t and W in any practical setting, it is hardly surprising that these equations are not statistically equivalent, and when fitted to the same data can produce different estimates of α and γ. Intuitively equation (1c) represents the most natural choice of dependent and independent variables, and in fact does produce physiologically more reasonable estimates for α and γ.

The model therefore proposes the following:

1. There exists a fixed anaerobic energy store α (called the anaerobic work capacity, AWC) which can be used sparingly over a long period of time, or lustily over a shorter period. When this store is fully depleted, the cyclist becomes exhausted and ceases exercise.

2. This store is supplemented by a continuous aerobic supply, the rate of which is bounded above by γ (called the critical power, CP). Clearly if the power demanded is $P \leq \gamma$, then the notion of exhaustion is vacuous.

Thus for any constant $P > \gamma$, equation (1c) and the other forms can be very readily deduced. This model formulation is referred to as the CP concept Hill [2], and has been extensively studied. The concept, although obviously a simplification, has withstood much of this investigation remarkably well, at least over a range of times from 2 to 20 minutes.

What has been frequently observed though, is that athletes cycling exactly at a power $P = \gamma$, have not been able to continue exercise for very long, 25 - 40 minutes, before becoming exhausted. Concomitantly their fuel stores have not been found to be fully depleted at exhaustion. Work-time data taken over a wider range of times, dips below $W = \alpha + \gamma t$ towards the origin for very short supramaximal exercise, and dips somewhat below for much longer times also. Clearly therefore this model needs to be reworked. This can be achieved from different perspectives (Morton [7]).

Equation (1c) is a rectangular hyperbola with the vertical asymptote estimable at $P = \gamma$, and the horizontal asymptote fixed at $t = 0$. If we relax this requirement, allowing both asymptotes to be estimable, we can write

$$t = \alpha/ (P - \gamma) + k \tag{2}$$

Fitting this to data results in higher estimates for α and lower estimates for γ, both of which are in the direction expected from improvements consequent on the problems described above. Estimates of k are found to be significantly less than zero. This means that equation (2) intersects the P axis at say P* when t = 0. This can be interpreted physiologically as a "maximum instantaneous power". Furthermore it can be shown that at exhaustion this model formulation allows for not all of the anaerobic work capacity to be depleted.

This last model property raises an intriguing question, for it can also be shown that at exhaustion after short supramaximal exercise, more of the anaerobic work capacity remains untapped, than at exhaustion after long duration lower intensity exercise. Could this be indicative of an endogenous self-preservation mechanism of the human organism?

One can reach equation (2) from a completely different starting point. Let us suppose that the maximum power that could be delivered at any instant depends on the state of the fuel resources at that instant. Specifically, if the anaerobic work capacity is fully charged, P* can be delivered, while if it is completely depleted, only γ can be delivered. Letting this maximum decline linearly from P* to γ as the anaerobic work capacity depletes from α to zero, equation (2) and all its interesting properties, can be derived. In so doing, a useful definition of exhaustion can be deduced; the point at which the maximum power that could be delivered equals the power required, *ie* the inability (just) to meet the power demand.

Figure 2 shows a comparison of equations (1c) and (2) fitted to the same data.

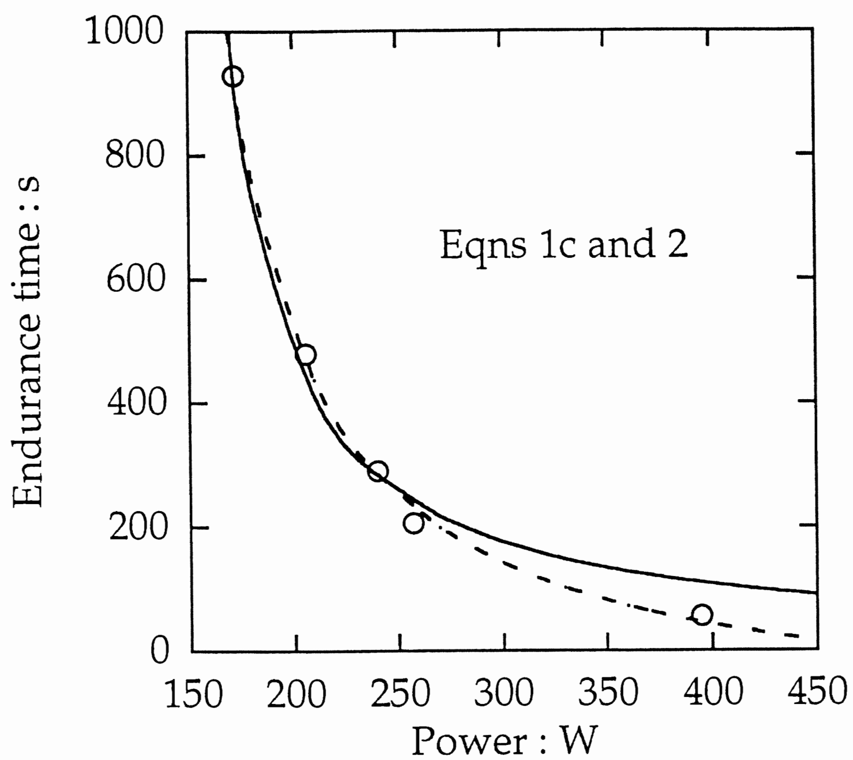


Figure 2: *The two and three parameter hyperbolas fitted to data from an exercising subject*

A more complex hydraulic anologue model of whole-body human bioenergetics has been proposed Morton [4] and [5], which makes an attempt to separate the ATP replenishment systems both according to the type of fuel used and the mechanism by which this is accomplished. Figure 3 is a diagrammatic representation of the model.

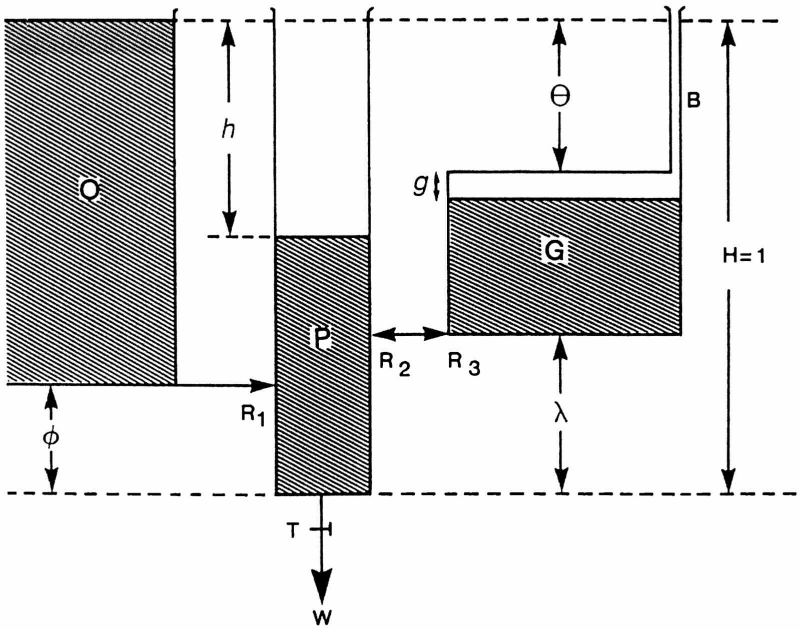


Figure 3: *The three component hydraulic model of human bioenergetics*

Vessel O, of "infinite" capacity, representing the oxidative energy source is connected to vessel P, representing the alactic (creatine phosphate) energy source, through a tube R_1. R_1 has a maximal flow M_o, known as the maximal oxygen uptake, frequently denoted VO_{2max} . Vessel P has an assumed height H = 1 arbitrary unit, a volume V_P and a cross sectional area A_P arbitrary units.

The height of the base of vessel O above the base of vessel P is denoted ϕ, and hence the constant height of fluid in vessel O above R_1 is $1-\phi$. A tap T, at the base of vessel P regulates the net outflow W from the system, where W represents the measured energy expenditure, power, or workload. Vessel G, representing the glycogen store, is connected at its base to P, by a tube R_2 at a height λ above the base of P. Vessel G has a finite fluid volume V_G and R_2 has a maximal flow M_G. The top of vessel G, except for a very narrow extension tube B, is at a level θ below the fluid level in vessel O. The fluid in B represents resting blood and tissue glycogen and does not contribute in measurable amount to the net flows in the system. Vessel G therefore has a height of $1 - \theta - \lambda$, and a cross sectional area A_G arbitrary units.

The model operates as follows. Suppose tap T is opened to allow a net outflow W. This induces a drop, *h*, in the level of fluid in vessel P. This in turn induces a flow from O to P through R_1. This flow, representing a rise in net oxygen consumption, is in accordance with the ratio of *h* to $1 - \phi$, equalling the maximum M_o when *h* equals (or exceeds) $1 - \phi$. If W is small, then *h* will reach an equilibrium position, no greater than θ. This corresponds to a steady state oxygen uptake, VO_{2ss}. If T is closed at any time, the level in P will return to its resting level, by virtue of a decreasing flow through R_1. This flow ceases when h equals zero, and corresponds to the repayment of the alactic oxygen debt.

If W is of sufficient magnitude, greater than a threshold value W_θ known as the anaerobic threshold (AT), then after a while h will exceed θ, in which case a net flow from G to P is then also induced. This flow, which represents glycogen depletion and the anaerobic production of lactic acid by the working muscles, is in accordance with the difference in levels between vessel G (an amount g below the top), and vessel P, with the level in G dropping also but lagging behind the level in P. If T is closed, P will be refilled, initially from both O through R_1 and from G through R_2. This continues only until the lag in levels between G and P has been eliminated. This brief phase represents partial repayment of the alactic oxygen debt by contracting an increased lactic oxygen debt, known as delayed or post-exercise lactate formation. Thereafter P refills by a decreasing flow through R_1 and G in turn refills by a flow through the return tube R_3. The flow through R_3 is also in accordance with the difference in levels between P and G. The maximal flow through R_3, M_R, is very much smaller than M_G or M_o. Ultimately both the lactic and alactic oxygen debts will be repaid. Once again, if W had not been too great, an equilibrium level with $h \leq 1 - \lambda$ and $VO_2 \leq M_o (1 - \lambda)(1 - \phi)$ could have been achieved, by which time the early lactate flow through R_2 would have ceased. If T is closed after equilibrium has been reached, there would be no delayed lactic acid formation and both P and G would be refilled immediately, through R_1 and R_3, respectively.

If W is of even higher magnitude, demanding an energy expenditure in excess of $M_o (1 - \lambda)(1 - \phi)$, then after a further while h will exceed $(1 - \phi)$. In this event, VO_2 will remain constant at VO_{2max} and the flow through R_2 will persist. Since G is of limited capacity, it will later become empty, and so too may P. The subject would then have depleted his energy stores, and would no longer voluntarily be able to maintain exercise at this level. Once T is closed, repayment of the lactic and alactic debts will be very similar to that described above, except for the absence of post exercise lactate formation and for the fact that initially VO_2 will be constant at VO_{2max} until such time as $h < 1 - \phi$.

Various assumptions about how the development of maximum power may depend on availability (or depletion) of fuel stores can be investigated. It turns out Morton [5] that an association between power limitation and the amount of fluid remaining in vessel G produces the most realistic predictions when compared to available experimental data for the two most common forms of W; W = constant and W = rt (the linear ramp).

More generally speaking, the mathematics of this model has given insight into the mechanisms of whole-body human bioenergetics, and a number of its earlier predictions have since been empirically verified. For example, a second exponential component to VO_2 starting about a minute or more into exercise was discovered empirically by Barstow et al [1]. Also the flow through R_2 can be linked to a 2-

compartment (muscle and blood) diffusion model due to Zouloumian and Freund [8] in order to relate predictions about blood lactate to the existing wealth of such data. Simulation studies suggest consistency with these observed data.

However, this hydraulic model is again a simplification of the real world. In particular it omits the oxidative consumption of fat, which would require another (large volume) compartment, F, and some regulatory device controlling allocation of oxidative mechanism between F and G. Some time I should get around to some work on this!

3. EXPERIMENTS

The simplest type of endurance experiments require the subject to exercise to exhaustion at a fixed power output on the cycle ergometer. Typically four or five minutes of light exercise is allowed as a warm-up before starting, and heart-rate and oxygen consumption are measured continuously during the experiment. Near the exhaustion point, subjects are usually exhorted verbally to continue as long as possible, and the end-point is usually reasonably precipitous. Immediately thereafter a cool-down period is allowed before the data collection is terminated. The power output and endurance time recorded form the basis for data such as displayed in Figure 2. Total work $W = P.t$ can be calculated for use in the other equations if required. In some cases, blood lactate data is also collected for anaerobic threshold determination.

It is common also for experimental sessions to involve work on a step wise incremental protocol. Typically the power output is raised by say 30 to 50 watts every third minute until exhaustion. With electronically controlled cycle ergometers a true ramp $P = rt$ can be closely approximated, with power adjustments of the order of 0.5 watt per minute. Equations (1) and (2) can be modified to study endurance data for a series of ramp tests with differing r (Morton [6]).

For the hydraulic model, due to its complexity, much more detailed experimentation is required to obtain a basis for estimation of its parameters.

Of course, exercise is not restricted to cycling. The ideas transpose easily to rowing, running, swimming, kyaking, weightlifting; or to any form of exercise where power output can be measured in some way.

4. APPLICATIONS

One application which readily springs to mind is human powered flight. Here the engineers have designed an aircraft with very precise requirements to get it airborne and maintain it aloft for some period of time. Getting this matched to the power of the human engine has taken some time to achieve. The two best known flights are the crossing of the English Channel by the Gossamer Albatross, and the retracing of the mythical flight of Daedalus and Icarus from the island of Santorini to the Greek mainland. Both were achieved by teams from the USA, though the latter did end in the breakers rather than on dry land!

Another application is the Colorado Speed Challenge, where a large prize is available for the first human powered bicycle to exceed 70 mph over a 200m distance. There are various rules to the contest of course, but the human engine has yet to meet the design requirements of the latest in bicycle technology. The prize is within reach, with over 68 mph being achieved two years ago. (Olympic cycle sprinters reach speeds of a little over 60 kph.)

5. REFERENCES

[1] Barstow T.J., Casabari R. and Wasserman K., O_2 uptake kinetics and the O_2 deficit as related to exercise intensity and blood lactate, *J.Appl.Physiol.,* **75**, 755-762, (1993).

[2] Hill D.W., The critical power concept, *Sports Medicine*, **16**, 237-254, (1993).

[3] Monod H., and Scherrer,J. The work capacity of a synergic muscular group, *Ergonomics*, **8**, 329-338, (1965).

[4] Morton R.H., A three component model of human bioenergetics, *J.Math.Biol.,* **24**, 451-466, (1986).

[5] Morton R.H., Modelling human power and endurance, *J.Math.Biol.,* **28**, 49-64, (1990).

[6] Morton R.H., Critical power test for ramp exercise, *Eur.J.Appl.Physiol.,* **69**, 435-438, (1994).

[7] Morton R.H., A 3-parameter critical power model, *Ergonomics*, **39**, 611-619, (1996).

[8] Zouloumian P. and Freund H., Lactate after exercise in man, *Eur.J.Appl.Physio.l,* **46**, 135-147 and 149-160, (1981).

FORMATION OF NETBALL TEAMS FOR A SERIES OF TRIAL MATCHES

David H. Noble[1]

Abstract

Until recently, in order to choose the best netball team to represent Victoria at primary school level, selectors organised a series of trial matches during which they could evaluate players' performances. The selectors wished the teams for these matches to be of roughly equal ability and to place players in their preferred positions wherever possible. In order to do this, players were required to nominate three positions, in order of preference, in which they would like to play. In addition, the organisers assigned each player a rank based on previous ability. These ranks were then used in combination with the expressed player/position preferences to manually form teams according to a set of rules. The organisers now have at their disposal a user-friendly computer program whose core is a linear programming assignment algorithm. This achieves their objectives in a fraction of the previous time and also enables them to cater for last-minute changes that often occur to the database.

1. INTRODUCTION

Until recently, the Victorian Primary Schools Sports Association (VPSSA) organised a series of netball trial matches twice a year in Melbourne aimed at selecting a team to represent the State at the Australian championships held once a year. Approximately 80 players from primary schools within Victoria were nominated by their schools for consideration by the selectors in these trial matches.

The following information was provided for each player:

name, school, height, rank and preferred positions.

Rank

Players were given a rank (from 1 to 4) by the organisers, usually on the basis of their performance at previously-held zonal (sub-State) trials. Rank 1 players were the best players.

[1]School of Mathematical Sciences, Swinburne University of Technology, Hawthorn Vic 3122

Preferred Positions

There are 7 positions in a netball team:

GS Goal Shooter,
GA Goal Attack,
WA Wing Attack,
C Centre,
WD Wing Defence,
GD Goal Defence,
GK Goal Keeper.

Each player was required to nominate three positions in order of preference.

2. FORMATION OF TEAMS

On the day of the trials, three rounds of matches were played. Prior to this investigation, the organisers formed teams manually using different sets of rules for each round, as described below:

In the first round:

1. Assign all rank 1 players their first preference if possible, otherwise their second preference and failing that their third.

2. Repeat for players of rank 2, then rank 3 and finally rank 4.

If the number of players participating in the trials was not a multiple of 7, those players who had not been assigned a position were classified as "emergencies" and were allocated across teams. They were interchanged with other players during the course of the matches.

In the second round, the rules were modified to take account of how well preferences had been satisfied in round 1. Players were assigned to their preferences in the following order:

1. those who ended up as emergencies in round 1,

2. those who were not placed in any of their preferred positions in round 1,

3. those who were allocated their third preference in round 1,

4. all remaining players in order of rank (rank 1 first) – but to a different position to that assigned in round 1.

Again, players not able to be assigned a position in this way bacame emergencies.

For the third and final round, the organisers selected the best 22 to 24 players based on their performances in rounds 1 and 2. These players then competed in a series of round-robin matches after which the team to represent Victoria was chosen. The programme described in this paper was not used for this final round of matches.

It will be noticed that the information on player height does not feature in any of the rules given above. This information was generally only used by the organisers when they could not otherwise decide which player should occupy a particular position. Certain positions (GS, GA, GD and GK) are generally allocated to the taller players.

Note also that it was possible that the set of preferences expressed by all the players could lead to a solution where a high ranking player was not assigned any of her three preferred positions (eg. if all top players preferred the same positions). However, this had not happened in practice. Should it have done so, the organisers were sufficiently confident that they would have been able to alter the expressed preferences so that an outcome, acceptable in the circumstances, could have been obtained. Players who prefer certain positions usually do not dislike playing in certain other positions.

Because the team formation process was carried out manually, essentially by trial and error, there was no guarantee that the best solution was obtained. Additionally, on the day of the trials, it often happened that one or two players did not turn up due to illness or other reasons, while other players, not previously notified to the organisers, turned up in the hope of inclusion. The organisers were therefore faced with a significant amount of manual re-adjustment just at a time when other organisational matters were at a peak. For these reasons, the availability of a user-friendly computer program that could produce teams that satisfied their requirements was a major help.

The core of this program is a linear programming assignment model which assigns players to positions for round 1 and round 2 in turn. However, in either round, this can result in a solution where there is an imbalance in the abilities of the various teams, ie. most of the good players can end up in the same team. Since the organisers used the matches to assess the ability of the players they did not wish this to happen and wanted teams to be of comparable strength. A smoothing algorithm which maintained the optimality of the solution was therefore added to reduce the imbalance between teams.

In the following, the formulation of the core linear programming assignment model (as applied to round 1) is explained first. A numerical example illustrating the working of the core model for a seven player - seven position problem is given. The adaptations required to handle the round 2 rules are then explained and this is followed by a description of the smoothing algorithm. The paper concludes with a short discussion of how alternative equally good solutions were generated. This additional facility was requested by the organisers once confidence in the initial program was achieved.

3. MODEL FORMULATION

In the general case, N players are assigned to $M = INT(N/7)$ teams with $E = N - 7*M$ emergencies left. The problem can be formulated as an unbalanced linear programming assignment problem which is balanced by the creation of E emergency positions. When $E = 0$, the problem is already balanced.

The formulation is as follows:

Number the players 1, 2, ... N.

Number the positions 1, 2, ... N with team k consisting of positions 7k – 6, 7k – 5, ..., 7k and the E emergencies being positions N – E + 1, N – E + 2, ..., N.

Then

$$\text{minimise } Z = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{ij} X_{ij}$$

where X_{ij} = 1 if player i is assigned to position j,
= 0 otherwise,

P_{ij} = a numerical score reflecting both the rank of the player and the preference of player i for position j (calculation of this score is described in more detail below)

subject to

$$\sum_{i=1}^{N} X_{ij} = 1 \text{ for } j = 1, 2, \ldots\ldots, N \tag{1}$$

$$\sum_{j=1}^{N} X_{ij} = 1 \text{ for } i = 1, 2, \ldots\ldots, N \tag{2}$$

Constraints (1) and (2) are the usual constraints found in the linear programming assignment problem. For this problem, they ensure the one-to-one assignment of players to positions.

4. CALCULATION OF NUMERICAL SCORES REFLECTING RANK AND PREFERENCES

The model formulated above allows players of all ranks to "compete" with players of all other ranks for positions. This is in contrast to the manual solution method which takes players one rank at a time.

It would be possible to apply the above model to each group of ranked players separately, starting at rank 1 and "removing" the positions assigned to these players from the problem before the next rank of players is assigned their position. However, it was decided that this was a bit cumbersome and so a set of numerical scores was developed for each player's preferred positions which would ensure, as far as possible, that the manual rules were observed.

It was also recognised by the organisers that some of the rules existed because of the manual nature of the team formation process and that, in some circumstances, these would not be strictly enforced as the manual process proceeded. It was agreed that, provided the underlying objectives of the rules were achieved by an automatic process, this would be acceptable.

The logic behind the development of the numerical scores is as follows:

Let the scores assigned to the 3 positions preferred by a player of rank r be

$$a_r\ ,\ a_r + d_r \text{ and } a_r + 2d_r.$$

It is assumed here that the "disappointment" experienced by a player when allocated her 2nd preference rather than her 1st is the same as when allocated her 3rd preference rather than her 2nd.

In addition, let the numerical score assigned to the 4 positions not expressed as preferences by a player of rank r be D_r, where D_r is some value of an order of magnitude greater than ar + $2d_r$, and is sufficiently high to dissuade allocation of an emergency position except "in emergencies".

In the Hungarian method of solution of assignment problems (Kuhn [1]), the first step is to subtract the minimum value in each row of the objective function matrix from all other values in that row. If we consider players as rows and positions as columns, the minimum value in each row will be a_r. Subtraction of this from the other values will result in values of 0, d_r and $2d_r$ for the preferred positions and $D_r - a_r$ for each of the positions not preferred. It is sufficient therefore to consider only what are appropriate values for the d_r and the set of differences $(D_r - a_r)$.

Taking the d_r first, it can immediately be seen that to ensure that a player of rank r is given her first preference before a player of rank r + 1, we must have

$$d_r > d_{r+1}\ , 1 \le r \le 3.$$

However, reference to the example illustrated in Table 1, where three players are competing for three positions reveals that, in fact, we must have $d_r > 2\, d_{r+1}$, since otherwise a solution other than the one desired would be obtained. In this example, player 1 (of rank r) is competing with players 2 and 3 (both of rank r+1). The desired assignments are indicated by asterisks (*).

To ensure player 1 is assigned position 1, we must have

$$3d_{r+1} < d_r + d_{r+1}$$

$$\text{ie. } d_r > 2\, d_{r+1} \tag{3}$$

Table 1. Player/position preference scores (* indicates desired assignments).

			Position		
		Rank	1	2	3
	1	r	*0	d_r	$2d_r$
Player	2	r + 1	0	*$2d_{r+1}$	d_{r+1}
	3	r + 1	0	$2d_{r+1}$	*d_{r+1}

For 3 players and 3 positions, there are a number of different scenarios that could exist for player/position preference scenarios. The one illustrated in Table 1 requires the largest d_r/d_{r+1} scaling factor. Extending such an investigation to N players and N positions becomes extremely complicated and in the end an arbitrary scaling factor for the d_r's was used, as will be explained at the end of this section.

Turning now to the determination of suitable values for the set of differences $(D_r - a_r)$, it would be attractive from a coding point of view to have $D_r = D \; \forall r$, for then the matrix of scores could be initialised to this value and only the scores corresponding to the preferred positions changed as the data is read in.

To ensure that a player of rank r is assigned at least her 3rd choice in preference to any player of rank r+1, we must have $D - a_{r+1} < D - (a_r + 2d_r)$, where the term on the left of the inequality reflects the fact that some players of rank r+1 could have given the position being contested as their first preference.

To achieve this, it is sufficient to set
$$a_{r+1} = a_r + 3d_r \, , r > 0,$$
$$a_1 = 0.$$

Thus, all that remains is to decide on the specific values of d_r to be used. If we wish, for neatness sake, to keep all the d_r values integer, it follows from (3) that the minimum values that they can take are $d_4 = 1, d_3 = 3, d_2 = 9, d_1 = 27$.

In fact, the values used in the analysis were

$$d_4 = 1, d_3 = 10, d_2 = 100, d_1 = 1000.$$

These values were chosen for two reasons. Firstly, it made examination of the solution easier and secondly it gave peace of mind in case there were any scenarios that had not been considered when investigating possible player/position preference score combinations. Examination of the solutions produced so far have not revealed any undesirable features.

The complete set of a_r and d_r values used in the model is shown in Table 2.

Table 2. Numerical scores used to generate the P_{ij} matrix.

	Preferred positions		
Rank	First	Second	Third
1	0	1000	2000
2	3000	3100	3200
3	3300	3310	3320
4	3330	3331	3332

5. NUMERICAL EXAMPLE

The following illustrates the use of the model to assign 7 players to 7 positions. Table 3 shows the ranks, preferred positions (in order) and the numerical scores assigned these positions for each of the 7 players. Positions which are not preferred are given a numerical score of 99999 but these values are not shown in the table. In the example, no player has a first preference for WA or C, while 2 players have a first preference for GS and 2 for GA.

Table 3. Matrix of player and position scores for example problem.

Player	Rank	Preferred positions	GS	GA	WA	C	WD	GD	GK
			Position scores						
1	1	GS, GA, WA	0	1000	2000	–	–	–	–
2	1	GA, C, GD	–	0	–	1000	–	2000	–
3	2	WD, GD, C	–	–	–	3200	3000	3100	–
4	2	GK, WD, WA	–	–	3200	–	3100	–	3000
5	3	GA, WA, C	–	3300	3310	3320	–	–	–
6	3	GS, C, GD	3300	–	–	3310	–	3320	
7	4	GD, C, GS	3332	–	–	3331	–	3330	–

The optimal solution is found and is shown in Table 4. Both the rank 1 players and both the rank 2 players get their first preferences. Both rank 3 players only get their 2nd preferences. This is because player 5's first preference (GA) and player 6's first preference (GS) have already been assigned to better players. Player 7, the only rank 4 player, is lucky. No-one else has given GD as a first preference and, although 3 players have given GD as a 2nd or 3rd preference, they have all been assigned a position higher in their order of preferences, so player 7 gets her first preference.

Should player 6's preferences be changed to GS, GD and C in that order, ie. GD is now her 2nd preference rather than C, then the model gives the GD position to player 6 in preference to player 7 who in turn is now given her 2nd preference (C).

Table 4. Optimal solution for example problem.

Player	Rank	Preferred positions	GS	GA	WA	C	WD	GD	GK
			Position scores						
1	1	GS, GA, WA	0	1000	2000	–	–	–	–
2	1	GA, C, GD	–	0	–	1000	–	2000	–
3	2	WD, GD, C	–	–	–	3200	3000	3100	–
4	2	GK, WD, WA	–	–	3200	–	3100	–	3000
5	3	GA, WA, C	–	3300	3310	3320	–	–	–
6	3	GS, C, GD	3300	–	–	3310	–	3320	
7	4	GD, C, GS	3332	–	–	3331	–	3330	–

6. ROUND 2

All that is required to be able to apply the model to round 2 is to change the ranks and preferences of the players to reflect the round 2 rules.

Thus, seven ranks are required for round 2 – reflecting the seven round 2 rules (round 2 rule 4 in reality consists of four rules, one for each of the original ranks) and those players assigned a round 2 rank of 4, 5, 6 or 7 will have one of their preferences removed (the position they were allocated in round 1).

Once these modifications have been made, the model can be re-applied to the new set of data.

7. Smoothing Algorithm

A measure of the collective ability of a team can be obtained by summing the ranks of the seven players allocated positions in that team (ie. excluding any emergencies) to obtain a total team rank. Without smoothing, unacceptably large variations can be obtained in the total team ranks.

Some thought was given to modifying the model in order to even out total team ranks by incorporating a quadratic term in the objective function to express the sum of squared deviations of total team ranks from the average. However, incorporation of such a non-linear term would have greatly increased computation time. An alternative method was therefore used whereby smoothing was performed once an optimal allocation of players to positions had been obtained. This smoothing algorithm did not change the position that a player was allocated but could change their team, thus maintaining the optimality of the solution. It has proved quite acceptable in finding solutions where total team ranks are either the same or within one of each other. It is described below.

Step 1: Calculate all team ranks.

Step 2: Identify the teams with the largest and smallest team ranks (teams A and B respectively).

Step 3: Calculate the difference, d, in the team ranks of teams A and B, and if $d \leq 1$ stop, otherwise

Step 4: Calculate $D = \min\{\text{int}(d/2), 3\}$

Step 5: Taking each position in teams A and B in turn, look for a rank difference equal to D. Here we are hoping to find a swap which will even out the team ranks of teams A and B. The maximun difference in rank that can occur in round 1 is 3 (this is replaced by a value of 6 when smoothing round 2 teams).

Step 6: If such a rank difference is found, swap the two players between teams A and B (retaining their positions), recalculate the team ranks of teams A and B and then return to Step 2.

If not, reduce D by 1 and, unless D = 0 (see below), return to Step 5.

It is possible, during the search for a suitable position on which to perform a swap, that the value of D reaches 0. This occurs, for instance, when 6 of the 7 positions in teams A and B have equal rank but there is a difference of 2 (or 3) in the rank of the 7th player. Thus, here d = 2 (or 3), the initial value of D is 1 and no position is found where this rank difference occurs. To overcome this problem, the algorithm replaces team A by the team with the next highest team rank and repeats from Step 3. A flow chart illustrating the working of the algorithm is shown in Fig. 1 and the stage-by-stage smoothing of a sample set of five teams is shown in Table 5.

8. ALTERNATIVE SOLUTIONS

From a strict linear programming point of view, there are a multitude of alternative optima for each round, since for each position in an n-team problem there will be n! ways in which the n players who have been allocated to this position in one optimal solution can then be allocated to teams.

However, the organisers did not wish alternative solutions in which players retained their positions (but swapped teams). They wanted alternative solutions in which players were given different positions. By experimentation, it was found that, if the data was sorted in different ways prior to feeding into the model, the sort of alternative solutions the organisers wanted could be generated. The programme handed over to the organisers allowed them to select one of four ways in which to feed the data to the model

in order of initial storage in the datafile,
in alphabetical order of first name,
in alphabetical order of second name,
in order of rank (and within rank, in original stored order).

As a rule, selecting all four options one after the other would result in two or three alternative solutions being generated. Typically, most of the player-position allocations in these solutions would be identical and contain only a small number of permutations of two or three player-position allocations.

Alternative solutions are generated for both rounds 1 and 2 in this way. Before calculating any solution for round 2, a method for calculating round 1 has to be selected, since the ranks and preferences used in round 2 depend on the solution chosen in round 1.

9. PROGRAM

The program delivered to the organisers was written in Visual Basic v3.0 to allow ease of input and editing of data. Forms were also created to allow selection of which round to provide solutions for and which alternative method to use to produce a solution. Command buttons were provided to allow output of the solution to a printer or a file.

10. CONCLUSION

As a result of this research, the organisers were provided with a user-friendly computer program which achieved their objectives in a fraction of the time that the manual method took. They could also use the program to cater for last-minute changes that occurred to the database.

REFERENCE

[1] Kuhn H.W., The Hungarian method for the assignment problem. Nav. Res. Logist. Q. 2, 83-97 (1955).

ACKNOWLEDGEMENT

My thanks to Mike Witomski, formerly Netball Convener for the VPSSA, who brought the problem to my attention, provided sample data for it and was always happy to provide feedback as the model was developed.

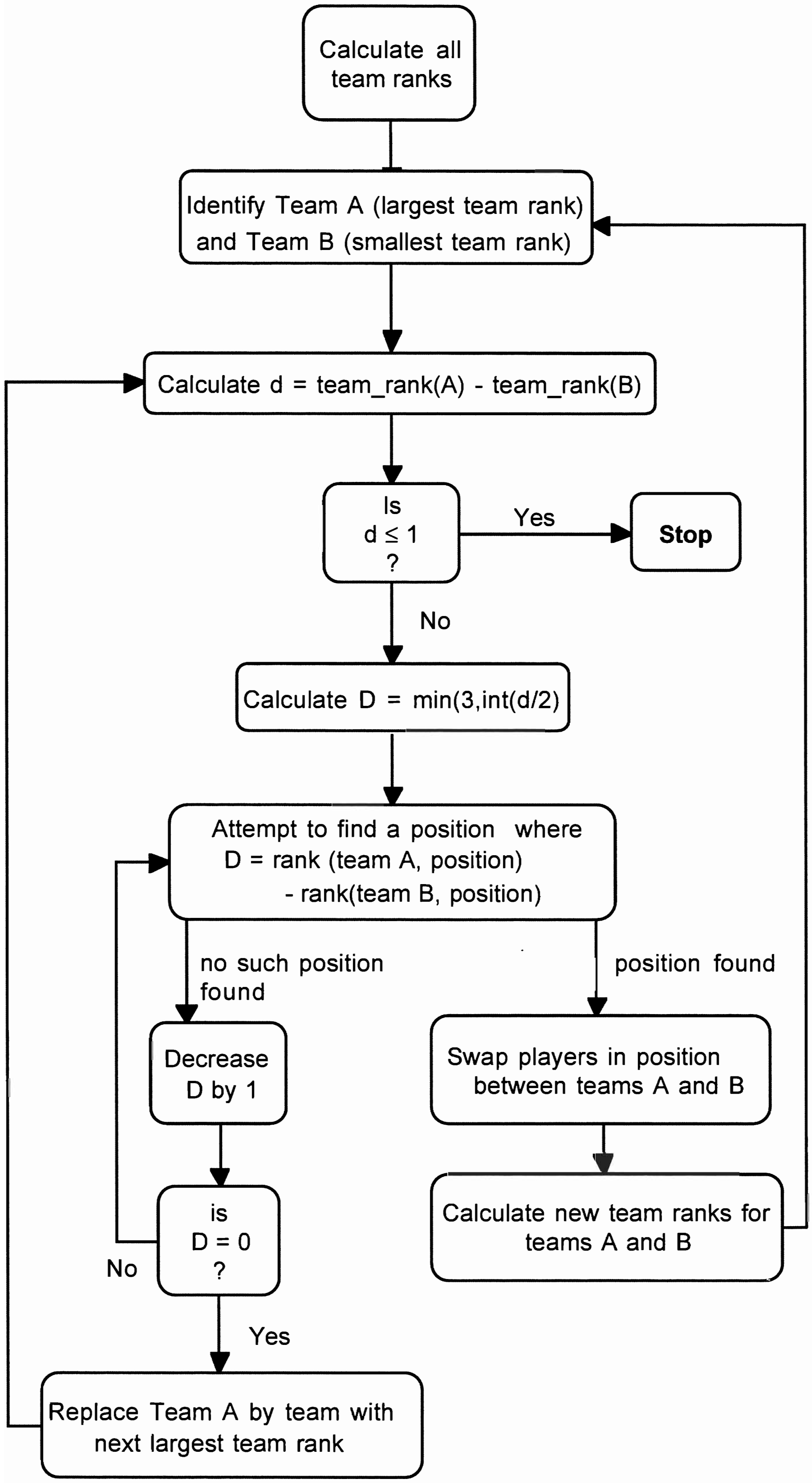


Figure 1. Flowchart illustrating the smoothing algorithm.

Table 5. Stage-by-stage illustration of the smoothing algorithm.

Initial solution.

Position	Team 1	Team 2	Team 3	Team 4	Team 5
GS	4	3	2	2	3
GA	3	2	1	4	4
WA	3	4	3	3	2
C	4	3	2	2	4
WD	3	1	4	3	1
GD	3	3	1	1	2
GK	2	4	2	3	4
Team rank	22	20	15	18	20

team A = team 1,
team B = team 3,
d = 7,
For D = 3, no position is found,
For D = 2, position GS is found,
Swap on position GS to form
new solution.

Stage 2 solution.

Position	Team 1	Team 2	Team 3	Team 4	Team 5
GS	2	3	4	2	3
GA	3	2	1	4	4
WA	3	4	3	3	2
C	4	3	2	2	4
WD	3	1	4	3	1
GD	3	3	1	1	2
GK	2	4	2	3	4
Team rank	20	20	17	18	20

team A = team 1,
team B = team 3,
d = 3,
For D = 1, no position is found
so team A = team 2,
For D = 1, position GA is found,
Swap on position GA to form
new solution.

Stage 3 solution.

Position	Team 1	Team 2	Team 3	Team 4	Team 5
GS	2	3	4	2	3
GA	3	1	2	4	4
WA	3	4	3	3	2
C	4	3	2	2	4
WD	3	1	4	3	1
GD	3	3	1	1	2
GK	2	4	2	3	4
Team rank	20	19	18	18	20

team A = team 1,
team B = team 3,
d = 2,
For D = 1, position GA is found,
Swap on position GA to form
new solution.

Stage 4 solution.

Position	Team 1	Team 2	Team 3	Team 4	Team 5
GS	2	3	4	2	3
GA	2	1	3	4	4
WA	3	4	3	3	2
C	4	3	2	2	4
WD	3	1	4	3	1
GD	3	3	1	1	2
GK	2	4	2	3	4
Team rank	19	19	19	18	20

team A = team 5,
team B = team 4,
d = 2,
For D = 1, position GS is found,
Swap on position GS to form
new solution.

Final solution.

Position	Team 1	Team 2	Team 3	Team 4	Team 5
GS	2	3	4	3	2
GA	2	1	3	4	4
WA	3	4	3	3	2
C	4	3	2	2	4
WD	3	1	4	3	1
GD	3	3	1	1	2
GK	2	4	2	3	4
Team rank	19	19	19	19	19

HOME ADVANTAGE IN BALANCED COMPETITIONS - ENGLISH SOCCER 1991-1996

Stephen R Clarke[1]

Abstract

This paper discusses the calculation of home advantage for individual clubs in any competition where each team plays each other at home and away. Simple formulae can be applied to the final ladder to produce home advantages for the individual teams. The method is used on a spreadsheet with data obtained from the internet. Home advantages of all clubs in the English soccer league from 1991-1996 are calculated and compared with previously published figures.

1. INTRODUCTION

Home advantage is often discussed in sporting circles but not often calculated. While the overall home advantage of the whole competition is often measured by the percentage of matches won by the home teams, this is not an appropriate measure for individual clubs. To correctly calculate the home advantage of individual clubs strength of opposition must be allowed for, and is often difficult to assess because the draw is not balanced. (AFL football is a case in point where in 1995 and 1996 16 teams play 22 rounds. Furthermore, ground sharing is common, and many matches are moved to the MCG to allow for anticipated large crowds. This results in other matches being moved to accommodate ground tenancy agreements). In such cases, mathematical models need to be fitted to estimate team ability and home advantage. Stefani and Clarke [1] do this for Australian Rules, Kuk [2] for English soccer and Harville and Smith [3] for American basketball. However in cases such as English soccer, where the draw is balanced so that each team plays each other team once at home and once away, the analysis can be much simpler. In this case it makes sense to separate home and away results, and this is the usual practice when presenting English soccer tables. In this case, Clarke and Norman [4] give a method that although equivalent to fitting a model to the individual match scores by least squares, can be applied using simple arithmetic to the final ladder. This method is applied to all results from 1991-92 to 1995-96, and results compared to previous results.

2. THE METHOD

The idea behind the method is best demonstrated by a simple example. Suppose there are 4 teams and the results of their matches are as in table 1. Margin totals for home and away matches are also shown. Note that as all results are given as the

---

[1]School of Mathematical Sciences, Swinburne University, Hawthorn Vic 3122

home team first, the column totals are in fact the negative of the respective team's away performance.

Table 1 Sample results

		Away Team				Total
		1	2	3	4	Margin
	1	–	3-2	4-0	6-1	10
Home	2	1-1	–	2-1	3-0	4
team	3	1-0	2-1	–	5-1	6
	4	1-3	1-2	0-2	–	-5
Total Margin		-1	1	3	12	15

We model the winning margin w_{ij} when home team i plays away team j as

$$w_{ij} = u_i + h_i - u_j + e_{ij}$$

where u_i is a measure of a team's ability, h_i is team i's home advantage and e_{ij} is random error. Since the u_i are relative, we can require they sum to zero. We wish to find the eight unknowns u_i , h_i . Clarke & Norman show that fitting a least squares model to the original results gives the same result as fitting the expected values to the marginal totals.

For example replacing the score in the above table with the expected margin we obtain

Table 2: Expected results

		Away Team				Total
		1	2	3	4	Margin
	1	–	$u_1+h_1-u_2$	$u_1+h_1-u_3$	$u_1+h_1-u_4$	10
Home	2	$u_2+h_2-u_1$	–	$u_2+h_2-u_3$	$u_2+h_2-u_4$	4
team	3	$u_3+h_3-u_1$	$u_3+h_3-u_2$	–	$u_3+h_3-u_4$	6
	4	$u_4+h_4-u_1$	$u_4+h_4-u_2$	$u_4+h_4-u_3$	–	-5
Total Margin		-1	1	3	12	15

Adding all results we get 3 ($h_1 + h_2 + h_3 + h_4$) = 15, so H = $h_1 + h_2 + h_3 + h_4 = 5$.

Using the difference between team 1's home and away performance (ie adding first column and first row) we get $H + 2h_1 = 9$ so $h_1 = 2$. For the other teams we get $h_2 = 0$, $h_3 = 2$, and $h_4 = 1$.

Now for team 1's home results we have $4u_1 + 3h_1 = 10$ so $u_1 = 1$. Similarly the ratings for the other teams are 1, 0, and -2.

Note that team 2 has no home advantage, even though it has better results at home than away. These are in fact due to the home advantage of the other teams.

In general, the formulae for N teams are

H = total home goal difference of all teams $/(N\text{-}1)$
h_i = (home goal diff for team i - away goal diff for team i - H)/(N-2)
u_i = (Home goal difference for team i - $(N\text{-}1)h_i$)$/N$

These formulae are very easy to apply, particularly for English soccer where home and away results are traditionally separated in the end of year ladder.

3. ENGLISH SOCCER 1991-1996

The final ladder results are archived on the internet from 1991 onwards. These were copies and read into an XL spreadsheet. This spreadsheet was created using the same form as the archive, with the above formulas used to calculate two additional columns containing the u_i and h_i. It was only necessary to cut and paste a year's results from the archive file to produce the us and hs. This was done for each division and year. The result for the premier division of 1995-96 is shown in Table 3.

Table 3: 1995-96 Premier division

Team	P	W	D	L	F	A	W	D	L	F	A	Pts	U	h
Manchester U	38	15	4	0	36	9	10	3	6	37	26	82	0.98	0.39
Newcastle U	38	17	1	1	38	9	7	5	7	28	28	78	0.40	1.11
Liverpool	38	14	4	1	46	13	6	7	6	24	21	71	0.54	1.16
Aston Villa	38	11	5	3	32	15	7	4	8	20	20	63	0.43	0.44
Arsenal	38	10	7	2	30	16	7	5	7	19	16	63	0.60	0.11
Everton	38	10	5	4	35	19	7	5	7	29	25	61	0.64	0.16
Blackburn R	38	14	2	3	44	19	4	5	10	17	28	61	-0.17	1.50
Tottenham H	38	9	5	5	26	19	7	8	4	24	19	61	0.72	-0.39
Nottingham F	38	11	6	2	29	17	4	7	8	21	37	58	-0.40	1.05
West Ham U	38	9	5	5	25	21	5	4	10	18	31	51	-0.22	0.44
Chelsea	38	7	7	5	30	22	5	7	7	16	22	50	0.14	0.27
Middlesbr.	38	8	3	8	27	27	3	7	9	8	23	43	-0.31	0.33
Leeds United	38	8	3	8	21	21	4	4	11	19	36	43	-0.42	0.44
Wimbledon	38	5	6	8	27	33	5	5	9	28	37	41	0.02	-0.34
Sheffield Wed	38	7	5	7	30	31	3	5	11	18	30	40	-0.15	0.11
Coventry C	38	6	7	6	21	23	2	7	10	21	37	38	-0.36	0.27
Southampton	38	7	7	5	21	18	2	4	13	13	34	38	-0.64	0.83
Manchester C	38	7	7	5	21	19	2	4	13	12	39	38	-0.95	1.11
Queens P R	38	6	5	8	25	26	3	1	15	13	31	33	-0.47	0.44
Bolton Wand	38	5	4	10	16	31	3	1	15	23	40	29	-0.38	-0.39
Totals		186	98	96	580	408	96	98	186	408	580	1042	0.00	9.05

Table 4: HAs of all teams 91-92 to 95-96

Team	91-92	92-93	93-94	94-95	95-96	Avge
Arsenal	0.67	-0.14	-0.56	-0.07	0.11	0.00
Aston Villa	0.87	0.61	0.39	0.08	0.44	0.48
Barnet	1.01	1.30	0.64	0.92	0.57	0.89
Barnsley	0.18	0.56	-0.09	1.02	0.44	0.42
Birmingham City	0.49	0.29	0.13	0.64	1.03	0.52
Blackburn Rovers	0.59	0.36	0.34	0.78	1.50	0.71

Blackpool	1.71	0.92	0.42	0.23	0.16	0.69
Bolton Wanderers	0.04	0.78	0.31	1.16	-0.39	0.38
Bournemouth	0.63	0.15	-0.22	0.14	1.07	0.35
Bradford City	-0.06	0.46	0.42	-0.36	0.84	0.26
Brentford	0.63	0.15	-0.40	0.32	0.57	0.25
Brighton & Hove A	0.41	0.37	0.64	0.45	-0.03	0.37
Bristol City	0.82	0.65	0.41	0.21	0.25	0.47
Bristol Rovers	0.96	-0.17	0.10	0.77	-0.30	0.27
Burnley	0.31	1.10	1.92	0.84	0.66	0.97
Bury	0.31	0.90	0.80	0.32	-0.15	0.44
Cambridge United	0.09	0.15	0.05	0.82	0.39	0.30
Cardiff City	0.46	0.30	0.64	0.32	0.80	0.50
Carlisle United	0.51	0.50	0.10	-0.23	1.52	0.48
Charlton Athletic	-0.54	0.15	0.86	0.52	-0.47	0.10
Chelsea	0.17	0.31	0.99	0.08	0.27	0.36
Chester City	0.04	0.33	0.20	0.00	0.80	0.27
Chesterfield	-0.09	0.20	0.30	-0.08	0.88	0.24
Colchester United	•	1.25	0.20	-0.13	0.48	0.45
Coventry City	0.32	-0.19	0.39	0.23	0.27	0.20
Crewe Alexandra	0.06	1.05	0.20	0.23	0.16	0.34
Crystal Palace	-0.13	0.31	0.13	-0.42	-0.15	-0.05
Darlington	0.26	-0.95	0.35	0.37	-0.43	-0.08
Derby County	-0.27	-0.40	0.95	0.66	1.08	0.40
Doncaster Rovers	-0.64	-0.25	0.10	-0.38	0.62	-0.11
Everton	0.42	-0.54	0.34	0.68	0.16	0.21
Exeter City	1.31	-0.22	1.01	0.17	0.16	0.49
Fulham	0.44	-0.08	-0.18	0.97	1.03	0.44
Gillingham	1.21	0.80	0.40	1.07	0.71	0.84
Grimsby Town	-0.04	0.15	0.13	0.71	0.44	0.28
Halifax Town	0.36	-0.75	•	•	•	-0.20
Hartlepool United	0.26	-0.08	0.51	0.97	1.03	0.54
Hereford United	0.86	0.65	0.70	0.72	0.39	0.66
Huddersfield Town	0.40	0.51	-0.27	0.41	1.21	0.45
Hull City	-0.10	0.69	0.32	1.00	0.38	0.46
Ipswich Town	0.46	0.41	-0.26	1.38	0.53	0.50
Leeds United	0.22	2.01	0.29	0.48	0.44	0.69
Leicester City	0.77	0.88	0.22	0.38	-0.38	0.37
Leyton Orient	0.63	1.37	1.23	0.91	1.07	1.04
Lincoln City	-1.09	0.50	-0.10	0.82	0.85	0.20
Liverpool	0.92	1.41	0.49	0.63	1.16	0.92
Luton Town	1.97	0.10	0.81	0.02	0.35	0.65
Maidstone United	0.26	•	•	•	•	0.26
Manchester City	0.72	-0.29	0.44	0.98	1.11	0.59
Manchester United	0.17	0.16	0.19	0.83	0.39	0.35

Mansfield Town	0.11	0.74	-0.10	0.12	-0.34	0.11
Middlesbrough	0.96	1.11	0.81	0.21	0.33	0.68
Millwall	-0.13	1.20	0.77	0.71	0.08	0.53
Newcastle United	1.09	0.92	1.34	1.13	1.11	1.12
Northampton	0.21	0.00	0.95	0.27	0.16	0.32
Norwich City	0.47	0.86	-0.81	0.98	-0.38	0.22
Nottingham Forest	0.37	-0.29	-0.23	-0.12	1.05	0.16
Notts County	0.17	1.24	1.18	0.21	0.29	0.62
Oldham Athletic	0.77	1.31	0.09	0.61	0.62	0.68
Oxford United	0.68	0.33	0.41	0.05	1.25	0.54
Peterborough United	0.76	0.20	0.81	0.00	0.97	0.55
Plymouth Argyle	0.55	0.60	-0.13	0.05	0.62	0.34
Port Vale	0.14	0.33	0.55	0.25	0.03	0.26
Portsmouth	1.55	1.47	0.36	0.16	0.17	0.74
Preston North End	0.85	0.24	1.00	0.72	-0.25	0.51
Queen's Park Rangers	-0.08	0.01	-0.06	0.43	0.44	0.15
Reading	0.13	0.96	0.01	-0.02	-0.15	0.19
Rochdale	0.51	0.50	0.65	0.92	-0.34	0.45
Rotherham United	0.06	-0.26	0.46	0.68	0.84	0.36
Scarborough	0.81	0.05	0.05	0.12	0.39	0.28
Scunthorpe United	1.31	0.75	0.65	0.32	-0.15	0.58
Sheffield United	0.07	0.81	0.69	0.39	-0.15	0.36
Sheffield Wednesday	0.42	0.06	0.99	-0.07	0.11	0.30
Shrewsbury Town	0.04	-0.05	-0.45	0.59	0.29	0.08
Southampton	-0.73	0.71	0.44	0.23	0.83	0.30
Southend United	0.55	0.92	0.18	1.02	0.76	0.69
Stockport County	0.90	1.01	0.69	0.45	-0.25	0.56
Stoke City	0.44	0.24	1.00	0.48	0.67	0.57
Sunderland	0.91	0.65	0.77	-0.39	0.44	0.48
Swansea City	0.90	0.56	1.14	-0.05	0.93	0.70
Swindon Town	0.37	0.42	0.34	0.39	-0.30	0.24
Torquay United	1.54	-0.10	-0.15	0.72	0.30	0.46
Tottenham Hotspur	-0.38	1.26	-0.46	-0.17	-0.39	-0.03
Tranmere Rovers	0.00	0.92	1.00	1.57	0.62	0.82
Walsall	0.21	-0.05	0.10	0.22	0.43	0.18
Watford	-0.41	-0.17	0.45	0.61	0.62	0.22
West Bromwich A	0.58	1.01	0.50	0.52	0.44	0.61
West Ham United	0.47	0.65	-0.26	0.63	0.44	0.39
Wigan Athletic	0.81	0.06	0.60	-0.28	0.66	0.37
Wimbledon	0.77	0.36	0.94	0.38	-0.34	0.42
Wolverhampton W	0.27	0.42	0.22	0.61	0.44	0.39
Wrexham	1.06	0.65	1.10	0.55	0.70	0.81
Wycombe Wanderers	•	•	0.25	0.50	0.20	0.32
York City	0.61	0.85	0.23	0.32	0.07	0.42

The columns for year, division, club, team rating and home advantage were then copied into a single spreadsheet and to SAS/JMP for further analysis. Clarke and Norman [4] gave the home advantage (HA) for all clubs for 1981-82 to 1990-91. Table 4 extends that through to 1995-96.

In general the results as reported by Clarke and Norman for the years 81-82 to 90-91 are repeated here. The average HA was 0.43 goal per match. The yearly HAs ranged from –1.1 to 2.0 goals per match, with about 18% negative. However the average HA over 5 years ranged from –0.11 to 1.1. Analysis shows that HA is not dependant on division, nor year. However the team effect is significant ($p = 0.0386$). Of the 10 clubs with the lowest HA, five are London clubs. Clearly the mean HA of 0.29 for the 13 London clubs is significantly lower than the mean HA of 0.44 for the 81 non-London clubs. A surprising fact was the lack of consistency in the HAs from one year to the next. The correlations between HA from one year to the next were very small or even negative, while the correlation between the average HAs obtained by Clarke & Norman and those here was only 0.15. This suggests that teams do not enjoy a large HA over many years, and that opponents may quickly counteract perceived HAs

4. CONCLUSION

Calculation of individual team home advantages is rarely undertaken as it usually involves complicated statistical fitting of mathematical models to individual results. However in the case of a balanced competition, this is not necessary and can be accomplished by simple calculations on the final ladder. This should be done on a range of sports, as the reasons for variations in HA can then be investigated.

REFERENCES

[1] Stefani R.T. and Clarke S.R., Australian Rules Football during the 1980's. *ASOR Bulletin,*, 10 (3) 11-15 (1991)

[2] Kuk A.Y.C., Modelling paired comparison data with large numbers of draws and large variability of draw percentages among players. *The Statistician*,, 44 (4) 523-528 (1995)

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

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

THE "LEVELLER SYSTEM" TENNIS TOURNAMENT

Pam Norton[1]

Abstract

The leveller system tennis tournament is one which allows all players to participate for its duration, and to find and play at their own level. The system is examined for its potential to be computer-generated. The question to be answered is: "What is the best method to generate the draw for a session given the results of previous sessions if any?" Preliminary analysis only has been done.

1. INTRODUCTION

The "leveller system" tournament was developed by Ken Snell, formerly of Mildura, to give country juniors as much competition tennis as possible. This type of tournament has now been run by Waverley and Districts Tennis Association (WDTA), in Melbourne's southeastern suburbs, over three separate school holidays, with more planned. The tournament is described as "a multi match tournament draw which allows all players, irrespective of age or ability, to participate for its duration". The original idea is for good players in a particular age group to be able to consistently play players of a higher age group, for girls to be able to play against boys and vice versa, and for all players to play some players of a similar standard. The "normal" standard junior tennis tournament is a single sex age group event, sometimes being a round robin. If it is a knockout event, then half the entrants have one match and have then finished. Round robin events usually last up to a day, and the participants get probably four or five matches. So in a leveller tournament every player gets to play from start to finish, regardless of their standard.

2. THE LEVELLER SYSTEM

The leveller system tournaments that have been run by WDTA have been held over five days, from Monday to Friday, during school holidays. They could well be marketed as a school holiday program! For the first four days singles matches are played, with doubles on the Friday. The entry says "partners will be allotted by the venue coordinator"; in practice, the venue coordinator asks the entrants to get their own partner and let him/her know.

The main idea of the leveller system is that there are one or two preliminary matches which sort the entrants into different groups, and then the players participate in a round robin within the group. Prizes are awarded to the winners of each group. For example, the simplest type would be a draw for eight people. The eight entrants

[1] Department of Mathematics, Monash University, Clayton, Victoria 3168

would play one match; the winners would go into the draw for the Group A round robin, and the losers to the Group B round robin. A draw sheet for such is given in Figure 1. The final position of each competitor is determined by the number of games obtained in the round robin, and all players in Group A will be positioned above all players in Group B.

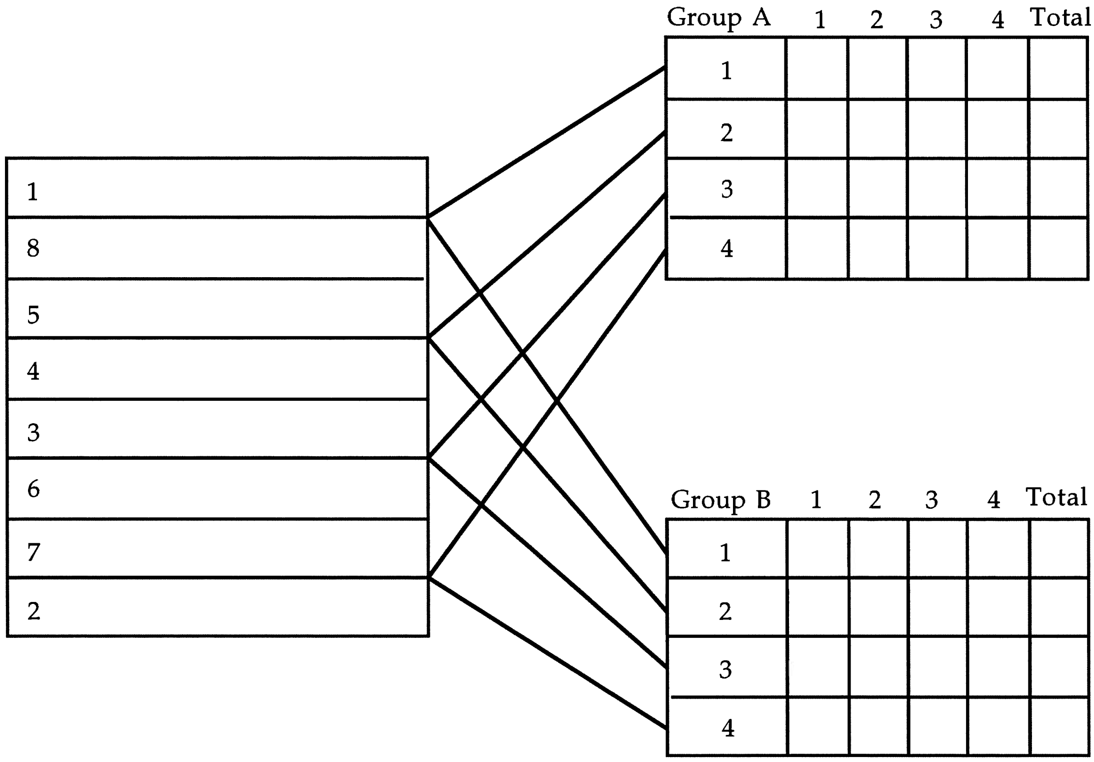


Figure 1: *Draw sheet for Leveller System, 8 entrants.*

The leveller system can be designed for any number of players; in practice, 40 players is close to the maximum. With forty players, there would be two preliminary rounds to sort the players into four groups of ten, and then there would be two round robins of five players each in each group, with the winners playing off. The four sorted groups correspond to win-win, win-loss, loss-win and loss-loss in the playoff rounds.

The first two days are defined as seeding days, whereby a specific cohort plays out a leveller system in order to produce a ranked list of players. Traditionally, the first day is a single sex competition day, with groupings depending on the numbers of entries, but probably similar to 12 and under girls, 13 and over girls, 12 and under boys and 13 and over boys. On the second day, the sexes are mixed, with a groups consisting of, for example, 11 and under boys and girls, 12 and 13 year old boys and girls, and the rest.

The third and fourth days are the days for the main singles events. The groupings here, for example, are 11 and under boys and 12 and under girls, 12 and 13 year old boys and 13 and 14 year old girls, and the rest.

3. Method of Seeding

There does not appear to have been any consistency in the way the original draws have been produced. There are two examples now given, involving 12 and 13 year old boys in the April 1996 and July 1996 tournaments. For the April 1996 tournament, it is clear that initially there were eight seeds in this group, with positions determined by information such as association, junior grade, senior grade, pennant grade, and whether a McDonald's or VCTA squad member. These seeds were unchanged for the first two days. The results of day 2 were used to seed for the main singles competition, consisting of 24 players. The final day's play, a doubles competition, was seeded on the results of the main singles competition. The seeds for each session's play and the positions at the end of each session are given in Figure 2. An interesting mathematical question is: how many such sessions, where the players are ranked at the end of one and the rankings given as seedings for the next, would need to be held to end up with a correctly ranked list?

For the July 1996 tournament, it would appear to have been a random draw on the first two days, with the results of these, particularly the first day, used to list four seeds only for the main competition, which consisted of 30 players. The positions at the end of each session's play are given in Figure 3.

13 year old boys

Seed, session 1	Position, session 1	Seed, session 2	Position, session 2	Seed, session 3	Position, final
	9		3	3	1
1	5	1	5	5	2
5	2	5	1	1	4
4	4	4	3	4	7
3	7	3	?		8
	15		15		14
2	10	2	7	8	10

12 year old boys

Seed, session 1	Position, session 1	Seed, session 2	Position, session 2	Seed, session 3	Position, final
	1		2	2	3
6	3	6	8		4
8	13	8	6	6	6
	12		14		9
	6		12		11
7	8	7	9		12
	11		9		13
	14		13		15

Figure 2: *Seeds for each session and positions at end of each session, April 1996. Both groups in same draw for duration. (The seventh seed on day 3 was a girl.)*

Position, session 1	Position, session 2	Seed, session 3	Final position
12 year old boys			
1	2	4	5
2	12	3	2
3	13		12
4	10		16
5	17		7
6	16		15
7	16		6
13 year old boys			
1	8	2	3
2	6		4
3	11		18
4	5		8
5	1	1	1
6	14		14
7	3		9
8	3		10
9	7		11
10	19		17
11	9		13
12	18		19

Figure 3: *July, 1996 tournament. Positions at the end of each sessions play, with seeds for final draw. Groups played separately on day 1, but together for day 2 and final draw.*

Under both systems, there is quite a bit of variation in position across the three days. In both cases, in the final session, three of the top four seeds finished in the first four positions. (Compare the men's singles at Wimbledon this year!)

4. ENTRANTS

The number of entrants by sex and age for each of the tournaments so far played is given in the following table:

	12 & Under Girls	13 & Over Girls	12 & Under Boys	13 & Over Boys	Total
October 1995	18	18	25	28	89
April 1996	10	19	24	27	80
July 1996	10	14	34	25	83

The number of girls is declining; this could be looked into.

As for consistency, of the entrants in the April 1996 tournament, 7 girls had played in the previous one and 14 boys. Of the entrants in the July, 1996 tournament, 11 girls had played in the previous one and 20 boys (all aged 13 and under). Considering the weather Melbourne turned on in April, this was quite good. (The en-tout-cas venues were washed out frequently that week, whilst at the plexipave venue all players had extensive practice in court mopping skills, and virtually all matches were shortened.) Seven boys and four girls have played in all three.

5. Some Observations

- The top players in each age group from the district are not participating in these tournaments. Perhaps they are off playing in other tournaments for points.

- For the players who do participate, they seem to enjoy the format, and like playing a variety of players. The players were not afraid to say "But I've already played her/him".

- There are often turnarounds in performance. A player will beat another one day, and lose to the same player the next.

- It is a far more social event than other tournaments. Players are there for the whole day, and when they are off the court, for example, they may be lined up to take their turn at table tennis. They make many new friends.

- Prizes are now given to the winners of each group in each session. A few players are thought to have thrown matches to give themselves a better chance at a trophy.

6. Conclusions

The algorithm for producing one day's draw from the previous day's results to produce the "best" tournament is not clear at this stage. The number of players seeded for each session is another question that must be answered. The greater the degree of seeding the less the potential variability in performance and in variety of players met.

EXERCISE PHYSIOLOGICAL, BIOMECHANICAL AND KINANTHROPOMETRIC PREDICTORS OF BICYCLE MOTOR CROSS IN YOUNG ADULTS: A PRELIMINARY STUDY

Steve Politi [1] and Ian T. Heazlewood [2]

Abstract

Bicycle motor cross (BMX) racing is conducted on a 380 - 400 metre dirt track. Every track is different and includes jumps, sharp bends and straights. There are up to eight competitors on the track at one time with the objective being to cross the finish line first. There has been minimal published systematic empirical research that addressed the relationship between exercise physiological, biomechanical and kinanthropometric constructs and BMX performance. The purpose of this study was to identify the exercise physiological, biomechanical and kinanthropometric constructs that may predict BMX finishing time. A multivariate approach was used to examine the effects of numerous potential predictor constructs simultaneously that predicted BMX finishing time. It was postulated that the multivariate approach would provide more meaningful interpretations of BMX performance. Nineteen male BMX competitors of varying ability levels between the ages of 16 and 35 years volunteered to participate in the study. A set of independent variables, that included age, power production, work output, torque production, flexibility, maximal oxygen consumption, height, weight, lean body mass, sitting height, arm span, chest circumference, thigh circumference and calf circumference were measured and statistically analysed against the dependent variable, BMX finishing time. Descriptive, bivariate and multivariate statistical methods were used to analyse the data. The results identified that lean body mass, power production and work output were the best predictors of BMX finishing time. These findings enabled a set of predictive equations to be developed which have important implications for talent identification and the development training programs in the sport of BMX.

1. INTRODUCTION

Bicycle motor cross (BMX) commenced in Santa Monica California in July 1969. A group of children not old enough to ride motorcycles, came up with the concept of BMX, which is now an organised sport in thirty countries (Scott [1]). BMX racing is conducted on a 380 - 400 metre dirt track. There is no standard design for a BMX track. Every track is different in some way, making each track a new challenge for riders. Each track must have an eight lane start gate, jumps, straights and bermed turns, to ensure that it is suitable for BMX racing. Up to eight competitors are on the course during race conditions, and it is not uncommon for accidents to occur. The course takes between 30 and 35 seconds to complete for elite competitors and up to 45 seconds to complete for the less capable competitors.

[1] Australian Catholic University, Sydney.NSW 2135

[2] Australian Catholic University, Sydney NSW 2135

The Australian Sports Research Centre (Draper [2]) identified the need for research in the sport of BMX. Hinspeter, outlined the need for research in relation to physiological characteristics of elite BMX competitors in the Sports Research Needs 1992 document (Draper [2]).

Research on BMX competition to date, has been confined to the areas of injuries, participation, bike technology, accidents, uniform and clothing, marketing and building tracks. However, minimal research has been published on BMX performance that originated from sport science or exercise science. Training guidelines have been outlined in books on BMX, such as the BMX Handbook by Spurdens [3], but there is minimal evidence to support these training practices.

The literature that relates motor tasks, such as sprint cycling, with the relevant human bioenergetic systems (Brooks and Fahey [4]; deVries and Housh [5]; Hahn [6]; Roberts [7]; Schell and Leelarthaepin [8]), indicates that sprint cycling events require the breakdown of adenosine triphosphate (ATP) and creatine phosphate (CrP). Events lasting one to ten seconds (200 metre sprint event) rely predominantly on ATP - CrP energy production. Longer events up to one minute (1000 metre time trial) rely on both the ATP - CrP system and anaerobic glycolysis for energy production. The ATP-CrP and anaerobic glycolysis systems, are collectively referred to as the anaerobic systems. It has been postulated (Brooks and Fahey [4]; deVries and Housh, [5]; Hahn [6]; Roberts [7]) that the aerobic energy system is important for recovery between repeated sprint activities and may have relevance for the sport of BMX. The anaerobic systems are also postulated to the most important in the development of maximal human torque, work and power (Hahn [6]; Roberts [7]; Schell and Leelarthaepin [8]). However in the sport of BMX, the relevance of the various bioenergetic systems and the relationship of torque, work and power with BMX performance have not been empirically substantiated.

Due to the lack of published research in the sport and exercise sciences that explains BMX performance, coaches and competitors are basing their training programs on what successful BMX competitors are doing in training, this training may or may not be appropriate. There is a great need to establish exercise physiological, biomechanical and kinanthropometric profiles of BMX competitors that are more objective and quantifiable. Quantifiable data, based on ratio level measurements melded with substantive mathematical-statistical modelling should provide a more complete insight into factors that determine and can predict BMX performance.

Research findings that provide this information will enable the development of a predictive model for BMX performance, a model that will allow athletes and coaches to identify their exercise physiological, biomechanical and kinanthropometric strengths and weaknesses and develop specific training programs that will improve the components of fitness that best predict BMX performance. This approach should ultimately result in better individual performances and a higher standard of BMX competition.

Research Problem

It appears that there is miniumal direct information from sport science or exercise science that explains BMX performance and that has focused on exercise physiological, biomechanical or kinanthropometric constructs as a potential set of predictors. Inferences can be made from the literature that analyses sprint track cycling from sport and exercise science perspectives. That is, will those factors that predict track sprint cycling be similar to those factors that predict BMX track performance? The purpose of this study is to determine those exercise physiological, biomechanical and kinanthropometric constructs that best predict performance in BMX racing.

Research Questions

1. Do the kinanthropometric constructs that predict sprint track cycling (as mentioned previously), predict BMX performance?

2. Do the exercise physiological predictors of sprint track cycling, predict BMX performance?

3. Do the biomechanical predictor, such as maximum torque, work and power, and maximal torque, work and power divided by body weight that predict sprint cycling, predict BMX performance?

4. Can mathematical-statistical models be derived, based on linear regression, path analysis and nonlinear regression, and using exercise physiological, biomechanical and kinanthropometric constructs, that adequately predict BMX performance?

Research Hypotheses

1. The kinanthropometric characteristics of sprint track cyclists will be similar to those of BMX competitors.

2. Significant exercise physiological predictors of sprint track cycling performance will also be significant predictors of BMX finishing time.

3. Maximal power output (watts), work capacity (kilojoules), peak torque (newton metres), relative power (watts/kilogram)and lean body mass (kilograms) will be significantly negatively correlated with BMX performance. These variables will be the most significant predictors of BMX performance as reflected by track times in competition.

4. There will be a significant negative correlation between grip strength and BMX finishing time.

5. There will be a significant negative correlation between circumference measurements of the chest, thigh and calf (centimetres) and BMX finishing time.

6. There will be no significant relationship between flexibility and BMX finishing time.

7. There will be a moderate negative correlation between maximal oxygen uptake ($V0_2$ max.) and BMX finishing time.

8. The multiple linear regression analysis will result in an accurate predictive equation being developed.

9. Path analysis will give a more holistic insight into the relationship between the predictive variable set and BMX performance time.

10. Nonlinear regression equations can be derived that predict significantly BMX performance.

2. METHODOLOGY AND INSTRUMENTATION

Laboratory and field-based tests were conducted in this study. A variety of instruments were used to gather the required data. Details of each test and each instrument used are presented in the following text. Descriptive, bivariate and multivariate quantitative data analyses were conducted to statistically analyse the raw data.

Research Design

The type of research design for the current investigation is referred to as applied research. According to Baumgartner and Strong [9], applied research is when researchers have identified a real problem and are interested in solving it. Hinspeter (in Sports Research Needs, [10]), identified the need for research to identify the exercise physiological and kinanthropometric factors that predictor BMX performance. The current study attempted to solve this real problem with the addition of biomechanical factors that may explain BMX performance. The research design employed for this investigation, was a nonexperimental and correlation predictive design. The researchers did not intervene in any way and measurements were taken of a group of individuals so that relationships could be determined among the set of variables measured (Spector [11]). The correlation and regression designs were used in this study to test specific hypotheses about relationships between the measured prediction variables and BMX performance, which served as the dependent variable (Cozby, Worden & Kee [12]).

The data that were analysed were ratio level quantitative data produced by the Cybex 340, the Morgan Gas Analyser and the Repco Cycle Ergometer power tests. In addition, measurements were conducted on flexibility, human girths, height, weight, limb lengths, adiposity and grip strength. The ratio level data, is the highest quality of data from the four levels of measurement (Rothstein [13]). All of the above laboratory-based measurements were conducted in the Exercise Physiology Laboratory of the Australian Catholic University, Mackillop campus, Sydney. Field based BMX performances, which were also quantitative in nature, (time in seconds to complete a BMX track in competition) were conducted at the Panthers BMX track, Penrith.

The researchers attempted to limit systematic error, and therefore improve internal validity in the laboratory environment, by using the same measurement tools, at the same location, at approximately the same time of day, under the same environmental conditions and using only one researcher to measure all subjects on each test. The same testing procedures were used for all tests on all subjects. All equipment was calibrated and checked for correct operation prior to testing.

The multivariate research approach was used in this study to examine the effect of numerous independent or predictor variables acting together on the dependent variable. The multiple linear regression analysis formulated a linear model that related the dependent variable to the set of independent variables (Norusis [14]). The path analysis was employed to examine the indirect and direct relationships between the independent variable set and the dependent variable (Asher [15]; Kim & Kohout [16]). Path analysis is based on linear structural equation modelling. A set of nonlinear regression equations was applied to assess which mathematical function best described and explained the relationship between the independent and dependent variables.

Sampling Procedures

"The extent to which the results of a study are representative of the intended population depends on sampling techniques," (Rothstein [13], p.76). For this investigation, it was imperative that the representative sample of BMX competitors were diverse in their ability levels, otherwise the results obtained would give a false impression of the BMX population. Also, to distinguish the spectrum of exercise physiological differences between elite and non-elite BMX competitors, it was necessary to recruit a wide range of ability levels.

Subjects were recruited by verbal communication with the President of the Panthers BMX Club, the BMX National Coaching Director (Scott [1]) and the NSW BMX State Coaching Director (White [17]). Visits to competitions conducted at the Panthers BMX Track for the Panthers intraclub competitions and interclub competitions facilitated the recruitment of participants who displayed a broad range of BMX abilities. BMX competitors were approached personally, details of the study were explained, they were given a copy of the informed consent and information letter, phone numbers were exchanged, and they were then contacted by phone and arrangements were made for a time and date to be tested. All the subjects who were approached volunteered to be part of the study.

Sample Selection

All subjects involved in this study were male and 16 years of age or older. No attempt was made to recruit younger BMX competitors because their incomplete physical development does not allow a comparison with the older elite competitors. Exclusion of females from the study was not intentional. Very few females participate in BMX in the 16 and older age groups, and those that do, live outside the Sydney metropolitan area. This situation made recruiting females for the study very difficult. Ability levels ranged from elite national champions to average club competitors.

The sample size of 19 subjects was considered reasonable when compared to other studies using a similar research design within the research domain of exercise physiology. A review of six studies using a similar study design, indicated that the sample size of the current research had the third highest number of subjects when compared to the six studies.

Testing Protocols

The testing sequence was as follows. Weight and height measurements, five minutes maximum warm-up on the cycle ergometer, followed by the 10 second cycle ergometer power test, a minimum of three minutes rest, 30 second cycle ergometer power test, flexibility measurements, girth measurements (circumferences of the chest, preferred calf and thigh), sitting height, arm span, grip strength (hand grip dynamometer), adiposity (skinfold calipers), isokinetic dynamometry (Cybex 340) and maximal aerobic power ($V0_2$ max. by the Morgan Gas Analyser).

3. RESULTS

Descriptive Analysis

Exercise physiological, biomechanical and kinanthropometric characteristics (mean, standard deviation and range) are summarised in Appendix 1. The descriptive statistics highlight the fact that a wide spectrum of ability levels for the BMX competitors utilised in the study.

Bivariate Analysis

The Pearson product moment correlation identified numerous significant correlations between the dependent variable and the independent variables. All significant Pearson product moment correlation coefficients and p-values are displayed in table 1.

Table 1. Significant Pearson Product Moment Correlation Coefficients with Finishing Time

Variable	Finish Time
Weight (kg)	r= -.5569
	p = .013
Grip Strength (kg)	-.5828
	p = .009
Thigh Circum (cm)	-.5517
	p = .018
Power/wt (W/kg) [10 sec]	-.7990
	p = .000
Max Power (W) [10 sec]	-.8125
	p = .000
Total Work (kj) [10 sec]	-.8181
	p = .000
Power 10 sec (W) [30 sec]	-.7581
	p = .000
Power 20 sec (W) [30 sec]	-.7319
	p = .000
Total Work (kj) [30 sec]	-.8158
	p = .000
Peak Torq (Nm)[flex 60]	-.6445
	p = .007
Peak Torq (Nm)[flex 180]	-.6452
	p = .007
Peak Torq (Nm)[flex 300]	-.5762
	p = .019

Variable	Finish Time
Ave power (W)[flex 300]	-.5680 $p = .022$
Peak Torq (Nm)[ext 60]	-.5332 $p = .033$
Peak Torq (Nm)[ext 180]	-.6400 $p = .008$
Peak Torq (Nm)[ext 300]	-.6216 $p = .010$
Ave Power (W)[ext 300]	-.7223 $p = .002$
Power/weight (W/kg) [ext 300]	-.7432 $p = .001$
20m Time (sec)	-.8233 $p = .000$
Half Track Time (sec)	-.9385 $p = .000$
Anaerobic Threshold (%)	-.5658 $p = .028$
Chest Circum (cm)	-.5682 $p = .011$
Power 1 sec (W) [10 sec]	-.7864 $p = .000$
Power 2 sec (W) [10 sec]	-.8556 $p = .000$
Power 3 sec (W) [10 sec]	-.7685 $p = .000$
Power 4 sec (W) [10 sec]	-.7670 $p = .000$
Power 5 sec (W) [10 sec]	-.8366 $p = .000$
Power 6 sec (W) [10 sec]	-.7828 $p = .000$
Power 7 sec (W) [10 sec]	-.7036 $p = .000$
Power 8 sec (W) [10 sec]	-.7314 $p = .000$
Power 9 sec (W) [10 sec]	-.8052 $p = .000$
Power 10 sec (W) [10 sec]	-.7534 $p = .000$
Power 1 sec (W) [30 sec]	-.5123 $p = .025$
Power 2 sec (W) [30 sec]	-.7314 $p = .000$
Power 3 sec (W) [30 sec]	-.7883 $p = .000$

Variables	Finish Time
Power 4 sec (W) [30 sec]	-.6631 $p = .000$
Power 5 sec (W) [30 sec]	-.8524 $p = .000$
Power 6 sec (W) [30 sec]	-.790 $p = .000$
Power 7 sec (W) [30 sec]	-.6791 $p = .000$
Power 8 sec (W) [30 sec]	-.7618 $p = .000$
Power 9 sec (W) [30 sec]	-.8108 $p = .000$
Power 11 sec (W) [30 sec]	-.7819 $p = .000$
Power 12 sec (W) [30 sec]	-.6818 $p = .000$
Power 13 sec (W) [30 sec]	-.7141 $p = .000$
Power 14 sec (W) [30 sec]	-.7293 $p = .000$
Power 15 sec (W) [30 sec]	-.7799 $p = .000$
Power 16 sec (W) [30 sec]	-.7342 $p = .000$
Power 17 sec (W) [30 sec]	-.7411 $p = .000$
Power 18 sec (W) [30 sec]	-.7569 $p = .000$
Power 19 sec (W) [30 sec]	-.7891 $p = .000$
Power 21 sec (W) [30 sec]	-.7260 $p = .000$
Power 22 sec (W) [30 sec]	-.6400 $p = .003$
Power 23 sec (W) [30 sec]	-.6906 $p = .001$
Power 24 sec (W) [30 sec]	-.6416 $p = .003$
Power 25 sec (W) [30 sec]	-.5253 $p = .021$
Power 26 sec (W) [30 sec]	-.4754 $p = .040$
Power/wt (W/kg) [30 sec]	-.8228 $p = .000$
Max Power (W) [30 sec]	-.8329 $p = .000$

Variable	Finish Time
Total Work (kj) [Flex 60]	-.5317 $p = .034$
Ave Power (W) [flex 60]	-.6689 $p = .005$
Ave Power (W)[Flex 180]	-.5800 $p = .019$
Peak Torq (%BW) [Ext 180]	-.6258 $p = .010$
Peak Torq (% BW) [Ext 300]	-.5218 $p = .038$
Total Work (kj) [Ext 60]	-.5292 $p = .035$
Total Work (kj) [Ext 180]	-.6063 $p = .013$
Total Work (kj) [Ext 300]	-.6164 $p = .011$
Total Work (% BW) [Ext 180]	-.5529 $p = .026$
Total Work (% BW) [Ext 300]	-.5665 $p = .022$
Ave Power (W) [Ext 60]	-.7236 $p = .002$
Ave Power (W) [Ext 180]	-.7553 $p = .001$
Ave Power (% BW) [Ext 60]	-.6857 $p = .003$
Ave Power (% BW) [Ext 180]	-.7894 $p = .000$
Ave Power (% BW) [Ext 300]	-.7349 $p = .001$
Ave Power (%BW) [Flex 60]	-.5850 $p = .017$
Total Work (% BW) [Flex/Ext Ratio 180]	.5791 $p = .019$
Total Work (% BW) [Flex/Ext Ratio 300]	.5578 $p = .025$
Ave Power (% BW) [Flex/Ext 180]	.6720 $p = .004$
Ave Power (% BW) [Flex/Ext 300]	.5430 $p = .030$
Lean Body Mass (kg)	-.6914 $p = .001$

Due to the fact that multicollinearity existed among some of the variables, the Pearson product moment correlation coefficients were used to select a subset of variables that could be used in the multivariate analysis. The following subset of biomechanical and kinanthropometric independent variables were selected. Biomechanical variables included, relative power (power/weight) in the 10 second test, absolute total work done in the 30 second test, absolute power in the 30 second test, Cybex measurements of absolute peak torque, relative peak torque and relative work. The only kinanthropometric variable included was lean body mass.

Multivariate Analysis

Regression equations were derived for the dependent variable (BMX performance time) based on the above set of predictor variables that were identified from the Pearson product moment correlations to be significantly correlated with finishing time.

The multiple linear regression analyses identified that relative power and absolute work from the independent variable set are the best predictors of BMX performance time. Table 2 shows that the highly significant beta values of these two variables are much closer to one than any of the other variables, indicating their predictive value.

Subsequent analyses of the interrelationships identified that lean body mass was highly predictive of relative power and absolute work. It was found that thigh circumference and chest circumference were not as predictive as lean body mass. Tables 3 and 4 show that lean body mass is highly predictive (beta value of one) of both relative power and absolute work.

To this point, multiple linear regression identified that relative power and absolute work are the most predictive variables for BMX performance time and that lean body mass is the most predictive anthropometric variable of relative power and absolute work.

Table 2. Block Multiple Regression Equation and Level of Significance for Constructs Predicting Finishing Time.

Independent Variables	B	SE B	Beta	T value	Sig value
Dorsiflexion	0.093	0.181	0.122	0.514	0.623
Peak Torq [Flex 60]	0.019	0.071	0.132	0.267	0.797
Total Work [10 sec]	-0.329	0.497	-0.403	-0.662	0.529
Max Power [30 sec]	-0.002	0.014	-0.113	-0.125	0.904
VO_{2max}	-0.062	0.095	-0.159	-0.647	0.539
Peak Torq [Ext 180]	0.059	0.053	0.479	1.110	0.304
Total Work (% BW)[Ext 300]	-0.059	0.064	-0.323	-0.929	0.384
Power/Wt [10 sec]	-1.161	0.916	-0.647	-1.267	0.246

Multiple R	0.893
R^2	0.797
Adjusted R^2	0.565
Standard Error (min)	2.835
Analysis of Variance	
DF = (8,7) F = 3.44	p-value = 0.061

Table 3. Block Multiple Regression Equation and Level of Significance for Constructs Predicting Relative Power.

Independent Variables	B	SE B	Beta	T value	Sig value
Thigh Circum	-0.197	0.186	0.373	-1.063	0.306
Chest Circum	-0.055	0.130	-0.138	-0.420	0.681
Lean Body Mass	0.335	0.122	1.031	2.749	0.016
Multiple R	0.675				
R^2	0.455				
Adjusted R^2	0.339				
Standard Error (min)	1.932				
Analysis of Variance					
DF = (3,14) F = 3.903	p-value = 0.032				

Table 4. Block Multiple Regression Equation and Level of Significance for Constructs Predicting Total Work.

Independent Variables	B	SE B	Beta	T value	Sig value
Thigh Circum	-0.108	0.222	-0.096	-0.485	0.635
Chest Circum	-0.049	0.156	-0.058	-0.314	0.759
Lean Body Mass	0.712	0.146	1.030	4.868	0.000
Multiple R	0.909				
R^2	0.827				
Adjusted R^2	0.790				
Standard Error (min)	2.315				
Analysis of Variance					
DF = (3,14) F = 22.319	p-value = 0.000				

Multiple linear regression revealed that when lean body mass and relative power are regressed against BMX performance time, relative power is more predictive of BMX performance time. When lean body mass and absolute work are regressed against BMX performance time, absolute work is more predictive of BMX performance time. Table 5 and table 6 display these results. As can be seen from the R squared value in table 5, and 70.8% of the explained variance for finishing time is accounted for by this model. The beta value of -1.089 in table 6 is inflating the explained variance, which suggests that the model is explaining and predicting the relationship better than it actually does. Due to the collinearity in this model it was not used in subsequent analyses. These findings resulted in path analyses being conducted to more fully explain the interrelationships between the best predictor independent variables and the dependent variable.

Table 5. Block Multiple Regression Equation and Level of Significance for Lean Body Mass and Relative Power Predicting Finishing Time.

Independent Variables	B	SE B	Beta	T value	Sig value
Power/Wt					
[10 sec]	-1.065	0.300	-0.600	-3.549	0.003
Lean Body Mass	0.186	0.095	-0.330	-1.954	0.068
A or Y intercept	65.685	4.866		13.498	0.000

Multiple R	0.841
R^2	0.708
Adjusted R^2	0.671
Standard Error (min)	2.350
Analysis of Variance	
DF = (2,16)F = 19.396	*p*-value = 0.000

Table 6. Block Multiple Regression Equation and Level of Significance for Lean Body Mass and Total Work Predicting Finishing Time.

Independent Variables	B	SE B	Beta	T value	Sig value
Total Work					
[30 sec]	-0.872	0.273	-1.089	-3.188	0.006
Lean Body Mass	0.169	0.192	0.299	0.878	0.393

Multiple R	0.825
R^2	0.681
Adjusted R^2	0.641
Standard Error (min)	2.455
Analysis of Variance	
DF = (2,16)F = 17.067	*p*-value = 0.000

A path analysis was conducted to establish the direct and indirect links between lean body mass, relative power and BMX finishing time. The results of this analysis are illustrated in figure 1, which represents the actual path diagram based on the interrelationships between lean body mass, power/weight ratio and total race time.

Figure 1 indicates that the indirect path of lean body mass to relative power output to BMX finishing time, is slightly more predictive than the direct path of lean body mass to BMX finishing time.

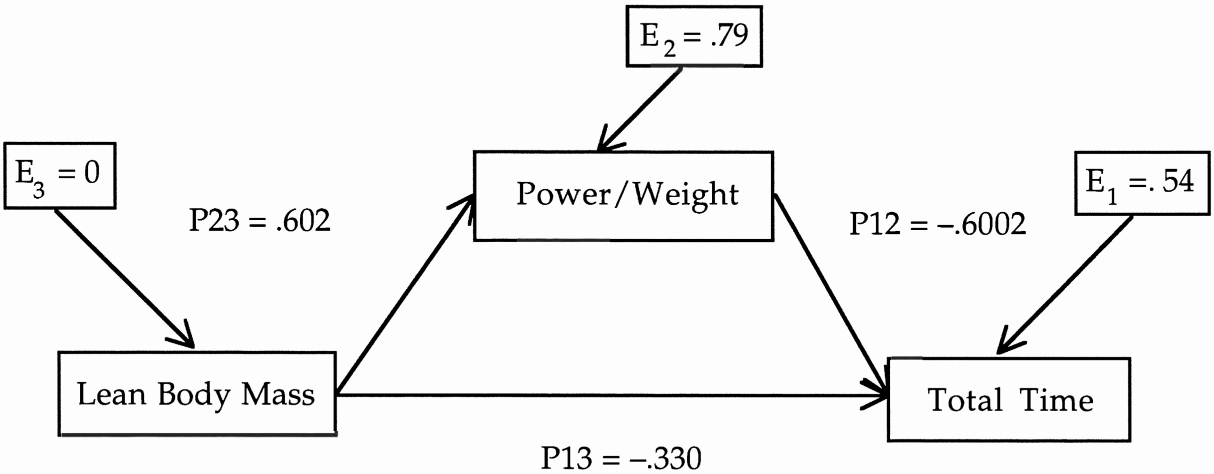


Figure 1: *The actual path diagram based on the interrelationships between lean body mass, power/weight ratio and total race time. Path coefficients were based on standardised beta coefficients. The symbols E_1, E_2 and E_3 represent all the residual causes of the measured variables for lean body mass, power/weight and total time.*

The path equations based on the model and path coefficients in Figure 1 can be decomposed and solved by the following equations (1) and (2).

$$\text{Total Effect} = \text{Direct Effect} + \text{Indirect Effect} + \text{Noncausal Effects} \quad (1)$$

$$\text{Original Covariation (correlation)} = r \quad (2)$$
$$= P_{13} + P_{23} \times P_{12} + 0 - .6914$$
$$= -.330 + .602 \times (-.6002) + 0 - .6914 = -.691$$

Direct Effect = P_{13} = —.330

Indirect Effect = $P_{23} \times P_{12} = .602 \times (-.6002) = -.361$

Noncausal Effect = 0

The multiple linear regression model can be used for prediction as well as explanation. The theoretical equation for prediction in the linear regression model is; Y = B1 X1 + B2 X2 + a, where Y = dependent variable, B = beta value, X = independent variable and a = constant. The predictive model developed is; Finishing time = –0.60 (power/weight ratio) + -0.33 (lean body mass) + 65.69. Table 7 represents the difference between the real and the predicted finishing times using the above equation. As can be seen the difference is small (calculated mean error 4.2%). Based on these results the predictive model developed can be used to estimate finishing time with confidence.

Table 7. The Difference between the Actual Finishing Time and the Predicted Finishing Time of BMX Competitors.

Subject	Actual Finishing Time (seconds)	Predicted Finishing Time (seconds)	Difference (seconds)
1	34.70	39.07	-4.37
2	42.10	39.04	3.06
3	33.00	34.95	-1.95
4	34.00	36.58	-2.58
5	36.19	34.85	1.34
6	41.00	40.55	0.45
7	46.26	44.12	2.14
8	35.53	37.35	-1.82
9	39.00	37.36	1.64
10	37.68	36.20	1.48
11	37.77	39.33	-1.56
12	37.78	41.08	-3.30
13	36.79	33.33	3.46
14	35.33	35.13	0.20
15	40.95	38.56	2.40
16	35.13	35.21	-0.08
17	46.67	46.33	0.34
18	32.00	33.64	-1.64
19	41.00	40.19	0.81

Nonlinear Regression

The nonlinear regression analysis first examined the relationship between power/weight ratio and track finishing time. This was assessed by visual inspection for linear or nonlinear trends by a scatterplot which is presented in figure 2. It was difficult to fit the relationship to a specific mathematical function by visual inspection, therefore, all potential mathematical functions were fitted and assessed by goodness-of-fit criteria. Table 8 indicates the goodness of fit criteria for each function, and includes the R^2 value (explained variance), degrees of freedom, the variance (F) ratio, and the coefficients for each mathematical equation. Note that in table 8 that; b_o= a constant; b1, b2 and b3 = regression coefficients; and Rsq = coefficient of determination or R^2.

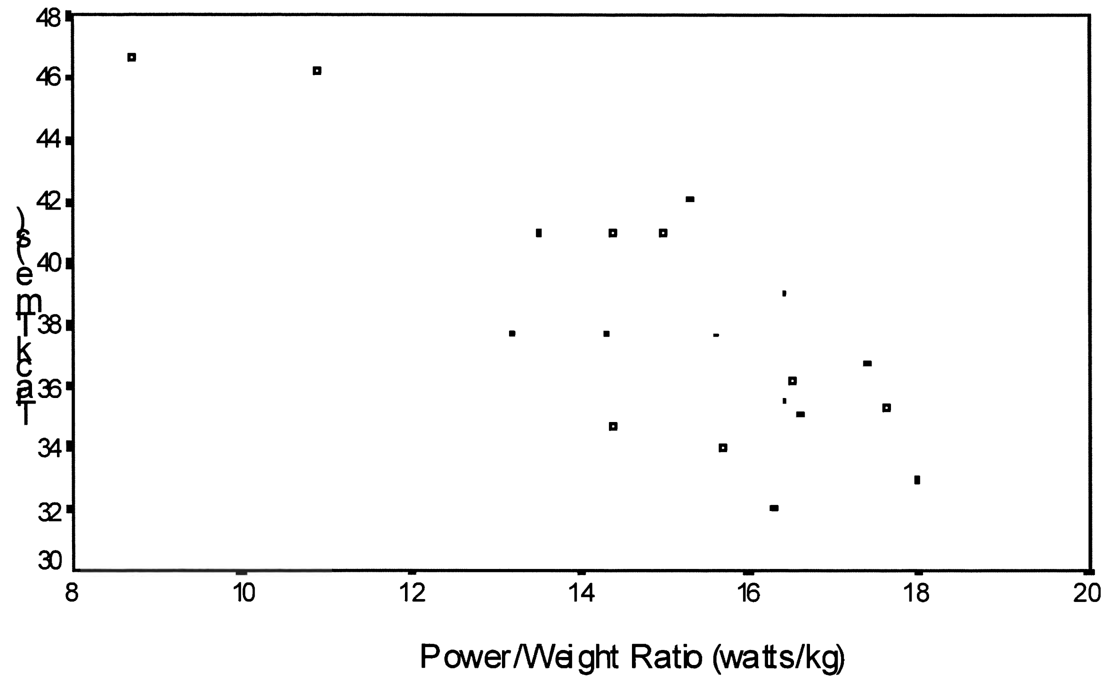


Figure 2: *The scatterplot relationship between power/weight ratio and track finishing time.*

Table 8. Indicates the Goodness-of-Fit for Each Function, and Includes the R^2-value (Explained Variance), Degrees of Freedom, the Variance (F) Ratio, and the Coefficients for Each Mathematical Equation.

Dependent Variable	Rsq	d.f.	F	sign.	b_0	b1	b2	b3	
Finish Time Function									
Linear	.638	17	30.00	.000	59.3972	–1.4174			
Logarithmic	.635	17	29.61	.000	88.1420	–18.561			
Inverse	.614	17	27.05	.000	22.4373	228.281			
Quadratic	.640	16	4.20	.000	63.2184	–1.9998	.0214		
Cubic		.640	16	14.21	.000	62.1894	–1.7369	ns	.0006
Compond	.610	17	26.56	.000	64.6115	.9651			
Power		.601	17	25.66	.000	132.023	–.4630		
Sigmoidal	.576	17	23.13	.000	3.2459	5.6690			
Growth	.610	17	26.56	.000	4.1684	–.0355			
Exponential	.610	17	26.56	.000	64.6115	–.0355			
Logistic	.610	17	26.56	.000	.0155	1.0361			

Table 8 indicates that the best fits were for the cubic and quadratic models with an R^2 = .640, p <.001, however all the functions have been identified as statistically significant with R_2 values ranging between .576 and .640. However, the cubic and quadratic functions (R^2=.640) are only marginally better than the bivariate linear function (R^2=.638). A more detailed analysis of the cubic method to fit the relationship is presented in table 9. This analysis indicates that the quadratic component in the equation was nonsignificant in its contribution to the nonlinear regression model (p-value = .8034). Figure 3 indicates the line of best fit for the cubic regression equation and the original values for finish time and power/weight ratio.

The predicted values for track finish times were calculated, as well as the residuals which are displayed in Table 10. The residuals are the difference between the actual score and the predicted score. A good predictive model will have small residuals and an R^2 that approaches unity (=1).

Table 9. The detailed nonlinear regression solution for the cubic model.

Dependent variable Finish Time   Method – CUBIC

Multiple R	.79987
R Square	.63979
Adjusted R Square	.59476
Standard Error	2.60822

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	2	193.32374	96.661868
Residuals	16	108.84511	6.802819

F-ratio = 14.20909   Significance of F = .0003

-------------------- Variables in the Equation --------------------

Variable	B	SE B	Beta	T	Sig. T
VAR00009	-1.736926	1.280266	-.979043	-1.357	.1937
VAR00009**3	.000556	.002181	184116	.255	8019
(Constant)	62.189446	11.670995		5.329	.0001

--------------- Variables not in the Equation ---------------

Variable	Beta In	Partial	Min Toler	T	Sig. T
VAR00009**2	–6.459211	-.065285	3.680E-05	–.253	8034

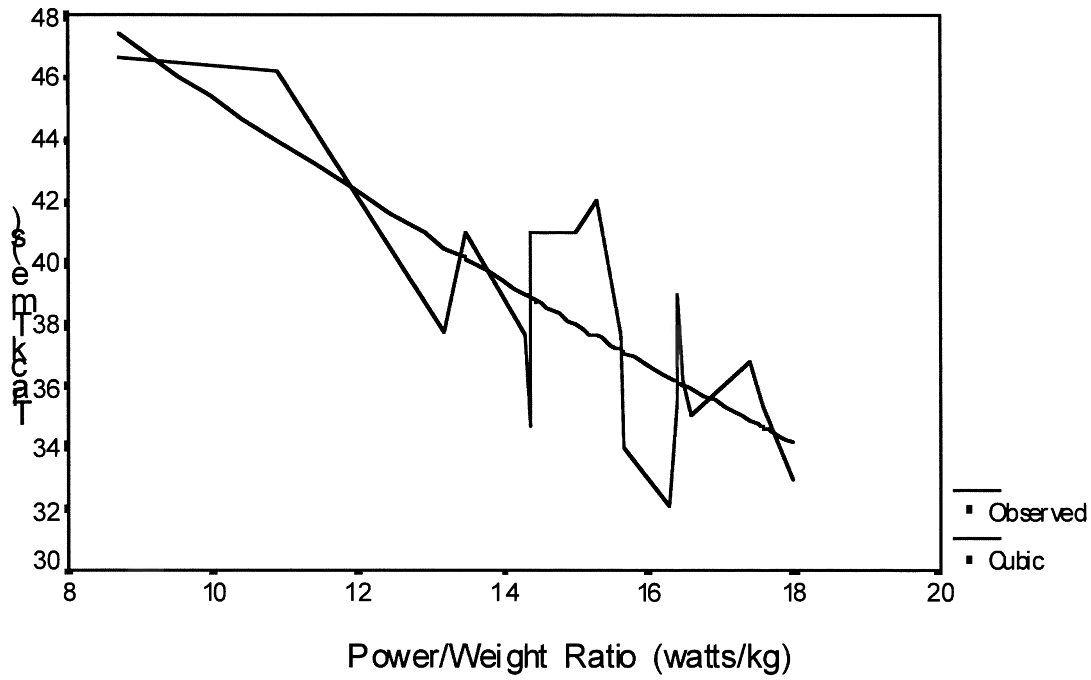


Figure 3: *The line of best fit using the cubic model (smooth curve) and the actual values for power/weight and track time (saw-tooth curve.*

Table 10. The predicted track time and residual for each participant.

Predicted Track Time (s)	Residual for Cubic Function (s)
38.14375	–3.44375
38.84245	3.25755
35.08638	–2.08638
35.93699	–1.93699
34.73058	1.45942
38.14375	2.85625
44.07690	2.18310
36.52015	–.99015
36.91775	2.08225
35.73246	1.94754
37.81535	–.04535
41.92132	–4.14132
34.66779	2.12221
35.54482	–.21482
39.99060	.95940
35.73246	–.60246
47.25153	–.58153
34.79122	–2.79122
41.03377	–.03377

4. DISCUSSION

Descriptive Analysis

When the results are compared to previous literature, McLean and Parker [18] found that for 35 elite male sprint track cyclists, the mean value of kinanthropometric variables measured was as follows: height 178 cm, weight 72.5 kg, age 22.6 years, sitting height 91.6 cm, chest circumference 95 cm, thigh circumference 55.7 cm, calf circumference 36.8 cm, total skin folds (eight sites) 56.3 mm and knee extension torque (isometric at 115 degrees) 235.2 Nm.

Mean values of kinanthropometric measures conducted in the present study were as follows: height 169 cm, weight 73.89kg, age 22.4 years, sitting height 90.46 cm, chest circumference 93.58 cm, thigh circumference 56.75 cm, calf circumference 37.83 cm, total skinfolds (four sites) 39.26 and knee extension torque (isokinetic 60 degrees/second) 214.38 Nm.

It can be seen from the means of the kinanthropometric measurements taken that BMX competitors and sprint track cyclists are very similar in relation to kinanthropometrics, except for height and the sum of skinfolds. The sprint track cyclists on average are nine centimetres taller than the BMX competitors. The young age of some of the subjects used in the current study could account for the height differences between BMX competitors and sprint track cyclists.

The differences in skinfold measurements between the two studies can be explained by noting that skinfold measurements were taken from a different number of sites. However, when the mean values of each study were converted to percent adiposity, the difference was negligible. This finding confirms hypothesis number one, that the kinanthropometric measurements of BMX competitors and sprint track cyclists will be similar.

It is possible to compare knee extension peak torque, however, caution needs to be taken, because two different measurement protocols were used and this could result in different findings. McLean and Parker [18] conducted their measurement during an isometric contraction at 115 degrees, whereas the BMX study conducted the same measurement during an isokinetic contraction at 60 degrees per second. However, both methods have measured essentially the same performance variable (peak torque extension). The isokinetic contraction at 60 degrees per second is a slow movement, and it is therefore possible to compare the two measurements to a certain degree.

The sprint track cyclists produced marginally greater force than the BMX athletes. Sprint track cyclists push a very large fixed gear and reach greater velocities than BMX competitors. The BMX athletes push a smaller fixed gear and rely on high cadence to gain maximum speed. This difference in the two events could explain the greater force production of the sprint track cyclists.

The remaining kinanthropometric variables measured, chest circumference, thigh circumference, calf circumference, weight, sitting height and age are very similar. This finding is interesting because of the samples used in each of the studies. The McLean and Parker [18] study was conducted on elite sprint cyclists, whereas the present study was conducted on a broad range of ability levels of BMX competitors.

The similarities were expected because of the similarities in the two events. However, the extent of the similarities was interesting because of the different samples used in both studies.

In the present study, the mean height of the BMX competitors was 169 cm which is below the average height for an adult male. The mean sitting height was calculated at 90.46 cm. Using sitting height, the mean leg length is therefore 78.54 cm which is relatively short in comparison to adult males. In the McLean and Parker study [18], the mean height was 178 cm and the mean sitting height was 91.6 cm, giving a mean leg length of 86.4 cm.

The BMX competitors have shorter leg lengths than the sprint track cyclists reported in the McLean and Parker [18] research. This is an interesting finding as it is consistent with a study conducted by Foley, Bird and White [19], where it was found that sprint track cyclists have shorter leg lengths than other cyclists, such as time trialists, pursuit cyclists and road cyclists. This finding and the findings in the present BMX study indicate that short leg length may be an advantage in sprint cycling events.

Foley et al. [19] suggested, that sprint cycling requires cyclists to generate large amounts of power very quickly. The athletes' utilisation of the relevant energy systems is very important in solving this problem successfully, however it appears that athletes with shorter legs can tolerate greater speed of movement. Power is equal to force multiplied by velocity. Therefore, generating higher leg speed (cadence) in cycling will result in higher power output if the force production remains constant. The fact that the McLean and Parker [18] study showed that sprint track cyclist have longer legs than BMX competitors indicates that because BMX competitors push an easier gear than the sprint track cyclists, BMX competitors may rely more on high movement speed to generate large amounts of power. In this situation shorter leg length would appear to provide a competitive advantage. This is only speculation and a controlled scientific research is required to confirm this interpretation.

Correlation Analysis

The purpose of the Pearson Product Moment correlation analysis was to identify a subset of independent variables that could be incorporated into a multivariate analysis, specifically multiple linear regression and path analysis. The correlation analysis identified 78 variables out of 116 variables (67%) that were significantly correlated with the dependent variable, that is BMX performance time. Many of the variables were essentially measuring similar constructs, this hypothesis was corroborated by conducting a confirmatory factor analysis. For example, each second of both the 10 second and 30 second cycle ergometer power tests, were loaded on a significant single factor. Also, many of the force production variables measured by the Cybex 340 were similar to those variables measured on the cycle ergometer tests as they loaded on the significant single factor. It was therefore expected that many predictor variables would be significantly correlated with the dependent variable, that is finish time, as well as with many other predictor variables.

The results of the correlation analysis showed that weight was significantly negatively correlated ($r=-.5569$) with BMX finish time. When weight was analysed in terms of body composition it was found that lean body mass had a significantly higher negative correlation (r=-.6914) with finish time than total weight. As lean body mass increased finishing time is reduced. Adiposity was not significantly correlated (r=.3908) with BMX performance time. These results indicate that weight is an important factor in BMX performance, however body composition is more important. If total weight includes a high percentage of adiposity performance will decrease, however if total weight has a large amount of lean body mass performance in BMX will improve.

In the Foley et al. [19] and McLean and Parker [18] studies, the results showed that the sprint cycle group were significantly more mesomorphic than the time trialists, pursuit and road cyclists. The authors concluded, that the high mesomorphic body type of the sprint cyclists is correlated with body strength and power, both factors are thought to be required for sprint cycling events. The results of the correlation analysis of the present BMX study confirm that a mesomorphic build (large amount of lean body mass) is beneficial to BMX performance, as BMX racing is essentially a sprint cycling event. This finding supports hypothesis number three which states that lean body mass will be one of the best predictors of BMX performance.

The correlation analysis identified that there was no significant relationship between height and BMX finishing time, sitting height and finishing time or between leg length and finishing time. This finding differs with the findings of the Foley et al. [19] study where sprint track cyclists had significantly shorter legs than endurance cyclists.

The correlation analysis identified that there was a significant negative correlation between grip strength and BMX performance time. This finding was not surprising when BMX racing technique was subjectively analysed, using a task analysis. Spurdens [3] stated that upper body strength is required in BMX because much of the impetus comes from the pull exerted on the bars of the bike for lifting the bicycle around bends and over jumps. The findings in the present study confirm hypothesis number three, that grip strength would be significantly negatively correlated with BMX finish time. Spurdens' [3] perceptions in relation to upper body strength and BMX performance has also been confirmed by the finding that grip strength had a significant relationship with finish time.

The results of the correlation analysis revealed that chest and thigh circumference were significantly negatively correlated with BMX finish time, however calf circumference has no significant association with BMX finish time. Hypothesis number four stated that chest, thigh and calf circumferences would be significantly negatively correlated with BMX finish time. The results of the correlation analysis support the relationship between chest and thigh circumference, however the relationship between calf circumference and BMX finish time has not been corroborated.

The McLean and Parker [18] study found that sprint cyclists had larger chest, thigh and calf circumferences than endurance cyclists. Mackova, Melichna, Havlickova, Platecha, Blahova and Semiginovsky [20] found that the diameter of fibres from the Vastus lateralis (thigh) was significantly greater in sprint cyclists than in non-cyclists.

The findings of the present BMX study also suggest that chest and thigh circumference are important in sprint cycling events. These findings are not surprising considering that it has been established that lean body mass and upper body strength are important factors in BMX performance. Muscle hypertrophy (large chest and thighs) would result in greater lean body mass and greater muscular strength (grip strength), which are all significantly associated with BMX finish time, a relationship that has been demonstrated in this research.

The fact that calf circumference was not found to be significantly associated with BMX finishing time in the present study, but was found to be significantly larger in sprint cyclists than endurance cyclists in the McLean & Parker [18] study, could possibly be attributed to the different techniques used in sprint track cycling and BMX competition. Apart from biomechanical differences between bicycles, BMX competitors stand up for the entire event with a non-fixed foot, therefore they only produce power in the down stroke of the pedal cycle, whereas sprint track cyclists remain seated with fixed feet which allows force to be applied for a large part of the pedal cycle. The different techniques when BMX pedalling is compared to sprint cycle pedalling, indicate that the leg muscles of the different cyclists are being stressed in different ways. The BMX pedalling technique, due to the standing position, is emphasising quadricep torque and power production, and the importance of thigh muscle development to BMX performance.

Hypothesis number six stated that there would be no significant correlation between flexibility and BMX performance time. The results supported this hypothesis. Four flexibility measurements were assessed and included sit & reach, dorsiflexion, plantar flexion and hamstring flexibility, none of which were significantly correlated with BMX finish time.

A search of the literature on flexibility and sprint cycling revealed that none of the studies conducted on sprint cyclists considered flexibility in their research designs. Spurdens [3] states that flexibility is important for BMX performance, as he believes flexibility helps reduce the chance of injury and increases the range of movement. Both factors should facilitate the movement pattern when cycling in BMX. This statement has not been confirmed by the present BMX study.

deVries and Housh [5] state that the need for flexibility varies with the athletic endeavour. In some activities it is more important than in others. The correlation analysis from the present BMX study indicates that flexibility is not an important predictor of BMX performance expressed as finishing time. The role of flexibility preventing injury in BMX probably requires a more controlled sport scientific study to elucidate the effects of flexibility on injury prevention in BMX participants.

It was hypothesised that there would be a moderate negative significant correlation between maximal oxygen consumption and BMX finishing time. The correlation analysis refuted clearly this hypothesis. Telford, Hahn, Pyne and Tumilty [21] found that there was no significant difference in maximal oxygen consumption between sprint track cyclists and road cyclists. Telford et al. [21] speculated that this result may due to the fact that track cyclists undertake a high volume of training in preparation for the track season. This result based on type of training is surprising considering the different energy systems required for sprint track cycling and road cycling.

The finding of the current BMX study indicates that maximal oxygen consumption is unrelated to finishing time in BMX competition. When the time span of a BMX event (30 - 40 seconds) and the dominant energy systems that would be utilised (ATP-CrP and anaerobic glycolysis) are considered, it makes perfect sense that maximal oxygen consumption is unrelated to finishing time in BMX competition. The researcher hypothesised that maximal oxygen consumption would be moderately negatively associated with BMX finishing time due to the BMX race day structure. The competitors are required to race numerous times on the same day before qualifying for the final. It was speculated that a moderate maximal oxygen consumption capacity would be an advantage in recovering between heats (motos) and thus allowing the individual to perform at a high intensity in all races on race day.

As was discovered by Dawson, Fitzsimons and Ward [22] in their study on repeated sprint ability, the aerobic energy system is closely associated with performance decrement in repeated sprint activities. However, the recovery time between BMX races in a given competition day may not be short enough to stress the aerobic system, a situation that diminishes the importance of aerobic ability in influencing BMX racing.

Many of the power variables measured were highly negatively correlated with BMX finish time. Hypothesis number three was supported, in that both absolute power and relative power were highly significantly negatively correlated with BMX finish time.

With reference to Telford's et al. [21] conclusion, that for optimal performance sprint cyclists must develop their maximal power and anaerobic energy systems to their maximum extent, the findings of the present study are not surprising. The amount of power produced for every second of the ten second power ergometer test was significantly negatively correlated with BMX finishing time. However, for the 30 second cycle ergometer test the final four seconds were not significantly correlated with finish time. This is interesting considering that a BMX event lasts between 30 and 40 seconds.

Faria [23] states that performance in sprint cycling events necessitates the breakdown of high energy compounds, such as adenosine triphosphate (ATP) and creatine phosphate (CrP). Sprint cycling events lasting approximately 10 seconds rely heavily on the ATP-CrP energy system, events lasting between 30 and 60 seconds rely heavily on the anaerobic glycolytic or lactic acid energy system.

Considering this information it would be expected that a BMX event would rely heavily on the lactic acid energy system. However, in the 30 second cycle ergometer test which measured the capacity of the lactic acid energy system, it was found that this system was not conditioned well enough to maintain maximal output for 30 seconds due to the significant decline in work output. Considering that a BMX event lasts for a minimum of 30 seconds this finding would not be expected.

However when a BMX race event is task analysed, it becomes apparent that the competitors do not pedal at maximum intensity for the entire duration of the event. In fact, at some stages of the race the competitors stop pedalling completely while going over jumps or around bends or when they are blocked by another competitor. These breaks allow recovery time between high intensity efforts. Therefore, the lactic acid system is not being stressed as much as would be expected when the time span of the event is considered. .

Telford et al. [21] found that sprint cyclists are able to produce more torque (force) and had higher anaerobic work capacities than road cyclists. The current BMX study found that both torque production and anaerobic work capacity variables were significantly correlated with BMX finishing time. These findings confirmed hypothesis number three that work capacity and force production would be highly significantly and negatively correlated with finish time.

Burke, Fleck and Dickson [24] concluded from their study that anaerobic metabolism contributes significantly to energy production in sprint track cycling competition. Mackova et al. [20] also concluded that anaerobic metabolism is important in sprint cycling events. The results of the correlation analysis of the present study also indicate that anaerobic metabolism is important in BMX competition, which is a sprint cycling event.

The fact that anaerobic constructs such as torque, work and power were so significantly and negatively correlated with finish time and that aerobic constructs such as maximal oxygen consumption were not significantly correlated with finish time, suggests that anaerobic power and work production are extremely important to successful BMX racing.

Regression Analysis

Regression analysis allowed a more complete analysis as to what variables were the most important to the BMX performance. In the regression analysis the variables were treated simultaneously, which allowed the researcher to identify a subset of independent variables that best predicted BMX finishing time. These results were then used to develop a predictive equation for BMX finishing time so that a causal relationship predicting performance could be identified.

The results of the regression analysis showed that lean body mass, total work achieved on the 30 second test (absolute) and relative power achieved on the 10 second test are the best predictors of BMX finishing time. These results corroborate hypothesis number three. Therefore, to be successful in BMX competition an individual requires a large amount of lean body mass, they need to be able to generate a large power output relative to their body weight in a short period of time (10 seconds) and they also need to be able to produce a large work output over 30 seconds.

Telford et al. [21] found that sprint track cyclists in comparison to road cyclists generated higher power output during a 10 second cycle ergometer test. The sprint track cyclists also produced larger work outputs over a 60 second cycle ergometer tests. Telford et al. [21] did not apply regression analysis to their results, however their findings indicate that power and work output are highly predictive of sprint cycling performance.

Foley et al. [19] and McLean & Parker [18] found that sprint track cyclists had high mesomorphic ratings (a large amount of lean body mass) in relation to other groups of cyclists such as time trialists, pursuit cyclists and road cyclists. Neither of these researchers applied regression analysis to their results, however the results of both studies indicate that lean body mass is predictive of sprint track cycling.

The above findings confirmed hypothesis number two that the best predictors of sprint track cycling performance will also be the best predictors of BMX finishing time. As was indicated in the results the predictive equation was reasonably accurate with a mean error of 4.2%. A good model that describes the relationship between the dependent variable and the predictor variables has been established due to the high multiple correlation (multiple R) and high value for the coefficient of multiple determination (R squared), small residuals and highly significant p-values. It should be highlighted that 70.8% of the variance was explained and the prediction of scores based on the multiple linear regression equation displayed low percentage error for the residuals. Specifically, the percentage error scores for each BMX cyclists varied from 0.1% to 11.1%.

The results of the multiple linear regression analysis and the development of the predictive equation have important implications for the development of BMX as a sport. The standard of competition could be improved by adopting two methods. The first method is by improving the performances of current athletes by strength and power training, and the second method is through talent identification, by measuring work and power production integrated with kinanthropometric profiling.

Current BMX competitors can be subjected to a series of biomechanical and kinanthropometric tests. The results could be assessed by the predictive equation to determine the BMX competitors strengths and weaknesses. For example, if it was identified that the competitor was unable to generate adequate power output in the ten second cycle ergometer test, the training program could be adjusted to improve power output, a factor which through the multiple linear regression analysis has been shown to be highly predictive of BMX finishing time.

A talent identification program could also be implemented into schools, as is the case with other sports, such as athletics. Field-based tests that measure the predictive biomechanical and kinanthropometric constructs could identify those who have the potential to be successful in the sport of BMX. Those that meet a set of selection criteria could be offered scholarships to train with a BMX development squad. Similar situations based on talent identification occur in athletics (Torbottom [25, 26]; Jones [27]).

Path Analysis

The results of the regression analysis identified the best predictors of BMX finishing time, however this information did not distinguish between direct and indirect causal paths of the independent variables (lean body mass, work output and relative power) influence on the dependent variable (BMX finishing time). Path analysis was used to determine the causal structure of an explanatory causal model by examining the direct and indirect effects of the independent variables on the dependent variable.

Separate path analyses were conducted on lean body mass, relative power and BMX finishing time and lean body mass, work output and finishing time. The results of the path analysis revealed that lean body mass directly effects both relative power and work output. However, lean body mass has significant indirect effect via relative power on BMX finishing time. Both relative power and work output directly influence BMX finishing time.

In other words relative power was identified as a mediating variable. Relative power and work output are dependent on lean body mass and BMX finishing time is dependent on work output and relative power. The path analysis identified a causal order and structure. Kinanthropometrics (lean body mass) directly influence wok output and relative power production which then directly influences BMX finishing time. However, lean body mass indirectly influences finishing time via the indirect path through relative power and work output.

The findings of the path analysis has important implications for training BMX competitors. Lean body mass needs to be increased to improve work output and relative power which will in turn decrease BMX finishing time both directly and indirectly. An increase in lean body mass that does not improve relative power and work output will not improve BMX performance time due to the finding that lean body mass has less significant direct influence on BMX finishing time.

Nonlinear Regression

The nonlinear regression analysis indicated that a marginally better regression equation can be derived than when using the linear regression approach. Many nonlinear models indicated a significant amount of explained variance between finish time and power/weight ratio. Both the quadratic and cubic models produced an identical amount of explained variance (64%). The application of nonlinear regression highlights the fact that there are a number of potential curves that can fit the data. However the researcher should endeavour to fit 'the best' model to the data, and not assume that the model will fit a linear function which is often the case in the exercise and sport sciences.

Summary of Discussion

The majority of hypotheses were corroborated by the findings of the current study.

1. The kinanthropometric constructs measured in the current study were found to be very similar to those found in sprint track cyclists.

2. The correlation analysis identified lean body mass as being significantly related to BMX finishing time.

3. Power, torque and work constructs were found to be highly predictive of finishing time.

4. The regression analysis revealed that lean body mass, power and work are the best predictors of finishing time.

5. Path analysis indicated the role of lean body mass influencing directly finishing time as well as the indirect path via power/weight ratio that also influences performance.

6. Nonlinear regression indicated that a number of mathematical models, for example quadratic and cubic models, may marginally improve the predictions of performance when compared to linear regression models.

REFERENCES

[1] Scott S., BMX is back, *Cycling*, June, 30-35, (1995)

[2] Draper J., Sports research needs. (Australian Sports Commission, Canberra, 1992).

[3] Spurdens D., BMX, (Sterling Publication Company, New York, 1984).

[4] Brooks G.A. and Fahey T.D., Exercise physiology human bioenergetic and its application, (Macmillan Publishing Company, New York, 1984).

[5] deVries H.A. and Housh T.J., Physiology of exercise: For physical education, athletics and exercise science, (WBC Brown & Benchmark Publishers, Iowa, 1994).

[6] Hahn A.G., Measuring physiological capacities in the laboratory and field. In F.S. Pyke (Ed.) Better coaching: Advanced coach's manual, (Australian Sports Commission, Canberra, 55-66, 1991).

[7] Roberts A.D., Physiological capacity for sports performance. In F.S. Pyke (Ed.) Better coaching: Advanced coach's manual, (Australian Sports Commission, Canberra, 43-54, 1991).

[8] Schell J. and Leelarthaepin B., Physical fitness assessment in exercise and sport science, (Leelar Biomediscience Services, Matraville, 1994).

[9] Baumgartner T.A. and Strong C.H., Conducting and reading research in health and human performance, (WBC Brown and Benchmark Publishers, Iowa, 1994).

[10] Hinspinter J., BMX. In J Draper (Ed.) Sports research needs, (Australian sports Commission, Canberra, 8, 1992).

[11] Spector P.E., Research designs, (Sage Publications, London, 1981).

[12] Cozby P.C., Worden P.E. and Kee D.W., Research methods in human development, (Mayfield Publishing Company, California, 1989).

[13] Rothstein A.L., Research design and statistics for physical education. (Prentice Hall Inc, New Jersey, 1985).

[14] Norusis M.J., SPSS-X: Advanced statistics guide, (SPSS Inc, Chicago, 1985).

[15] Asher H.B., Causal Modeling, (Sage, London, 1983).

[16] Kim J.O. & Kohort F.J. (1975). Special topics in general linear models. In H.N. Nie, C.H. Hull, J.C. Jenkins, K. Steinbrenner and D.H. Bent (Eds.) SPSS, (McGraw-Hill, Sydney. 368-397, 1975).

[17] White G., Personal Communication. Training and conditioning. (18th July, 1995).

[18] McLean B. and Parker A., An anthropometric analysis of elite Australian track cyclists, *Journal of Sports Sciences*, 7, (3), 247-255 (1989).

[19] Foley J.P., Bird S.R. and White J.A., Anthropometric comparison of cyclists from different events, *British Journal of Sports Medicine*, 23, (1), 30-32 (1989).

[20] Mackova E., Melichna J., Havlickova L., Placheta Z., Blahova D. and Semiginovsky, B., Skeletal muscle characteristics of sprint cyclists and nonathletes, *International Journal of Sports Medicine*, 7, (5), 295-297 (1986).

[21] Telford R.D., Hahn A.G., Pyne D.B. and Tumilty D.M., Strength, anaerobic capacities and aerobic power of Australian track and road cyclists, *Excel*, 6, (4), 20-22 (1990).

[22] Dawson B., Fitzsimons M. and Ward D., The relationship of repeated sprint ability to aerobic power measures of anaerobic work capacity and power, *Australian Journal of Science and Medicine in Sport*, 25, (4), 88-92 (1993).

[23] Faria I.E., Applied physiology of cycling, *Sports Medicine*, 1, (3), 187-201 (1984).

[24] Burke E.R., Fleck S. and Dickson, T., Post competition blood lactate concentrations in competitive track cyclists, *British Journal of Sports Medicine*, 15, (4), 242-245 (1981).

[25] Torbottom D., NSW academy track and field high performance squads, *Athletics*, 5, (1), 11-14 (1995).

[26] Torbottom D., AWB talent 2000 squad wins gold, *Athletics*, 5,(1), 17-19 (1995).

[27] Jones M., Despite weather Oz squad camp a success, *Athletics*, 5, (1), 15-16 (1995).

APPENDIX I

Table of Means, Standard Deviations and Ranges for all Variables in the Study.

Variable	M	SD	Range
Weight (kg)	73.89	7.6	62.7 - 88.3
Height(cm)	174.5	4.53	168 - 181
Grip Strength (kg)	59.58	7.3	50 - 76
Sitting Height(cm)	90.46	3.0	84 - 95
Arm Span(cm)	181.5	7.3	169 - 194
Thigh Circum(cm)	56.75	4.5	47 - 67
Sit & Reach(cm)	7.26	7.54	-7 - 26
Adiposity (mm)	39.26	14.9	23.5 - 82
Power/wt (W/kg) [10 sec]	15.06	2.3	8.7 - 18
Max Power (W) [10 sec]	346.4	35.2	285 - 420
Total Work [kj] [10 sec]	9.56	2.1	5 - 12.8
Power 10 sec (W) [30 sec]	974.1	229.9	511 - 1401
Power 20 sec (W) [30 sec]	794.26	152.9	554 - 1142
Power 30 sec (W) [30 sec]	524.32	173.8	38 - 826
Total Work (kj)[30 sec]	25.04	5.1	14.9 - 35
Peak Torq (Nm)[flex 60]	143.44	30.1	89 - 195
Peak Torq (Nm)[flex 180]	116.5	21.6	77 - 162
Peak Torq (Nm)[flex 300]	89.69	17.7	65 - 132
Ave power (W)[flex 300]	278.81	62.3	178 - 449
Power/weight (W/kg) [flex 300]	378.81	60.3	287 - 516
Endurance Ratio (%) [flex 300]	64.5	7.0	55 - 76
Endurance Ratio (%) [ext 300]	68.57	8.9	54 - 83
Peak Torq (Nm)[ext 60]	214.38	46.6	141 - 287
Peak Torq (Nm)[ext 180]	154.56	37.1	93 - 218
Peak Torq (Nm)[ext 300]	109.19	22.3	69 - 151
Ave Power (W)[ext 300]	313.75	81.8	197 - 453
Power/weight (W/kg) [ext 300]	423.75	75.6	281 - 539
VO2 Max (ml/kg/min)	60.29	10.7	38.23 - 77
20m Time (sec)	2.87	0.3	2.46 - 3.6
Half Track Time (sec)	21.33	2.1	18.85 - 26.2
Finish Time (sec)	38.05	4.1	32 - 46.67

Variable	M	SD	Range
Anaerobic Threshold (%)	80.66	12.1	61 - 96
Chest Circum (cm)	93.58	6.0	86 - 105
Calf Circum (cm)	37.83	2.1	35 - 41
Dorsiflexion (Deg)	79.18	6.5	67 - 94
Plantaflexion (Deg)	146.68	19.5	72 - 165
Hamstring Flex (Deg)	156.1	12.0	130 - 170
Power 1 sec (W) [10 sec]	613.68	229.4	196 - 884
Power 2 sec (W) [10 sec]	994.05	267.3	478 - 1366
Power 3 sec (W) [10 sec]	1066.68	254.9	545 - 1501
Power 4 sec (W) [10 sec]	1084.32	249.9	543 - 1540
Power 5 sec (W) [10 sec]	1109.63	243.3	536 - 1449
Power 6 sec (W) [10 sec]	1076.84	210.7	596 - 1371
Power 7 sec (W) [10 sec]	1047.27	238	546 - 1490
Power 8 sec (W) [10 sec]	1029.68	227.9	601 - 1407
Power 9 sec (W) [10 sec]	1011.26	242.7	534 - 1422
Power 10 sec (W) [10 sec]	1003.21	231.5	600 - 1304
Power 1 sec (W) [30 sec]	555.42	274.5	37 - 970
Power 2 sec (W) [30 sec]	905.53	312.6	326 - 1374
Power 3 sec (W) [30 sec]	1058.53	290.5	458 - 1523
Power 4 sec (W) [30 sec]	1108.00	338.4	482 - 1949
Power 5 sec (W) [30 sec]	1099.89	264.1	521 - 1457
Power 6 sec (W) [30 sec]	1088.00	237.6	520 - 1520
Power 7 sec (W) [30 sec]	1046.12	234.5	543 - 1419
Power 8 sec (W) [30 sec]	1058.79	261.7	550 - 1516
Power 9 sec (W) [30 sec]	1008.89	212.5	574 - 1348
Power 11 sec (W) [30 sec]	941.53	200.3	599 - 1341
Power 12 sec (W) [30 sec]	918.21	221.2	532 - 1332
Power 13 sec (W) [30 sec]	911.12	194.7	565 - 1371
Power 14 sec (W) [30 sec]	893.21	166.4	562 - 1164
Power 15 sec (W) [30 sec]	874.26	185.7	555 - 1245
Power 16 sec (W) [30 sec]	853.05	196.1	572 - 1279
Power 17 sec (W) [30 sec]	830.00	178.3	505 - 1175
Power 18 sec (W) [30 sec]	825.74	152.8	604 - 1189
Power 19 sec (W) [30 sec]	815.74	154.1	514 - 1200
Power 21 sec (W) [30 sec]	761.37	156.5	473 - 1048
Power 22 sec (W) [30 sec]	755.79	137.6	554 - 1042
Power 23 sec (W) [30 sec]	718.12	163.3	468 - 1103
Power 24 sec (W) [30 sec]	724.37	165.9	492 - 1128
Power 25 sec (W) [30 sec]	679.00	173.3	401 - 1084

Variable	M	SD	Range
Power 26 sec (W) [30 sec]	643.37	159.4	276 - 1003
Power 27 sec (W) [30 sec]	616.47	159.1	168 - 902
Power 28 sec (W) [30 sec]	598.95	177.1	97 - 874
Power 29 sec (W) [30 sec]	590.26	175.1	52 - 841
Power/wt (W/kg) [30 sec]	14.98	2.6	8.5 - 18.6
Max Power (W) [30 sec]	1116.32	258.8	564 - 1540
Total Work (kj) [Flex 60]	177.31	44.6	96 - 241
Total Work (kj) [flex 180]	136.94	29.2	92 - 174
Total Work (kj) [flex 300]	104.50	22.34	63 - 147
Total Work (% BW) [flex 60]	234.25	44.87	137 - 295
Total Work (% BW) [Flex 180]	186.5	33.7	136 - 234

Table of Means, Standard Deviations and Ranges for all Variables in the Study.

Variable	M	SD	Range
Total Work (% BW) [Flex 300]	142.13	24.55	101 - 180
Ave Power (W) [flex 60]	105.69	22.7	66 - 142
Ave Power (W)[Flex 180]	225.19	41.1	158 - 292
Peak Torq (%BW) [Flex 60]	193.19	31.2	127 - 245
Peak Torq (% BW) [Flex 180]	157.56	22.3	110 - 193
Peak Torq (% BW) [Flex 300]	121.81	16.5	92 - 151
Peak Torq (% BW) [Ext 60]	293.63	42.9	201 - 363
Peak Torq (%BW) [Ext 180]	209.31	35.1	134 - 259
Peak Torq (% BW) [Ext 300]	149.13	21.8	106 - 179
Total Work (kj) [Ext 60]	216.38	50.7	135 - 305
Total Work (kj) [Ext 180]	158.00	39.1	98 - 231
Total Work (kj) [Ext 300]	114.94	26.6	70 - 161
Total Work (% BW) [Ext 60]	293.06	49.2	192 - 386
Total Work (% BW) [Ext 180]	213.38	6.6	140 - 292

Variable	M	SD	Range
Total Work (% BW)	14.98	2.6	8.5 - 18.6
[Ext 300]	155.25	23.5	112 - 193
Ave Power (W) [Ext 60]	136.13	28.6	88 - 187
Ave Power (W) [Ext 180]	267.50	71.7	154 - 377
Ave Power (% BW) [Ext 60]	184.45	25.6	134 - 229
Ave Power (% BW) [Ext 180]	360.94	67.8	227 - 438
Ave Power (% BW) [Ext 300]	421.06	76.1	281 - 539
Ave Power (%BW)[[Flex 60]	147.81	26.4	94 - 207
Ave Power (%BW) [Flex 180]	299.56	53.1	185 - 388
Ave Power (% BW) [Flex 300]	363.63	84.6	142 - 516
Flex/Ext Ratio (60)	65.69	7.5	50 - 78
Flex/Ext Ratio (180)	76.13	11.2	59 - 103
Flex/Ext Ratio (300)	81.94	9.3	68 - 105
Total Work (% BW) [Flex/Ext Ratio 60]	79.63	9.7	63 - 100
Total Work (% BW) [Flex/Ext Ratio 180]	87.88	15.3	65 - 126
Total Work (% BW) [Flex/Ext Ratio 300]	91.69	14.7	69 - 132
Ave Power (% BW) [Flex/Ext Ratio 60]	77.44	8.6	58 - 94
Ave Power (% BW) [Flex/Ext 180]	86.69	16.2	63 - 130
Ave Power (% BW) [Flex/Ext 300]	90.75	14.5	69 - 127
Age (years)	22.39	2.4	16 - 36
Adiposity (%)	15.54	4.4	9 - 25
Lean Body Mass (kg)	62.38	7.3	53.55 - 79

THE NATIONAL STATISTICAL INITIATIVE

Ian Smith[1]

Abstract

Reliable, well organised and accessible statistics on the Sport and Recreation Industry are mostly unavailable in Australia. This situation has hindered effective planning and development and limited the ability of government and non-government organisations to formulate sound strategic policies. The Sport and Recreation Ministers Council (SRMC) has recognised the need for better data by establishing a Statistical Working Group (SWG). SWG's aim is to improve the definition, range and quality of statistics for the industry. To meet this aim, information is needed based on widely agreed statistical definitions and needs.

SWG sought regular, reliable and independent data collections, that could only be provided by the Australian Bureau of Statistics (ABS). However, before ABS data collections could be unlocked, changes were needed to ABS definitional frameworks. Thus, the first task was to develop a *National Sport and Recreation Industry Statistical Framework*, which defines the industry and its data needs, and establishes long term plans of how these needs should be met by ABS. The Framework identifies statistical needs of each industry segment from the creation of a service or product through to its end use.

Following wide consultation, SWG also developed reports on employment in the National Sport and Recreation Industry Occupational Structure and a snapshot of the industry in Available Data an Sources for the Sport and Recreation Industry. The three reports have been endorsed by SRMC and presented to ABS for implementation. Other SWG initiatives include:

- the National Sport and Recreation Industry Directory and Database (NSRIDD) – which provides computer access to annual updates of all ABS collected data on the industry. NSRIDD is located in each state and territory;

- ABS sport and recreation surveys in areas such as a participation, consumption, income, expenditure, employment, facilities and services; and

- liaison with ABS through its National Culture Recreation Statistical Unit (NCRSU) which assists SWG to seek inclusion of the industry in relevant ABS classifications and collections.

In summary, SWG seeks to develop greater understanding based on reliable data; establish trends in performance; identify priority needs and actions to maintain the industry, and prospects for new facilities and programs; analyse implications of changes in the level, mechanisms and distribution of public and private funding; and develop reliable statistics consistent with those of other nations, for performance comparisons.

---

[1]Office of Recreation Development , Department of the Environment, Sport and Territories, GPO Box 787, Canberra ACT 2601

1. INTRODUCTION AND NEED FOR INDUSTRY STATISTICS

Many Australians are in one way or another involved with the provision or consumption of services and products in the fields of sport and recreation. These fields are not only important to individuals. It has been established the services provided and consumed are a major component of the national economy.

An analysis of information from the Australian national accounts collected by the national statistical collection agency, the Australian Bureau of Statistics (ABS), indicates sport and recreation contributed a highly conservative gross product of $4.2 billion in 1987/88 (which is only about half the actual size as expenditure by government, the corporate sector and those Australians living in institutions was not included). It has been found that local government spent more on sport and recreation annually than either of the other two spheres of government (Australia has a federal system which includes government at the national, state (and two territories which equate to states) and local levels).

In 1987/88, sport and recreation compared favourably with other industries such as agriculture, forestry and fishing ($12.9 billion); electricity, gas and water ($10.9 billion); food, beverages and tobacco ($9.8 billion); cultural services ($7.8 billion); base metal products ($5.1 billion); chemical, petroleum and coal products ($4.7 billion); transport equipment ($3.8 billion); wood and wood products ($2.6 billion); clothing and footwear ($1.7 billion); and textiles ($1.3 billion).

The Sport and Recreation Industry is a major employer even on the basis of the limited readily available data (the quality of which varies). The ABS was only able to provide quality data for a few industry segments. Nevertheless, it has been estimated that the industry employed at least 140 000 persons in 1993/94. The provision of services for the industry had the highest employment (33 000). Jobs for males and females varied greatly with males dominating (71 per cent) in amusement and passive recreation, and females (56 per cent) in gambling.

The importance of sport and recreation to the national economy, and its potential for creating employment, is not widely recognised. This lack of recognition is because the industry has changed in only a few decades from a service industry largely managed by volunteers, to the economic status it enjoys today, where full time staff are increasingly replacing volunteers.

With leisure time growing for the community, continuing efforts are being made to stimulate excellence and participation in the sporting and recreational life of the nation. Such efforts have led to increasing demands for sound statistical information for planning and development action, and decision making by government and industry at all levels. Statistics are used as a basis for policy planning and programme development, and to monitor the impact of existing policy. Governments at all levels are major users of such statistics. Other major users include non-government organisations, industry, unions and academia who use statistics for funding submissions, research, etc.

Reliable, well organised and accessible statistics on the Sport and Recreation Industry are mostly unavailable in Australia. This situation has hindered effective planning and development and limited the ability of government and non-government organisations to formulate sound strategic policies. It is difficult to raise the economic profile of an industry that is so poorly served by statistics.

2. THE NATIONAL STATISTICAL INITIATIVE

In September 1993, the Sport and Recreation Ministers Council (SRMC) recognised the need to rectify this situation by establishing a Statistical Working Group (SWG). SRMC consists of Commonwealth, State and Territory Government Ministers responsible for sport and recreation in Australia. The objective of the Council is to coordinate the development of sport and recreation.

The aim of SWG is to improve the definition, range and quality of statistics for the Sport and Recreation Industry through the National Statistical Initiative (NSI). SWG roles include advising on the statistical needs of the industry, and initiating studies and the collection of sport and recreation data. This paper outlines major SWG activities including development of reports, initiation of surveys by the national statistical collection agency (the Australian Bureau of Statistics (ABS), how liaison is undertaken with the ABS and the development of computer software to access the available data collected by the ABS.

SWG has representatives from the Commonwealth (Department of the Environment, Sport and Territories, Australian Sports Commission and Australian Bureau of Statistics), each relevant State and Territory Government Agency and local government. With the assistance of a consultant, the Office of Recreation Development has undertaken most of the SWG work.

At its first meeting, SWG considered how to improve the definition, range and quality of industry statistics. It was agreed information was needed based on detailed statistical definitions and needs, that had been widely endorsed by potential users throughout the industry. It was further agreed that regular, reliable and independent data collections were required, which could only be provided by the national statistical collection agency, the Australian Bureau of Statistics (ABS).

Sport and recreation is not well served by ABS data collections, partly as sport and recreation is a new service industry which has only emerged in the past few decades. It is also because the industry started with low economic importance and has only gradually achieved its current status of major importance. Other industries with such status have long enjoyed the advantages of regularly collected ABS data. SWG appreciated that before ABS data collections could be unlocked, changes were needed to ABS definitional frameworks.

Guidance was sought from the experience of other countries and the advice of international agencies. In this regard, the Framework for Cultural Statistics (FCS), developed after more than a decade of review and international consultation by Unesco and the UN Statistical Office, was found to be particularly useful. In the late 1980's, this Framework was also used as a basis for a similar task on behalf of the Cultural Ministers Council (CMC) for the Cultural Services Industry.

CMC has a similar role and composition to the Sport and Recreation Ministers Council (SRMC). The objective of the Commonwealth, State and Territory Government Ministers responsible for culture in this Council is to coordinate the development of cultural services in Australia. CMC established a Statistical Advisory Group (SAG) to undertake its National Statistical Initiative.

3. FRAMEWORK DEVELOPMENT

At the time work commenced by SWG, an appropriately detailed Australian statistical classification did not exist for sport and recreation. Thus, the opportunity was taken to capitalise on the FCS work by adapting it to meet the particular needs of Australia. A key criterion used by Unesco for FCS was:

> *an appropriate statistical definitional framework should be an integrated whole, including both the social and economic aspects of cultural phenomena e.g. production, distribution and consumption of cultural goods and services.*

By adopting this holistic approach, a broad framework could be constructed to which all kinds of known information could be mapped. Because the most robust statistical systems, both internationally and in Australia, are those linked to economic activities, the national sport and recreation Framework was developed with a strong link to ABS economic, as well as social, ABS statistical collections.

The first task was to develop a *National Sport and Recreation Industry Statistical Framework,* acceptable to national and international statistical agencies, which defined the industry and its data needs, and established long term plans of how these needs should be met by the ABS. Criteria used to develop the first edition of the Australian Framework included that it should be:

- relevant to the needs of, and usable by, potential users, interest groups and government at all levels;
- an integrated whole, including both the social and economic aspects of sport and recreation e.g. the production, distribution and consumption of sport and recreation goods and services;
- free standing, and all components are recognisable as being part of the Sport and Recreation Industry (ensuring all aspects of the industry are included, sport and recreation activities are clearly and comprehensively defined, and that every activity can be clearly and unambiguously classified);
- all Sub-Sectors of the Framework are to be recognised as economically or socially significant in the Australian community;
- possible to collect economic and demographic statistical information for all components (including data on income, expenditure, employment, participation and consumption);
- compatible with national, and where appropriate international, statistical definitional frameworks and collection systems (so that meaningful comparisons can be made with data from ABS collections, from other countries or international agencies);
- flexible enough to accommodate changing needs of the industry and of users of sport and recreation statistics in the economic and social environment; and

- oriented towards the development of a viable Sport and Recreation Industry that recognises the importance of active and passive participation in sport and recreation to the quality of life of all Australians.

Features of the Framework

Key features of the first edition of the Australian Framework are:

- the industry is divided into five Sectors and 39 Sub-Sectors (the Sectors are Organised Sport, Active Recreation, Amusement and Passive Recreation, Gambling and Services to Sport and Recreation). There are also a further six Sectors and 22 Sub-Sectors of interest to sport and recreation in other industries (the Sectors in other industries are Construction, Retail and Wholesale Trade, Manufacturing, Finance and Insurance, Agriculture, Forestry and Fishing and Cultural-Services);

- each Sector and Sub-Sector is defined, along with its primary activities;

- statistical data needs are identified for each Sector and Sub-Sector under four functional elements i.e. participants/providers, services/products, organisations/facilities and consumers; and

- a long term statistical plan is provided for each Sector and Sub-Sector.

Hence, the Framework is a two dimensional matrix divided hierarchically by Sector and Sub-Sector (see Appendix 1). The second dimension is four functional elements to cover data needs in each industry segment from the creation of a service or product through to its end use (see example at Appendix 2), as follows:

- **Participants/Providers** - participants are those who participate in sport or recreational activities. The data need includes the frequency, duration, nature and level of such participation (whether as a player, coach, official, referee, etc.). The Framework covers all those involved on their own, in groups, clubs, teams or organisations whether on a professional, amateur or, simply, on a recreational basis. For providers, the data need includes the number, occupation and average income of the full, part time or casually employed providers of sport and recreation activities and services i.e. the managers, professionals, clerks, tradespersons, etc.;

- **Services/Products** - of interest are the products, goods and services produced by sport and recreation organisations. Included are the administration, coaching and other services of sport and recreation bodies to organise, promote and sponsor activities. The basic measure of this element is its economic value in dollar terms. Also included are services provided externally to each Sub-Sector by other industries such as television broadcasts of sport by the Communications Industry. Provision of other goods and services are covered under other industries of interest to sport and recreation. Such services include facilities construction, and the manufacture, import/export, wholesale, retail and distribution of goods and services for areas of sport and recreation interest e.g. magazines, television, radio, clothing, equipment, business services, etc.;

- **Organisations** - whilst many recreational activities are undertaken informally by individuals, families or groups, all organised sports and some recreational services are provided by organisations whether commercial, government, subsidised or voluntary. The role of these organisations includes provision of facilities and venues, organisation, presentation and promotion of events, training and development of participants. The basic measures of this element are the income and expenditure of organisations. Also included where appropriate, are the venues/facilities provided, a basic measure of which is the number and capacity of venues/facilities; and

- **Consumers** - this element identifies those in the community, and their characteristics (age, gender, etc.), who attend sporting events, listen to radio and watch television on sport and recreation events. It also includes those who attend amusement parks, horse and dog races, casinos or who buy lottery tickets. Also included are books, videos, magazines and other material purchased by the general public, and goods and services purchased by participants. Consumers of these products and services are not to be confused with active participants as defined above.

For each Sub-Sector, information in the Framework consists of:

- **a definition** - sport and recreation activities covered in the Sub-Sector which are precisely defined in line with ABS requirements. Where an organisation engages in activities belonging to more than one Sub-Sector, the organisation is included under its major activity. Organisations undertaking largely administrative or management functions ancillary to a specific Sub-Sector are generally included in that Sub-Sector;

- **statistical data needs under each of the four functional elements** - which covers the detailed data need from the creation of a service or product through to its end use for each Sub-Sector; and

- **statistical plan** - this provides a long term plan for collecting data to meet the national statistical needs of each Sub-Sector and functional element. Each plan ultimately seeks the use of an appropriate ABS collection.

4. THE OCCUPATIONAL STRUCTURE

SWG has also developed a report on employment in the *National Sport and Recreation Industry Occupational Structure.* This document was developed for consideration by the Sport and Recreation Ministers Council, as a basis for data collections and for a submission to the review of the Australian Standard Classification of Occupations (ASCO - currently being finalised). This skill based definitional framework is used by the Australian Bureau of Statistics as the national standard for production and analysis of labour force statistics, human resource management, educational planning, listing of job applications and vacancies, provision of occupational information and vocational guidance. The Structure will also contribute to the development of an industry training plan.

How employment in the industry might be defined is suggested in the Structure in terms of skill requirements and a range of descriptors for each industry occupation.

Its purpose is to establish specifications for the collection of national employment statistics by the Sport and Recreation Industry and Australian Bureau of Statistics. Again it is compatible with national and international statistical systems, and is based on the 1986 Unesco and UN Statistical Office *Framework for Cultural Statistics,* but adapted to meet Australian needs.

In ASCO, a **job** is a set of tasks performed by one individual in any given establishment. An **occupation** is a set of jobs with identical sets of tasks. It is recognised that every job is a little different. In practice, an occupation is a collection of jobs sufficiently similar in their main tasks to be grouped together for classification purposes. Most occupations in the Structure meet the other ABS criteria for inclusion as a distinct occupation in ASCO. These criteria include:

- the occupation is consistent with the criteria identified for ASCO and the International Standard Classification of Occupations (ISCO);

- a minimum of 300 people are employed full time and the proportion of the national population serviced by the occupation is large (so the occupation was clearly significant in the domestic labour force and national economy);

- a strong user demand exists for statistics in the occupation;

- the occupation is readily codeable in ABS collections i.e. incumbents would identify their occupation by that title and description; and

- the occupation is delineated on the basis of a clearly unique skill specialisation e.g. a waiter/waitress in a sport and recreation club is a unique skill specialisation. But it is not unique when compared to similar occupations in hotels, restaurants or cafes. Even if ASCO did not already have such an occupation, it would not be possible to amend ASCO to cover waiters/waitresses in these clubs. Fortunately, it will be possible to obtain this information from the ABS Population Census, in due course, by using the occupation and industry identifiers for sport and recreation clubs.

The Structure is in two sections, Section 1 - occupations within the industry and Section 2 - occupations in other industries of interest to sport and recreation. Like the Framework, it is a two dimensional matrix of vertical components (which divides the industry into major groups and individual occupations - see Appendix 3) and horizontal components (that covers a range of job descriptors - see example at Appendix 4). Key features of the first edition of the Structure are:

- it is a skill based structure using the concept of skill level to classify occupations into Major Groups and skill specialisation to delineate occupations below the Major Group level. The 1986 ASCO classification system is used (although the revised ASCO is likely to follow a similar system, changes are likely to improve the hierarchy of skill levels) as follows:

  Major Group 1 Managers and Administrators
  Major Group 2 Professionals
  Major Group 3 Para-Professionals
  Major Group 4 Tradespersons
  Major Group 5 Clerks

Major Group 6 Salespersons and Personal Service Workers
Major Group 7 Plant and Machine Operators and Drivers
Major Group 8 Labourers and Related Workers;

- within these groups, 146 separate occupations were identified - 100 for the five core Sport and Recreation Sectors and the remainder for activities in other industries of interest to sport and recreation;

- each occupation has been described on the basis of six indicators as follows:

  - **definition** of the specific occupation;

  - **tasks** which include the primary set of tasks required to undertake the occupation;

  - **specialisation** of the title of each of job classified within the occupation;

  - **skill levels** for formal education, on the job training and work experience;

  - **specialist skills** in a field of knowledge, tools and equipment or special techniques or skills used, materials, data, information or people worked on or with and goods or services including information produced;

  - **job descriptors** which may contain descriptors such as aptitudes, environmental conditions, physical abilities, educational and vocational development, labour market factors, interests and competencies (criteria to be used to develop these descriptors have yet to be established and so the descriptors are still to be developed); and

- within each Major Group, ASCO classifies like occupations into minor groups and unit groups principally on the basis of skill specialisation. Each occupation is identified under an appropriate ASCO Major and Unit group.

It is proposed that the Framework and Structure be periodically reviewed (every ten years or so - minor revisions might be undertaken more frequently, by the Australian Bureau of Statistics in consultation with appropriate Sub-Sector organisations of the industry, if reliable statistics indicate changes are needed). It is to be recommended that the Sport and Recreation Ministers Council agree the Australian Bureau of Statistics be requested to implement the long term statistical plans of each Sub-Sector in the Framework and take account of the Structure in undertaking data collections on industry employment.

5. AVAILABLE DATA AND SOURCES

To support justification of industry proposals to the ABS, and for policy development and implementation purposes, a third report was prepared on *Available Data and Sources for the Sport and Recreation Industry*. This publication identifies data currently available to meet the statistical needs of the Framework and establishes short term plans on how the industry might seek to overcome the deficiencies until the ABS can implement the long term plans of the Framework.

Consultation was mainly conducted with over 120 national sport and recreation organisations, who were consulted on three occasions. Government agencies and other industry representatives, including academia, were also widely consulted in each state and territory. The three reports are to be submitted to SRMC for endorsement before being presented to the ABS for implementation and are available in AGPS Bookshops.

6. OTHER SWG WORK

Other work being co-ordinated by SWG in the period 1993/96 includes:

- *National Sport and Recreation Industry Directory and Database (NSRIDD)* - which provides computer access to all industry data collected by ABS e.g. from the Population Survey Monitor (PSM), Population Census, Service Industry, Household Expenditure, Time Use, Manufacturing, Wholesale, Retail and Labour Force Surveys, and Foreign Trade and Public Accounts Statistics, etc.. NSRIDD is located in Commonwealth, State and Territory Sport and Recreation Government Agencies and may be accessed by all Australians. The database is to be updated annually. Although ABS information is still limited, NSRIDD Year Two had 30 MB of data, and the Year three version is likely to have about 100 MB of tables at the national, state/territory and national/state levels;

- from August 1993, quarterly ABS surveys on sport and recreation under the Population Survey Monitor (PSM). Each PSM covers about 2000 households nationally (for an annual sample of over 20 000 persons). The sample was doubled in August 1995 (annual sample of over 40 000 persons). Each survey has core questions for household residents aged 15 years and over. Other questions are on sport and recreation participation/services/consumption activities (some apply to residents aged 7 to 14 years). The ABS trained interviewers ask questions for 7 to 14 year olds from an adult. The 1993/94 PSM results on sport and recreation participation, expenditure, reasons for non-participation, spectators, school sport, etc., are in NSRIDD Year Two;

- Service Industry Surveys of the Sport and Recreation Industry to obtain information on the income, expenditure, employment and services provided by a minimum of 30 per cent of establishments in the relevant industry Sub-Sectors. Results of these surveys should be available by the end of 1996; and

- liaison with the ABS through the recently established National Culture Recreation Statistical Unit (NCRSU) located in the ABS. This unit assists SWG to seek inclusion of the industry in appropriate ABS classifications and other standards, relevant ABS collections and the ABS Database.

In summary, the National Statistical Initiative is designed to develop greater industry understanding based on reliable data; establish trends in performance; identify priority needs and actions to maintain the industry, and prospects for new facilities and programmes; analyse implications of changes in the level, mechanisms and distribution of public and private funding; and develop reliable statistics consistent with those of other nations, to compare our performance.

APPENDIX 1

SECTORS AND SUB-SECTORS OF AUSTRALIAN NATIONAL SPORT AND RECREATION INDUSTRY STATISTICAL FRAMEWORK

The Sport and Recreation Industry

Organised Sport

- Australian Football
- Basketball
- Baseball
- Cricket
- Golf
- Hockey
- Indoor Cricket
- Lawn Bowls
- Netball
- Rugby League
- Rugby Union
- Soccer
- Squash
- Tennis
- Other Team Sports
- Individual Sports (not elsewhere classified)

Active Recreation

- Aerobics, Gymnastics and Other Fitness
- Air Based Active Recreation
- Boating and Yachting
- Camping and Caravanning
- Horse Riding
- Martial Arts
- Snow Skiing
- Swimming
- Ten Pin Bowling
- Water Safety and Underwater Diving
- Other Active Recreation

Amusement and Passive Recreation

- Amusement Parks and Venues
- Horse and Dog Racing
- Motor Sports
- Other Passive Recreation

Gambling

- Casinos
- Lotteries
- Wagering

Services to Sport and Recreation

Adaptive Sport and Recreation
Government and General Sport and Recreation Organisations
Sports Medicine and Exercise Science
Sport and Recreation Education
Other Sport and Recreation Clubs

Sport and Recreation in Other Industries

Construction

Swimming Pool Construction
Other Sport and Recreation Facilities Construction

Retail Trade

Boat and Marine Equipment Retailing
Camping and Bushwalking Goods and Equipment Retailing
Caravan Retailing
Cycling Goods and Equipment Retailing
Fishing Goods and Equipment Retailing
Gardening Supplies and Equipment Retailing
Golf Goods and Equipment Retailing
Shooting Goods and Equipment Retailing
Snow and Water Skiing Goods and Equipment Retailing
Underwater Dive/Surfing Goods and Equipment Retailing
Other Sport and Recreation Goods, Equipment and Clothing Retailing

Wholesale Trade

Amusement and Gambling Equipment Wholesaling
Other Sport and Recreation Goods, Equipment and Clothing Wholesaling

Manufacturing

Amusement and Gambling Equipment Manufacturing
Boat and Marine Equipment Manufacturing
Caravan and Camping Equipment Manufacturing
Surfboard/Canoe/Small Water Craft Manufacturing
Other Sport and Recreation Goods, Equipment and Clothing Manufacturing

Finance and Insurance

Sport and Recreation Insurance

Agriculture, Forestry and Fishing

Horse Breeding

Cultural Services

The Cultural Services Industry

APPENDIX 2

EXAMPLE FOR ONE SUB-SECTOR OF AUSTRALIAN NATIONAL SPORT AND RECREATION INDUSTRY STATISTICAL FRAMEWORK

Sub- Sector: Basketball **Sector: Organised Sport**

Definition

This Sub-Sector consists of units or establishments (clubs, associations or other bodies) whose primary activities are the operation and maintenance of venues and facilities, organisation of matches and provision of associated services including coaching, promotion and publicity for the game of basketball, including junior variations such as Mini Basketball, whose rules are defined by the Federation Internationale de Basketball Amateur (FIBA).

Included in this Sub-Sector are basketball players, providers and consumers of the above services, and venues, facilities, licensed clubs, kiosks, and associated amenities for players, spectators and public at local, club, district, interstate, national and international levels, whether professional or amateur.

Activities in establishments where basketball is a secondary activity, such as in schools and in other industries of the national economy, are excluded from this Sub-Sector with the following exceptions. Based on a request from Basketball Australia, some limited information is to be sought in this Sub-Sector on basketball in schools and the contribution of services from selected other industries of special interest.

Participants /Providers

Participants are professional, amateur, recreational and school basketball players, coaches, referees, development officers, officials, etc..

Data requirements for participants are the number, gender, age, education and income level, ethnicity, occupation, frequency, duration and type of participation, and reasons for non-participation or discontinuing participation, and participation in major events.

Providers are defined as those employed in Sub-Sector for whom pay as you earn (PAYE) tax is paid in the following major occupations (minor occupations are shown in Appendix 2 of Framework and volunteers at Appendix 3):

- Basketball Centre Managers/Administrators;
- Coaches Development Officers (Basketball);
- Other Professional Sportspersons (includes Professional Basketballers);
- Referees;
- Sports Trainers; and

- Other Sport and Recreation Venue/Facility Curators.

Data requirements for providers are the number employed (FTE) and income annually in each occupation.

Services

Services provided internally by organisations in Sub-Sector are categorised as:

- coaching, training, etc., talent identification/development of participants;
- staging and participating in events/competitions;
- provision of material, to promote basketball;
- organisation of sponsors for teams, clubs, venues, facilities and games; and
- administrative services.

Services provided externally to Sub-Sector by other industries categorised as:

- coaching, training, etc., staging and participating in matches at schools and tertiary education establishments;
- television broadcasts (number/duration);
- provision/maintenance of grounds/other facilities by local government;
- videos, books, magazines and other material on basketball; and
- equipment and clothing (balls, shoes, singlets, etc.) used by participants.

Data requirements are the annual services provided and cost (national and state) in each category.

Organisations

Organisations are clubs and other national, state and district associations that operate basketball venues/facilities, and organise matches, categorised on a national, state and local basis as:

- number of organisations and teams;
- paid staff and volunteers;
- income from government, ticket sales, social/licensed clubs, membership fees, sponsors and advertising, and other (e.g. media rights);
- expenditure on salaries and wages, goods and services, promotion, advertising, maintenance and other (including travel);
- capital expenditure on venues or facilities, goods/equipment; and
- venues and courts (indoor and outdoor).

Data requirements are the number of organisations, staff and venues, and income and expenditure annually.

Consumers

People who consume the services provided internally by Sub-Sector organisations are categorised as:

- attendance by the general public at international, interstate and other matches (including frequency of attendance);
- non-attendance at the above matches (with reasons); and
- attendance at coaching classes.

People consuming services provided externally by other industries to Sub-Sector organisations, categorised as:

- basketball on television (number of viewers);
- purchase by participants and the general public of videos, books, magazines and other material on basketball (national and state annual cost); and
- purchase by participants of basketball equipment and clothing (e.g. balls, shoes, uniforms, etc. - national and state annual cost).

Data requirements are annual number in each category, and cost of purchases.

Statistical Plan

The long term plan to meet the main data needs of this Sub-Sector is as follows:

Participants/Providers - obtain data for Participants from ABS Population Survey Monitor and Labour Force Surveys and for Providers from ABS Population Census and Service Industry Surveys.

Services - obtain from ABS Service Industry Surveys.

Organisations - as for Services.

Consumers - obtain from ABS Population Survey Monitor, and Labour Force Supplementary and Household Expenditure Surveys.

Note: The above Australian Bureau of Statistics (ABS) collections are only the major means of meeting the statistical needs of Sub-Sector. Relevant information will also be available from many other ABS collection programmes which are too numerous to detail here. As with other Sub-Sectors of the national economy, ABS collections may rely on appropriately collected data from Sub-Sector organisations in some instances.

APPENDIX 3

OCCUPATIONS IN AUSTRALIAN NATIONAL SPORT AND RECREATION INDUSTRY OCCUPATIONAL STRUCTURE

THE SPORT AND RECREATION INDUSTRY

1. **Managers and Administrators**
   Amusement Park and Venue Managers/Administrators
   Basketball Centre Managers/Administrators
   Bookmakers
   Caravan Park/Campsite Managers/Administrators
   Casino Managers/Administrators
   Football Managers/Administrators
   Fitness Centre/Gymnasium Managers/Administrators
   Horse and Dog Racing Managers/Administrators
   Indoor Cricket Centre Managers/Administrators
   Motor Sports Speedway/Circuit Managers/Administrators
   Multi-Purpose Sport and Recreation Centre Managers/Administrators
   Squash Centre Managers/Administrators
   TAB Agency Managers/Administrators
   Ten Pin Bowling Centre Managers/Administrators
   Other Sport and Recreation Venue/Centre/Association/Unlicensed Club Managers/Administrators
   Lawn Bowls Club Managers/Administrators
   Golf Club Managers/Administrators
   Boat/Yacht Club Managers/Administrators
   Other Sport and Recreation Licensed Club Managers/Administrators
   Public Policy Managers (Sport and Recreation)
   Boat Shop Managers/Administrators
   Camping and Leisure Shop Managers/Administrators
   Cycling Shop Managers/Administrators
   Fishing Shop Managers/Administrators
   Gardening Shop Managers/Administrators
   Gun Shop Managers/Administrators
   Underwater Dive/Surf Shop Managers/Administrators
   Other Sport and Recreation Shop Managers/Administrators
   Wholesalers (Sport and Recreation Goods and Equipment)
   Swimming Pool Builders
   Professional Builders (Sport and Recreation Facilities)

2. **Professionals**
   Sports Medicine Practitioners
   Sports Medicine, Fitness and Recreation Counsellors
   Sports Physiotherapists
   Sports and Exercise Science Professionals
   Sport and Recreation Officers
   Veterinarians (Horse and Dog Racing)
   Aviation Instructors (Sport and Recreation)

Fitness Instructors/Advisers
Other Sport and Recreation Teachers/Instructors
Accountants (Sport and Recreation)
Announcers/Commentators (Sport and Recreation)
Journalists/Reporters (Sport and Recreation)
Lawyers (Sport and Recreation)
Marketing/Public Relations Officers (Sport and Recreation)
Naval Architects (Sporting and Leisure Boats)
Photographers (Sport and Recreation)
Technical Producers (Sport and Recreation)

3. **Para-Professionals**
Coaches/Development Officers (Basketball)
Coaches/Instructors (Canoeing)
Coaches/Instructors (Equestrian/Horse Riding)
Coaches/Instructors (Football)
Coaches/Instructors (Gymnastics)
Coaches (Martial Arts)
Coaches/Instructors/Patrollers (Snow Skiing)
Coaches (Swimming)
Coaches (Tennis)
Coaches/Instructors (Underwater Diving)
Outdoor Education Instructors and Adventure Leaders
Fitness/Aerobics Leaders
Sports Trainers
Teachers/Instructors (Swimming)
Other Coaches (Sport and Recreation)
Community Workers (Sport and Recreation)
Recreational Charter Boat Skippers
Jockeys
Professional Footballers
Professional Golfers
Other Professional Sportspersons
Horse and Dog Racing Officials
Swimming Pool Superintendents
Driving Instructors (Sport and Recreation)
Television Equipment Operators (Sport and Recreation)

4. **Tradespersons**
Animal Trainers (Sport and Recreation)
Casino Inspectors
Casino Pit Bosses
Chefs/Cooks (Sport and Recreation Venues/Clubs/Facilities)
Farriers (Sport and Recreation)
Gardeners (Sport and Recreation Facilities)
Head Gardeners (Sport and Recreation Facilities)
Apprentice Gardeners (Sport and Recreation Facilities)
Greenkeepers
Head Greenkeepers
Apprentice Greenkeepers
Horse Trainers (Sport and Recreation)

Horse and Dog Racing Curators/Attendants
Other Sport and Recreation Venue/Facility Curators
Caravan Repairers/Technicians
Ski Lift Operators/Maintenance Workers
Slot Machine Repairers/Technicians
Ten Pin Bowls Centre Maintenance Workers
Canvas Goods Makers (Sport and Recreation)
Horse Breakers (Sport and Recreation)
Landscape Gardeners (Sport and Recreation Facilities)
Motor Sports Mechanics
Sailmakers
Snooker, Billiard or Pool Table Makers
Sport and Recreation Boat Builders and Repairers
Supervisors Sport and Recreation Boat Builders
Surfboard/Sailboard/Canoe Builders and Repairers
Swimming Pool Construction Workers

5. **Clerks**
Administrative Assistants (Sport and Recreation Organisations)
Bookmaker Clerks
Data Processing Operators (Sport and Recreation Organisations)
Inquiry Clerks (Sport and Recreation Organisations)
Receptionists (Ten Pin Bowls Centres)
Receptionists (Sport and Recreation Organisations)
Telephone Betting Operators

6. **Salespersons and Personal Service Workers**
Bar Attendants (Sport and Recreation Clubs/Venues)
Supervisors Bar Attendants (Sport and Recreation Clubs/Venues)
Croupiers/Dealers
Massage Therapists
Office Cashiers (Sport and Recreation Organisations)
TAB Agency Clerks
Ticket Sellers (Sport and Recreation Organisations)
Waiters and Waitresses (Sport and Recreation Clubs/Venues)
Supervisors Waiters and Waitresses (Clubs/Venues)
Other Personal Service Workers (Sport and Recreation)
Sales Representatives (Fishing Equipment)
Sales Representatives (Games and Toys)
Sales Representatives (Other Sport and Recreation Goods)
Sales Assistants (Bicycles)
Sales Assistants (Camping Equipment)
Sales Assistants (Caravans)
Sales Assistants (Fishing Equipment)
Sales Assistants (Gardening Supplies)
Sales Assistants (Games and Toys)
Sales Assistants (Gun Shops)
Sales Assistants (Underwater Dive/Surf Goods)
Sales Assistants (Other Sport and Recreation Goods)

7. **Plant and Machine Operators and Drivers**
   Stationary Engine and Boiler Operators (Sport and Recreation Facilities)
   Stationary Plant Operators (Sport and Recreation Facilities)

8. **Labourers and Related Workers**
   Animal Attendants
   Caretakers (Sport and Recreation Centres/Venues)
   Cleaners Sport and Recreation Centres/Venues)
   Door Attendants (Sport and Recreation Centres/Venues)
   Entertainment Ushers (Sport and Recreation Centres/Venues)
   Recreational Charter Boat Deckhands
   Garden Labourers (Sport and Recreation Centres/Venues)
   Horse Stable Hands/Strappers
   Kitchen Hands (Sport and Recreation Centres/Venues)
   Professional Lifeguards
   Security Officers (Sport and Recreation Facilities)
   Ticket Takers (Sport and Recreation Centres/Venues)
   Stud Hands (Horse Breeding)
   Storepersons (Sport and Recreation Centres/Venues)

Sport and Recreation in Other Industries

1. **Managers and Administrators**
   Finance Managers (Sport and Recreation)
   Personnel Managers (Sport and Recreation)
   Executive Producers (Sport and Recreation)
   Purchasing Managers (Sport and Recreation)
   Training Managers (Sport and Recreation)
   Information Systems Managers (Sport and Recreation)
   Research Managers (Sport and Recreation)
   Horse Stud Managers
   Canteen Managers (Sport and Recreation)
   Resort Managers (Sport and Recreation)
   Hostel Managers (Sport and Recreation)
   Bar Managers (Sport and Recreation)

2. **Professionals**
   Analytical Chemists (Horse and Dog Racing)
   Food Technologists (Sport and Recreation)
   Nutritionists (Sport and Recreation)
   Anatomists (Sport and Recreation)
   Physiologists (Sport and Recreation)
   Medical Laboratory Scientists (Sport and Recreation)
   Biomedical Engineers (Sport and Recreation)
   Urban and Regional Planners (Sport and Recreation)
   Water Treatment Engineers (Sport and Recreation)
   Electronics Engineers (Gambling Equipment)
   Air Conditioning Engineers (Sport and Recreation)
   Aeronautical Engineers (Sport and Recreation)
   Occupational Therapists (Sport and Recreation)

Physical Education Teachers (Sport and Recreation)
Remedial Teachers (Sport and Recreation)
University Lecturers (Sport and Recreation)
University Tutors (Sport and Recreation)
Computer Systems Programmers (Lotteries and TAB)
Computer Systems Analysts (Lotteries and TAB)
Database Administrators (Lotteries and TAB)
Librarians (Sport and Recreation)
Management Consultants (Sport and Recreation)
Graphic Designers (Sport and Recreation)
Occupational Psychologists (Sport and Recreation)
Sociologists (Gambling)
Market Research Analysts (Sport and Recreation)

3. **Para-Professionals**
Fisheries Technical Officers (Sport and Recreation)
Artificial Insemination Technicians (Horse Breeding)
Electronics Engineering Associates (Gambling Equipment)
Mechanical Engineering Associates (Sport and Recreation)
Lighting Designers (Sport and Recreation)
Community Development Officers (Sport and Recreation)
Youth Workers (Sport and Recreation)
Lighting Supervisors (Sport and Recreation)
Sound Technicians (Sport and Recreation)
Child Care Coordinators (Sport and Recreation)
Referees
Umpires
Linesmen/Women
Handicappers (Racing)

4. **Tradespersons**
Maintenance Fitters (Sport and Recreation)
Aircraft Maintenance Engineers (Sport and Recreation)
Automotive Electricians (Sport and Recreation)
Air Conditioning Mechanics (Sport and Recreation)
Broadcasting Technicians (Sport and Recreation)
Computer Technicians (Sport and Recreation)
Motor Cycle Mechanics (Sport and Recreation)
Outboard Motor Mechanics (Sport and Recreation)
Vehicle Body Makers (Sport and Recreation)
Bakers (Sport and Recreation)
Supervisors Plant Nurseries (Sport and Recreation)
Nurserymen/Women (Sport and Recreation)
Tree Surgeons (Sport and Recreation)
Cabinetmakers (Billiard Tables)
Shipwrights (Sport and Recreation)
Loftsmen/Women (Sport and Recreation)
Apprentice Boat Builders (Sport and Recreation)
Tailors (Sport and Recreation)
Dressmakers (Sport and Recreation)
Leathergoods Makers (Sport and Recreation)

5. **Clerks**
Office Secretaries (Sport and Recreation)
Office Typists (Sport and Recreation)
Word Processing Operators (Sport and Recreation)
Bookkeepers (Sport and Recreation)
Account Clerks (Sport and Recreation)
Payroll Clerks (Sport and Recreation)
Filing Clerks (Sport and Recreation)
Dispatch Clerks (Sport and Recreation)
Stock Clerks (Sport and Recreation)
Purchasing Clerks (Sport and Recreation)
Telephonists (Sport and Recreation)

6. **Salespersons and Personal Service Workers**
Insurance Agents (Sport and Recreation)
Office Cashiers (Sport and Recreation)
Child Care Attendants (Sport and Recreation)

7. **Plant and Machine Operators and Drivers**
Delivery Drivers (Sport and Recreation)
Air Conditioning Plant Operators (Sport and Recreation)
Plastics Machine Operators (Sport and Recreation)
Wood Processing Machine Operators (Sport and Recreation)
Fabric Machine Operators (Sport and Recreation)
Textile Sewing Machinists (Sport and Recreation)

8. **Labourers and Related Workers**
Supervisors, Cleaners (Sport and Recreation)
Concrete Workers (Swimming Pools)
Supervisors, Guards and Security Officers (Sport and Recreation)

APPENDIX 4

EXAMPLE FOR ONE OCCUPATION OF AUSTRALIAN NATIONAL SPORT AND RECREATION INDUSTRY OCCUPATIONAL STRUCTURE

Occupation: Sport and Recreation Officers **Major Group: Professionals**

Definition

Sport and Recreation Officers plan, organise, coordinate and promote the use of sport and recreation events, facilities and programmes.

ASCO Groupings

Major - Professionals.

Minor - Miscellaneous Professionals.

Unit - Other Professionals.

Tasks

Study and analyse recreation needs and resources.

Develop and implement recreation management policies for government and community, the young, aged and disabled groups.

Offer technical and professional advice to authorities and others concerned with providing recreation facilities.

Organise and supervise use of recreation facilities, programmes, activities and special events. Prepare reports and assist in budget/policy formulation.

Skill Levels

Formal education - recreation management diploma.

On job training - three years.

Work experience - three to five years after completing formal education.

Specialisations

Recreation Advisers.
Activities Officers.
Community Recreation Officers.
Sport or Recreation Development or Promotions Officers.
Specialist (e.g. disabled, aged, youth) Sport or Recreation Officers.
Sport or Recreation Events Organisers.

Specialist Skills

Field of knowledge - managing sport and recreation activities.

Skills used - coordination, media promotion, marketing, public relations.

Works with - local government, community groups.

Services produced - community sport and recreation activities and events.

Job Descriptors

Part II of the Occupational Structure is to be developed later but is likely to contain information on seven other job descriptors i.e. aptitudes; environmental conditions; physical abilities; educational and vocational development; labour market factors; interests; and competencies.

DOING MATHS STANDING ON YOUR HEAD

Alex McNabb[1]

Abstract

A summary is presented of the mechanical constraints governing the motion of the centre of gravity of a gymnast and the rate of rotation around this balance point. The role of elastic interactions between the gymnast and the equipment is studied for basic single point interactions and applied to vaulting, tumbling and circles around a bar. Optimal control type strategies focussing on the strength and flexibility constraints and switching points from one constraint boundary to another are suggested as useful domains for coaches to explore.

1. INTRODUCTION

I watched the Atlanta Olympics gymnastic events with amazement and in some events almost a sense of disbelief. It is difficult to define a gymnasts objective on any one piece of apparatus except to say he or she wishes to be awarded the highest score, and even the judging rules keep changing over the years. Marks are assessed on performance, aesthetic appeal, conformity to certain vague and changing compositional requirements and a criterion called degree of difficulty. The degree of difficulty is assessed from an historical perspective and so spectacular new moves can gain bonus points in this department before imitators have an opportunity to relegate them to the too-easy basket.

For any operation with a well defined objective involving a precisely specified system, the best results are solutions of an optimal control problem. These solutions are usually strategies for guiding the operational parameters around their constraints in the systems parameter space. For most linear and many non-linear systems, the optimal strategies are paths along the constraint boundaries with precisely defined switches from one boundary to another. For example, to drive a car from stationary at A to stationary at B in the shortest time, the obvious solution is maximum acceleration from A to C and maximum deceleration from C to B. If the switch point C has been chosen correctly, the car will just come to a halt at B in the shortest possible time.

Although a gymnastic routine has no precisely-defined objective, it will be a sequence of "tricks" and many of these will have an objective involving say a flight path attaining greatest height or perhaps involving as many rotations about some axis as possible. It is therefore useful for a gymnastic coach to be aware of this constraint-boundary aspect of optimal control strategies. The important constraints in the gymnasium are firstly the gymnasts internal constraints, fitness, flexibility and

---

[1]Department of Mathematics, Tamaki Campus, University of Auckland, New Zealand

strength, secondly the external constraints of the performance environment, such as the dimensions and kinematic properties of the equipment, and finally, the laws of mechanics, which ultimately determine what is or isn't possible.

Some of these the coach can change and since the outcome of an optimal strategy is affected by such changes it is useful to have some idea as to which changes are possible and the sensitivity of the outcome to these changes. It is even desirable to have an "optimal strategy" for accomplishing the changes themselves.

It is obvious from a historical study of gymnastic performances of Olympic champions that fine tuning gymnastic equipment has had a great effect on performance and composition of exercises. The catapult-like response of modern horizontal bars and women's high and low bars allows release and catch moves like grand circle - fly away above bars - regrasp and continue in grand circle, or one arm grand circles which were talked about fifty years ago in jest as gymnastic dreams or nightmares. Sprung floors, beat boards, even sprung beams and box tops have raised performance expectations to mind-boggling levels. The coach and gymnast must of course accommodate their expectations to the equipment available. There is no way in which the laws of mechanics can be changed, but a good understanding of these laws and the equipment properties will help to find optimal performance strategies and qualitatively assess the outcome sensitivity to the internal constraints of strength and flexibility in appropriate limbs.

I know a coach is expected to be a gymnast, a weight lifter, a mind reader, a doctor, psychologist, an accountant, a club manager, a politician and a general dogs body, but asking a coach to be an engineer, physicist and mathematician too might be a bit much.

In this paper, the nature of the laws of mechanics and basic gymnast-equipment interaction are discussed and the study is directed towards a rule-of-thumb approach for finding optimal performance strategies. Newton's laws focus attention on the position and motion of the gymnast's centre of gravity or balance point, rates of rotation around this and the forces and moments generated by the gymnast's interaction with the environment. In most situations this interaction is a single or two point elastic interaction involving swinging from a bar or bouncing a beat board or sprung floor and the main focus of this paper is a simple model for these "one point" elastic interactions.

2. NEWTONIAN CONSTRAINTS

For the purpose of mechanical analysis, the gymnast may be considered as a system of particles of mass m_i, at position $r_i(t)$ at time t governed by internal interactive forces maintaining the integrity of the system and driven by external forces such as gravity and equipment contact forces. If F_{ij} is the force of interaction of the j th particle on particle i for $j \neq i$, and F_{ii} is the sum of the external forces on particle i, then from Newton's 2nd and 3rd laws

$$m_i \ddot{r}_i = \sum_j F_{ij} \text{ and } F_{ij} + F_{ji} = 0 . \tag{2.1}$$

If $\bar{r} = \sum_j m_i r_i \Big/ M$, $M = \sum m_i$, is the centre of gravity of the system then

$$M\ddot{\bar{r}} = \sum_i \sum_j F_{ij} = \sum_i F_{ii} = F_e \tag{2.2}$$

where F_e is the sum of the external forces on all the particles.

This is the classical derivation (Bullen [1]) of the well known and powerful result that the centre of gravity $\bar{r}$ moves as a particle of mass M equal to the total body mass driven by the external forces only.

This result gives rise to some useful coaching "rules of thumb". Make sure the centre of gravity is in the right place at all time. Relinquish contact with the environment as late as possible and re-establish contact as early as possible. Flexibility may be the major constraint limiting late release or early contact.

If we write $r_i = \bar{r} + s_i$ so that $\sum m_i s_i = 0$ and define h_G the moment of momentum of the body about the centre of gravity as $h_G = \sum_i s_i \wedge [\mathrm{s_i}]$, then

$$h_G = \frac{d}{dt}\sum_i s_i \wedge \left[m_i\left(\dot{\bar{r}} + \dot{s}_i\right)\right] = \sum_i s_j \wedge \left[m_i\left(\ddot{r}_i\right)\right] = \sum_i \sum_j s_i \wedge F_{ij} \,. \tag{2.3}$$

Now from Newton's third law $F_{ij} + F_{ji} = 0$ and $\left(s_i - s_j\right) \wedge F_{ij} = 0$ which is the sum of the internal moments about G the centre of gravity.

We can write $\dot{\mathrm{s}}_\mathrm{i}$ as a vector $\omega_i \wedge s_i$ normal to s_i plus a vector $k\, s_i$ parallel to s_i. In fact if $\omega_i = (s_i \wedge \dot{s}_i)/(s_i \cdot s_i)$ we have

$$\dot{s}_i = \frac{\left(s_i \wedge \dot{s}_i\right)}{\left(s_i \cdot s_i\right)} \wedge s_i + \frac{\left(s_i \cdot \dot{s}_i\right)}{\left(s_i \cdot s_i\right)} s_i = \omega_i \wedge s_i + ks \;:$$

Hence $h_G = \sum_i s_i \wedge \left[m_i \omega_i \wedge s_i\right] = \sum_i m_i (s_i \cdot s_i)\ \omega_i - \sum_i m_i (s_i \cdot \omega_i)\ s_i$

$$= \sum_i I_{G\,i}\ \omega_i \equiv I_G \bar{\omega}_G \tag{2.4}$$

if we define $\bar{\omega}_G$ as the mean value from equations (2.4)

If the body were momentarily frozen in shape, I_G would be its moment of inertia about G and $\bar{\omega}_G$ would be the spin vector describing its rotation rate about the centre of gravity. The gymnast has some control over I_G and how much depends on strength in various muscle grasps and flexibility. He or she can therefore effect a change in $\bar{\omega}_G$ but the behaviour of h_G is governed entirely by the external moment M_{Ge} about G.

The external forces and moments derive from gravity and the apparatus interactions. Gravity produces no contribution to M_{Ge} since

$$\sum s_i \wedge m_i \underline{g} = \left(\sum m_i s_i\right) \wedge \underline{g} = 0. \quad (2.5)$$

The external forces driving the centre of gravity are $-Mg$ and F_e from the apparatus interaction. The gymnast has no control over M or g but can affect F_e in many ways.

3. FLIGHT PHASES

In the simple case of free flight when $F_e = M_{Ge} = 0$ the gymnast has no influence over the path of his or her centre of gravity G, and no influence on h_G. However it is possible to change I_G within a range $I_{G\min}$ to $I_{G\max}$ determined by the gymnasts flexibility (and strength, since some positions of minimum moment of inertia require great strength to hold when ω is large). This gives control over $\overline{\omega}_G$ within the range $h_G/I_{G\max}$, $h_G/I_{G\min}$. All the kinetic energy and moment of momentum requirements for the flight must therefore be met before release. During the flight phase, the motion of the centre of gravity is parabolic and its maximum altitude and horizontal velocity is predetermined by conditions at the last release point. The components of the vector h_G along any line of fixed orientation through G stays constant, but the rotation rate about the line changes as the inertial matrix I of the gymnast about G is changed by changing body shape. Tight tucks produce rapid rotation and full body extension reduces the rotation rate to a minimum. Landing techniques are used to reduce h_G to zero in a fashion. Steps after landing, arm waving and secondary steps and jumps are considered to warrant major deductions. In general, the flight phase can accommodate a single or double somersault in many ways but again, for aesthetic appeal, it is usually performed in "bang-bang" fashion with fastest rotation and tightest tuck earliest in the flight followed by full extension and secondary control measures into the landing.

By changing the inertia matrix non-symmetrically with respect to the spinal axis of the gymnast, it is possible to re-orient the body so that somersaulting becomes twisting or a combination of both, and visa versa. This does not produce any change in h_G . Twisting effects can be augmented by twisting motion on launch, but on the floor and the vaulting horse this means having to live with the same twisting motion in the landing since the take-off and the landing positions are similar. It is difficult to aesthetically kill this twisting motion while landing and hazardous to the ankles. It is safer to use other twisting techniques.

4. SINGLE POINT ELASTIC INTERACTIONS IN A VERTICAL PLANE.

Many gymnastic moves such as circles on a bar, tumbling moves like hand springs or somersaults, and take-offs from a beat board involve the hands or feet interacting with an elastically-responding bar, sprung floor or beat board and can be treated as motion in a vertical plane normal to the line of contact of the hands or feet with the equipment. The symmetry of the manoeuvre and equipment with respect to this plane allows the contact between the gymnast and the apparatus to be treated as a single-point interaction, and the motion as two dimensional.

Suppose this point of contact P responds elastically to the interaction stresses F_P helping to drive the manoeuvre. If the point P is at 0, the origin of our plane of motion when the interaction stresses are zero, and at (X,Z) when the horizontal and vertical components of F_P are F_X, F_Z, then we will assume,

$$F_X = -k_X X\,,\ F_Z = -k_Z Z\,. \tag{4.1}$$

For a bar we can take $k_X = k_Z$ and for a sprung floor and beat board, k_X will be assumed very large so that X is negligible and the elastic displacements are essentially vertical.

We can regard the move strategy as a process of controlling the position of G, the gymnast's centre of gravity relative to P and controlling I, the moment of inertia of the gymnast about G during each phase of the move in order to produce some desired best output at the end of the phase.

One or both of these changes can be brought about by tucks or pikes. It is strength which governs the speed and effectiveness of these changes and during such switches the spectrum of switching strategies may be considered as limited by the strength constraints. Let x,z be the coordinates of G, R denote the distance from P to G as in figure 1, and θ the angle this line makes with the horizontal, so that

$$x = X + R\cos\theta,\ \ z = Z + R\sin\theta \tag{4.2}$$

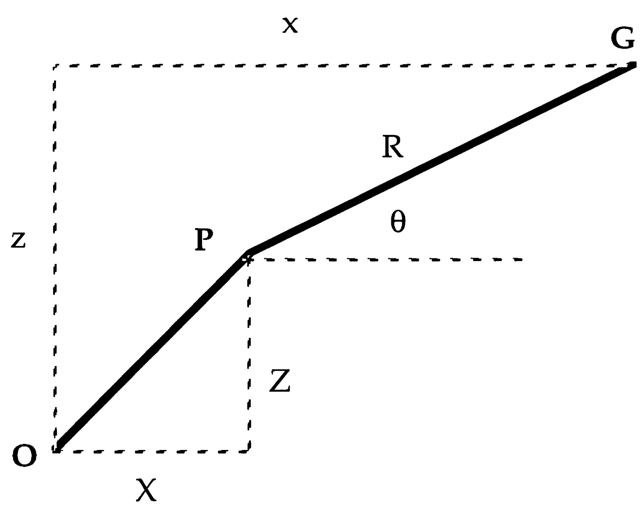


Figure 1.

The external forces driving G are, $-k_x X$ horizontally and $-k_z Z - Mg$ vertically and hence

$$M\ddot{x} = -k_x X,\ \ M\ddot{z} = -k_z Z - Mg \tag{4.3}$$

Optimal strategies will often involve rotation phases where the moment of inertia I_G and R are as large or small as possible and the gymnast rotates as a rigid body. Many circle moves on the bar involve a rigid body rotation starting with G vertically above it. When G has rotated 180° and is at the bottom a fast switch is made to bring G closer to the bar and reduce I+MR^2 as much as possible. This position is then held till G reaches its highest point over the bar, when a switch is made to return to the starting position. During such rotation phases

$$\begin{aligned} I\theta &= k_z ZR\cos\theta - k_x XR\sin\theta \\ &= -M\left[\ddot{Z} + R\left(\cos\theta\,\ddot{\theta} - \sin\theta\,\dot{\theta}^2\right) + g\right]R\cos\theta + M\left[\ddot{X} - R\left(\sin\theta\,\ddot{\theta} - \cos\theta\,\dot{\theta}^2\right)\right]R\sin\theta \end{aligned} \tag{4.4}$$

so that

$$[I + MR^2]\ddot{\theta} = -M[\ddot{Z} + g]R\cos\theta + M\ddot{X}R\sin\theta \tag{4.5}$$

If the X and Z displacements are small and $\ddot{X}$ and $\ddot{Z}$ are neglected we get the pendulum equation for θ,

$$[I + MR^2]\ddot{\theta} = -MgR\cos\theta \tag{4.6}$$

In general, we have a coupled system of 3 second-order ordinary differential equations in θ, x and z of the form

$$\begin{aligned} I\ddot{\theta} &= kR(z\cos\theta - x\sin\theta) \\ M\ddot{x} &= -k(x - R\cos\theta) \\ M\ddot{z} &= -Mg - k(z - R\sin\theta) \end{aligned} \tag{4.7}$$

These lead to the interesting question as to the "best" choice of k for the bar. This would depend on the particular move to be performed but if it were tuned to a frequency that lifted the gymnast during the quarter circle from the bottom of a grand circle and fell away under the gymnast in the second quarter circle, it is not hard to imagine a strong pike-in at the bottom of the swing providing enough lift to consider attempting a release and somersault over the bar to regrasp.

For small pendulum motion under the bar we get unstable amplification when k =mg/R, so that the frequency of the elastic vibrations coincide with the frequency of the pendulum motion.

If we define a new dependent variable y, so that

$$z = y - \frac{Mg}{k} \tag{4.8}$$

the equations of motion become

$$\begin{aligned} \frac{d}{dt}\left(I\dot{\theta}\right) &= -MgR\cos\theta + kR(y\cos\theta - x\sin\theta) \\ M\ddot{x} &= -k(x - R\cos\theta) \\ M\ddot{y} &= -k(y - R\sin\theta) \end{aligned} \tag{4.9}$$

During phases of rigid body rotation, the moment of momentum of the gymnast about the new origin or mean of points of static equilibrium (x=y=0) is given by

$$h^* = I\dot{\theta} + M(x\dot{y} - y\dot{x}) \tag{4.10}$$

and so

$$\frac{dh^*}{dt} = -MgR\cos\theta \tag{4.11}$$

The change in h* is maximised during a circle of the bar when R is a minimum for cos θ positive and a maximum for cosθ negative. This gives a general strategy for bar circles however flexible the bar might be.

The system of equations (4.9) leads to the consideration of two limiting cases, the first for small k, the "bungy-hori-bar" case, and the second corresponding to large k, approaching the behaviour of the "museum relics" called horizontal bars that haunted YMCA's and school gymnasiums in my youth. For small k, we expand θ, x, y as power series in k. The zero-order equations are,

$$M\ddot{x}_0 = M\ddot{y}_0 = 0 \tag{4.12}$$

with the solution $x_0 = y_0 = 0$, and

$$\frac{dh_0}{dt} = \frac{d\left(I\dot{\theta}_0\right)}{dt} = -MgR\cos\theta_0 \tag{4.13}$$

In this zero order approximation, G is fixed at O* and the contact point P (of the hands) moves around the point O* at a distance R.

When k is large, a series solution in powers of ε = 1/k may be sought from the equations

$$\begin{aligned} &\frac{d}{dt}\left(I\dot{\theta}\right) = -MgR\cos\theta - MR(\ddot{y}\cos\theta - \ddot{x}\sin\theta) \\ &\varepsilon M\ddot{x} = -(x - R\cos\theta) \\ &\varepsilon M\ddot{y} = -(y - R\sin\theta) \end{aligned} \tag{4.14}$$

If $\theta = \theta_0 + \theta_1 + ..,\ x = x_0 + x_1 + ..,\ y = y_0 + y_1 + ..$ the zero order equations are,

$$x_0 = R_0\cos\theta_0,\ y_0 = R_0\sin\theta_0,\ \frac{d}{dt}\left(I_0\dot{\theta}_0\right) + Mgx_0 + M(x_0\ddot{y}_0 - y_0\ddot{x}_0) = 0 \tag{4.15}$$

In this case of an inflexible bar, the point G circles the bar and the hands at P remain almost fixed in position.

For the spring board problem we consider the feet at P in board contact and $\theta = \frac{\pi}{2} + \phi$ where ϕ is small. We can also take k_x as very large and ignore X displacements. The gymnast runs, brings his feet together onto the board and injects as much energy into the board as possible using a leg punch and arm lift at bottom of the board deflection to increase the deflection to a maximum. He or she then rides the rebound as a rigid body. If the beat board is regarded as an instrument for diverting kinetic energy of motion into flight, then a gymnast running twice as fast would have the potential to rise four times higher. Hence speed of run up must be a crucial factor in vaulting.

A vault usually requires rotation as well as height in flight, and so it is interesting to explore the moment of momentum delivered by the board. Another important factor in designing a board is the k factor since a slow rebound limits the horizontal speed of entry into the board. In fact an effective entry into the board requires contact to be made ahead of the centre of gravity. The spring-board equations involving Z and ϕ are

$$\begin{aligned} &M\frac{d^2}{dt^2}[Z + R\cos\phi] = -Mg - kZ \\ &\frac{d}{dt}[(I + MR^2)\dot{\phi}] = MR\sin\phi[g + \ddot{Z}] \end{aligned} \tag{4.16}$$

so that

$$\left(I + MR^2\right)\phi - MR^2 \sin\phi \cos\phi \ \phi_2 = kRZ \sin\phi \tag{4.17}$$

With ϕ assumed small, the equations for ϕ and Z become

$$M\ddot{Z} + kZ = -Mg, \quad \frac{d}{dt}\left[(I + MR^2)\dot{\phi}\right] = -kRZ\phi + MR^2\dot{\phi}^2\phi \ . \tag{4.18}$$

At the start of the launch phase, $\dot{Z} = 0,\ Z = -Z_0$ and $\phi = \phi_0,\ \dot{\phi} = \dot{\phi}_0$.

The increase in moment of momentum generated by the board during the time interval $(0,t_1)$ of the launch phase is

$$\int_0^{t_1}\left[-kRZ + M(R\dot{\phi})^2\right]\phi dt \tag{4.19}$$

where Z increases from $-Z_0$ to 0. If we assume a mean value for kz of -Mg and a mean value of Mv_0^2 for $M(R\dot{\phi})^2$ where v_0 is the speed of the gymnast over the beat board, this integral is of the order

$$[MgR + Mv_0^2]\int_0^{t_1}\phi dt \tag{4.20}$$

This estimate gives an idea of the effect of v_0 and the changing orientation of ϕ during the take-off phase.

REFERENCES

[1] Bullen K.E., An introduction to the theory of mechanics, (Science Press, Sydney, 1971).

THE USE OF MATHEMATICAL MODELS TO PREDICT ELITE ATHLETIC PERFORMANCE AT THE OLYMPIC GAMES

Ian T. Heazlewood and Gavin Lackey

Abstract

Prediction of future athletic performances is a common practice and is usually based on plotting the changes in world records over time. This approach is probably unrealistic as the world record does not necessarily reflect how the best eight athletes, that is the event finalists, have performed. Data based on Olympic final performances may reflect a more realistic representation of changing athletic performances over time. Prediction of the mean result in the finals for certain athletic events at the 1996 Olympic Games and future Olympic Games were able to be made after determining the regression equation 'of best fit' that matched closely the mean results of previous Olympic finals. The events considered in the mathematical statistical models were the 100m, 400m, Long Jump and High Jump for both men and women. The coefficient of determinations for the regression equations ranged from 0.6589 for the Men's 100m to 0.9467 for the Women's High Jump. In all cases the regression equations derived were nonlinear mathematical functions. For example the Men's 100m was an inverse function and the Women's 100m was a cubic function. It was suggested that greater accuracy of predictions may be possible if other factors that contribute to athletic performance, such as altitude, wind speed and direction, were able to be included in the equations.

1. INTRODUCTION

There is a familiar saying that 'records are made to be broken'. This implies that there may be no limit to human performance. In spite of this, there is an intuition or belief that there is some finite limit as to how fast a human can run or jump (Edwards and Hopkins [1]; Morton [2]). Exercise physiology is concerned with the improvement in human capacity and the adaptations that occur with training (deVries and Housh [3]). Such improvement can lead to the optimisation of human performance. Herb Elliott states in Gordon's [4] book, 'Australia and the Olympic Games' that "it is the inspiration of the Olympic Games that drives people not only to compete but to improve...". How much improvement is possible in terms of human athletic performance is a question that will be addressed in this research.

The prediction of future athletic performance by humans is a recurring theme during Olympic years, as well forming the basis for some stimulating 'crystal ball gazing' in some of the learned sports science journals. A number of researchers have attempted to predict future performances by deriving and applying a number of mathematical statistical models based on past performances in athletics.

Prendergast [5] applied the average speeds of world record times to determine a mathematical model for world records. The records or data used in the analysis spanned a 10 year period. Following his analysis, Prendergast [5] raised the question of whether any further improvements can be expected or if the limits of human performance have been reached. Olympic records in many events appear to be harder to break as each Olympiad passes, while records in many other events have been consistently broken. For example, in the past 16 Olympiads the Men's Long Jump record has only been broken 4 times and the current record, 8.90m, stands from 1968. This record was set at altitude. The altitude factor is frequently cited as a major factor contributing to this astonishing performance, even though Hay [6] has calculated that altitude only contributed marginally to Beamon's long jump performance. Alternatively, in the Women's 400m, only held from 1964 onwards, the record has been broken at every games with the exception of 1992. How far a human can jump and how quickly he or she can cover a given distance are questions that form the basis of most athletic competition.

With 1996 being the Centenary of the modern Olympic Games, is it reasonable to expect that many Olympic Records may be broken as the elite/high performance athletes reach for the utmost of human potential?

Edwards and Hopkins [1] applied linear regression analysis to World Records in numerous male athletic events to predict future athletic performance. They found a high linear correlation ranging from 0.956 to 0.991 in these events. By extrapolating from the linear equation for each event, the researchers were able to predict with a degree of accuracy some World Record times of future years. However, when additional extrapolation was applied, a prediction of zero time for each event was an outcome of this model. At some time in the future, according to this model, some races would be run in negative time. While at this stage the record times predicted may be quite accurate over short time periods when extrapolated into the near future, it is obvious based on the predictions of zero time and negative times that this straight line regression theory cannot apply to athletic events.

Athletic performance is the result of the interaction of many physiological, biomechanical and psychological factors (Hahn [7]). Many factors may be used to predict athletic performance. Physiological factors that may be tested in order to attempt performance include VO_2 max., respiratory ventilation threshold, oxygen transport and fuel utilisation (Noakes [8]). The testing of these factors is quite involved and the results obtained may be complicated by prior experience in the tests and also may not correlate well with the actual physiological processes that occur in the athletic event (Åstrand [9]).

Previous research has looked at the performance of athletes and attempted to analyse this with respect to maximal aerobic power, the capacity of anaerobic metabolism and the reduction in peak power that occurs with the increase of the length of the race (Péronnet and Thibault [10]). This type of research, whilst arriving at predictions that are believed to be extremely accurate, incorporates many factors into the model and the data needed for the calculations requires in-depth measurement of these factors.

Morton [2] used mathematical modelling to determine the ultimate time in which athletic events may be completed. Factors used in these calculations include the propulsive force per unit mass, availability of oxygen in the muscles and the energy equivalent of the available oxygen. Predictions such as 9.15 seconds for the Men's 100m, when compared to the present record of 9.86 seconds, suggest there is still improvement possible.

In summary, a number of approaches have been applied to understand and predict human athletic performance based on mathematical and statistical models, usually time series analysis. Some have proved reasonably successful and others have produced predictions that are clearly inappropriate.

2. RESEARCH PROBLEM

The previous research that attempted to predict future athletic performance was based on linear regression, a statistical-mathematical methods which assumes that a linear relationship exists between the independent (predictor variable) and the dependent (criterion variable). Human changes over time in this context were assumed to be linear in nature. There are many biological changes with time that do not or might not conform to a mathematical function that is linear (Arya and Larder [11]). If the relationship is curvilinear, an attempt to fit a linear model to the data may in fact be inappropriate. If the relationship is nonlinear there are a number of potential mathematical functions that may provide a better fit to the data and subsequently a more accurate predictive model. Such mathematical functions could be logarithmic, inverse, cubic, compound, sigmoidal and so on. Finding the most appropriate model should improve the prediction of future human athletic performance when mathematical extrapolation is applied.

3. RESEARCH QUESTIONS

A number of research questions can be generated that will address the predominant theoretical issues. These are:

1. What is the best mathematical-statistical model that will best fit the data for each event?

2. Does one model fit all events?

3. Will the mean score based on the top eight (finalists) better reflect actual changes in human performance that have occurred in the past 70 years?

4. Will the models be identical for the different genders for each event?

5. Will the models be gender specific as well as event specific?

6. Will the mathematical-statistical models enable the accurate prediction of future performance?

7. How much variance will be explained in the different models for each events for each gender?

8. Will the derived mathematical-statistical models result in absurd predictions that have occurred in previous research of this type?

4. RESEARCH HYPOTHESES

While prior research has used many physiological factors to predict performance, the aim of this paper is to predict performances at the 1996 Olympic Games and future

Olympic Games using only easily accessible data, that is, the results of past Olympic finalists. The task of this descriptive research is to determine if there are significant mathematical relationships between the mean of the finalists of different Olympic events over the years. Further to this, it is generally hypothesised that the significant relationship can be used in equations to accurately predict the average times run, lengths and heights of jumps in specific Olympic events. The specific events that will be predicted are the men's and women's 100m, the men's and women's 400m, the men's and women's long jump, and the men's and women's high jump.

The following hypotheses will be formally stated as the positive or alternate hypotheses. The null hypotheses are assumed to be true only for the tests of statistical significance, as statistical tests essentially test the validity of the null hypotheses.

1. The different events will have different models of best fit as they will be dependent on different human factors such as speed, power, endurance, speed endurance, strength and motor co-ordination that are represented by different mathematical-statistical functions.

2. The different genders will have different models of best fit as a result of the different number of years that females have been competing in the different events and as a result of females only recently adopting more stressful physical training programs that will influence the response curves (training and competition performance) over time.

3. The mathematical-statistical models will be both event and gender specific.

4. The mean score for the Olympic finalist will more realistically reflect human performance and inferred ability in each event than previous predictive models.

5. The different mathematical models derived will provide a more accurate prediction of future performances both for short time and long time periods.

6. More of the variance will be explained and a better model fit will occur based on nonlinear mathematical functions.

7. The likelihood of absurd mathematical predictions will be reduced or non-existent.

5. METHOD AND STATISTICAL ANALYSIS

Results from each Olympiad from 1924 - 1992 were collected and the average result of each final was calculated (refer to Appendix 1 for the raw data). The eighth Olympic Games in Paris 1924 was chosen as the starting point for the collection of results, as up to this Olympiad the results were only available for the first 3 places in each event. The events selected were the 100m, 400m, High Jump and Long Jump for both men and women. The women's athletic program was restricted at the Olympic Games when compared to the men's program, and therefore different events were included in the Olympics as the Summer Games evolved. For example, women's track and field commenced at the 1928 Olympics and women's results from 1928 - 1992 were taken for the 100m and High Jump, from 1948 onwards for Long Jump and from 1964 - 1992 for the 400m. The Olympic performances which constituted the set of data were taken from records held by the Australian Olympic Committee [12] and other published sources (Wallechinsky [13]).

The means for each event in each Olympic Year were calculated and then included as a data set for analysis by the statistical software Statistical Package for the Social Sciences (abbreviated as SPSS program version 6.1, SPSS Inc, [14]; Norusis [15]) in order to derive regression equations for each event. Eleven different mathematical regressions functions were available that enabled curve estimation. The eleven functions were linear, logarithmic, inverse, quadratic, cubic, compound, power, sigmoidal (abbreviated as S), growth, exponential and logistic (Norusis [15]). The regression equation that produced the best fit for each event, that is, produced the highest coefficient of determination (abbreviated as R^2), was then determined using the eleven equation approach. The specific criteria to select the regression equation of best fit were the magnitude of R^2, the significance of the analysis of variance alpha or p-value and the residuals. The residuals are the difference between the actual value and the predicted value for each case.

The equation of best fit was then selected and using this equation a prediction of the mean result for each event was calculated. At this stage, graphs representing the means for each event in each Olympiad were also generated in addition to predicted means using the appropriate regression equation.

6. RESULTS

The means of Olympic finalists in each of the selected events included in the study are shown in Table 1. Table 2 contains the current Olympic Records and the year in which each record was set.

Table 1. Mean Result of Finalists of the Different Olympic Events.

Year	Men's 100m	Women's 100m	Men's 400m	Women's 400m	Men's Long Jump	Women's Long Jump	Men's High Jump	Women's High Jump
1924	10.82	NC	48.35	NC	7.09	NC	1.89	NC
1928	10.93	12.30	48.37	NC	7.42	NC	1.90	1.52
1932	10.48	12.06	47.53	NC	7.22	NC	1.95	1.59
1936	10.57	11.93	47.13	NC	7.69	NC	1.97	1.58
1948	10.43	12.10	47.47	NC	7.34	5.53	1.94	1.60
1952	10.43	11.87	46.62	NC	7.26	5.91	1.97	1.61
1956	10.60	11.77	47.32	NC	7.46	5.98	2.09	1.67
1960	10.28	11.42	45.45	NC	7.86	6.18	2.09	1.70
1964	10.31	11.66	45.84	54.09	7.63	6.41	2.13	1.76
1968	10.07	11.26	45.11	52.50	8.17	6.54	2.17	1.79
1972	10.33	11.45	45.16	51.79	8.01	6.59	2.18	1.86
1976	10.21	11.23	45.06	50.64	8.01	6.56	2.20	1.89
1980	10.39	11.19	45.24	50.42	8.18	6.90	2.27	1.92
1984	10.24	11.29	44.76	49.98	8.10	6.72	2.30	1.95
1988	10.34	10.99	44.51	50.30	8.18	6.91	2.35	1.96
1992	10.10	10.96	44.47	49.79	8.22	6.85	2.33	1.95

NC = Not Contested.

Timed events recorded in seconds and jump events in metres.

Table 2. Current Olympic Records.

Event	Men's 100m	Women's 100m	Men's 400m	Women's 400m	Men's Long Jump	Women's Long Jump	Men's High Jump	Women's High Jump
Olympic Record	9.92	10.54	43.50	48.61	8.90	7.40	2.38	2.03
Year	1988	1988	1992	1988	1968	1988	1988	1988

Figures 1 - 8 on the following pages show graphical representations, generated by the SPSS program, of the means of the finalists in each event with a line of best fit. In each graph the mean result in seconds or metres is shown on the Y-axis and the years are shown on the X-axis. It can be seen from these figures that the performances in each event have generally been improving at each Olympic Games. Performances that stand out from the others, such as those in the 100m and Long Jump for both men and women at the 1968 Mexico City Olympic Games and the Men's Long Jump in 1936, can be observed in these figures.

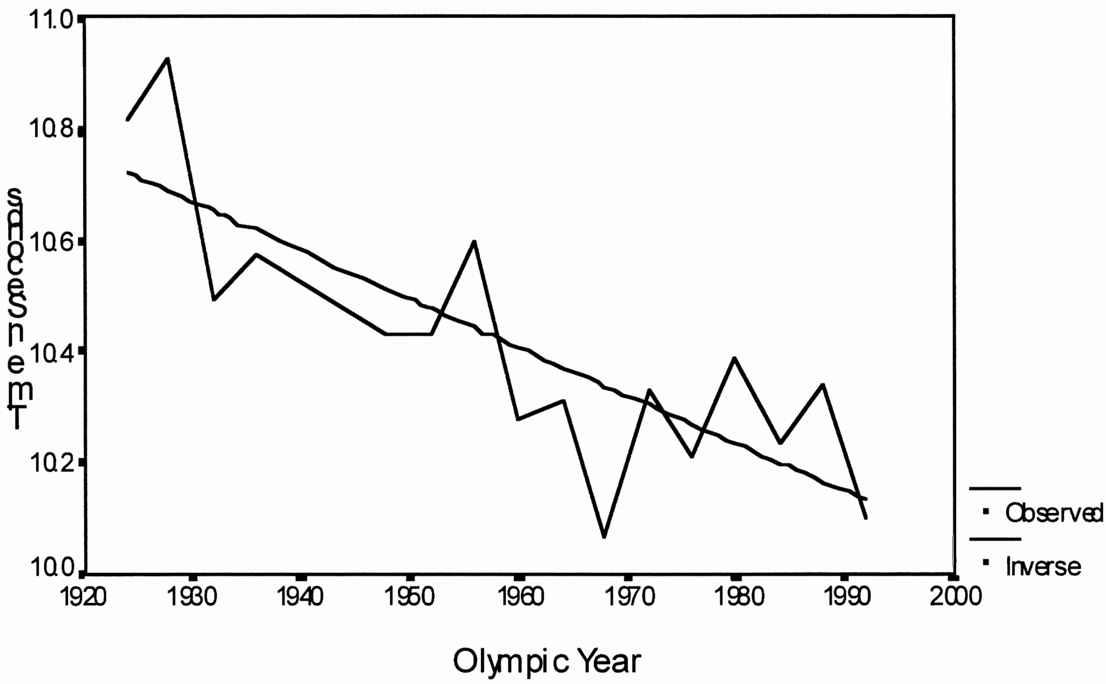


Figure 1: *Graph of the means of the Men's 100m with line of best fit.*

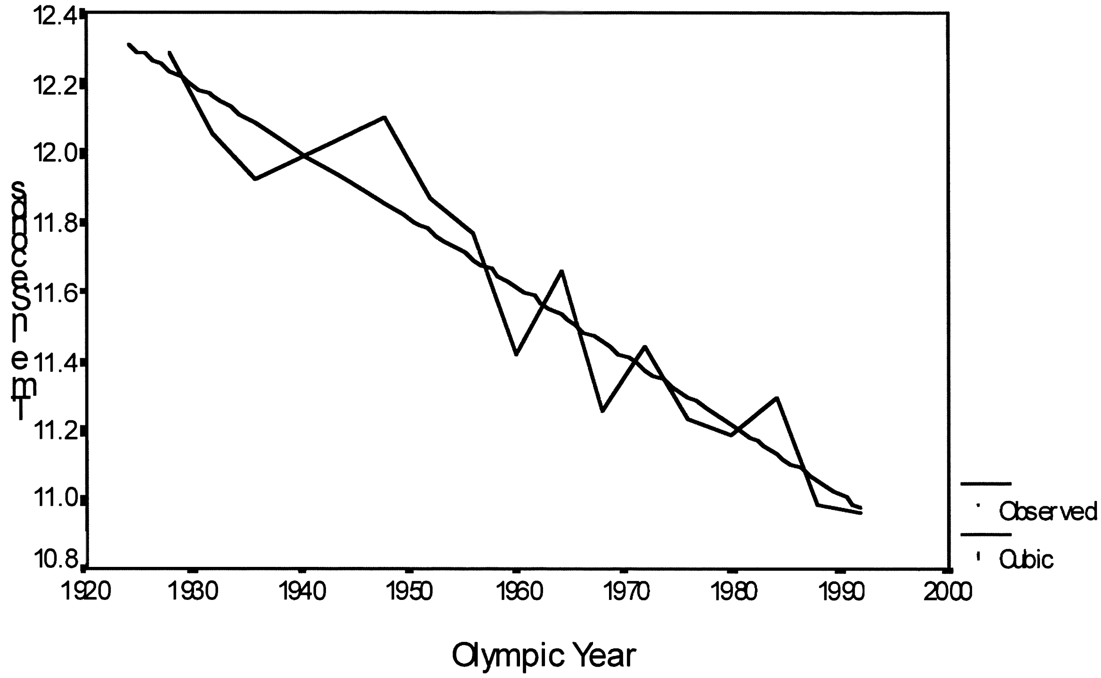


Figure 2: *Graph of the means of the Women's 100m with line of best fit.*

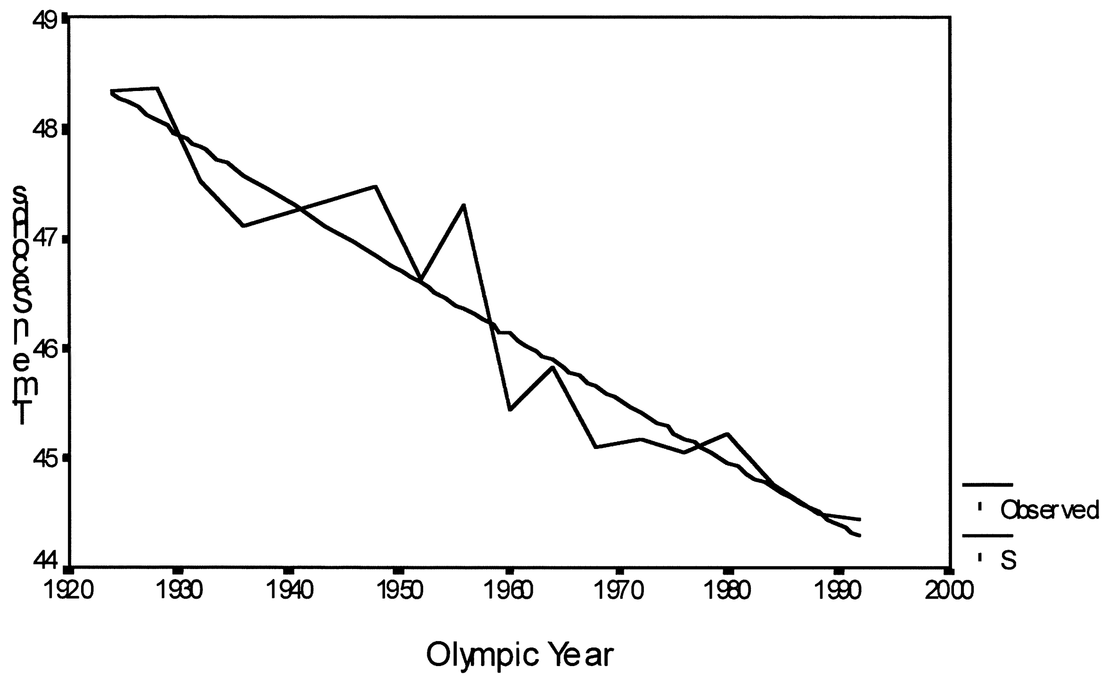


Figure 3: *Graph of the means of the Men's 400m with line of best fit.*

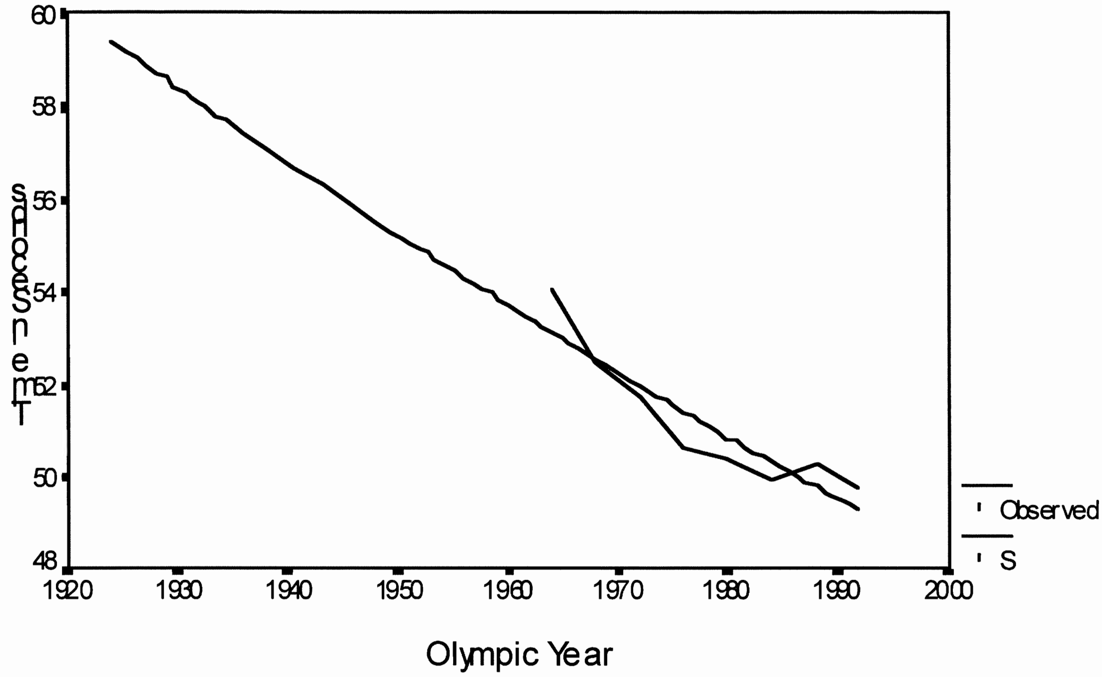


Figure 4: *Graph of the means of the Women's 400m with line of best fit.*

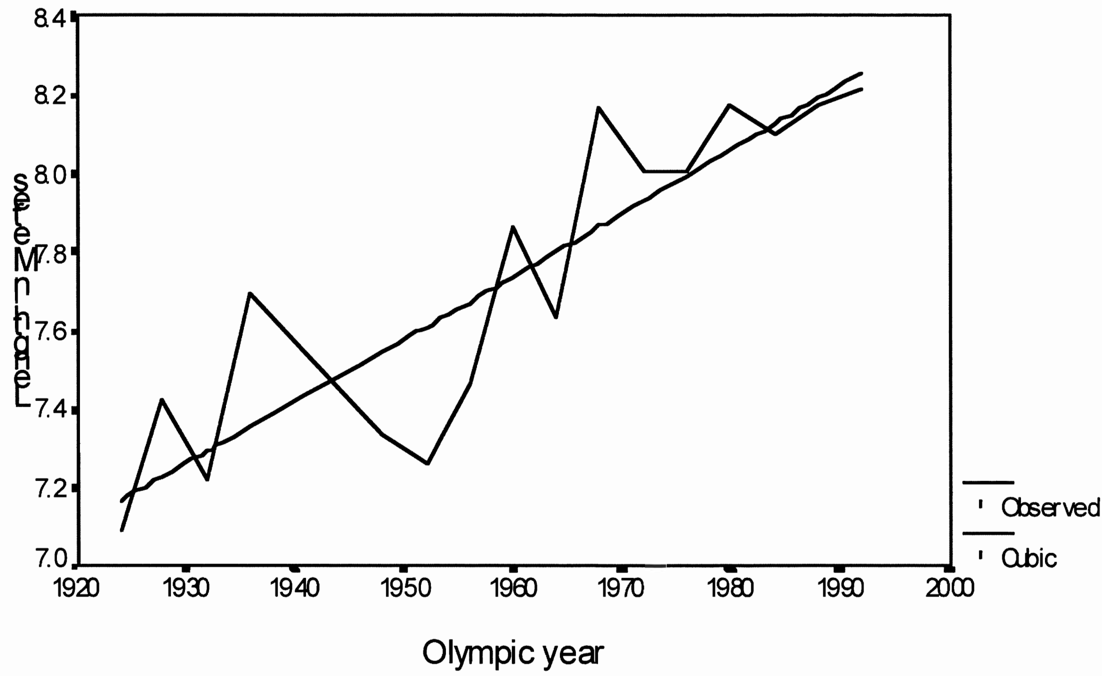


Figure 5: *Graph of the means of the Men's Long Jump with line of best fit.*

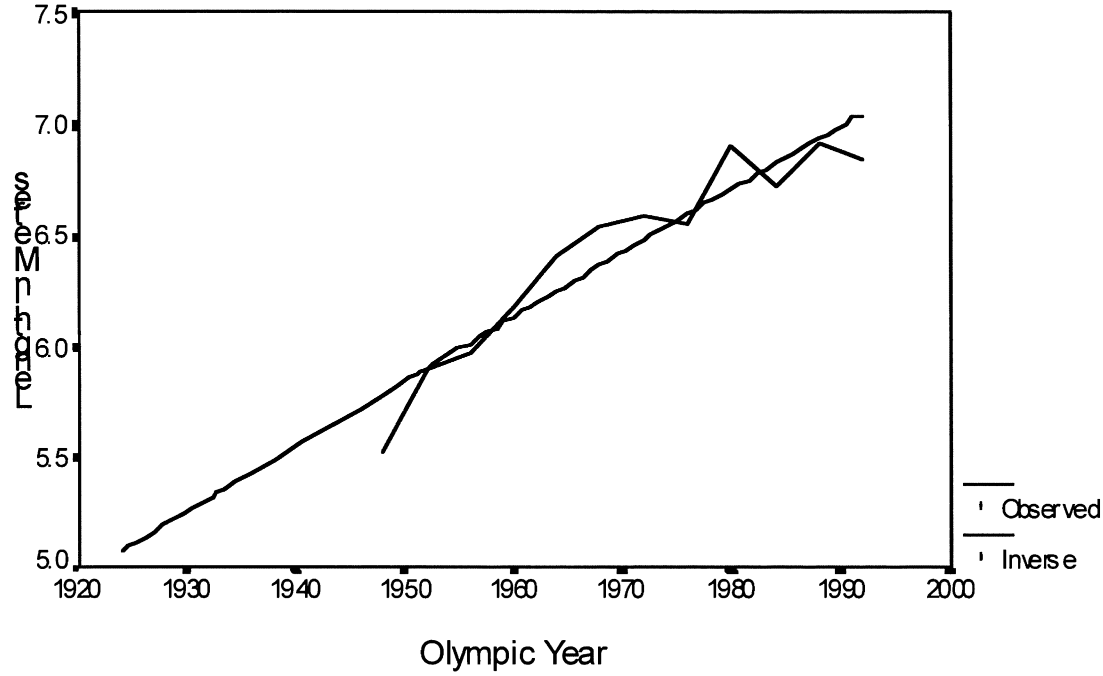


Figure 6: *Graph of the means of the Women's Long Jump with line of best fit.*

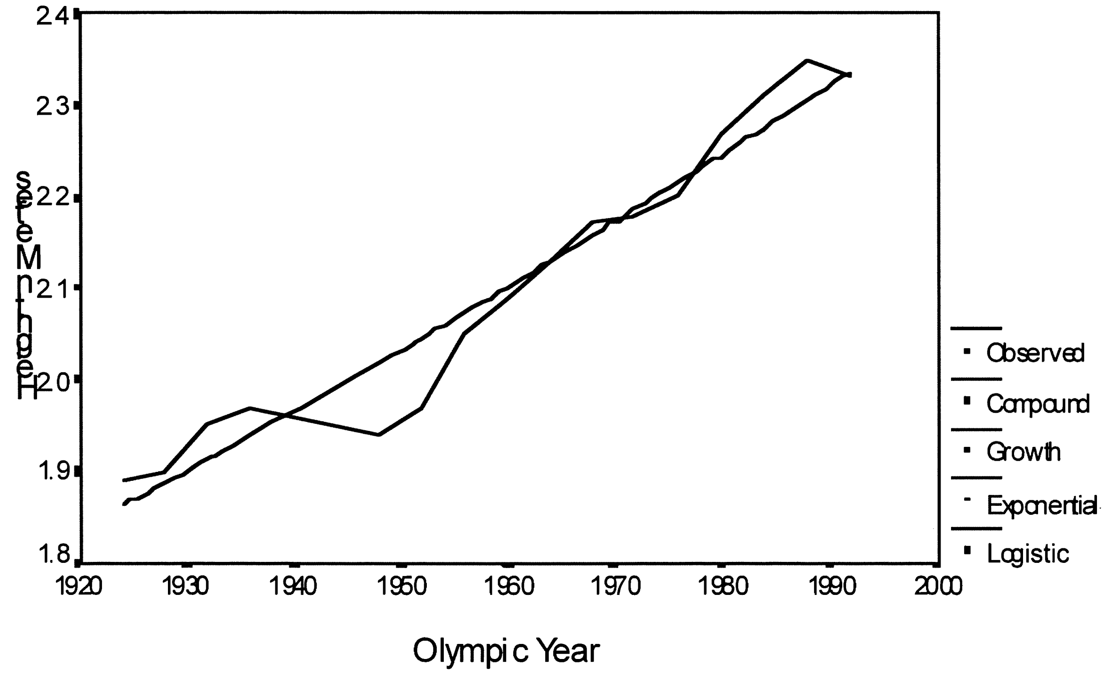


Figure 7: *Graph of the means of the Men's High Jump with line of best fit.*

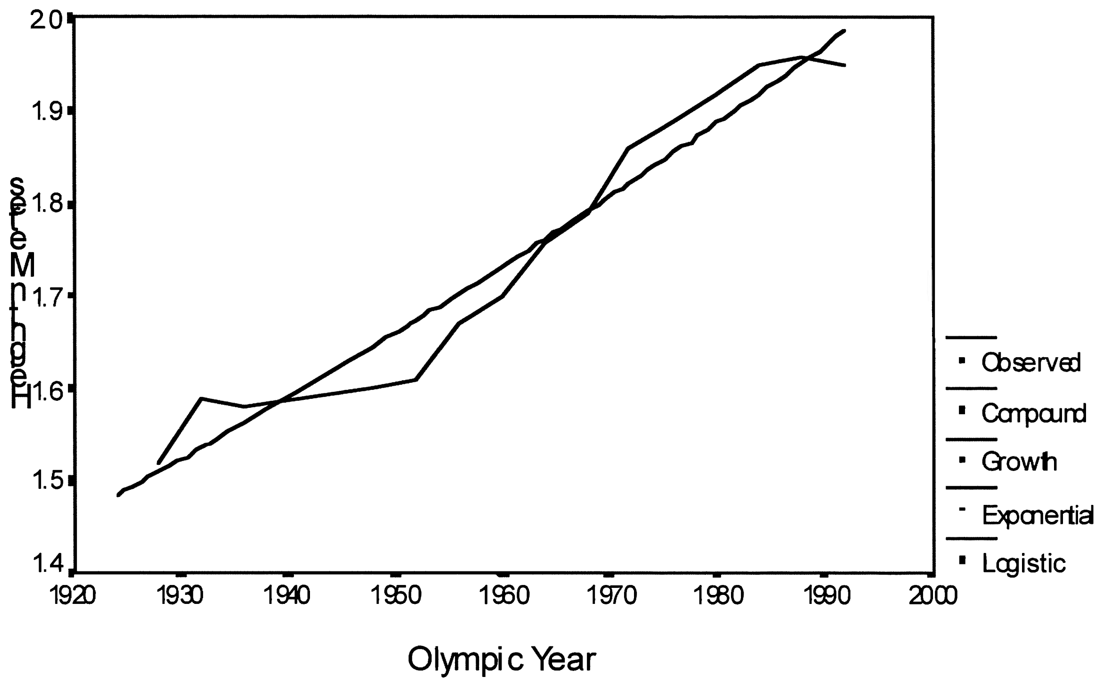


Figure 8: *Graph of the means of the Women's High Jump with line of best fit.*

The mathematical functions that resulted in the line of best fit for each event as well as the coefficients of determination (R^2) for each event are displayed in Table 3.

Table 3. Events, Mathematical Functions, Equations and R^2 Values.

Event	Regression Type	Equation *	R^2
Men's 100m	Inverse	$Y = b_0 + (b_1/t)$	0.6589
Women's 100m	Cubic	$Y = b_0 + b_3 t^3$	0.90175
Men's 400m	S	$Y = e^{(bo + b1/t)}$	0.90676
Women's 400m	S	$Y = e^{(bo + b1/t)}$	0.84251
Men's Long Jump	Cubic	$Y = b_0 + b_3 t^3$	0.77735
Women's Long Jump	Inverse	$Y = b_0 + (b_1/t)$	0.89389
Men's High Jump	Compound Logistic Exponential Growth	$Y = b_0(b_1)^t$ $Y = b_0 e^{b1t}$ $Y = e^{bob1t}$	0.94390 0.94390 0.94390 0.94390
Women's High Jump	Compound Logistic Exponential Growth	$Y = b_0(b_1)^t$ $Y = b_0 e^{b1t}$ $Y = e^{bob1t}$	0.94665 0.94665 0.94665 0.94665

*Where b_0 = a constant
b_1, b_3 = regression coefficients
t = year
y = mean result for each event

Table 4 shows the regression equation and the predicted mean for each event. The mean of the 1992 Olympic finalists is shown for comparison, the 1996 predicted performance and the calculated percentage improvement from 1992 to 1996 is also represented.

Table 4. Event, Regression Equation, 1992 Mean, 1996 Predicted Mean and Percentage Improvement.

Event	Regression Equation	1992 Mean	1996 Prediction	% Improve
Men's 100m	$Y = -6.627 + 33387.44/1996$	10.10s	10.099s	0.99
Women's 100m	$Y = 24.52 - 1.71x\ 10^{-9}x1996^3$	10.96s	10.92s	0.93
Men's 400m	$Y = e^{(1.318 + 4925.11/1996)}$	44.47s	44.06s	0.99
Women's 400m	$Y = e^{(-1.4 + 10554.33/1996)}$	49.79s	48.80s	0.98
Men's Long Jump	$Y = -2.76 + 1.39 \times 10^{-9}\ x1996^3$	8.22m	8.29m	1.01
Women's Long Jump	$Y = 63.004-111458.44/1996$	6.85m	7.16m	1.04
Men's High Jump	$Y = 0.002971 \times 1.003356^{1996}$	2.33m	2.38m	1.02
	$Y = 0.002971 \times e^{0.003347x1996}$		2.37m	
	$Y = e^{-5.818741+0.003347x1996}$		2.37m	
Women's High Jump	$Y = 0.000385 \times 1.0043^{1996}$	1.95m	2.02m	1.04
	$Y = 0.000385 \times e^{0.004291x1996}$		2.02m	
	$Y = e^{-7.861459+0.004291x1996}$		2.02m	

Based on the data and the derived equations, it can be seen in Table 5 that it is possible to calculate a year when the 100m events for both men and women will, according to the prediction equations, be run in zero time.

Table 5. Calculations of zero time for men's and women's 100m events.

Event	Regression Equation	Zero Time Equation	Year
Men's 100m	$Y = -6.627 + 33387.44/t$	$0 = -6.627+33387.44/x$ *	5038
Women's 100m	$Y = 24.52 - 1.71x\ 10^{-9}x\ t^3$	$0 = 24.52-1.71x10^{-9}x\ t^3$	2429

* Where t = year

Y = time in seconds

Table 6 displays the calculated mean result of finalists in each event using the appropriate regression equations derived.

Table 6. Calculated Means Using the Derived Nonlinear Regression Equations.

Year	Men's 100m	Women's 100m	Men's 400m	Women's 400m	Men's Long Jump	Women's Long Jump	Men's High Jump	Women's High Jump
1924	10.73*		48.32		7.14**		1.87	
1928	10.69	12.26	48.06		7.20		1.90	1.51
1932	10.65	12.19	47.81		7.26		1.92	1.53
1936	10.62	12.11	47.56		7.33		1.95	1.56
1948	10.51	11.88	46.82		7.51	5.79	2.03	1.64
1952	10.48	11.80	46.58		7.58	5.90	2.06	1.67
1956	10.44	11.72	46.34		7.64	6.02	2.08	1.70
1960	10.41	11.64	46.10		7.71	6.14	2.11	1.73
1964	10.37	11.57	45.86	53.19	7.77	6.25	2.14	1.76
1968	10.34	11.49	45.63	52.61	7.83	6.37	2.17	1.79
1972	10.30	11.41	45.40	52.04	7.90	6.48	2.20	1.82
1976	10.27	11.33	45.17	51.48	7.96	6.60	2.23	1.85
1980	10.24	11.25	44.94	50.93	8.03	6.71	2.26	1.88
1984	10.20	11.17	44.72	50.39	8.10	6.83	2.29	1.92
1988	10.17	11.08	44.50	49.85	8.16	6.94	2.32	1.95
1992	10.13	11.00	44.28	49.32	8.23	7.05	2.35	1.98
1996	10.10	10.92	44.06	48.80	8.29	7.16	2.38	2.02

* Timed events in seconds

** Jump events in metres

Table 7 displays the predicted performances for each event for past and future performances from 1924 to 2296 based on the nonlinear regression equations. It is important to note that the performances extrapolated back to 1924 performances for women appear realistic, however the predictions of future performance indicate that women will surpass the men in the 100m in 2060, in the 400m in 2072 and in the Long Jump and High Jump in 2196.

DISCUSSION

As can be seen in both Table 1 and the figures, the average result of the finalists has been, in general, improving each Olympic year. How far this will continue is still an open question that will be based on improvements as a result of human evolution, enhanced training methods, innovative technology and biochemical manipulation via performance enhancing substances (not just drugs).

The regression type that provided the highest coefficient of determination (R^2) value was used as this is an indication of the proportion of the variability that is accounted for or explained by the X value (Norusis [15]). In each case, the X value represents the year and the Y value represents the mean result of the finalists in each event under consideration. By accounting for the greatest amount of explained variance, a more accurate prediction was possible.

The coefficient of determination for the Men's 100m (0.6589) is the lowest of the events tested. Therefore, this event will have the least accurate prediction of the events tested. The predicted mean for the finalists of this event at the Atlanta Olympics is 10.099 seconds. If a more accurate equation could be developed, perhaps using other factors in addition to the mean results of previous Olympics, a higher coefficient of determination might be derived and predictions of future performance could be made with more certainty.

Table 7. The Predicted Performances for Men's and Woman's 100m, 400m, Long Jump and High Jump for Past and Future Olympics.

Year	100M	100W	400M	400W	MLJ	WLJ	MHJ	WHJ
1924	10.73	12.34	48.32	59.48	7.14	5.07	1.87	1.48
1928	10.69	12.26	48.06	58.81	7.20	5.19	1.90	1.51
1932	10.65	12.19	47.81	58.14	7.26	5.31	1.92	1.53
1936	10.62	12.11	47.56	57.49	7.33	5.43	1.95	1.56
1948	10.51	11.88	46.82	55.59	7.51	5.79	2.03	1.64
1952	10.48	11.80	46.58	54.98	7.58	5.90	2.06	1.67
1956	10.44	11.72	46.34	54.37	7.64	6.02	2.08	1.70
1960	10.41	11.64	46.10	53.78	7.71	6.14	2.11	1.73
1964	10.37	11.57	45.86	53.19	7.77	6.25	2.14	1.76
1968	10.34	11.49	45.63	52.61	7.83	6.37	2.17	1.79
1972	10.30	11.41	45.40	52.04	7.90	6.48	2.20	1.82
1976	10.27	11.33	45.17	51.48	7.96	6.60	2.23	1.85
1980	10.24	11.25	44.94	50.93	8.03	6.71	2.26	1.88
1984	10.20	11.17	44.72	50.39	8.10	6.83	2.29	1.92
1988	10.17	11.08	44.50	49.85	8.16	6.94	2.32	1.95
1992	10.13	11.00	44.28	49.32	8.23	7.05	2.35	1.98
1996	10.10	10.92	44.06	48.80	8.29	7.16	2.38	2.02
2000	10.07	10.84	43.84	48.29	8.36	7.27	2.42	2.05
2004	10.03	10.76	43.63	47.78	8.43	7.39	2.45	2.09
2008	10.00	10.68	43.41	47.28	8.49	7.50	2.48	2.12
2012	9.97	10.59	43.20	46.79	8.56	7.61	2.51	2.16
2016	9.93	10.51	42.99	46.31	8.63	7.72	2.55	2.20
2020	9.90	10.43	42.78	45.83	8.70	7.83	2.58	2.24
2024	9.87	10.34	42.58	45.36	8.77	7.94	2.62	2.28
2028	9.84	10.26	42.37	44.89	8.83	8.04	2.65	2.32
2032	9.80	10.17	42.17	44.44	8.90	8.15	2.69	2.36
2036	9.77	10.09	41.97	43.99	8.97	8.26	2.73	2.40
2040	9.74	10.00	41.77	43.54	9.04	8.37	2.76	2.44
2044	9.71	9.92	41.58	43.10	9.11	8.47	2.80	2.48
2048	9.68	9.83	41.38	42.67	9.18	8.58	2.84	2.52
2052	9.64	9.74	41.19	42.24	9.25	8.69	2.88	2.57
2056	9.61	9.66	41.00	41.82	9.32	8.79	2.91	2.61
2060	9.58	9.57	40.81	41.41	9.39	8.90	2.95	2.66

Table 7. continued

Year	100M	100W	400M	400W	MLJ	WLJ	MHJ	WHJ
2064	9.55	9.48	40.62	41.00	9.46	9.00	2.99	2.70
2068	9.52	9.40	40.43	40.59	9.53	9.11	3.03	2.75
2072	9.49	9.31	40.24	40.20	9.60	9.21	3.07	2.80
2076	9.46	9.22	40.06	39.80	9.68	9.31	3.12	2.84
2080	9.42	9.13	39.88	39.42	9.75	9.42	3.16	2.89
2084	9.39	9.04	39.70	39.03	9.82	9.52	3.20	2.94
2088	9.36	8.95	39.52	38.66	9.89	9.62	3.24	2.99
2092	9.33	8.86	39.34	38.29	9.97	9.73	3.29	3.05
2096	9.30	8.77	39.16	37.92	10.04	9.83	3.33	3.10
2100	9.27	8.68	38.99	37.56	10.11	9.93	3.38	3.15
2196	8.58	6.41	35.19	30.15	11.96	12.25	4.66	4.76
2296	7.91	3.82	31.92	24.45	14.06	14.46	6.51	7.31

Note: Year = Olympic year
100M = men's 100m (s)
100W = women's 100m (s)
400M = men's 400m (s)
400W = women's 400m (s)
MLJ = men's long jump (m)
WLJ = women's long jump (m)
MHJ = men's high jump (m)
WHJ = women's high jump (m)

The most accurate prediction should be obtained in the Women's High Jump due to the extremely high explained variance ($R^2 = 0.94665$). The predicted mean result in the Women's High Jump in 1996 is 2.02 metres, a result that will be known in the very near future.

A comparison of the data in Table 1 and Table 6, which includes the 1924 to the 1996 Olympic Games, indicates that many of the predicted means are quite similar to the actual means, especially in the Men's and Women's High Jump. This is an expected outcome due to the high coefficients of determination.

The smallest improvement, according to the prediction, is expected in the Women's 100m where the improvement from 1992 to 1996 is only 0.93%. This would appear to indicate that performances in the Women's 100m will plateau from 1992 to 1996. The largest improvement should occur in the 1996 Women's Long Jump and High Jump. In the Long Jump, the 1992 mean is expected to be improved by 1.04%, that is 31 centimetres. In the High Jump, an improvement of 7 centimetres can be expected. More improvement may be expected in the Women's Long Jump and High Jump before future performances in these events plateau.

Using the extrapolations shown in Table 4 and comparing these to the Olympic records in Table 2, it is predicted that an Olympic Record will be set in the Men's

High Jump in 1996, as the predicted mean of the finalists in Atlanta is equal to the current Olympic Record (2.38m). This is not to say, however, that this is the only event in which an Olympic Record will be set in 1996.

Table 7 displays the predicted results from 1924 to 2296 for both men and women. Some of the performances for women have been extrapolated into the past. It must be remembered that the Women's 100m and High jump were introduced into the Olympic program in 1928, the Women's Long Jump in 1948 and the Women's 400m in 1964. The predicted performances based on the extrapolations appear to be realistic. If the trends are extrapolated to the year 1996 to 2100 in four year intervals, the predictions of times and distances become interesting, as female performances are predicted to surpass males in the 100m in the year 2060 and in the 400m in the year 2072. This trend was replicated with the long jump and high jump where the cross-over time (where the women will surpass the men) is the year 2196.

The possible reasons for these paradoxical predictions are the greater percentage improvements in women's performances over the past 30 years when compared to the men's performances. These greater improvements in women's performances are the result of more recent inclusions of some Olympic events; the adoption by women of more demanding training programs; the use of performance enhancing drugs, especially androgenic anabolic steroids which have a more significant performance enhancing effect in females (admitted by coaches and athletes, as well as having been proven with female sprinters and female swimmers by IAAF and FINA drug testing); the improvements in track and shoe technology; and the increase in years of training as recognised by the increasing age of world record holders (IAAF [16]). It is very interesting to note that recent sprint performances for women have been well below current world records whereas the men have recently broken the 100m and 200m records, and run close to the 400m record.

A further extension of this work may involve the inclusion of other factors that may enhance the predictive equations. These other factors could include athlete physiology factors, such as the effects of competing at altitude, in high humidity and high temperature, and biomechanic factors, such as the influence of running surfaces, wind speed and direction. Elements such as the altitude of previous venues and of the present venue are easily obtained, much more so than any internal physiological factors responding to changes in the environment.

From the research domain of biomechanics, it has been shown that running economy is greater at altitude than at sea level due to less air resistance and a decreased energy cost (Kreighbaum and Barthels [17]; deVries and Housh [3]). To overcome air resistance on even a calm day, sprinting at 10 metres per second requires an extra 7.8% of energy requirements (deVries and Housh [3]). Linthorne [18] reports that sprinters run faster at altitude than at sea level competitions. The improvement in time is due to a lesser air density and hence a lower aerodynamic drag force applied to the athlete. The advantage gained at altitude has been shown using mathematical models to be proportional to the altitude at which the event is held.

Wind speed and direction is also a factor in events such as Long Jump and Triple Jump where run-up speed is important to the performance. Wind speed and direction may not be as important in events that involve one or more laps of the track as the head wind on one side of the track is a tail wind on the other side. However, the disadvantage of a head wind is greater than the advantage gained from a tail wind of the same magnitude (Linthorne [18]).

In the domain of exercise physiology the use of physiological factors in the prediction of results, such as the mean of finalists as in this study, requires the knowledge of the results of the testing and measurement of each athlete in the finals. This information at this stage of the research process is well beyond the scope of this study.

These predictions have been made with the use of mathematical models. The true accuracy of these predictions will be determined in the finals of these events at the Atlanta Olympics in July and August 1996. The equations derived and the predictions made using these are more realistic than those made in 1978 by Edwards and Hopkins in that there is no point in which an event will be run in zero time with the exception of the 100m running events. While the equations used by Edwards and Hopkins [1] had high coefficients of determination (0.956 to 0.991), the calculations made in this study should prove to be more accurate over time as the derived equations of best fit were all nonlinear in nature. The prediction of a zero time is of course totally unrealistic and, with time and additional factors included in the calculations, more realistic equations and predictions may be developed.

The aim of this research was not to reduce athletic contests to mathematical formulae but to determine if these formulae could be used to extrapolate a prediction of the mean result of the finalists in Atlanta and in future Olympic Games. It is hoped that the challenge of athletics will still exist within the athlete, and, if these predictions turn out to be accurate, similar models may be developed which will allow both coach and athlete to know the performances required to make the finals at any particular future Olympic events.

The feasibility exists for similar work to be performed with many other events. These events could possibly include other athletic events, swimming, cycling and other events that are measured by the tape or the clock. This technique, if the predictions turn out to be close to the actual mean result of the Olympic finalists, may also be useful in determining when it might be expected that elite athletes break a certain barrier, or to predict when athletes may not show any further progression in an event over time.

These models have been derived in an attempt to predict the results of selected events at the 1996 Olympic Games as well as future Olympic Games both in the near and distant future. While a prediction for any future Olympiad can be made using these formulae, to remain effective the mean results of future Olympic finalists must be included in the calculation to refine the appropriate regression equations.

REFERENCES

[1] Edwards D. and Hopkins W., Targets for the future, *Modern Athlete and Coach.* 17 (3), 34-36 (1979).

[2] Morton R. H., The supreme runner: A theory of running and some of his physiological attributes, *The Australian Journal of Science and Medicine in Sport.* 16 (2) 26-28 (1984).

[3] deVries H. and Housh T., Physiology of exercise for physical education, athletics and exercise science, (Brown and Benchmark Publishers, Wisconsin, 1994).

[4] Elliott H., Foreword by Herb Elliott in H. Gordon. Australia and the Olympic Games, (University of Queensland Press. St. Lucia, 1994).

[5] Prendergast K., What do world running records tell us?, *Modern Athlete and Coach.* 28 (1) 33 -36 (1990).

[6] Hay J.G., The biomechanics of sports techniques, (Prentice Hall, Englewood Cliffs, 1993).

[7] Hahn A.G., Measuring physiological capacities in the laboratory and field. In F. S. Pyke Better coaching: Advanced coach's manual, (Belconnen, Australian Coaching Council, 1991).

[8] Noakes T., Implications of exercise testing for prediction of athletic performance: A contemporary perspective, *Medicine and Science in Sports and Exercise.* 20 (4), 319-330 (1988).

[9] Åstrand P-O. and Rodahl, K., Textbook of work physiology. Physiological bases of exercise, (McGraw Hill, New York, 1977).

[10] Peronnet F. and Thibault G., Mathematical analysis of running performance and world running records, *Journal of Applied Physiology*, 6 (7) 453-465, (1989).

[11] Arya J.C. and Lander R. W., Mathematics for the biological sciences, (Prentice Hall, Engelwood Cliffs, 1979).

[12] Phillips C., Australian Olympic Committee, Personal Communication, (March, 1996).

[13] Wallechinsky D., The complete book of the Olympics, (Penguin Books, New York, 1988).

[14] SPSS Inc., SPSS for windows: Version 6.1, (Chicago, SPSS Inc, 1995).

[15] Noruis M., SPSS for windows: Base systems user's guide release 6.0, (Chicago, SPSS Inc, 1993).

[16] IAAF (International Amateur Athletics Federation), Results and statistics, (IAAF Publications, Monte Carlo, 1996). Web Site address http://www.iaaf.org/rdb/index.html.

[17] Kreighbaum E. and Barthels K., Biomechanics: A qualitative approach for studying human movement, (Macmillan Publishing Co., New York, 1990).

[18] Linthorne N., The effect of wind on 100m sprint times, *Journal of Applied Biomechanics*, 10 (2) 110 - 131 (1993).

APPENDIX I

Results of finalists and mean of results at each Olympic Games.

Table A1. Men's 100m. Results in seconds.

Year	1st	2nd	3rd	4th	5th	6th	7th	8th	Mean
1924	10.6	10.7	10.8	10.9	10.9	11.0			10.82
1928	10.8	10.9	10.9	11.0	11.0	11.0			10.93
1932	10.3	10.3	10.4	10.5	10.6	10.8			10.48
1936	10.3	10.4	10.5	10.6	10.7	10.9			10.57
1948	10.3	10.4	10.4	10.4	10.5	10.6			10.43
1952	10.4	10.4	10.4	10.4	10.5	10.5			10.43
1956	10.5	10.5	10.6	10.6	10.7	10.7			10.60
1960	10.2	10.2	10.3	10.3	10.3	10.4			10.28
1964	10.0	10.2	10.2	10.4	10.4	10.4	10.4	10.5	10.31
1968	9.95	10.0	10.0	10.1	10.1	10.1	10.1	10.2	10.07
1972	10.14	10.24	10.33	10.36	10.40	10.40	10.46		10.33
1976	10.06	10.08	10.14	10.19	10.25	10.27	10.31	10.35	10.21
1980	10.25	10.25	10.39	10.42	10.43	10.44	10.46	10.49	10.39
1984	9.99	10.19	10.22	10.26	10.27	10.29	10.33	10.35	10.24
1988	9.92	9.97	9.99	10.04	10.11	10.11	12.26		10.34
1992	9.96	10.02	10.04	10.09	10.10	10.12	10.22	10.26	10.10

Table A2. Women's 100m. Results in seconds.

Year	1st	2nd	3rd	4th	5th	6th	7th	8th	Mean
1928	12.2	12.3	12.3	12.4					12.30
1932	11.9	11.9	12.0	12.2	12.3				12.06
1936	11.5	11.7	11.9	12.0	12.2	12.3			11.93
1948	11.9	12.2	12.2						12.10
1952	11.5	11.8	11.9	11.9	12.0	12.1			11.87
1956	11.5	11.7	11.7	11.8	11.9	12.0			11.77
1960	11.0	11.3	11.3	11.4	11.5	12.0			11.42
1964	11.4	11.6	11.6	11.6	11.7	11.7	11.8	11.9	11.66
1968	11.0	11.1	11.1	11.1	11.3	11.4	11.5	11.6	11.26
1972	11.07	11.23	11.23	11.32	11.38	11.41	11.45	12.48	11.45
1976	11.08	11.13	11.17	11.23	11.24	11.31	11.32	11.34	11.23
1980	11.06	11.07	11.14	11.16	11.16	11.28	11.32	11.34	11.19
1984	10.97	11.13	11.16	11.25	11.39	11.40	11.43	11.62	11.29
1988	10.54	10.83	10.85	10.97	10.97	11.00	11.26	11.49	10.99
1992	10.82	10.83	10.84	10.86	10.88	11.10	11.15	11.19	10.96

Table A3. Men's 400m. Results in seconds.

Year	1st	2nd	3rd	4th	5th	6th	7th	8th	Mean
1924	47.6	48.4	48.6	48.8					48.35
1928	47.8	48.0	48.2	48.4	48.8	49.0			48.37
1932	46.2	46.4	47.4	48.2	48.2	48.8			47.53
1936	46.5	46.7	46.8	46.8	47.8	48.2			47.13
1948	46.2	46.4	46.9	47.2	47.9	50.2			47.47
1952	45.9	45.9	46.8	47.0	47.0	47.1			46.62
1956	46.7	46.8	47.0	47.0	48.1	48.3			47.32
1960	44.9	44.9	45.5	45.6	45.9	45.9			45.45
1964	45.1	45.2	45.6	45.7	46.0	46.0	46.3	46.8	45.84
1968	43.86	43.9	44.4	45.0	45.3	45.4	45.4	47.6	45.11
1972	44.66	44.8	44.92	45.13	45.31	45.59	45.68		45.16
1976	44.26	44.40	44.95	45.04	45.24	45.40	45.57	45.63	45.06
1980	44.60	44.84	44.87	45.09	45.10	45.55	45.56	46.33	45.24
1984	44.27	44.54	44.71	44.75	44.75	44.93	45.35		44.76
1988	43.87	43.93	44.09	44.55	44.72	44.94	44.95	45.03	44.51
1992	43.50	44.21	44.24	44.25	44.52	44.75	45.10	45.18	44.47

Table A4. Women's 400m. Results in seconds.

Year	1st	2nd	3rd	4th	5th	6th	7th	8th	Mean
1964	52.0	52.2	53.4	54.4	54.6	55.2	55.4	55.5	54.09
1968	52.0	52.1	52.2	52.5	52.7	52.7	52.8	53.0	52.50
1972	51.08	51.21	51.64	51.86	51.96	51.99	52.19	52.39	51.79
1976	49.29	50.51	50.55	50.56	50.65	50.90	50.98	51.66	50.64
1980	48.88	49.46	49.66	50.07	50.17	51.33	51.35	52.4	50.41
1984	48.83	49.05	49.42	49.91	50.25	50.37	50.45	51.56	49.98
1988	48.65	49.45	49.9	50.16	50.72	51.12	51.17	51.25	50.30
1992	48.83	49.05	49.64	49.69	49.93	50.11	50.19	50.87	49.79

Table A5. Men's Long Jump. Results in metres.

Year	1st	2nd	3rd	4th	5th	6th	7th	8th	Mean
1924	7.44	7.27	7.26	7.07	6.99	6.92	6.89	6.86	7.09
1928	7.73	7.58	7.40	7.39	7.35	7.32	7.32	7.29	7.42
1932	7.64	7.60	7.45	7.41	7.39	7.15	6.66	6.43	7.22
1936	8.06	7.87	7.74	7.73	7.73	7.67	7.41	7.34	7.69
1948	7.82	7.55	7.54	7.45	7.27	7.07	7.03	7.00	7.34
1952	7.57	7.53	7.30	7.23	7.16	7.14	7.10	7.02	7.26
1956	7.83	7.68	7.48	7.44	7.36	7.30	7.28	7.27	7.46
1960	8.12	8.11	8.04	8.00	7.69	7.68	7.66	7.58	7.86
1964	8.07	8.03	7.99	7.60	7.44	7.34	7.30	7.26	7.63
1968	8.90	8.19	8.16	8.12	8.09	8.02	7.97	7.94	8.17
1972	8.24	8.18	8.03	8.01	7.99	7.96	7.91	7.75	8.01
1976	8.35	8.11	8.02	8.00	8.00	7.89	7.88	7.82	8.01
1980	8.54	8.21	8.18	8.13	8.13	8.10	8.09	8.02	8.18
1984	8.54	8.24	8.24	8.16	7.99	7.97	7.87	7.81	8.10
1988	8.72	8.49	8.27	8.08	8.08	8.00	7.92	7.89	8.18
1992	8.67	8.64	8.34	8.11	8.08	8.04	7.98	7.87	8.22

Table A6. Women's Long Jump. Results in metres.

Year	1st	2nd	3rd	4th	5th	6th	7th	8th	Mean
1948	5.695	5.60	5.575	5.57	5.545	5.495	5.38	5.35	5.53
1952	6.24	6.14	5.92	5.90	5.81	5.81	5.75	5.74	5.91
1956	6.35	6.09	6.07	5.89	5.88	5.85	5.85	5.82	5.98
1960	6.37	6.27	6.21	6.19	6.16	6.11	6.11	6.01	6.18
1964	6.76	6.60	6.42	6.40	6.35	6.24	6.24	6.23	6.41
1968	6.82	6.68	6.66	6.48	6.47	6.43	6.40	6.40	6.54
1972	6.78	6.77	6.67	6.52	6.51	6.49	6.48	6.46	6.59
1976	6.72	6.66	6.60	6.59	6.59	6.54	6.39	6.38	6.56
1980	7.06	7.04	7.01	6.95	6.87	6.83	6.71	6.71	6.90
1984	6.96	6.81	6.80	6.78	6.77	6.67	6.53	6.44	6.72
1988	7.40	7.22	7.11	7.04	6.73	6.62	6.60	6.55	6.91
1992	7.14	7.12	7.07	6.76	6.71	6.68	6.66	6.62	6.85

Table A7. Men's High Jump. Results in metres.

Year	1st	2nd	3rd	4th	5th	6th	7th	8th	Mean
1924	1.98	1.95	1.92	1.88	1.88	1.85	1.85	1.83	1.89
1928	1.94	1.91	1.91	1.91	1.91	1.88	1.88	1.88	1.90
1932	1.97	1.97	1.97	1.97	1.94	1.94	1.90	1.90	1.95
1936	2.03	2.00	2.00	2.00	1.97	1.94	1.94	1.94	1.97
1948	1.98	1.95	1.95	1.95	1.95	1.90	1.90	1.90	1.94
1952	2.04	2.01	1.98	1.98	1.95	1.95	1.95	1.90	1.97
1956	2.12	2.10	2.08	2.06	2.03	2.00	2.00	2.00	2.05
1960	2.16	2.16	2.14	2.14	2.09	2.03	2.03	2.03	2.09
1964	2.18	2.18	2.16	2.14	2.14	2.09	2.09	2.09	2.13
1968	2.24	2.22	2.20	2.16	2.14	2.14	2.14	2.12	2.17
1972	2.23	2.21	2.21	2.18	2.18	2.15	2.15	2.15	2.18
1976	2.25	2.23	2.21	2.21	2.18	2.18	2.18	2.18	2.20
1980	2.36	2.31	2.31	2.29	2.24	2.24	2.21	2.21	2.27
1984	2.35	2.33	2.31	2.31	2.29	2.29	2.29	2.27	2.31
1988	2.38	2.36	2.36	2.36	2.34	2.34	2.31	2.31	2.35
1992	2.34	2.34	2.34	2.34	2.34	2.31	2.31	2.28	2.33

Table A8. Women's High Jump. Results in metres.

Year	1st	2nd	3rd	4th	5th	6th	7th	8th	Mean
1928	1.59	1.56	1.56	1.51	1.48	1.48	1.48	1.48	1.52
1932	1.657	1.657	1.60	1.58	1.58	1.58	1.55	1.49	1.59
1936	1.60	1.60	1.60	1.58	1.58	1.55	1.55	1.55	1.58
1948	1.68	1.68	1.61	1.58	1.58	1.58	1.55	1.55	1.60
1952	1.67	1.65	1.63	1.58	1.58	1.58	1.58	1.58	1.60
1956	1.76	1.67	1.67	1.67	1.67	1.67	1.67	1.64	1.67
1960	1.85	1.71	1.71	1.71	1.68	1.65	1.65	1.65	1.70
1964	1.90	1.80	1.78	1.74	1.71	1.71	1.71	1.71	1.76
1968	1.82	1.80	1.80	1.78	1.78	1.78	1.76	1.76	1.79
1972	1.92	1.88	1.88	1.85	1.85	1.85	1.85	1.82	1.86
1976	1.93	1.91	1.91	1.89	1.89	1.87	1.87	1.87	1.89
1980	1.97	1.94	1.94	1.91	1.91	1.91	1.91	1.88	1.92
1984	2.02	2.00	1.97	1.94	1.94	1.94	1.91	1.91	1.95
1988	2.03	2.01	1.99	1.96	1.93	1.93	1.93	1.90	1.96
1992	2.02	2.00	1.97	1.94	1.94	1.91	1.91	1.91	1.95

MODELLING DISTANCE PREFERENCE IN THOROUGHBRED RACEHORSES

William F. Benter[1], George J. Miel[2] and P. Diane Turnbough[3]

Abstract

A computer model is presented for empirically estimating racehorse performance relative to race distance. The method uses least-squares curve fitting based on the horse's race history. The model yields a quantitative estimate of expected horse performance at given distances suitable for inclusion in a computerized handicapping model. Constraints are employed during the fitting procedure to improve predictions in cases with limited available data.

1. INTRODUCTION

In recent years various methods have been proposed for predicting the outcome of thoroughbred horse races from fundamentals. *Fundamentals* in this case refers to attributes of the horse, jockey, or race which help predict horse performance. These methods have generally taken the form of either multiple regression or multinomial logit models Brecher [6], Bolton [5], Benter [4]. In such formulations, horse performance is estimated as a function of several predictor variables chosen to reflect those attributes of the horse, jockey, or race which are thought to have significance in predicting race outcome. While much has been written about the general form that such models should take, little attention has been given to the selection and formulation of the predictor variables.

As horse racing is a highly complex phenomenon, it is difficult to specify a model *a priori*. An alternative is to obtain a large database of past races and perform a trial and error empirical search for the best model. In this approach, the modeler begins with the notion that a certain attribute such as the "horse's fitness", the "jockey's ability", or "weight to be carried" is likely to be related to horse performance, and then formulates a predictor variable which provides a measure of that attribute. The candidate variable is then tested for predictive significance over a large sample of races via multiple regression or some other appropriate statistical method.

The predictive power of a given model is limited by the extent to which the predictor variables capture the real underlying determinants of race outcome. While many of the likely determinants are simple data explicitly available in the racing program, (e.g. weight to be carried by the horse in today's race) others are abstract constructs (e.g. the "current fitness" of the horse). Searching for predictive constructs and

---

[1]Happy Valley Syndicate, Hong Kong

[2]University of Nevada, Las Vegas, Nevada, USA

[3]University of Nevada, Las Vegas, Nevada, USA

finding the best quantitative formulation of them are perhaps the most difficult aspects of modelling horse performance. We attempt to address herein one such predictive construct.

Races for thoroughbreds may be run at distances as short as 0.4 km to over 5 km. It is an established tenet of horse racing lore that most horses possess an attribute which could be called "distance preference", that is, each horse tends to perform better at some distances and worse at others. In estimating how distance will effect a horse's performance in an upcoming race, a handicapper might consider the age, sex, breeding, etc. of the horse, but he will principally consider how the horse has performed in the past in races at different distances. In this paper, we discuss the underlying mechanisms of distance preference and describe a quantitative method for estimating it using only past performance data.

2. DISTANCE PREFERENCE

Mathematical models, based on principles of physics and physiology, have been used to predict performance of human runners. The Hill-Keller model Keller [8] which uses the laws of motion and of balance of linear momentum with a simplified representation of the runner's energy, yields a good overall representation of world records from 60 yds. to 10 km. Using a completely different approach, the Péronnet-Thibault model is empirically-based, relating running performance to metabolic and energy processes Péronnet [10]. While this model also yields good correlation with world record performances, being strictly empirical, it gives little insight into the underlying physical processes. Whether based on fundamental analytic laws or on empirical data fitting, mathematical models of running rely on accurate measurements of physiological and other parameters.

It is tantalizing to think, in view of the ready availability of computers, that mathematical models of human running could be extrapolated to thoroughbred racing. Whereas techniques for making diverse physiological measurements are well-established for human subjects, even if hypothetically such techniques were to be modified and regularly carried out on thoroughbreds, one would not expect to see the results published in the racing program. For this reason, mathematical models akin to those mentioned earlier are not likely to be useful for predicting the outcome of horse races. However, it is informative to consider at a high level the analytical background behind the concept of distance preference.

2.1 Physiological Considerations

Our model of distance preference considers the physiological interplay between the anaerobic and the aerobic pathways of energy conversion. Our reasoning uses human physiology Noble [9] with the understanding that the basic mechanism as related to our model is similar for horses.

At the cellular level, energy for muscular activity is provided by synthesis of adenosine triphosphate (ATP) either anaerobically (without oxygen) or aerobically (with oxygen). The anaerobic pathway, rapid but of short duration, is the primary source of energy for sprinters. Maximal exertion during sprinting results in accelerated glycolysis and increased lactate production, quickly followed by impaired performance. During an extended race, the replenishment of energy relies on the transport of oxygen to the active cells. In the aerobic pathway, the rate of oxygen availability is a limiting factor in the conversion of energy. Oxygen consumption rises with the intensity of exercise and the ability to sustain consumption over time depends on the interaction of race duration and intensity. Although independent as physiological systems, the aerobic and anaerobic pathways of energy conversion are highly interrelated within the activity context. Even at extremes of endurance and sprint activities, both pathways are in use. For any one horse at any one time during a race, we envision the two systems acting in concert, with one predominating.

As we will show below, such considerations lead us to conjecture that the individual distance d versus preference p relationship for a thoroughbred can be idealized as a unimodal curve $p = f(d)$, monotonically increasing to a most preferred distance D^* and then decreasing thereafter. Hence for a given interval $[D_0,D_1]$ of distances that the horse will race in, the portion of the idealized curve over $[D_0,D_1]$ is respectively increasing if $D_1 \le D^*$, decreasing if $D^* \le D_0$, and otherwise has a maximum at D^* inside $[D_0,D_1]$. We thus wish to estimate the curve under the constraint that it be unimodal and that it have no local minimum inside the interval $[D_0,D_1]$ of interest. In order to justify this assumption, we have studied the concept of distance preference within the framework of the Hill-Keller mathematical model of running Keller [8]. For reasons of brevity, a description of the resulting analysis is inappropriate here. However, it is informative to discuss informally the analytical background behind the assumption.

Consider the four variables

d	distance (m)
V	maximum attainable average velocity (m/s)
E_0	initial energy level (J/kg)
σ	rate of energy replenishment (J/kg s)

and two corresponding curves, a $V = V(d)$ curve on the (d,V)-plane and a curve ε on the (σ, E_0)-plane. The function $V(d)$ represents record performances for races at distance d. The points on the curve ε denote the two types of energy attributes, σ for aerobic and E_0 for anaerobic, corresponding to the points on the $V(d)$ curve. As distance increases, the function $V(d)$ is at first increasing (for sprints when E_0 is dominant) and then decreasing (for longer distances when σ takes over). There is a bijection $d = \delta(\sigma, E_0)$ between points (σ, E_0) on ε and the distance scale d, such that $V(d)$ corresponds to the pair of energy values $\delta^{-1}(d) = (\sigma, E_0)$ on ε. The optimal energy curve ε is monotone decreasing. On that curve, E_0 is high and σ is low for sprint distances, while the opposite holds for long distances.

For a horse with energy pathways represented by a pair of values (σ', E_0'), consider the optimal average velocity $v(d)$ that the horse can attain for a race of distance d. (The Hill-Keller theory allows characterization of the function $v(d)$ if two additional physiological parameters are specified.) Necessarily, the inequality

$$v(d) \le V(d)$$

holds for every d. If the point (σ', E_0') is on the curve ε then for the distance $d' = \delta(\sigma', E_0')$ we have $v(d') = V(d')$ and $v(d) < V(d)$ whenever $d \neq d'$. In this case, for a race of distance d', the horse is capable of reaching record performance. If the point (σ', E_0') is not on the optimal energy curve ε then the curve v(d) is strictly below the record curve V(d).

The horse's preferred distance is a d value which minimizes the difference

$$\psi(d) = V(d) - v(d).$$

Thus, at the preferred distance, the horse's best average velocity is closest to the record average velocity. In essence, the assumption that distance preference be unimodal means that the curve v(d) cannot approach minimally the curve V(d) over more than one region on the d scale because of the interplay between the physiological parameters σ and E_0. We can quantify distance preference as a composite function,

$$p(d) = \phi \circ \psi(d),$$

where $\phi(x)$ is a strictly decreasing function for $x \geq 0$, so that p is greatest when ϕ is smallest. This means that the horse is capable of performing best for distances d which yield high values of p. To illustrate, if we take $\phi(x) = e^{-kx}$, $k > 0$, and if (σ', E_0') is on the optimal energy curve ε then $p(d) \leq 1$ with equality only when $d = d'$. If $(\sigma', E_0') \notin \varepsilon$ then $p(d) < 1$ for every d and for a distance interval $[D_0, D_1]$ we again expect p(d) to be unimodal with a maximum inside that interval.

2.2 Objectives and Limitations

The goal of the modelling procedure is to be able to estimate, for any given race distance, the effect that the distance will have on a particular horse's expected performance. The data available with which to make this estimate is the record of the horse's past races. In each of these races the distance of the race is known, and we can use some measure such as finishing position or elapsed time, to quantify the horse's performance. This will give us a set of data points from which we can estimate the distance/preference relationship.

Two main difficulties arise in attempting to estimate this relationship. The first is that the observed performance in each of a horse's past races is not only a function of distance, but also of many different factors such as fitness, type of surface, jockey skill, etc., as well as a number of unknowable random factors. Observations of an individual horse's performances over time at races of identical distance show that there is considerable variability in performance which is not related to distance. For the purposes of our modelling procedure, we will assume that this state of affairs amounts to the following: Each measured past performance of a horse contains some 'signal' about its distance preference, but is corrupted by a large component of Gaussian 'noise'.

The second difficulty is that the number of past races may be very small. The high level of random noise in each observed past performance dictates that a large sample size is necessary to make significant statistical estimations. Unfortunately, even by the end of a horse's career the total number of past races is barely adequate for making confident estimates, and the model will be required to make estimates for horses at all experience levels, including those with only a few past races. Our goal then is to derive a robust algorithm that produces meaningful estimates based on sparse, noisy, and heterogeneous data.

3. MODELLING PROCEDURE

The available data consists of a series of distance/performance observations corresponding to each of a horse's past races. The performance measure used by the authors is an adjusted finish position which places each performance on a scale of 0.5 to -0.5. We need not describe this measure in detail since the specifics of it are not important to the following discussion.

The first step in the calculations is to normalize these past performances so that the mean performance is zero. This is done because the desired output of the distance preference model is an estimate of only the effect that distance will have on the horse's performance, which will be added to other variables within the context of a comprehensive computer handicapping model. At this stage, the data can be represented as points on a distance/performance plane.

The second step is to perform a constrained parabolic least squares fit through this data. (See Figure 3 below) A second order polynomial was chosen for reasons of simplicity, robustness, ease of control and because the parabolic shape is adequate for representing the idealized distance preference relationship mentioned earlier. In view of the sparse and very noisy data, fitting with higher order polynomials or splines (Bartels [3]) is not warranted, being harder to control and more prone to following random aberrations in the data.

3.1 Fictitious Data Points

A least squares parabolic fit requires at least one observation at each of three separate distances. However, even when a horse has raced at only two separate distances, there is meaningful information to extract about its preference. To deal with such cases, we employ the device of adding fictitious data points to the horse's record before the parabolic fitting. The placement of these fictitious data points is always at intervals along the distance-axis. These fictitious data points, which we expediently call *tack points*, tend to flatten the parabola toward the distance axis, namely, toward zero preference when there is insufficient data to infer otherwise.

The tack points serve a much larger purpose than simply allowing a solution in cases of less than three observations: they also tend to prevent the least squares procedure from simply "fitting the noise" in cases with limited data. As an example, consider the case of a horse with four races as represented in Figure 1. The solid line shows the shape of the curve fitted using only the real data. The dashed line shows the shape of the curve fitted to the data including three tack points.

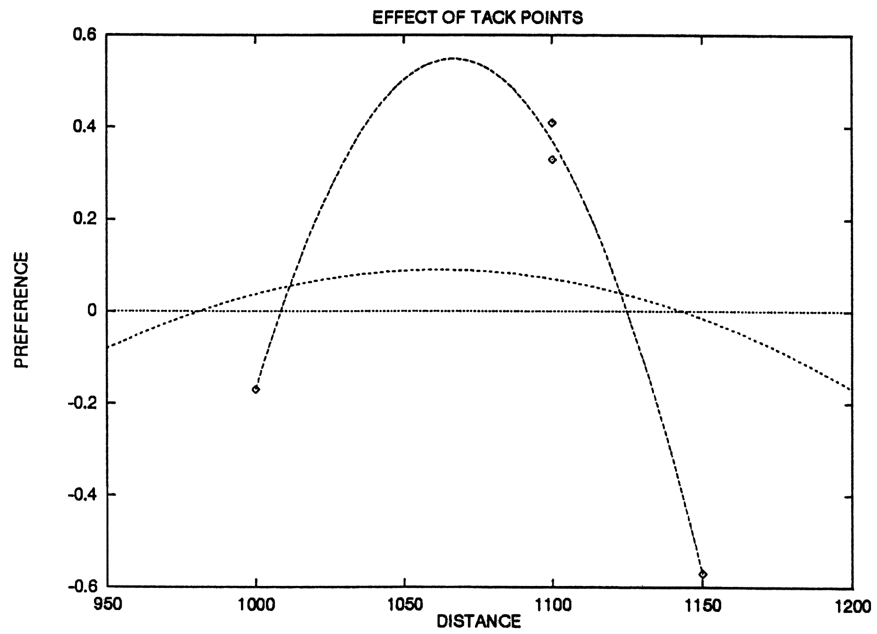


Figure 1: Effect of three tack points

Given the assumption that the random noise in each observation accounts for much more of the variance than the effect we are trying to model, the second curve appears to be a more plausible estimate of a horse's distance preference. In a sense the tack points are a way of introducing our prior knowledge that the likely magnitude of a horse's distance preference is small relative to the random variation in performance. Varying the number and placement of the tack points allows the modeler considerable flexibility in fine-tuning the model. Adding more tack points makes the curve more inflexible, leading to more conservative estimates of distance preference. Large estimated distance preferences would only occur when a horse had many races exhibiting a clear pattern.

Via trial and error, we have found that a scheme using three tack points on the distance axis, one positioned slightly to the left of the distance of the shortest race in which the horse has participated, one slightly to the right of the longest race, and the last at the midpoint between the other two, produces good estimates. A convenient feature of the *tack points* approach is that as the number of real observations increases the effect of the tack points diminishes. The shape of the fitted curve evolves smoothly as a horse's career progresses, from being dominated by the artificial data in the early stages, to closely following the real data as the number of races becomes large.

3.2 Disallowed shapes

Even with the use of the tack points, in certain cases, particularly when the horse has had few races, the shape of the fitted curve will violate our original assumption that the curve be unimodal. We presume that this is the result of random noise in the observations. In order to disallow these shapes while still retaining as much of the information as we can from the data, we use a special procedure to adjust the shape of the polynomial.

A frequent practice in computer aided design (CAD) is the interactive use of a graphics terminal in order to adjust parameters defining a curve until the curve has an acceptable shape. In 1962, Bézier and Casteljau, while working on CAD systems for the French car companies Renault and Citroën respectively, developed independently a mathematical method for doing this. Since the Renault software was described in the open literature by Bézier, the underlying theory often bears his name. The resulting methodology has since evolved and is now established as a major tool in computer graphics (Bartels [3]).

In its simplest form, a *Bézier curve* over an interval [a, b] is a polynomial of degree ≤ n, represented as a linear combination

$$p(s) = \sum_{i=0}^{n} \beta_i B_i^n(s)$$

of n+1 *Berstein polynomials*

$$B_i^n(s) = \binom{n}{i} \frac{(b-s)^{n-i}(s-a)^i}{(b-a)^n}$$

Each coefficient β_i is associated with a *control point* on the plane,

$$Q_i(t_i, \beta_i), \quad t_i = a + i(b-a)/n$$

The abscissae t_i of these n+1 control points are equally spaced with $t_0 = a$ and $t_n = b$.

It is known that if the control points are monotonic, convex or concave so is the resulting curve p(s). The curve always passes through the end control points, $p(a) = \beta_0$ and $p(b) = \beta_n$, and it is contained in the convex hull of the n+1 control points. Since this convex hull is often the same as the region enclosed by the *Bézier polygon*, obtained by connecting successive control points and finally connecting Q_n back to Q_0, changing the value of a coefficient β_i, which is equivalent to moving the corresponding control point Q_i up or down, has a direct and intuitive effect on the function p(s). In practice, the user adjusts the β_i values until the plot of p(s) acquires an acceptable shape.

For our purpose of estimating the distance preference of a horse, we use a Bézier curve of degree n = 2 over the unit interval [0, 1],

$$p(s) = \beta_0 B_0^2(s) + \beta_1 B_1^2(s) + \beta_2 B_2^2(s) \tag{1}$$

where

$$B_0^2(s) = (1-s)^2, \quad B_1^2(s) = 2s(1-s), \quad B_2^2(s) = s^2$$

The corresponding three control points are

$$Q_0{:}(0, \beta_0), \quad Q_1{:}\left(\frac{1}{2}, \beta_1\right), \quad Q_2{:}(1, \beta_2)$$

and the following properties are valid:

1. The parabola passes through the two points Q_0 and Q_2.

2. The curve lies within the triangle with vertices Q_0, Q_1, and Q_2.

3. There is a local minimum inside the interval (0, 1) if and only if $\beta_1 < \min(\beta_0, \beta_2)$.

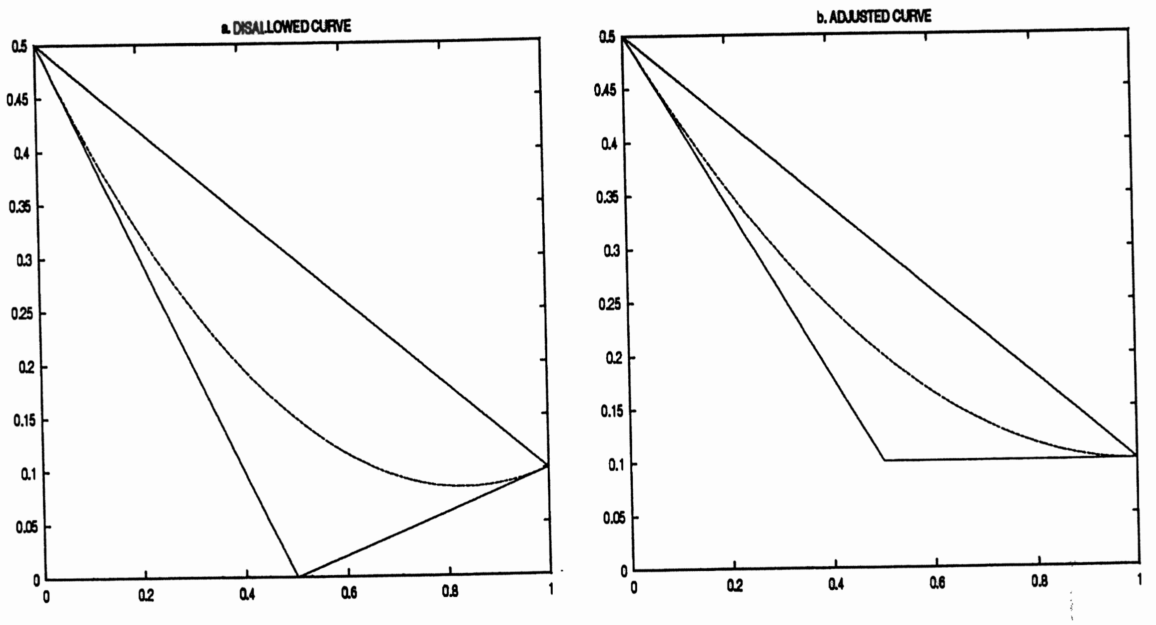


Figure 2: Effect of the Bézier adjustment

From physiological considerations described earlier, we do not allow a distance preference curve to have a local minimum inside the distance interval of interest. Property 3, illustrated in Figure 2a, provides a simple means of adjusting the shape of inadmissible curves defined over the unit interval. In a computer program, the single overwrite statement

$$\beta_1 := \max(\beta_1, \min(\beta_0, \beta_2)) \tag{2}$$

accomplishes the desired task. If $\beta_1 \geq \min(\beta_0, \beta_1)$ the parabola has no local minimum inside (0, 1) and the parabola is left unchanged. Otherwise, in order to flatten the local minimum, we move the middle control point Q_1 upward until it is aligned with the lowest of the two other control points Q_0 and Q_2; see Figure 2b.

3.3 Algorithm

The following procedure combines the above adjustment with conventional least squares fitting of a parabolic curve through m data points (d_i, p_i), $1 \leq i \leq m$.

Step 1. Choose $D_0 \leq \min d_i$ and $D_1 \geq \max d_i$, let $\Delta = D_1 - D_0$ normalize the distances d_i to the unit interval,

$$s_i = (d_i - D_0)/\Delta, \qquad 1 \leq i \leq m,$$

and fit a parabola

$(s) = \gamma_0 + \gamma_1 s + \gamma_2 s^2$

through the m points (s_i, p_i) by solving the 3×3 system of normal equations,

$$\begin{bmatrix} m & \Sigma s_i & \Sigma s_i^2 \\ \Sigma s_i & \Sigma s_i^2 & \Sigma s_i^3 \\ \Sigma s_i^2 & \Sigma s_i^3 & \Sigma s_i^4 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} \Sigma p_i \\ \Sigma p_i s_i \\ \Sigma p_i s_i^2 \end{bmatrix}$$

Step 2. Execute the following statements to possibly adjust the least squares parabola:

$$\beta_1 := \max\left(\gamma_0 + \frac{1}{2}\gamma_1, \min(\gamma_0, \gamma_0 + \gamma_1 + \gamma_2)\right)$$
$$\gamma_2 := 2\gamma_0 + \gamma_1 + \gamma_2 - 2\beta_1$$
$$\gamma_1 := -2\gamma_0 2\beta_1$$
$$\gamma_0 := \frac{1}{m}\sum_{i=1}^{m}\left(p_i - \gamma_1 s_i = \gamma_2 s_i^2\right)$$

Step 3. Find the coefficients of the corresponding curve

$$p = f(d) = c_0 + c_1 d + c_2 d^2$$

for the actual distances over $[D_0, D_1]$:

$$c_2 := \gamma_2 / \Delta^2$$
$$c_1 := \frac{\gamma_1}{\Delta} - 2c_2 D_0$$
$$c_0 := \gamma_0 - c_1 D_0 - c_2 D_0^2$$

The last step is a straight-forward consequence of the transformation

$$p\left(\frac{d - D_0}{\Delta}\right) = f(d)$$

The second step alters the least squares parabola of the first step if there is a local minimum inside the unit interval.

The correspondence between the standard form (3) of the least squares parabola and its Bézier representation (1) is given by

$$\beta_0 = \gamma_0, \qquad \gamma_0 = \beta_0,$$
$$\beta_1 = \gamma_0 + \frac{1}{2}\gamma_1, \qquad \gamma_1 = -2\beta_0 + 2\beta_1,$$
$$\beta_2 = \gamma_0 + \gamma_1 + \gamma_2, \qquad \gamma_2 = \beta_0 - \beta_1 + \beta_2,$$

The first statement in Step 2 corresponds to (2) in the Bézier representation. If $\beta_1 = \gamma_0 + \frac{1}{2}\gamma_1$ then the next three statements leave all coefficients $\gamma_2, \gamma_1, \gamma_0$ unchanged. Otherwise, two types of adjustments take place. First, the flattening of the local minimum, created by moving the middle control point upward, alters the values of γ_2 and γ_1 but by itself leaves γ_0 unchanged. However, the second adjustment, carried out by the last statement in Step 2, evaluates the best least squares estimate of γ_0 for the parabola with new values of γ_2 and γ_1 resulting from the first adjustment. The second adjustment, a vertical translation of the Bézier-reshaped parabola, occurs only when γ_1 and γ_2 have been altered.

Several observations are in order. The reshaping of the parabola, via movement of the middle control point, is meant to occur, if at all, in cases of limited data. The scheme was designed under the working assumption that the *shape* of the distance preference curve is as important as its *fit* through the (extremely noisy) data. It turns out, as we will show later, that this assumption is borne out statistically. Instead of applying a sophisticated numerical method for least squares estimation with inequality constraints (e.g., see section 28 in Björck [2]), we opted for the above scheme for reasons of robustness, simplicity, and easy experimentation within the overall handicapping system.

4. DISCUSSION

The estimation procedure described above presumes that thoroughbred horses possess significant distance preferences and that these can be estimated from their past races. The extent to which this is true cannot be determined without empirical tests on actual data. In what follows, we show statistically that horses do possess this trait and that the procedure for estimating it described above is quite effective.

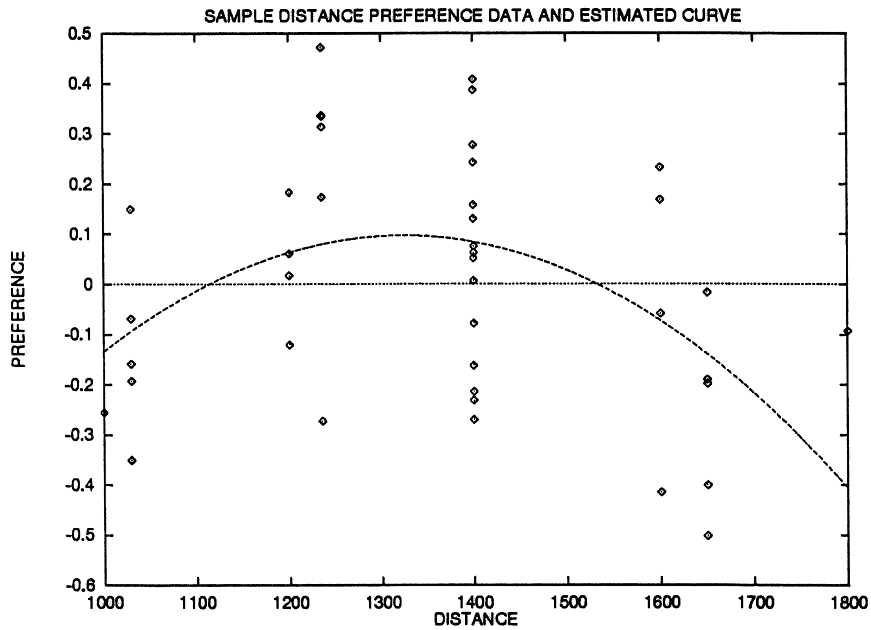


Figure 3: Estimated curve for a horse with 41 races

Figure 3 shows the data points and fitted curve for a horse with a relatively large number of races. The tack points, (not shown in the figure) have little effect on the shape of the estimated curve. No Bézier adjustment was necessary in this example because the distance preference curve does not have a local minimum.

4.1 Empirical Validation

Having found the coefficients of the distance preference parabola, we are now able to make specific predictions of horse performance at given race distances. With access to a large database of horse performances, statistical tests of the validity of these predictions can be made. The authors have carried out such tests on a database of 4,602 races run in Hong Kong during the years 1986-1996. The sample contained 47,404 individual horse performances with full prior career records available for each. Consider the following two linear regressions:

$$E\{Y\} = b_0 + b_1 X_1$$

$$E\{Y\} = b_0 + b_1 X_1 + b_2 X_2$$

where the response variable Y and the two predictor variables X_1 and X_2 are:

Y = normalized finish position in today's race
X_1 = average normalized finish position in past races
X_2 = distance preference estimate

The results of these regressions are summarized in the following two tables.

	coefficient	t-ratio
b_0	0.000	0.0
b_1	0.598	62.0

	coefficient	t-ratio
b_0	0.000	0.0
b_1	0.585	51.8
b_2	0.710	21.6

The r^2 values of these two regressions are 0.075 and 0.084 respectively. The large value of the t-ratio achieved by the distance preference estimate X_2 indicates the high significance of this variable and the increase in r^2 between the two regressions demonstrates its large positive contribution to the prediction of horse finishing position.

Trial and error experimentation by the authors have shown that the predictive power of the model is very sensitive to the number of tack points used and their placement. The optimal number of tack points seems to be three or four. Further tests both with and without the Bézier adjustments however show that this refinement had only a statistically insignificant effect on the accuracy of the estimations in this data sample.

4.2 Further Improvements

While the procedure described above evidently extracts a significant amount of predictive information from the available data, what fraction this represents of the amount achievable by the hypothetical 'optimal' model is not known. The authors believe that further improvements could produce significant gains. Other estimation procedures, perhaps employing curves other than the parabola used above, may yield improvements. Also, incorporating the assumption that a horse's distance preference is not necessarily constant over time may have merit.

This model represents the result of a practical effort to produce quantitative estimates of one attribute of thoroughbred racehorses. Its principal strengths are its simplicity and robustness. Inclusion of a sufficient number of tack points, and the Bézier adjustment procedure to prohibit disallowed shapes, ensure that the model will always produce a reasonable estimate of distance preference even in cases when little or no data is available. This reliability makes it suitable for inclusion in a fully computerized handicapping program which does not require human supervision.

REFERENCES

[1] Ainslie T., Encyclopedia of Thoroughbred Handicapping, (William Morrow and Company, New York, 1978).

[2] Björck A., Least Squares Method, chapter in Handbook of Numerical Analysis, Vol 1, edited by P.G. Ciarlet and J.L. Lions, (Elsevier/North Holland, New York, 1990).

[3] Bartels R., Beatty J., and Barsky B., An Introduction to Splines for use in Computer Graphics and Geometric Modeling, (Morgan Kaufmann, Los Altos, CA, 1987).

[4] Benter W., Computer based horse race handicapping and wagering systems: a report, on pp. 183-198 in Efficiency of Racetrack Betting Markets, edited by D.B. Hausch, V.S.Y. Lo, and W.T. Ziemba, (Academic Press, New York, 1994).

[5] Bolton R.N. and Chapman R.G., Searching for positive returns at the track: a multinomial logit model for handicapping horse races, *Management Science*, Vol 32, 1040-60 (1986).

[6] Brecher S.L., Beating the Races with a Computer, (Software Supply, Long Beach, CA, 1980).

[7] Jones G., Horse Racing Logic, (Vantage Press, New York, 1989).

[8] Keller J.B., A theory of competitive running, *Physics Today*, Vol 26, 42-47 (1973).

[9] Noble B.J., Physiology of Exercise and Sport, (Times Mirror/Mosby College Publishing, St. Louis, 1886).

[10] Péronnet F. and Thibault G., Mathematical analysis of running performance and world records, *J. Applied Physiology*, Vol 67, p 453 (1989).

[11] Pritchard W.G., Mathematical models of running, *SIAM Review*, Vol 35, 359-379 (1993).

[12] Pritchard W.G. and Pritchard J.K., Mathematical models of running, *American Scientists*, Vol 82, 546-553 (1994).

BETTING STRATEGIES IN HORSE RACES

Ravi Phatarfod[1]

Abstract

In this paper we consider the strategies a gambler may employ in situations such as horse races. We assume that the gambler knows which horses have odds which are favourable to him, that he wants to bet in such a way as to have a given positive expectation of win on a race, and that he wishes to minimise the probability of loss of his finite capital. We show that the best strategy is to bet on all the horses whose odds are favourable so as to minimise the probability of loss on a race. We further show that in order to achieve the last objective it is advisable to have a bet on a horse with fair odds, and at times on a horse with unfavourable odds, in addition to a bet on a horse with favourable odds.

1. INTRODUCTION

The use of probability theory to investigate gambling strategies is not new; see, Feller [1], Dubins and Savage [2], Epstein [3], Breiman [4], Rotando and Thorp [5], to name just the more recent contributions. Feller showed that if a gambler's objective is to increase his initial capital of b by an amount a (a < b), and the game is unfavourable to him, then for even money games, and with a view to minimising the probability of his ruin, his initial bet should be for the amount a, and if he loses that game, his next bet should be for the amount 2a etc; if however the game is a favourable one, he should bet as small an amount as possible. Rotando and Thorp showed that if the game is for even money and the game is a favourable one, i.e. p the probability of win is greater than $\frac{1}{2}$, then to maximise the exponential rate of growth of the gambler's capital, the gambler should bet the fraction p – q (where q = 1 – p of his capital at every stage of the play. Thus, by and large, the strategies explored so far are the ones the gambler may employ from game to game in a sequence of games. In this paper we consider the strategies a gambler may employ within each game. Horse racing provides the most common example of this situation. We shall discuss the subject matter in the context of horse racing, although the conclusions reached are valid in other contexts also. Within each game (i.e. a horse race), there are a number, (usually about ten to twenty) of mutually exclusive betting propositions (i.e. horse to win). We shall assume that the gambler knows the probability of each horse winning the race, and is offered odds (or prices) about these, so that he can divide the race field into the three categories of favourable bets (i.e. those for whom the expectation is positive), fair bets (i.e. those for whom the expectation is zero) and unfavourable bets (i.e. those for whom the expectation is negative). We assume that there is at least one favourable bet in a game and that the gambler is able to bet on a sequences of such games. We also assume that he has a large but finite capital and is

[1]Department of Mathematics, Monash University, Clayton Victoria 3168

playing against an infinitely rich adversary, and that he wants to bet in such a way as to produce a given (positive) expectation of gain per race. The question is: How should he bet if his objective is to minimise the probability of his ruin, i.e. the exhaustion of his capital?

For example suppose in a race there are, among others, three horses A, B, C.

The odds on offer against them winning the race are 5/1, 3/1 and 3/1 and the probabilities of them winning are 0.2, 0.3 and 0.3 respectively. The question is: What bets should the gambler take? Should he bet on the horse with the most favourable odds? Or, should he hedge his bets, i.e bet on all the horses with favourable odds?

We shall show that the probability of his ruin is a decreasing function of the probability of win on a race (with the same expectation of gain), so that in the situation above, rather than bet only on horse A, he should bet on all the three horses A, B, C. Indeed, if the gambler is in the fortunate position of being offered such bets in a succession of races, then with a capital of 80 dollars, betting 8 dollars on A above (to give an expectation of gain of $1.60 per race) makes the probability of his ruin equal to 0.4662, whereas betting amounts 2, 3 and 3 dollars on A, B and C respectively (to give the same expectation of $1.60 of gain on the race) would make the probability of his ruin considerably smaller, namely 0.00013.

On the surface the difference between the two probabilities is striking. However, we need not seek too far to see the reason for this difference. The probability of ruin depends very heavily on the results of initial games, and the probability of getting a succession of losses when the loss probability per game is 0.8 is far greater than what it is when the loss probability per game is only 0.2. So, although there is not much to choose between the two alternatives when the capital is infinite, the difference between the loss probabilities per game has a telling consequence when the capital is finite. In practical terms the latter alternative has another advantage over the former. In practice, a gambler is more like to be able to correctly assess that in a particular race, the probability is 0.8 that the winner would come from one of the three horses A, B and C than correctly apportion probabilities of win to individual horses.

If now, we suppose that the situation is such that there is only one favourable bet in the race, e.g. *A* as above and he is offered a fair bet on horse D at even money in the same race, then betting 8 dollars each on both A and D would (if such situations were available in a succession of races) be preferable to betting 8 dollars on A alone as the probability of ruin now is 0.37729, somewhat less than 0.4662 for betting on A alone. More surprisingly, the probability of ruin for the combination of A and an unfavourable bet in the same race (albeit only marginally unfavourable) is less than what it would be for the single bet on A. These conclusions go counter to our intuition that the gambler must avoid, at all cost, fair and unfavourable bets.

The preceding discussion was by necessity a hypothetical one; in practice a gambler intending to bet on a succession of races in a season, or a life-time, would meet a large variety of betting propositions. What the above tells us is that, a gambler must consider only races where there is at least one bet which is favourable, and should bet such that the probability of loss on a race is minimum, and moreover to satisfy this objective, even take fair and unfavourable bets.

2. PRELIMINARIES

Let X be the net gain made by the gambler on a game, be it with one bet or more than one bet and let $E(X) = \mu > 0$. Let us assume we have a sequence of games, for each of which the net gain has a distribution, the same as that of X. Then we have a sequence {X} of independent and identically distributed random variables, and $S_n = \sum_1^n X_1$ is the net gain after n games. If the gambler's capital is b > 0 {Sn}, (n = 0, 1, 2, 3, …) is a random walk process with an absorbing barrier at –b, the gambler's ruin corresponding to the absorption of the random walk at –b. A convenient tool used to derive the probability of absorption in random walks is Wald's Identity (Wald [6]) which is as follows:

Let N be the first time the random walk is absorbed, and let P(z) be the probability generating function (p.g.f.) of X. Then,

$$E\left[z^{S_{n'}} P(z)^{-N}\right] = 1 \tag{1}$$

We shall also need the following lemma: For a random variable X with E(X) > 0, there is a unique value z_1, $(0 < z_1 < 1)$ such that $P(z_1) = 1$.

Putting $z = z_1$ in (1) and ignoring the overshoot over –b, we have

$$P(R) \approx z_1{}^b \tag{2}$$

Let's now consider the scenario described in the introduction, where we have three horses A, B and C with odds 5/1, 3/1 and 3/1 and probabilities of win 0.2, 0.3 and 0.3 respectively. Suppose, we bet the amounts 2, 3 and 3 dollars on A, B and C. X now takes the values 4 and –8 with probabilities 0.8 and 0.2 respectively, so that we have $P(z) = 0.8z^4 + 0.2z^{-8}$. Now, the probability of ruin for this case with b = 80, is the same as when we have b = 20 and X takes the values 1 and –2 with probabilities 0.8 and 0.2. The resulting p.g.f. is $P(z) = 0.8z + 0.2z^{-2}$, and taking P(z) = 1, yields $z_1 = 0.6404$. The approximate value of P(R) obtained from (2) is 0.00013. Table 1 gives the approximate values of P(R) obtained from (2) with b = 80, for various possible combinations of bets for the game (all with the same expected gain on the game). We notice that P(R) decreases as q the probability of loss in the game decreases.

Table 1 P(R), the probability, of the gambler's ruin with a capital of 80 units

BettingCombination	8 units onA only	8 units onB (or C) only	3.2 units on A, 4.8 units on B (or C)	4 units each on B and C	2 units on A, 3 units each on B, C
Prob. of loss on a game, q	0.8	0.7	0.5	0.4	0.2
P(R)	0.4662	0.2769	0.0591	0.0173	0.00013

In the general theory given in the rest of the paper, we shall assume that we have either a unit bet on the game (Section 3), or else the primary bet is of unit amount (Sections 4 and 5). In this case, the net gain is unlikely to assume integer values; however, with b large and the loss per game restricted to under 2, the overshoot is small and may be neglected, and P(R) is approximately given by (2). The problem therefore reduces to comparing the values of z_1 obtained for various betting strategies. The z_1 is the unique positive solution ($0 < z_1 < 1$) of $P(z) = 1$, where the definition of p.g.f. is extended to include the case of non-integer valued random variables.

3. BETTING ON FAVOURABLE BETS

For betting only on favourable bets, we can bet amounts on each horse, so that the net gain is the same. Let p be the combined probability of a win and q(= 1 – p) the probability of a loss. We shall assume the total bet is one unit; thus c the overall odds we are getting is such that $cp - q = \mu > 0$. The net gain is the random variable X assuming values c and –1 with probabilities p and q respectively.

Here $$P(z) = pz^{(\mu+1)/p-1} + qz^{-1}. \tag{3}$$

It is fairly easy to show that (for constant μ) the root $z_1 (0 < z_1 < 1)$ of $P(z) = 1$ decreases as p increases from 0 to 1.

To see how z_1 depends on μ and *p*, we let $z_1 = 1 - \delta$, and use Taylor series expansion for the expression in (3) with z replaced by z_1 and solve for δ. We obtain

$$\delta = \frac{2\mu p}{(1+\mu)(q+\mu)}.$$

Thus, (for a constant value of μ) as p increases, δ increases, and since $P(R) = z_1^b$, P(R) decreases. Table 2 gives the variation of the values of z_1 obtained from (3) with respect to values of q for $\mu = 0.05, 0.1, 0.2$, and it can be seen that as q decreases, z_1 decreases. Indeed, since the ruin probability is an exponential function of z_1, the effect on the ruin probability is very much greater.

Table 2 z_1, values for values of q and μ for favourable games.

μ \ q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.05	0.5145	0.7163	0.8156	0.8744	0.9132	0.9407	0.9612	0.9770	0.9897
0.10	0.3458	0.5640	0.6966	0.7846	0.8469	0.8932	0.9291	0.9576	0.9808
0.20	0.2183	0.4101	0.5556	0.6667	0.7535	0.8229	0.8795	0.9265	0.9662

4. BETTING ON AN ADDITIONAL FAIR BET

Suppose now there are one or more favourable bets in a race with total probability of win p and expectation $\mu > 0$, so that the odds offered are given by $c = (1 + \mu)/P - 1$. Let us assume we have an additional fair bet in the race with probability of win p' i.e. the odds are q'/p', where $q' = 1 - p'$. Whatever amount x we have on the fair bet leaves our expectation of the net gain unchanged. Betting a unit amount on the favourable bet and an amount x units on the fair bet, the net gain X has the distribution

X	$c - x$	$xq'/p' - 1$	$-1 - x$
Pr	p	p'	$1 - p - p'$

To minimise the probability of ruin by betting on this game (and successive such games), we should take that value of x such that the root z_1 $(0 < z_1 < 1)$ of the resultant equation $P(z) = 1$, i.e.

$$pz^{c+1} + p'z^{x/p'} - z^{1+x} + 1 - p - p' = 0 \tag{4}$$

takes its minimal value.

Putting $z = 1 - \delta$ and using Taylor series expansion for the left hand side of (4), we obtain

$$\delta = \frac{2\mu}{c(c+1)p - 2x + x^2 q'/p'}$$

For given c and p, the maximum value of δ is obtained when $x = p'/q'$. Substituting the value of x in (4) we solve for z. Table 3 gives the z_1 values obtained for $\mu = 0.1$, 0.2, $p = 0.1, 0.2, 0.3, 0.4, 0.5$ and $p' = 0.1, 0.2, 0.3, 0.4$. The required z_1 values are those which correspond to the case $u > 0$. The corresponding z_1 values for the same p with only favourable bets is given in each case. From the table we note, for example, that for $\mu = 0.2$, the single favourable bet at $p = 0.2$ gives $z_1 = 0.9265$. However, with the additional fair bet at $p' = 0.5$, yields $z_1 = 0.9071$, so that with $b = 40$, P(R) decreases from 0.04723 to 0.02026.

We note that as p' increases the value of z_1 decreases. So if an additional fair bet is to be taken, the shorter the odds the better it is. Ideally, of course, we should take a fair bet at $p' = 1 - p$, but quite obviously, in practice this would be unattainable, as it would mean there is no loss, only a possibility of a gain.

5. Betting on an Additional Unfavourable Bet

Suppose now we have, as before, a favourable bet (or many favourable bets) of one unit with total probability of win p, and expectation $\mu > 0$. Let us assume we have in addition an unfavourable bet with probability of win p', and let us assume the odds offered are $q'/p' - u$ to 1, so that the unfavourability factor is $u > 0$. It is obvious that to maintain the expectation of the total transaction of the two bets to μ, we need to increase the amount on the favourable bet. Let $x > 0$ be the increase of the amount on the favourable bet and y the amount on the unfavourable bet. Since the loss on the unfavourable bet (due to its unfavourability) has to be compensated by the extra gain on the favourable bet, we have the relation $p'uy = \mu x$. The net gain X on the game has the probability distribution

X	$c(1+x)-y$	$(q'/p'-u)y-(1+x)$	$-1-x-y$
Pr.	p	p'	$1-p-p'$

so that the equation P(z) = 1 becomes

$$pz^{(1+x)(1+c)} + p'z^{(1/p'-u)y} - z^{1+x+y} + 1 - p - p' = 0 \tag{5}$$

As before, to find the values of x and y so that the root z_1 $(0 < z_1 < 1)$ of (5) takes its minimum value, we let $z = 1 - \delta$, expand the left hand side of (5) by a Taylor series expansion and solve for δ in terms of μ, p, p', c, x, y and u. The maximum value of δ is obtained when

$$y = \frac{-(c+1)(2c+1)p\,p'u/\mu + 2 + p'u(1-\mu)/\mu}{2\left[pp'^2u^2(1+c)^2/\mu^2 + (1-p'u)^2/p' - (\mu+p'u)^2/\mu^2\right]}$$

Using the value of y and x given by $x = p'uy/\mu$ in (5) we solve for z. The values of z_1 for $\mu = 0.1, 0.2$, $p = 0.1, 0.2, 0.3, 0.4$, $p' = 0.1, 0.2, 0.3, 0.4, 0.5$ and $u = .01, .05$ are given in Table 3. The asterisks * in the table correspond to the cases where the values of x and y are negative, and are therefore not admissible. The symbols # correspond to the cases where $p + p' = 1$; these cases are obviously not realizable in practice.

From Table 3 we note for example that for $\mu = 0.1$, the single favourable bet at $p = 0.4$ yields $z_1 = 0.8932$, and with an additional fair bet at $p' = 0.5$ (the case $u = 0$) yields $z_1 = 0.7510$. However, if the additional bet is an unfavourable bet with $p' = 0.5$ and $u = 0.05$, we have $z_1 = 0.8410$, so that with $b = 20$, the values of P(R) for the three cases above are 0.1045, 0.0033, and 0.0313 respectively.

Table 3 z_1 values for values of p, p′, μ, u for additional fair and unfavourable bets.

	p = 0.1 $\mu = 0.1$ $(z_1 = 0.9807)$	$\mu = 0.2$ $(z_1 = 0.9662)$
u p'	0.1 0.2 0.3 0.4 0.5	0.1 0.2 0.3 0.4 0.5
0	.9805 .9802 .9799 .9793 .9785	.9658 .9653 .9647 9638 .9626
0.01	.9806 .9804 .9803 .9802 .9802	.9658 .9655 .9651 .9647 .9643
0.05	.9807 .9808 * * *	.9660 .9660 * * *
	p = 0.2 $\mu = 0.1$ $(z_1 = 0.9575)$	$\mu = 0.2$ $(z_1 = 0.9265)$
u p'	.1 .2 .3 .4 .5	.1 .2 .3 0.4 0.5
0	.9565 .9550 .9530 .9500 .9451	.9248 .9225 .9193 .9146 .9071
0.01	.9565 .9554 .9541 .9524 .9500	.9249 .9229 .9203 .9168 .9115
0.05	.9569 .9568 .9571 * *	.9252 .9242 .9237 .9236 .9240
	p = 0.3 $\mu = 0.1$ $(z_1 = 0.9291)$	$\mu = 0.2$ $(z_1 = 0.8795)$

u \ p'	.1	.2	.3	.4	.5	.1	.2	.3	0.4	0.5
0	.9259	.9216	.9153	.9051	.8858	.8747	.8682	.8588	.8438	.8165
0.01	.9261	.9224	.9173	.9097	.8959	.8749	.8689	.8605	.8476	.8246
0.05	.9267	.9250	.9240	.9236	.9241	.8754	.8712	.8667	.8612	.8530
	p = 0.4									
	$\mu = 0.1$ $\left(z_1 = 0.8932\right)$					$\mu = 0.2$ $\left(z_1 = 0.8229\right)$				
u \ p'	.1	.2	.3	.4	.5	.1	.2	.3	0.4	0.5
0	.8860	.8755	.8587	.8277	.7510	.8124	.7974	.7741	.7330	.6404
0.01	.8863	.8766	.8620	.8357	.7709	.8126	.7983	.7766	.7389	.6537
0.05	.8872	.8809	.8738	.8636	.8410	.8133	.8018	.7863	.7615	.7057
	p = 0.5									
	$\mu = 0.1$ $\left(z_1 = 0.8469\right)$					$\mu = 0.2$ $\left(z_1 = 0.7535\right)$				
u \ p'	.1	.2	.3	.4	.5	.1	.2	.3	0.4	0.5
0	.8316	.8075	.7642	.6627	#	.7327	.7011	.6473	.5343	#
0.01	.8319	.8092	.7693	.6762	#	.7329	.7023	.6508	.5425	#
0.05	.8332	.8157	.7886	.7280	#	.7339	.7071	.6645	.5754	#

REFERENCES

[1] Feller W., An introduction to probability theory and its applications, Vol. 1., J. (Wiley, N.Y. 1967)

[2] Dubins L.E and Savage L.J., How to gamble if you must, (McGraw Hill, N.Y., 1965)

[3] Epstein R.A., The theory of gambling and statistical logic, (Academic Press, San Diego, 1977).

[4] Breiman L., Optimal gambling systems for favourable games. Fourth Berkely Symposium on Probability and Statistics, I., 65-78, (1961)

[5] Rotando L.M. and Thorp E.O. , The Kelly Criterion and the stock market, *Amer. Math. Monthly*, 99, 922-931, (1992)

[6] Wald A., *Sequential Analysis, J.* (Wiley, N.Y., 1947)

STATISTICAL ANALYSIS OF HORSE RACING DATA

Kuldeep Kumar[1], S. Ganesalingam[2] and Siva Ganesh[2]

Abstract

The principles of statistics can be applied to almost every sphere of modern life; sports are no exception. It is hypothesized in horse racing that the odds of winning the horse is reflected in the volume of bets placed up to the start of the race. We have tested this hypothesis in this paper. We also looked at the effect of various variables such as the weight carried by the horse, the time taken and the barrier effects on the finishing position of the horse. It appears that the betting remains almost constant for winning horses while the loosing horses exhibit an increasing trend.

1. INTRODUCTION

Horse racing is one of the oldest of all sports and is mentioned in the Olympic Games of Greece over the period 700 – 40 BC. Charles II (reigned 1660 – 85) became known as "the father of English turf" and inaugurated the King's Plates, races for which prizes were awarded. His articles for these races were the earliest national rules. The horses raced were six year old and carried 168 pounds, and the winner was the first to win two 4 - mile heats. Similar references are found in North America and France around the 16th and 17th centuries. The beginning of the modern era of racing is generally considered to have been the inauguration of the English Classical races; the St. Leger in 1776, the Oaks in 1779, and the Derby in 1780. (The other historical details about horse racing can be found in Encyclopedia Britannica Vol. 6 *Micropaedia.*)

The winning odds of a horse are offered by bookmakers. In horse racing the odds reflect the volume of bets placed on each horse. These vary with time, as bets are placed, up to the start of the race. According to Dixon [3] betting markets exist for many sports and give a novel motivation for development of statistical models for sporting events. The use of scientific improvement of the model, with the purely incidental by-product of winning money from the bookmakers, should encourage statisticians to consider more complex models for sporting events. Predicting the winning horse has always been a million dollar question not only for the gambler or common man but also for the statistician. Yitsak and McCartney [7] developed the prediction model in horse racing and White and Dattero [6] used vector forecasts to predict horse racing outcomes.

---

[1]School of Information Techology, Bond University, Gold Coast Qld 4229

[2]Department of Statistics, Massey University, Palmerston North, New Zealand

In this paper we have used the data collected from horse races on the Gold Coast, Australia. The data consists of amount of mony bet on each horse at 12 occasions up to the start of the race and the placing of the horse in the race. The hypothesis we wish to test here is whether the placing of the horse is dependent on the betting money placed on that horse. In other words if the money bet on the horse (over the time slice from 1 to 12), increases for each time interval, the chance of the horse winning the race will be higher than if the money bet remains constant for each time interval.

In Section 3 we have described the methodology and in Section 4 we have analysed the data. Results and conclusions are also give in Section 4.

2. DESCRIPTION OF THE DATA

Data has been collected from various races at the Gold Coast, Australia in 1995. The data consists of race numbers, placing of the horse, the original money in $ bet on each horse along with other variables observed on each horse such as; finishing position, time taken to complete the race, weight, age, amount of money won, barrier position and starting price.

3. METHODOLOGY: REGRESSION AND REPEATED MEASURES ANALYSIS

Let us introduce briefly the statistical techniques we employed in this study. The logistic procedure fits linear logistic regression models for binary or ordinary response data by the method of maximum likelihood.

Binary response variables, for example, success, failure, as well as ordinal response variables for example in a horse race "first place, second place and third place", or in a medical context "none, mild, severe" arise in many fields of study. Logistic regression analysis is often used to investigate the relationship between the response probability and explanatory variables. A thorough discussion of binary response model methodology is given in Cox and Snell [2]. Several texts that discuss logistic regression are Agresti [1], Freeman [4] and Hosmer and Lemeshaw [5].

For the Binary response type model we let the response y of an experimental unit or an individual take on one of the two possible values, denoted for convenience by 1 and 0, for example y = 1 if a horse wins otherwise y = 0. Suppose X is a vector of explanatory variables and p = Pr[y = 1 given X] is the response probability to be modelled. The linear logistic model has the form:

$$\log\mathrm{it}(p) = \log\left(p/(1-p)\right) = \alpha + \beta' X$$

where α is the intercept parameter and β is the vector of slope parameters and β' is tis transpose.

The logistic model shares a common feature with a more general class of linear models where a function $g = g(\mu)$ of the mean response variable is assumed to be linearly related to the explanatory variables. Since the mean μ implicitly depends on the stochastic behaviour of the response, and the explanatory variables are assumed fixed, the function g provides the link between the random component and the deterministic component of the response variable. For this reason, Nelder and Wedderburn (1972) refer to $g(\mu)$ as a link function.

In our study, this logistic regression approach was first considered in the context of the horse race data and the findings are not encouraging but these are highlighted in Section 4.

Further in order to see how the pattern of betting has changed from the original position to slice 12, we have fitted the time series regression model:

$$y_t = \beta_0 + \beta_1 t + \varepsilon$$

where, $t = 0$ (origin), 1,2,......,12
y_t = betting money at time t
ε denotes the random error.

If β_1 is positive it will imply that over the period of time the betting money has gone up and it will also tell the amount. If betting is constant β_1 will be zero. and if betting on the particular horse has gone down the value of β_1 will be negative. Hence β_1 will be a good indicator of the change in betting pattern. If the hypothesis that the batting for the winning horse goes up over the period of time, then β_1 should be correlated with the placing of the horse.

Repeated measures analysis on the other hand deals with analysing data consisting of a number of relatively short non-stationary time series in which the trends, $\mu(t)$, say, are of direct interest. This situation arises in experiments which involve comparisons amongst trends associated with different treatments. It also applies to growth studies. We refer to data of this kind as repeated measurement data. The horse race data where the bets are placed on a horse at different points of time, is a typical example of a repeated measurement data.

We shall make the pragmatic assumption that the nature of the random variation is the same in all the individual series. Formally, we assume that the i^{th} series is generated by a random process $\{y_i(t)\}$ such that:

$$y_i(t) = \mu_i(t) + z_i(t)$$

where $\{z_i(t)\}$ is a stationary random process, with the same structure for all series but realised independently for each.

4. Results and Conclusions

Repeated measures analysis of variance results are given in Table 1 and the corresponding plots of interaction means are given in Fig 1.1 and 1.2. It can be seen from Table 1 that overall there are no noticeable differences between the "races" and between the "winners" and the "losers". However, the evidence for the latter conclusion is marginal (*p-value* = 0.1071). When we look at Table 1 again, the univariate tests of hypothesis for within subjects effects indicate that there are noticeable differences between "bets" at different times before the race (*p-value* =0.0001). Further the interaction between the "bets" and "races" is significant (*p-value* = 0.0272).

Table 1: Repeated Measures Analysis of Variance

Tests of Hypotheses for Between Subjects Effects

Source	DF	SS	MS	F	p-value
RACE_NO	2	17249.18	8624.59	0.77	0.4705
WIN	1	30546.96	30546.96	2.72	0.1071
Error	38	426121.03	11213.71		

Tests of Hypotheses for Within Subject Effects

Source	DF	SS	MS	F	p-value
BETS	12	495.39	41.28	3.95	0.0001
BETS*RACE_NO	24	415.21	17.30	1.66	0.0272
BETS*WIN	12	326.64	27.22	2.61	0.0023
Error(BETS)	456	4763.11	10.44		

It is interesting further to see in Table 1, that the interaction between "winning" and "betting" is highly significant (p = 0.0023). This is evident in Fig 1.1 and Fig 1.2 in that the mean betting is almost constant for the winning horses while the loosing ones exhibit an increasing trend of betting.

Further analysis carried out with the data corresponding to race 1, shows that our expectation is not clearly supported. No definite trend in betting was observed over the 13 time periods starting from the origin to slice 12, the last betting time just before the race. However a constant nature of betting is noted with the winning horses.

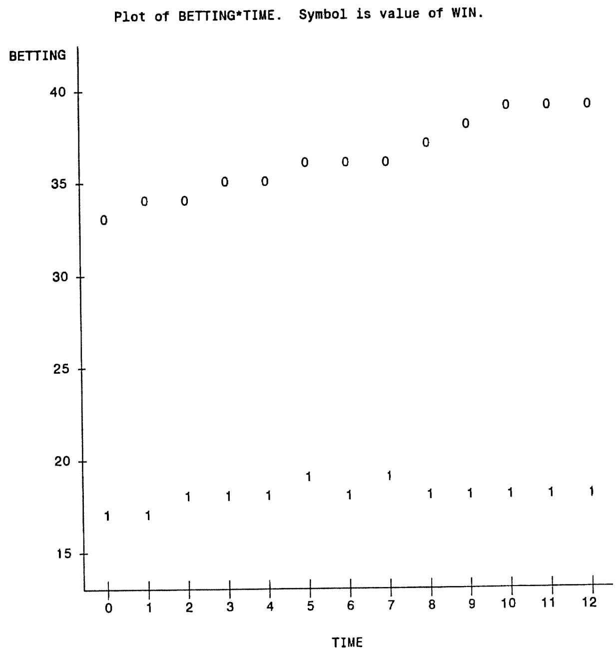


Figure 1.1: *Means of "betting" (1-winners, 0-losers)*

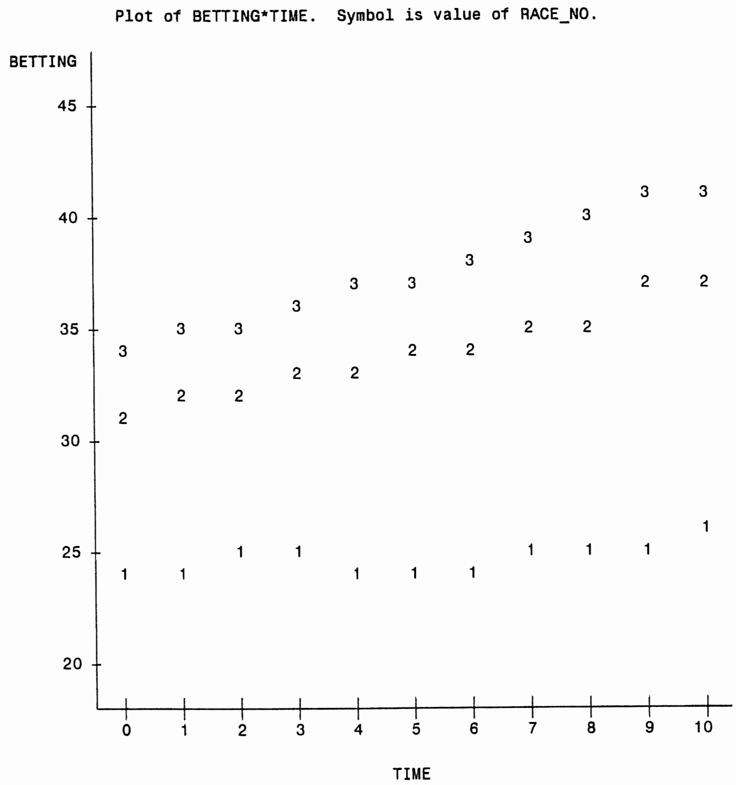


Figure 1.2: *(1- horse finishes 1st; 2- horse finishes 2nd; 3- horse finishes 3rd)*

A multiple regression analysis was also carried out on the data. It was noted that the 'finish position' of a horse is predictable. The corresponding regression equation is given as:

$$\text{Finish Position} = 5.31 - 0.0056*\text{weight} + 0.722*\text{margin} - 0.126*\text{horse age} + 0.0562*\text{barrier} + 0.0165*\text{start price}.$$

It appears that most of these coefficients are significantly different from zero at the 10% level. However the variable "margin" appears to be the best predictor of the finish position. To be more precise we need more data on each horse in the past races so that the current performance of each horse can be correctly predicted. The best next predictor is the starting price. The above prediction equation explains almost 70% of the total variation and hence can be used for future prediction of the finish position of a horse taking part in race. The logistic approach proves to be inappropriate as the data set 1 is highly correlated and prompted us with the repeated measure approach.

However, these results may not be valid for other races. It will be interesting to analyse many more races before any definite conclusions can be drawn.

ACKNOWLEDGMENT

The authors are very grateful to Mr. John Rose of ITS, Bond University for providing the data.

REFERENCES

[1] Agnesti A., Analysis of ordinal categorical data, New York, (John Wiley, 1984)

[2] Cox Dr. and Snell E.J., The analysis of Binary Data, 2 edition, London, (Chapman & Hall, 1989)

[3] Dixon M., Beating the bookmakers, RSS News, Vol.23, No.9, pp1-2, (1996)

[4] Freeman D.H. Jr., Applied Categorical Data Analysis (New York, Marcel Decker, 1987)

[5] Hosmer D.W. Jr. and Lemeshaw, S., Applied Logistic Regression, (New York John Wiley, 1989)

[6 White E.M., and Dattero, R., Combining Vector forecasts to predict thoroughbred horse racing outcomes, *International Journal of Forecasting*, Vol 8 p595 - 611 (1992).

[7] Yitshak D. and McCartney D.L., Prediction models in Horse racing, ASA proceedings of business and Economic Statistics Section, p508-512, (1981)

INFORMATION-THEORETIC FOOTBALL TIPPING

David L. Dowe[1], Graham E. Farr[1], A. John Hurst[1] and Kevin L. Lentin[1]

Abstract

This paper describes a football tipping competition based on the estimation of probabilities of victory (rather than simply an estimation of which team will win), and its connection with information theory and gambling. It also describes a football tipping competition for which entrants tip the mean and standard deviation they perceive on the margin of the game. The optimal long-term strategy in both competitions is the minimisation of the expected Kullback-Leibler distance from the "true" probability (distribution) to that tipped by the entrant.

1. A PROBABILISTIC TIPPING COMPETITION

> "The Master said, Yu, shall I teach you what knowledge is? When you know a thing, to recognise that you know it, and when you do not know a thing, to recognise that you do not know it. That is knowledge."
>
> — *Analects of Confucius* (transl. by Arthur Waley), Book II, No. 17.

It is a shortcoming of many statistical, "machine learning" and human predictive methods that, while they might offer predictions, they are not always willing to associate a probability or degree of certainty with the prediction.

This paper reports on an attempt to improve this situation in a familiar predictive activity: *football tipping* . This domain provides an excellent setting in which to test and illustrate ideas from statistics, inductive inference and information theory, while at the same time being well known to many people outside these fields.

Football tipping competitions, in which participants try to predict, week by week, which team will win each game of football played in some league, have been popular in Australian workplaces for a long time. Some such competitions require that participants predict further information such as winning margins, and have various ad hoc rules for rewarding accuracy of predictions. Our own Department's tipping competition has traditionally been of this type. Certainly predicting a team's winning margin does say something about how sure you are that the team will win, but it would be good to have your degree of certainty described more directly, and to have the accuracy of your prediction rewarded in a mathematically sound way.

---

[1]Department of Computer Science, Monash University, Clayton Victoria 3168

With these remarks in mind and in an attempt to add interest to our departmental tipping competition, a new competition was started (as noted in earlier work [4, 6, 7], this competition began with the Round 3 matches of the 1995 Australian Football League (AFL) season, which began with the match on Saturday 15 April 1995) in our Department in which participants must estimate, for each game, not just the winning team, but its probability of winning (Determining probabilties is also of interest to both the gambler and the bookmaker. See also §5 where we discuss combining tips for a "weighted" prediction.). The first suggestion to have a probabilistic competition came from Jon Oliver (as pointed out in [6]). Below we describe how a probabilistic tipping competition works. As we shall see, it is closely related to the application of information theory to certain gambling situations, and finds parallels in other fields where quantifying the information cost of a prediction is important.

For such a competition to be meaningful, predictions must be properly rewarded. The rewards must be such as to encourage people to nominate their actual estimates of the winning probabilities. The traditional scoring method of one point for a correct prediction, zero otherwise, will not work.

The reward function we use is as follows. If a tipster assigns probability p to a win by team A, then the score for the tipster on that game is

$$\begin{aligned} &1+\log_2 p, && \text{if } A \text{ wins;} \\ &1+\log_2(1-p), && \text{if } A \text{ loses.} \end{aligned} \tag{1}$$

This score is measured in bits. If the true probability $\Pr(A \text{ wins}) = \pi$, then a tipster's expected score is $1+\pi\log_2 p+(1-\pi)\log_2(1-p)$, which is maximised by the estimate $p=\pi$ (elementary calculus, or appeal to the Information Inequality [1, Theorem 2.6.3]).

The use of such logarithmic functions to reward predictions goes back at least to J. Kelly [10] and (we believe) to I. Good and P. Dawid. Recent applications which might be of interest include [15, p. 20] and [5, p. 3].

The maximum expected score which a tipster can achieve is therefore

$$1+H(\pi,1-\pi), \tag{2}$$

where H is entropy. The tipster of course does not know this as he does not know π. The tipster *believes* that his expected score is

$$1+H(p,1-p)\ . \tag{3}$$

The score obtained from (1), for a single game, is accumulated over all the games in a season to give a player's total score.

Some elementary observations may be helpful at this point. Consider a single game, and let X be the team that actually wins the game. The score given by (1) is positive if the tipster predicted that $\Pr(X \text{ wins}) > 1/2$ (i.e., if the tipster predicted the winning team correctly), negative if the tipster predicted $\Pr(X \text{ wins}) < 1/2$, and 0 if the tipster thought each team had an equal chance (or was simply expressing complete prior ignorance by nominating probabilities of $1/2$ each). The maximum possible score from a game is 1 bit, when the tipster predicts a winner with certainty and is correct. However being too certain can be dangerous. A tipster who predicts a winning probability of 1, and is wrong, gets a score of $-\infty$ and essentially becomes incapable of ever winning the competition.

Note that the constant 1 in (1) is not necessary for comparing tipsters, but is included for a few reasons. Firstly, the sign of the expression then indicates whether the tipster got the winning team right or not, and 0 points corresponds to complete ignorance (or maximum uncertainty) as described above. Secondly, having no constant at all would keep everyone's score negative throughout the competition, which seemed likely to put people off.

In most football codes, a *draw* is also a possible outcome. In Australian Rules football, a game is a **draw** if the two teams finish with identical scores. (There is no possibility of draws due to bad weather, as in cricket.) About 1.1% of Australian Rules football games end in draws (which makes them much rarer than in other codes), and there is little interest for tipsters in predicting draw probabilities separately for each game. It is therefore desirable to stick to the usual two outcomes (win/loss), and augment the reward function to take account of draws. If for example Alice tips $\Pr(X \text{ wins}) = 0.9$ and Bob tips $\Pr(X \text{ wins}) = 0.6$, and the outcome is a draw, then clearly Bob was nearer the mark and, we argue, should be rewarded accordingly (as opposed, for example, to just ignoring draws and giving everyone the same score for them). We suggest that there is enough randomness in football to regard a draw as a game which could have gone either way, equiprobably. This suggests that a tipster who nominates $\Pr(X \text{ wins}) = p$ should, in the event of a draw, be given

$$1 + (1/2)\log_2 p + (1/2)\log_2(1-p) \quad \text{bits.} \tag{4}$$

The basic ideas of probabilistic tipping are easily extended to cope with more than two outcomes of interest. Tipsters are required to nominate probabilities for each outcome, and their reward is the $\log_2$ of the probability which they assigned to the actual outcome plus $\log_2$ of the number of possible outcomes (assuming that prior ignorance is best modelled by the assignment of equal probabilities to all outcomes).

We remarked earlier that a tipster *expects* to gain (sweeping under the carpet and disregarding the possibility of draws)

$$1 + p_i \log_2 p_i + (1-p_i)\log_2(1-p_i) \tag{5}$$

bits from a game i in which he nominated a probability of p_i for one of the teams (say, the home team). A tipster may do worse than he expects, in which case we might say that his predictions were bolder, or more *positive* , than they should have been. On the other hand, he may do better than he expects, in which case his predictions were more cautious, or *negative* . Either way, he will on average fall short of the maximum expected score. The tipster's score, alone, will not reveal whether this shortfall is due to his being consistently too positive, or too negative, or neither. (If neither, then his positive errors cancel out his negative errors, on average, and he does about as well as he expected, although still less than the maximum expected score. The only way to attain the maximum expected score in the long run is to always predict the probabilities correctly, i.e. to never make a positive or negative error.)

It is easy, however, to work out whether a tipster is being consistently positive or negative in their tips. Let the actual result of game i, for the team for which the tipped probability was p_i, be r_i, where r_i = 1, 0, 1/2 for a win, loss, draw respectively. Then the actual score gained from game i is

$$1 + r_i \log_2 p_i + (1 - r_i)\log_2(1 - p_i)$$

bits. In order to measure the *positivity* of the tipster, we take the difference between the score they expected and the score they obtained, for each game, and sum over all the games (denoting the result by $\mathrm{pos}(\mathbf{p})$ to indicate dependence on the probability estimates p_i):

$$\mathrm{pos}(\mathbf{p}) = \sum_i (p_i - r_i)\log_2(p_i / (1 - p_i)) . \tag{6}$$

Positive indicates too bold, negative indicates too cautious, and zero indicates that neither positive nor negative tendencies predominate, and that the tipster may be regarded as a good judge of their own predictive ability. For example, a tipster who was completely ignorant of football and aware of that fact would assign probabilities of 1/2 to all teams, would score poorly (in fact, zero), but would be given a positivity measure of zero which would indicate the fact that they did precisely as well as they expected.

It is hoped to implement this positivity measure in future competitions, to give participants some automatic feedback on their own judgment (related to this, Kevin Korb has suggested the possibility of *calibrating* the tipsters.) Certainly it lends itself to easy calculation on a week-by-week basis.

2. GAUSSIAN TIPPING

The probabilistic tipping competition described above can be thought of as a minimum expected Kullback-Leibler distance competition. If the true probability of (e.g.) Geelong defeating St. Kilda is π, then the Kullback-Leibler distance (see, e.g., [1, p. 18]) from the true distribution $(\pi, 1-\pi)$ to the estimated distribution $(p, 1-p)$ is

$$(\pi \log_2 \pi + (1-\pi)\log_2(1-\pi)) - (\pi \log_2 p + (1-\pi)\log_2(1-p)) . \tag{7}$$

In effect, we wish to put the appropriate probability on both outcomes. (See §4, on fully-invested gambling.)

For the Gaussian competition (which began [7, 4] with the Round 1 games of 1996, commencing on Friday 29 March 1996) rather than reward the tipster by a constant plus the logarithm of the probability they assigned to the correct outcome, we now reward the tipster with a constant plus the logarithm of the probability they assigned to the *margin* of the game.

The tipster predicts a Gaussian (or Normal) distribution for the margin of the game, specifiying this with a predicted mean, $\hat{\mu}$ and a predicted standard deviation, $\hat{\sigma}$. The tipster's score for such a prediction is

$$10+\frac{1}{\hat{\sigma}\sqrt{2\pi}}\int_{x-1/2}^{x+1/2} e^{-(1/2)((x-\hat{\mu})/\hat{\sigma})^2}\,dx \tag{8}$$

Once again, we have a constant, the motivation for which is that a tipster expressing prior ignorance, by nominating a distribution of maximum entropy, should get a score of 0. While it is possible in principle for a tipster to nominate arbitrarily large $\hat{\sigma}$, we take the view that anyone who is in this competition at all brings with them (or will soon acquire) at least some vague idea of the approximate sizes of margins that occur in practice. (The greatest winning margin in a typical year, consisting of about 180 games, seems to be around 130 to 140 points, and a margin of over 200 points is (we believe) unknown in the 100-year history of the competition.) We do not attempt to quantify this level of ignorance exactly, and the exact value of the constant is not critical. It seems to us, from experience, that a value of 9 is about right, while 10 is a little generous (but is readily accepted as it makes scores larger). Someone who does not submit tips for a game is still given 0 bits for it.

This competition, like the simpler one described in §1, is about minimising a Kullback-Leibler distance, and both are relevant to problems of scientific prediction, gambling and combining opinions of different "experts" (of which more below, in §§4,5).

3. EXPERIENCE WITH THE COMPETITION

The probabilistic tipping competition was run for the first time [6] in 1995, and is being run again with many more participants this year.

The introduction of the tipping competition had a number of interesting effects on participants. Some found that it actually revived an interest in football that had begun to wane. In the early days of the competition, a number of participants tended to be too bold in their tips, with probabilities too close to 1, so that a couple of "upset" results led to major setbacks. Perhaps football followers generally are too sure of their predicitions, and do not realise the full extent of the uncertainty associated with results of games. Another amusing observation in the first few weeks was the poor performance of computer scientists specialising in information theory, compared with some of the general staff with no knowledge of that field [7]. As the season progressed, tipsters began to be a little more restrained in their estimates. However, very near the end, some participants, needing large scores to catch up, began to get reckless and nominate probabilities very close to, and in a few cases equal to, 1. This strategy resulted in some spectacular (and sometimes infinite) slides.

Some participants tried to improve their performance by seeking help outside their own intuition. Considerable secrecy usually attended these efforts. Kevin Korb (and, later, at least one other) experimented with variants of the Elo rating scheme used by chess players (see, e.g., [8; 9, pp.183–205; 11, §9.1]). Others looked up the odds for each game given in the newspaper every Friday, and used corresponding probabilities. Experience so far suggests that an Elo-type scheme performs well (indeed better than most tipsters), but that using published odds does much better. In fact the participants who used the odds have consistently led our competitions by significant margins.

To be fair to the Elo-type schemes, however, they used very little information: just outcomes of games this season, some rough estimate of the initial standings of the teams at the start of the season (which is not a critical factor), and a simple means of taking some account of home ground advantage. No information about players, injuries, weather and so on was used. The success of such methods, when pitted against knowledgeable football supporters, is thought-provoking. Presumably it would be possible to do better with more sophisticated rating schemes. We suggest that such methods might also applied in other sports, such as cricket.

The Gaussian competition was, not surprisingly, more difficult for people to grasp. By Round 4 (1996), the designer (Dowe), the programmer (Hurst) and most of the participants (or those that hadn't dropped out) almost seemed to have the hang of it. While predicting the margins is natural and interesting enough, giving an estimate $\hat{\sigma}$ of the standard deviation seemed to many of us to be just a matter of finding out, through experience, what was the right sort of figure to use (very roughly, about 40), and then applying that figure to each game, with only minor variations.

4. CONNECTION WITH GAMBLING THEORY

Suppose a gambler places bets on football games with a bookmaker. The gambler starts with an amount of money S_0, and we assume that she is **fully invested**, i.e., bets all her money on each game. The gambler must simply decide how to apportion her funds between the two possible outcomes of each game. Supose that team A is to play team B, and that the true probabilities of winning are $\pi_A = \Pr(A \text{ wins})$ and $\pi_B = 1 - \pi_A = \Pr(B \text{ wins})$. The bookmaker offers odds of z_A-to-1 against A winning and z_B-to-1 against B winning. Let p_A and $p_B = 1 - p_A$ be the proportions of her wealth which the gambler bets on A and B respectively. We regard the game as a random experiment.

Consider what happens in the long run. Suppose the game A versus B is repeated n times, and let $X_j \in \{A,B\}$ denote the winner of the j-th game (so that $X_1, X_2, \ldots, X_n$ are independent identically distributed random variables). Let r_A and r_B be the proportions of these games won by each of the two teams.

Suppose the gambler's wealth prior to the j-th game is S. She loses the bet placed on the loser, but the amount $p_{X_j} \cdot S$ placed on the winner earns a payment of $(z_{X_j} + 1)$ times that amount from the bookmaker. The gambler's wealth after the game is $(z_{X_j} + 1)p_{X_j} \cdot S$, and has grown by a factor of $(z_{X_j} + 1)p_{X_j}$.

Over the whole sequence of games, the gambler's wealth increases by a factor of

$$\prod_{j=1}^{n} \left(z_{X_j} + 1\right) p_{X_j}$$

which is easily shown to be equal to

$$2^{n(r_A \log_2(p_A(z_A+1))+r_B \log_2(p_B(z_B+1)))}.$$

As observed in [1, Theorem 6.1.1], this converges in probability as $n \rightarrow \infty$ to

$$2^{n(\pi_A \log_2(p_A(z_A+1))+\pi_B \log_2(p_B(z_B+1)))}.$$

It is straightforward to prove, once again using the Information Inequality, that this is maximum when $p_A = \pi_A$ and $p_B = \pi_B$, that is, when the proportions which the gambler bets on each team equal the true probabilities of winning [1, Theorem 6.1.2]. The gambler's task is then to come up with the best possible estimates of the true probabilies. It is interesting to note that this aim takes no account of the odds offered by the bookmaker.

The fully invested football gambler is thus in exactly the same position as the probabilistic football tipster. Proportions of wealth gambled correspond to winning probabilities tipped. An increase in the logarithm (base 2) of the gambler's wealth, following bets on some game, corresponds to the tipster's score for that game. (They are not exactly equal, but differ only by a constant which depends only on the odds. This constant is thus the same for all gamblers/tipsters, and thus has no bearing on comparing their performance.) In each case, the best strategy is to estimate the true probability as closely as possible. (In gambling, the strategy of apportioning one's wealth into bets so that the proportions are estimates of the true probabilities is known as **proportional gambling**.)

Our information-theoretic football tipping competition is thus a simulation (since real money was not used) of fully invested gambling.

The information-theoretic approach to this and other gambling situations was pioneered by Kelly [10] and is discussed in [1, Chapter 6].

5. COMBINING TIPSTERS TO MAKE A BETTER TIPSTER

Minimum Message Length (MML) and related inductive inference work indicate that predictions are best made by combining theories, roughly in proportion to their posterior probability [12–14]. Results of Solomonoff [12] indicate that whether one wishes to maximise expected predictive accuracy or minimise expected Kullback-Leibler distance, one can do no better than to weight and combine all relevant theories. We note in passing the MML theory comes relatively close to minimising the expected Kullback-Leibler distance [3].

This result of weighting and combining theories applies not only to general scientific prediction problems in inductive and statistical inference. We also observed it to be the case in the 1995 football season [6], where we endeavoured to combine the opinions of different "experts".

Two of many possible ways of combining our "expert" tips are (i) to average them and (ii) to somehow weight them, where the better tippers are given larger weights than the less successful tippers. The "Average" tipper and a "Weighted" tipper were both entered in the 1995 information-theoretic competition. The weightings used in the "Weighted" tipper were based on our interpretation of Solomonoff's early ideas [12], where we gave each tipper a relative weighting of 2 raised to the power of their current bit score.

The "Average" tipper performed very well in 1995, leading 11 human tippers and trailing only 3 human tippers after Round 21. The "Weighted" tipper was a clear winner in 1995, but it is not yet clear whether or not this was because it was inadvertently given a slight but nonetheless unfair advantage. It is desired that we have the "Average" and (correct) "Weighted" tippers retrospectively for the 1996 information-theoretic competition, or at least for the 1997 season. The mathematics has also been done to put "Average" and "Weighted" tippers in the Gaussian competition in the following way: we would weight the various tippers as described above, and then fit them with the Gaussian distribution of minimum Kullback-Leibler distance. More complicated methods of combining predictors are not discussed here.

6. CONCLUSIONS AND FUTURE WORK

Information-theoretic football tipping, such as discussed in this paper, is interesting in several respects. Although based on technical ideas, it is readily understood and enjoyed by non-technical people. We are not aware of another football tipping competition like it, although it is very closely related to fully invested gambling (and other predictive activities). The success of tipping strategies based on bookmakers' odds is another illustration of how hard it is for the gambler to make a buck. The use of strategies based on rating systems merits further investigation. In teaching, information-theoretic football tipping can be used to demonstrate principles from probability and information theory.

The "old", 1995, WWW site was

`http://www.cs.monash.edu.au/~kevinl/footy.html`

and the new (1996) WWW site is at

`http://www.cs.monash.edu.au/~ajh/footy/results/`.

ACKNOWLEDGEMENTS

We thank the many participants in our Department's information-theoretic football tipping competitions for their interest and comments.

REFERENCES

[1] Cover T.M. and Thomas J.A., *Elements of Information Theory*, (Wiley, New York, 1991).

[2] Cover T. M. and Ordentlich E., Universal portfolios with side information, *IEEE Trans. on Information Theory*, **42** (2) 348–363, (March 1996).

[3] Dowe D.L., Baxter R.A., Oliver J.J. and Wallace C.S., Point estimation using the Kullback-Leibler loss function and MML, *Proceedings Kullback Memorial Research Conference*, Washington D.C., U.S.A., May 1996 (to appear).

[4] Dowe D.L., Hurst A.J., Lentin K.L., Farr G., Oliver J.J., Probabilistic and Gaussian football prediction competitions — Monash, to appear (in both), *Artificial Intelligence in Australia (AIIA) Research Report* , (1996), and also in *Computer Science Association Newsletter* (Australia), (1996).

[5] Dowe D.L. and Krusel N., A decision tree model of bushfire activity, *Technical Report 93/190* , Department of Computer Science, Monash University, (1993).

[6] Dowe D.L. and Lentin K.L., Information-theoretic footy-tipping competition — Monash, *Computer Science Association Newsletter* (Australia) 55–57 (1995).

[7] Dowe D.L., Lentin K.L., Oliver J.J. and Hurst A.J., An information-theoretic and a Gaussian footy-tipping competition, *FCIT Faculty Newsletter*, (Faculty of Computing and Information Technology, Monash University) **1** 2–6 (June 1996).

[8] Elo A.E., *The USCF Rating System* , U.S. Chess Federation Publication, (1966).

[9] Harkness K., Official Chess Handbook , New York: David McKay, (1967).

[10] Kelly J., A new interpretation of the information rate, *Bell Sys. Tech. Journal* **35** 917–926 (1956).

[11] Sadovskii L.E. and Sadovskii A.L., *Mathematics and Sports* , American Mathematical Society, (1993).

[12] Solomonoff R.J., A formal theory of inductive inference I, II, *Information and Control* **7** 1–22 and 224–254 (1964).

[13] Wallace C.S. and Boulton D.M., An information measure for classification, *Computer Journal* **11** (2) 185–194 (1968).

[14] Wallace C.S. and Freeman P.R., Estimation and inference by compact coding, *J. Royal Statist. Soc. Ser. B* **49** (3) 240–265 (1987).

[15] Wallace C.S. and Patrick J.D., Coding decision trees, *Machine Learning* **11** 7–22 (1993).

THE CAR JUMP

Neville de Mestre[1] and Maurice N Brearley[2]

Abstract

This paper looks at the dynamical feasibility of performing a car long jump greater than 100 metres using a straight flat approach of 400 metres and a suitably inclined ramp for the take-off. Two possibilities for the impact zone are considered.

1. INTRODUCTION

An investigation will be made of the feasibility of performing a long jump of about 100 metres by car using a straight flat 400 metre approach strip, an inclined ramp for take-off and a bed of sand for landing on (see Figure 1).

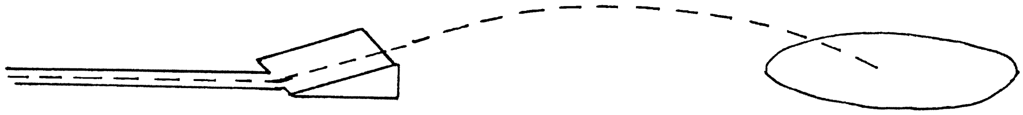


Figure 1

The motivation for the investigation arose through an approach by Stuart "Fireball" Campbell to one of us (Neville de Mestre) to make calculations for the design of a suitable ramp to enable him to break the world long jump record for a car; at present 96.4 metres. The event was to be staged at Cable Ski World on the Gold Coast during Easter 1995 with the added stimulus of being performed through a "hoop of fire" at the top of the ramp. The rights to televise the attempt were to be marketed.

Earlier long jumps with the proposed car had been performed in Sydney using a dirt ramp and on the Gold Coast using a ramp manufactured at Movieworld. Both jumps covered distances much shorter than 100 metres and were used as preliminary trials for the main attempt. Campbell was adamant that there would be no trials with the new ramp; it was to be a one-off attempt.

[1]School of Information Technology, Bond University, Gold Coast Qld 4229

[2]85 Dandarriga Drive, Clifton Springs Vic 3222

Various reasonable assumptions were made. It will be shown that air resistance can be neglected in the main; its effect on the length of the jump will be shown in Appendix A to be small. In the absence of any details of the mass distribution of the car, assumptions will be made to enable its dynamical characteristics to be calculated.

2. THE CAR

The vehicle to be used was a 1972 Ford Falcon (Model XY) with all the seats, except the driver's removed. The driver's seat was centre mounted in a safety cage to maintain symmetrical left and right weight distribution, and was placed halfway towards the back of the cabin to reduce the likelihood of injury from dashboard collapse. Only two gallons of petrol were to be in the car at the beginning of the approach run to improve safety considerations. The estimated maximum speed of the car on the flat approach is 176 kilometres per hour (110 miles per hour).

With all its modifications the total mass (M) of the car and driver was 1842 kilograms.

It is necessary for dynamical calculations to know the position of the centre of mass of the car, and also its moment of inertia about the rear axle.

The distance (d) from the rear to the front wheels is 2.8 metres, also known as the wheelbase.

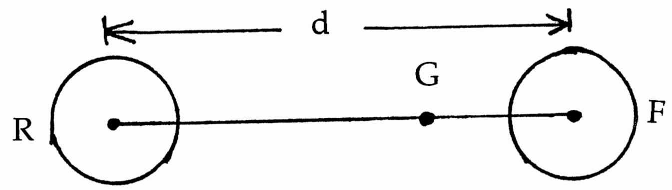


Figure 2: *Rear wheels (R) and front wheels (F) of car*

No details are available on how the mass of the car is distributed but the presence of the engine and gear box ensures that it will be heavier towards the front. The mass will be assumed to increase linearly along the line RF, and the centre of mass G will be taken to lie on RF. Suppose λ denotes the mass per unit length along RF, then

$$\lambda = Kx,$$

where x is the distance from R along RF and K is a constant. Then

$M = \int_0^d Kx\,dx = \frac{1}{2}Kd^2$ and so $K = 2M/d^2$. By taking moments of mass about R,

$$M(RG) = \int_0^d \lambda x\,dx = \int_0^d Kx^2\,dx = Kd^3/3$$

which yields $RG = 2d/3$ as the position of the centre of mass of the car.

The moment of inertia I_R of the car about its rear axle is given by

$$I_R = \int_o^d \lambda x^2 \, dx = \int_o^d Kx^3 \, dx = Kd^4/4 = \frac{1}{2}Md^2 \approx 7220\left(\text{kg m}^2\right)$$

3. THE TAKE-OFF RAMP

The Movieworld Ramp is 7.2 metres long with a wooden surface and its height at the take-off point is 1.8 metres. To give the car springs time to settle down after the change of slope encountered on entering the ramp it would be helpful to have a longer ramp. This has two disadvantages: car speed would be lost during progress up a long ramp, and the height of the take-off point would be greater, adding to the height attained by the car during its flight and thus increasing the hazard of the landing.

It is therefore recommended that the present ramp design be used, but with the central supporting columns lengthened slightly to eliminate the change of slope that occurs in the middle of the ramp. If this is done, the angle of inclination of the ramp will be (see Figure 3)

$$\alpha = \sin^{-1}(H_1 - H_2)/L$$

where H_1 = height of take-off point in metres = 1.8

H_2 = height of start of ramp in metres = 0.3

L = length of ramp in metres = 7.2

Then $\alpha = 12°$.

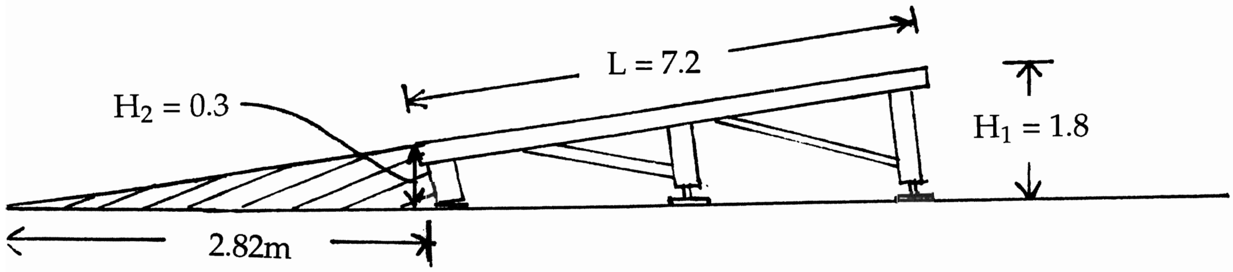


Figure 3: *Side elevation of take-off ramp*

To permit the springs of the car to adjust gently to the change in slope from 0° to 12°, a region of packed earth will be constructed at the beginning of the ramp. The top profile of this packed earth will be a parabola

$$y = Ax^2$$

Therefore

$$\frac{dy}{dx} = 2Ax$$

For continuity of slope at the beginning of the ramp

$$2Ax = \tan(12°)$$

Also

$$Ax^2 = 0.3$$

Hence elimination of A yields

$$\frac{x}{2} = \frac{0.3}{\tan 12°}$$
$$x = 2.82$$

and so

$$A \approx 0.0377$$

Thus the packed-earth entry to the ramp should begin 2.82 metres from the ramp and have a parabolic profile given by $y \approx 0.0377x^2$.

4. Velocity at the Top of the Ramp

It will be assumed that the car achieves its top speed on reaching the beginning of the packed earth entry to the ramp. Its speed on reaching the top of the ramp will now be calculated.

Suppose that V_0(=48.9) denotes the top speed of the car in ms^{-1}, V_1 denotes the speed of the car as its back wheels leave the ramp, P is the maximum power generated by the engine and M is the mass of the car and driver (1842 kilograms).

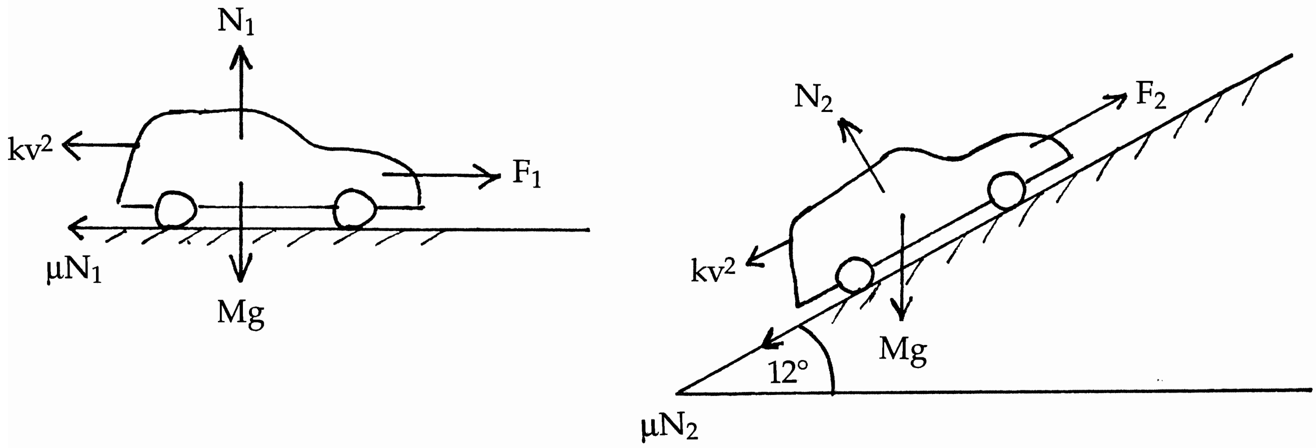


Figure 4: *Force diagrams for the flat and the ramp*

On the flat at top speed, the Principle of Linear Momentum for the car yields (see Figure 4)

$$F_1 - \mu_1 N_1 - kV_0^2 = 0$$
$$N_1 - Mg = 0$$

Hence

$$F_1 = \mu_1 Mg + kV_0^2$$

where F_1 is the force exerted by the engine at full power on the flat, μ_1 is the coefficient of friction between the tyres and the ground (0.05 for a dirt road, Radovanovic [1]) and k is the air resistance coefficient. Now it is well known from fluid mechanics that

$$k = \frac{1}{2}\rho S C_D$$

where ρ denotes the density of air (1.22 kg m^{-3}), S denotes the cross sectional area $(\approx 2m^2)$ and C_D denotes the drag coefficient (0.5 for a Ford Falcon, Radovanovic [1]). Thus $k \approx 0.61$.

When the car is travelling up the ramp the Principle of Linear Momentum yields (see Figure 4)

$$F_2 - \mu_2 N_2 - Mg\ \sin(12^\circ) - kv^2 = M\frac{dv}{dt}$$
$$N_2 - Mg\ \cos(12^\circ) = 0$$

where F_2 is the force exerted by the engine at speed v under maximum power, and μ_2 denotes the coefficient of friction on the ramp (μ_2 = 0.015 for a wooden surface, Radovanovic [1]) But

$$P = F_2 v = F_1 V_0$$

and so

$$M\frac{dv}{dt} = \frac{F_1 V_0}{v} - \mu_2 Mg\ \cos(12^\circ) - Mg\ \sin(12^\circ) - kv^2$$

Thus

$$\frac{dv}{dt} = g\left(\frac{\mu_1 V_0}{v} - \mu_2\ \cos(12^\circ) - \sin(12^\circ)\right) + \frac{k}{M}\left(\frac{V_0^3 - v^3}{v}\right)$$

Since $k/M = 3.3 \times 10^{-4}$ and v will be close to V_0 at the top end of the ramp, the last term is neglected and

$$\begin{aligned}\frac{dv}{dt} &\approx g\left(\frac{\mu_1 V_0}{v} - \mu_2\ \cos(12^\circ) - \sin(12^\circ)\right)\\ &= \frac{23.99}{v} - 2.18\end{aligned}$$

Thus

$$v\frac{dv}{ds} = \frac{23.99 - 2.18}{v}$$

where s is the distance up the ramp. This differential equation with $v = V_0$ when s=0 has the solution

$$s = -0.23v^2 - 5.05v - 121.10\ \ln(2.18v - 23.99) + 1331.48$$

When the car reaches the top of the ramp s = 1.8 cosec (12°) = 8.66 and $v = V_1$. Hence

$$8.66 = 1331.48 - 0.23\,V_1^2 - 5.05V_1 - 121.10\ln(2.18V_1 - 23.99)$$

which has the solution $V_1 = 48.6$. The neglected term $k\left(V_0^3 - v^3\right)/(Mv) \approx 0.013$ which is small compared with the terms kept. Thus the speed of the car at the top of the ramp is $48.6ms^{-1}(\approx 175km / hr)$.

But the speed at the top of the ramp is only one aspect of the car's velocity. It would also be useful to know the angle of elevation of the car immediately it leaves the ramp.

Two factors may cause this to be different from the 12° slope of the ramp. Firstly, the springs of the car may be in either a compression or expansion phase at the top of the ramp. Various combinations of compression and expansion for the rear and front springs could either increase the angle of elevation of the car or reduce it. Secondly, when the front wheels leave the ramp, the weight of the car will start to rotate the car towards the ground and so by the time the rear wheels have cleared the ramp the direction of travel of the car will be less than 12° to the vertical. It will be shown later that this amounts to a difference of only 0.43°.

Without being able to measure the spring effects, let the assumption be made that the two effects cancel out and the car takes off with an angle of elevation 12°.

5. The Distance Travelled in the Jump

In calculating the horizontal distance travelled, it is adequate to treat the car as a particle (that is, a point mass). In a later section when the attitude of the car is investigated it will be necessary to regard it as a body of finite size.

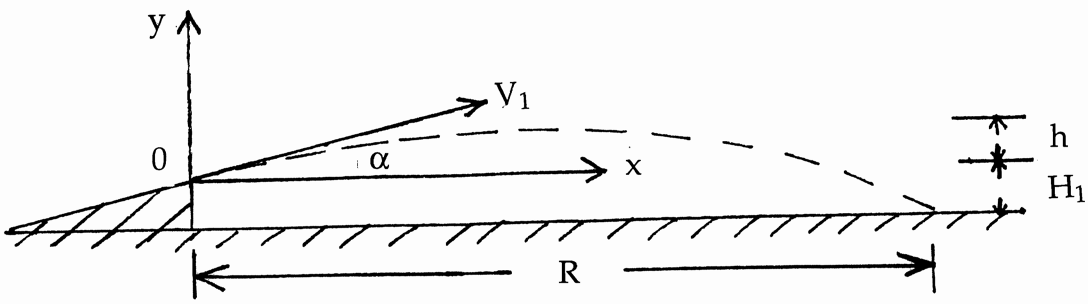


Figure 5: *Path of the car through the air*

With axes 0xy as shown in Figure 5, the equations governing the flight of the car are:

$$\ddot{x} = 0, \qquad \ddot{y} = -g,$$
$$\dot{x} = V_1 \cos\alpha, \qquad \dot{y} = V_1 \sin\alpha - gt,$$
$$x = V_1 t \cos\alpha \qquad y = V_1 t \sin\alpha - \frac{1}{2} g t^2,$$

where t is the time after leaving the top of the ramp and a dot above a symbol denotes differentiation with respect to t. Detailed calculations show that the effect of air resistance is small enough for it to be ignored in this part of the analysis also (see Appendix A).

The total time of flight of the car, t_1 say, is given by

$$-H_1 = V_1 t_1 \sin\alpha - \frac{1}{2} g t_1^{\,2}$$

where it is assumed that the car will reach the ground in a sandpit at ground level, or in some catching device (see Appendix B).

Thus

$$t_1 = \frac{V_1 \sin\alpha + \sqrt{V_1^2 \sin^2\alpha + 2gH_1}}{g}$$

is the only positive solution, and with $H_1 = 1.8$, $V_1 = 48.6$, $\alpha = 12°$ gives $t_1 \approx 2.25$ (seconds)

The horizontal distance travelled in metres from the top of the ramp is

$$\begin{aligned} R_1 &= V_1 t_1 \cos\alpha \\ &= 105.8 \text{ (metres)} \end{aligned}$$

This is slightly above the objective of 100 metres which is desirable as it allows for possible failure to attain the predicted speed or angle of elevation at the top of the ramp.

It is also not clear what the rules for measurement of a car long jump are. Clearly the measurement will start from the vertical line through the top of the ramp. The centre of mass of the car will be ahead of this top end of the ramp, adding about $\frac{2}{3}$ of the wheelbase ($\approx$ 1.8m) to the jump.

On the other hand the measuring point at the far end of the jump is probably where the car first strikes the ground. If the rear of the car flips over the front, there is no measuring problem (but the driver may have other problems). As it is shown in a later section, it is most unlikely that the car would fall back into the sandpit as human long-jumpers sometimes do, because the car will be travelling at much too high a speed for this possibility to occur. Therefore another quarter of a wheelbase ($\approx$ 0.7m) can be added to the jump distance because of the position of the centre of mass of the car at landing and the angle of striking (see the next section).

To find the height reached during the jump the condition $\frac{dy}{dt} = 0$ is used to show that the time t_2 to the highest point is

$$t_2 = (V_1 / g)\sin\alpha,$$

The maximum height above the take-off point 0 is thus

$$\begin{aligned} h &= \frac{1}{2}(V_1^2 / g)\sin^2\alpha \\ &= 5.2 \end{aligned}$$

Therefore the maximum height of the car during its flight above the ground is $h + H_1 = 5.2 + 1.8 = 7.0$ metres. This is quite a drop!

6. ROTATION OF THE CAR DURING FLIGHT AND LANDING

When the front wheels of the car leave the top of the ramp, the car begins to rotate about the rear wheels under the influence of the moment exerted by its weight.

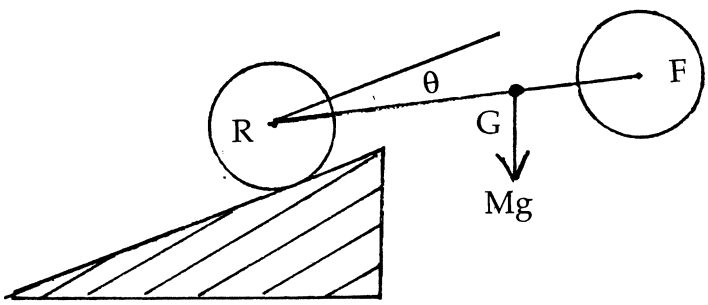


Figure 6: *Front wheels clear of ramp*

Let θ denote the angle of rotation of RG in the time t since the front wheels left the ramp.

The Principle of Angular Momentum about A yields

$$I_R\ddot{\theta} = Mg(RG)\cos(\alpha - \theta)$$

If it is assumed that θ can be neglected compared with α during the whole time that the rear wheels remain on the ramp, then substitution for I_R and RG produces

$$\ddot{\theta} = (4g\ \cos\alpha)/3d = 4.57\ (\text{rad/s}^2)$$

Thus

$$\dot{\theta} = 4.57\ t\ (\text{rad/s})$$

and

$$\theta = 2.285\ t^2\ (\text{rad})$$

The time between the front and the rear wheels leaving the ramp is (to a very close approximation)

$$t = d / V_1 = 2.8 / 48.6 = 0.0576(\mathrm{s})$$

So when the rear wheels leave the ramp,

$$\dot{\theta}_1 = 4.57 \mathrm{x} 0.0576 = 0.263(\mathrm{rad/s})$$
$$\theta_1 = 7.58 \mathrm{x} 10^{-3}(\mathrm{rad}) = 0.43^0$$

This small calculated value θ_1 substantiates the earlier claim that θ may be neglected compared with α in the Principle of Angular Momentum and also that it could be easily cancelled by the effect of the springs.

The calculated value $\dot{\theta}_1$ of the angular velocity of the car when it leaves the ramp will persist unchanged throughout the subsequent flight through the air since air resistance is being neglected and the only force acts through the centre of gravity. The angle of forward rotation of the car during its flight is therefore

$$\theta_2 = \dot{\theta}_1 t_1 = 0.263 \mathrm{x} 2.25 = 0.592\ (\mathrm{rad}) \approx 34^0$$

On landing, the nose down angle of the car will therefore be

$$\theta_2 - \alpha \approx 34^0 - 12^0 = 22^0$$

This value must be regarded as approximate since it depends to a certain degree on the assumed nature of the mass distribution of the car.

Since air resistance is neglected, the speed on landing is approximately $48.6\mathrm{ms}^{-1}$ and the direction of the velocity vector is 14.3° below the vertical. When this is combined with the angle of the car (22°) when it strikes the ground, the rotational behaviour of the car upon striking the sandpit can be determined.

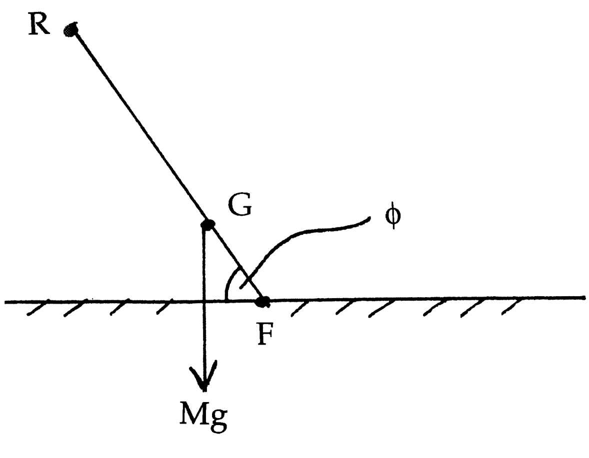


Figure 7: *Rotational motion just after impact*

Again using the Principle of Angular Momentum (with moments about F to eliminate the moments of the impact forces) it is seem that

$$I_F\ddot{\phi} = -Mg(d/3)\cos\phi$$

where ϕ is the angle made by the car with the horizontal as shown in Figure 7. Now

$$I_R = I_G + M\left(\frac{2d}{3}\right)^2$$

and so $I_G = Md^2/18$. Hence

$$I_F = I_G + M\left(\frac{d}{3}\right)^2 = Md^2/6$$

Thus

$$\ddot{\phi} = -(2g/d)\cos\phi$$

which on integration with respect to ϕ yields

$$\frac{1}{2}\dot{\phi}^2 = -(2g/d)\sin\phi + C$$

Now at impact

$\dot{\phi} = (3V_1/d)\sin(\theta_2 - \theta_1) \approx 6.98$ (rad / sec) while θ = 22° = 0.384 (rad), and so C = 26.96. When $\phi = \pi/2$ (radians) it is seen that $\dot{\phi} = 2\sqrt{26.96 - 7.01} = 8.93$ and so the car will begin to rotate end-over-end in its direction of motion along the sandpit.

7. CONCLUSIONS

The analysis indicates that a jump of 100 metres is feasible with the equipment described. One small area of doubt is the behaviour of the car on its springs after encountering the change of slope at the start of the ramp. Another is the final nose down angle of the car, the value of which depends on the distribution of mass within the car; an estimate has had to be made on this matter.

The speed on landing, and the height from which the car descends (7.0m), mean that the jump is a very hazardous stunt. The danger could be reduced if a nylon mesh barrier was placed at the landing zone. In Appendix B a suggested form of such a barrier is described.

Fortunately, sponsorship money was not forthcoming for the attempt and it was cancelled.

REFERENCES

[1] Radovanovic G., Ford Motor Co, Geelong (Private Communication, 1994).

[2] Brearley M.N., The long jump at Mexico City, Function Magazine, Monash University, Vol 3, Part 3, 16-19 (1979).

Appendix A: The Effect of Air Resistance on the Length of the Jump

To determine the reduction in the range R of the jump caused by air resistance, use will be made of a result obtained by Brearley [2]. That article investigated the effect of air resistance on the long jump of an athlete, and is relevant because the approximations used in it are the same as those applicable to the car jump, being based on the face that the angle of take-off is small. It was shown by Brearley [2] that the reduction (δR_1) in the range R_1 is

$$\delta R_1 = \rho S C_D R_1^2/(4M)$$

Using the values listed earlier for the car travelling through air, it is seen that $\delta R_1 \approx 1.85$ (metres). Thus air resistance will reduce the calculated length by less than two metres, which justifies the claim that it may be safely neglected in the calculations.

Appendix B: A Landing Safety Net

A nylon rope mesh, securely attached to a strong steel frame, would help to cushion the shock of landing. Figure 8 shows a suggested form for such a net.

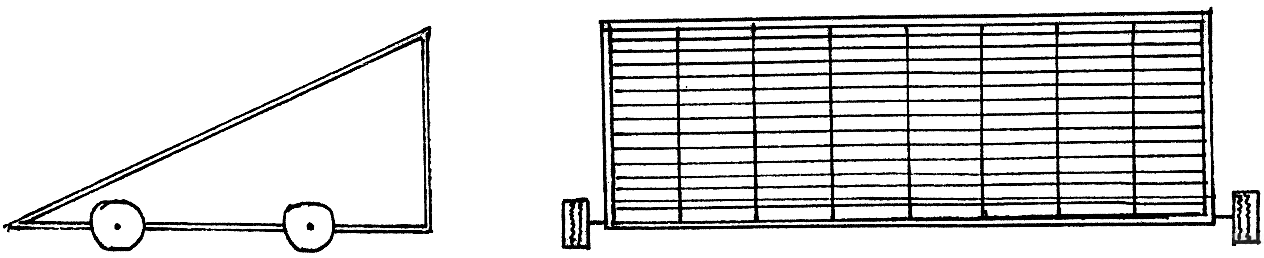


Figure 8: *Side and front elevation sketches of a proposed nylon mesh landing net*

In Figure 8, the double lines represent substantial steel members, the single lines represent thick nylon ropes. The wheels permit the safety net to run forward under the impact of the car, and also enable the barrier to be moved by means of a tractor or mobile crane.

However if the net wheels moved along ground level, the impact point of the car would be above ground level and the distance of the jump would be reduced because of the reduction in the height of fall to impact.