Fourth Conference on:

MATHEMATICS AND COMPUTERS IN SPORT

Held at Bond University, Queensland, Australia
13th July to 15th July, 1998

Sponsored jointly by

ANZIAM

and

School of Information Technology, Bond University

Edited by Neville de Mestre, Professor of Mathematics
Kuldeep Kumar, Associate Professor of Statistics
School of Information Technology, Bond University

CONTENTS (Vol 4)

GENERAL

This is the fourth conference on Mathematics and Computers in Sport, and they have all been held at Bond University. Previous conferences were in 1992, 1994 and 1996. Once again researchers have come with many different backgrounds in sport applications.

This conference features at least one paper where computers have made a direct and substantial contribution to the outcome. On the other hand, at least one other paper seems to have direct application to coaching techniques.

We would like to thank the three invited speakers John Croucher (Macquarie), Tony Lewis (West of England) and Chris Harman (Southern Queensland) for their willingness to contribute to the success of the conference, as well as to all other participants for their contributions.

We would also like to thank Michelle Brown who has been indispensable in formatting these conference proceedings and helping us with many other tasks pertaining to the smooth running of the conference.

Neville de Mestre
and
Kuldeep Kumar
Conference Directors, 1998.

PROGRAM

Monday 13 July

8.50 am	Opening by Grant Hackett
9.00 am	J.S. Croucher - "Using computers and scientific method to determine optimal strategies in tennis"
10.00 am	I.L. Collings and N.J. de Mestre - "A refined aerodynamic model for low trajectory flight"
10.30 am	*Morning Tea*
11.00 am	M.N. Brearley and N.J. de Mestre - "The effect of oar flexing on rowing performance"
11.30 am	I. Skinner - "A geometrical argument defining an order relation for all-round performances in cricket"
12.00 noon	S. Innes and S. Sugden - "One day cricket scoring programme"
12.30 pm	*Lunch*
2.00 pm	I. Heazlewood and G. Lackey - "Mathematical models that predict performance decline in elite veteran athletes 30-90 years in the sprint, distance and jump events"
2.30 pm	G. Lackey and I. Heazlewood - "The use of mathematical models to predict elite swimming performance"
3.00 pm	G. Christos - "The AFL finals: It's more than a game"
3.30 pm	J. Cross - "Scheduling"

Tuesday 14 July

9.00am	T. Lewis - "Developments in the Duckworth-Lewis (D/L) method of target resetting in one day cricket matches"
10.00 am	D. Dyte - "Constructing a plausible cricket simulation using available real world data"
10.30 am	*Morning Tea*
11.00 am	S.R. Clarke - "How fair is the AFL draw?"
11.30 am	N. Tomecko and J.A. Filar - "Player Assignments in Australian rules football"
12.00 noon	G. Christos - " It's just not cricket"
12.30 pm	*Lunch*
2.00 pm	T. Swartz and B. Da Silva - "Issues in cricket and golf"
2.30 pm	K. Kumar and J.P. Verma - "Sport and Physical Fitness: An application of factor analysis for developing a test battery"
3.00 pm	R.J. Neal, S. McLean, W. Daniel and P. Myers - "In vivo predictions of ACL stress and strains during sidestepping"
3.30 pm	Mathsport Meeting

Wednesday 15 July

9.00 am	C. Harman - ""Who's on first!" "What?" "What's on second!" and how 'What' got there on an optimal baserunning path"
10.00 am	G.L. Cohen - "The Subject: Mathematics in Sport"
10.30 am	*Morning Tea*
11.00 am	A. Patterson and S.R. Clarke - "An analysis of team statistics in Australian rules football"
11.30 am	W.H. Cogill - "The mathematics of bicycling: Part II"
12.00 noon	N.J. de Mestre - "Optimising the shot put"

PARTICIPANTS

Emeritus Professor Maurie BREARLEY	Clifton Springs
Dr George CHRISTOS	Curtin University of Technology
Dr Stephen CLARKE	Swinburne University
Dr John COGILL	University of New South Wales
Associate Professor Graeme COHEN	University of Technology, Sydney
Dr Ian COLLINGS	Deakin University
John CORMACK	Charles Sturt University
Dr Jim CROSS	University of Melbourne
Professor John CROUCHER	Macquarie Univerity
Professor Neville DE MESTRE	Bond University
Dr Basil DE SILVA	RMIT University
Mr David DYTE	Swinburne University
Mr Paul ECKERMANN	Adelaide University
Dr Richard FISHER	Peter MacCallum Cancer Institute
Dr Siva GANESH	Massey University, NZ
Mr Jonathan GODFREY	Massey University, NZ
Dr Chris HARMAN	University of Southern Queensland
Dr Ian HEAZLEWOOD	Australian Catholic University
Mr Sean INNES	Bond University
Dr David JENKINS	CSIRO
Associate Professor Kuldeep KUMAR	Bond University
Dr Gavin LACKEY	Australian Catholic University
Mr Tony LEWIS	University of West of England, UK
Mr Gary MCCOY	Bond University
Dr Robert NEAL	University of Queensland
Mr Ted O'KEEFFE	Macquarie University
Mr Andrew PARROTT	Adelaide University
Mr Andrew PATTERSON	Swinburne University
Dr Iain SKINNER	University of New South Wales
Dr Steve SUGDEN	Bond University
Dr Tim SWARTZ	Simon Fraser University
Ms Nikoleta TOMECKO	University of South Australia

USING COMPUTERS AND SCIENTIFIC METHOD TO DETERMINE OPTIMAL STRATEGIES IN TENNIS

John S. Croucher[1]

Abstract

This study traces the development of strategies for playing tennis during the past forty years, from simple probability models to the more sophisticated techniques involving modern technology. A comparison is made between various serving strategies and probabilities of success are calculated from different scorelines, with the importance of each point played also being considered. Data recording techniques using computers and video-recorders are discussed and an illustration made using actual information from a Wimbledon Men's Final.

1. INTRODUCTION

Tennis has filled the research literature over the years with an abundance of mathematical papers attempting to analyse the game from a variety of perspectives. Many papers are theoretical in nature and based on arguable assumptions, while others draw conclusions using data recorded from actual matches. In recent years, the advent of sophisticated recording techniques using computers has enhanced the way in which tennis can be analysed, with every facet of the game now being examined in minute detail.

This paper traces the development of tennis research over the past twenty-five years, including some of the classical findings that are referenced time and again in modern studies. Only the essential mathematical equations will be shown, with the emphasis being on the conceptual results. The interested reader is encouraged to explore the appropriate reference for a fuller explanation.

2. SERVICE STRATEGIES

An important area of study in tennis is the effectiveness of a player's serve. In professional tennis in particular, players are expected to win their service games and failure to do so results in a 'break' of their serve that can lead to the loss of the set. It is not surprising that much research has been directed at analysing the tennis serve, not only service strategies but also ways in which the return of serve can neutralise a server's advantage.

---

[1] School of Economic and Financial Studies, Macquarie University NSW 2109

Strong and Weak Service Strategies

In one of the earliest papers of its type, Kemeny and Snell [1] used a Markov chain approach to model a single game of tennis. Some time later Hsi and Burych [2] presented a probability model for games involving two players in which they used the game of tennis as one of their illustrations and calculated the probability that one player wins a single set of 'classical' (i.e. non tie-breaker) tennis. As is the case with much of the theory-based research, it was assumed that the probability that a player wins a point on serve is constant throughout the match and that the points played are independent. While these assertions may not be entirely accurate in all cases, they certainly permitted the development of a multitude of mathematical results that provided some interesting, if debatable, conclusions. To provide some ammunition in support of these assumptions, Pollard [3] examined 5503 points played in 35 championship matches and showed that neither of these assumptions could be rejected on statistical grounds

Two opponents on the professional tennis circuit are nearly certain to meet each other several times during the year under relatively similar situations (although, presumably different playing surfaces were also taken into account). In this way, it was claimed (George [4]) that the required probabilities could be estimated with 'reasonable precision'. George [4] also used these probabilities to uncover a service strategy which maximised the probability of a player winning a point on serve. Although a simple attempt to do this was made two years earlier by Gale [5], it is the paper by George that is considered by many to be pioneer work in the field and is worthy of some discussion here.

The rules of tennis give a player two chances to make a proper service on each point, George correctly claimed and that most experienced players approach the first serve differently from the second, which is only used if the first serve is faulted. As a result, tennis players usually possess two types of serve. One, labelled the 'strong' serve, is traditionally used on the first service. This serve generally gives a high probability of winning the point to the server, given that the serve is good. The second serve, labelled as the 'weak' serve, is usually reserved for the service following an initial fault. Although this serve gives a reduced probability of winning the point given that is good, it has the advantage of a higher probability of actually being good. The relative efficacy of the two serves can vary markedly from player to player.

A probability model for winning a service point was developed for this situation using the following definitions:

$P(A)$ = the probability of the server winning the point (Event A)

$P(S)$ = the probability of a non-faulted strong serve (Event S)

$P(W)$ = the probability of a non-faulted weak serve

$P(A \mid S)$ = the conditional probability that the server wins the point if the serve is strong and not faulted

$P(A \mid W)$ = the conditional probability that the server wins the point if the serve is weak and not faulted

$P(AS)$ = the probability of the player serving a non-faulted strong serve *and* winning the point
= $P(A \mid S)P(S)$

P(AW)　= the probability of the player serving a non-faulted weak serve *and* winning the point

= P(A | W)P(W)

It follows that, if a player follows the usual strategy of using an initial strong serve followed by a weak serve if the initial serve is faulted, then:

$$P(A) = P(A \mid S)P(S) + P(A \mid W)P(W)[1 - P(S)] \qquad (1)$$

However, the above strategy is not the only one available, since there are four possible combinations of sequences of strong and weak serves. The probabilities for each of these combinations are shown in Table 1.

Table 1

Probability model of winning according to service strategy

Strategy	First serve	Second serve	P(A)
1	Strong	Weak	P(AS) + P(AW)[1 – P(S)]
2	Strong	Strong	P(AS)[2 – P(S)]
3	Weak	Weak	P(AW)[2 – P(W)]
4	Weak	Strong	P(AW) + P(AS)[1 – P(W)]

Since it is reasonable to assume that $P(A \mid S) \geq P(A \mid W)$, it follows that Strategy WS will always be inferior to Strategy SW and should never be used. If we define:

$$R = [1 + P(S) - P(W)]^{-1} \quad \text{and} \quad Z = P(AW)/P(AS)$$

then, from Table 1, it can be shown that:

Strategy SW is optimal if　$1 \leq Z < R$
Strategy SS is optimal　if　$Z \leq 1 < R$
Strategy WW is optimal if　$1 \leq R < Z$

Unfortunately for George, in the early 1970s no detailed data were kept routinely on professional matches, and so he was obliged to personally record information from two matches to illustrate his point. (This was one of the first attempts at data recording techniques that are considerably more sophisticated today.) The first of these CBS Tennis Championship matches played on clay was held on 27 August 1971 between John Newcombe (Australia) and Arthur Ashe (USA). The second match was the final played next day between Newcombe and Rod Laver (Australia). Newcombe won the first match 6-4, 7-5 while Laver won the final 6-2, 6-4. Hence, serving data were collected for two of Newcombe's matches (204 service points) and one each for Laver (85 service points)and Ashe (86 service points).

Based on the estimates from these games, it was found that 1 < Z < R in all cases, and so Strategy SW was optimal for all three players. However, only in Laver's match with Newcombe was the evidence reasonably strong, with his P(A) = 0.66 for Strategy SW while only being 0.54 for Strategy SS. For Ashe v Newcombe, the P(A) values for Strategy SW were 0.56 and 0.63 for the two players, respectively, while for

Strategy SS they were 0.53 and 0.60, respectively. In the final, Newcombe's P(A) was 0.52 using Strategy SW and 0.50 using Strategy SS.

A full account of the probabilities in all matches can be found in George's paper, but what made this research of particular interest was its combination of theoretical work with at least a crude attempt to record actual data to produce an interesting result. Indeed, if enough data were collected on modern players, Strategy SS using a second strong serve could well be an option that should be used more often depending on the opponent.

Several years later, George's results were employed (King and Baker [6]) to rank the strategies using more extensive data, this time from ten matches among world-class women players during the 1976 Virginia Slims tournament played on indoor courts. Once again the authors recorded the data by hand. Interestingly, they reached a similar conclusion to George in that they found the usual strong serve – weak serve strategy was by no means *clearly* superior to the other strategies. Indeed, they concluded that by adopting the conventional SW Strategy, players actually reduce their match-winning chances if they do not consider maximising their own strengths relative to those of their opponents.

As an example, a player strong in volley execution may be wise to use a SS strategy throughout a match since they have an increased chance of success by coming into the net following a strong serve. For players using the conventional SW strategy, improvement of rallying strength following a successful strong service ($P(A \mid S)$) was clearly the most potent factor in improving point-winning probability. In all cases, such improvements would have been much more effective than improving proportions of good first services $P(S)$.

Service Strategies within the Context of a Match

The importance of the effect of improving service point-winning probability in relation to the probability of winning the match was already well-known at the time of these papers, (Carter and Crews [7], Hsi and Burych [2] and Weinberg et al. [8]). In a more recent analysis of serving strategies (Croucher [9]) video replays were used to investigate the serving performance of the two Men's Wimbledon finalists in 1994, namely Pete Sampras (a right-handed player from the USA) and Goran Ivanisevic (a left-handed player from Croatia). In this match, Sampras was heavily favoured (as the No. 1 seed) while Ivanisevic (the No. 4 seed) was only given an outside chance. The match was notable for its unusual scoreline of 7-6, 7-6, 6-0 to Sampras. (These two players met again in the 1995 semi-final where Sampras won a much closer contest 7-6, 4-6, 6-3, 4-6, and 6-3.) This paper illustrates how match analysis can be used to develop strategies for play, and some of the relevant details of the contest are discussed below.

The total points won in each set by each player are shown in Table 2. Overall, 206 points were played of which 118 (57%) were won by Sampras and 88 (43%) by Ivanisevic. However, Sampras won 24/31 (77%) of the points in the final set to emphasise his dominance at the finish.

Table 2

Points won (Sampras v Ivanisevic, 1994 Wimbledon Men's Singles Final)

Player	Set 1	Set 2	Set 3	Total
Sampras	50	44	24	118
Ivanisevic	44	37	7	88
Total	94	81	31	206

Overall, Sampras won 76 (75.2%) of his 101 service points while Ivanisevic won 63 (60%) of his 105 serves. Interestingly, 25 (40%) of Ivanisevic's winning service points were due to an ace, emphasising his reliance on a strong first serve. Croucher [9] then examined the match video-tape in more detail to uncover any strengths or weaknesses in serving patterns. In particular, service points were classified according to whether they were an ace, double fault, and made from the backhand or forehand court. The results are shown in Table 3, including from which court the service was made which was a significant factor since the proportions of points won in each case varied markedly.

From a tactical point of view, with Sampras being right-handed and Ivanisevic being left-handed, serving from the forehand court or backhand court may not bear the same significance for each player. In particular, Sampras won 80% of points he served from the forehand court but only 70% from the backhand court, while Ivanisevic won 68% from the forehand court but only 52% from the backhand court. Hence for both players, serving from the forehand court brought the largest success, although Ivanisevic served nearly two-thirds of his aces from the backhand court. However, when he couldn't produce an ace from there, he won only 11 of the remaining 27 points (41%), thus placing an extremely heavy reliance on his producing a clean winning first serve.

Table 3

*Percentage of service point outcomes for Sampras v Ivanisevic
(1994 Wimbledon Men's Singles Final)*

Server	Court served from	Points won		Points lost	
		Ace	Other	Double fault	Other
Sampras	Forehand	14.8%	66.7%	5.5%	13.0%
	Backhand	19.2%	51.1%	4.2%	25.5%
Ivanisevic	Forehand	17.0%	50.9%	0.0%	32.1%
	Backhand	30.8%	19.2%	5.8%	44.2%

One of the major deficiencies in Ivanisevic's service game was his poor record in winning points on his second serve from the backhand court. In fact, of his 19 serves from that court he manage to win only 6 points (32%) compared to winning 13/21 (62%) of second serves from the forehand court. Interestingly, in his first five service games of the match, Ivanisevic lost the second point played (all served from the backhand court) in each game (every time trying to serve to Sampras' backhand)

until finally winning the second point of his sixth service game with an ace out wide to the backhand. Ivanisevic should not have persisted with such a poor tactic throughout the match. A superior strategy would have been to try a stronger second serve from the backhand court and randomly mixing the direction by occasionally serving to Sampras' forehand.

The combination of service court and direction of serve for each player is summarised in Tables 4 and 5. Table 4 shows Sampras' outstanding success rate for serving to Ivanisevic's forehand by winning all 16 points from the forehand court and 7 out of 8 from the backhand court for an overall success rate of 96%. This was in marked contrast to his efforts at serving to Ivanisevic's backhand in winning 27/38 (71%) serves he made from the forehand court and 26/39 (67%) made from the backhand court. Given his relatively lower success rate serving to Ivanisevic's backhand it would have been a much better strategy for Sampras to serve more often to his Ivanisevic's forehand, clearly his weak point. In fact, all of Sampras' five double faults were attempts to serve to Ivanisevic's backhand.

Table 4

Service points won-lost by Sampras
(1994 Wimbledon Men's Singles Final)

To From	Ivanisevic's forehand	Ivanesivec'sba ckhand	Total
Forehand Court	16 – 0	27 – 11	43 – 11
Backhand Court	7 – 1	26 – 13	33 – 14
Total	23 – 1	53 – 24	76 – 25

From Table 5 it can be seen that Ivanisevic had equal success serving to Sampras' forehand (59% success) or backhand (60% success). However, there were significant differences in combinations of serves where he did poorly when serving to Sampras' forehand from the forehand court and Sampras' backhand from the backhand court in winning only 25/53 (47%) of points. This contrasted sharply to when he reversed the direction and served to Sampras' backhand from the forehand court where he won 38/52 (73%) of points.

Second serves were also a major factor in the match with Ivanisevic winning only 19/40 (48%) while Sampras won 31/51 (61%). Curiously, Ivanisevic made only two attempts to serve to Sampras' forehand when making a second serve from the backhand court (both times early in the second set) and in doing so won one and lost one. The remaining 17 such serves he directed at Sampras' backhand for a poor success rate of only 29%.

Table 5

Service points won-lost by Ivanisevic
(1994 Wimbledon Men's Singles Final)

To From	Sampras' forehand	Sampras' backhand	Total
Forehand Court	8 – 8	27 – 9	35 – 17
Backhand court	11 – 5	17 – 20	28 – 25
Total	19 – 13	44 – 29	63 – 42

3. PROBABILITY OF SUCCESS

As well as considering the merits of various service strategies, there has been much research over the years on the theoretical probability of winning tennis matches, although much of this has not related to actual match data. Apart from the papers by Carter and Crews [7] and Hsi and Burych [2] in which they both formulated games of tennis as mathematical model where the probability of winning a point is constant for each player, Fischer [10] presented a sophisticated analysis in which he gave probabilities of winning both a set (both tie-break and advantage) and a match in terms of the probability of winning an individual game. Once again the independence of points is assumed, and the author uses an 'average' probability of a player winning an individual point, irrespective of serving or receiving. Since these probabilities are almost certainly different in practice, the results are of limited value.

1. While most papers developed probabilities for success from the commencement of a match, Croucher [11] considered these probabilities from various starting points. That is, he calculated the probabilities that a server would win a game from each of the sixteen possible scorelines. Table 6 presents these values given the probability p of winning an individual point on serve ranging between 0.30 and 0.80. (These probabilities should be sufficient for most players –in a study of professional tennis players (Pollard [13]), it was found that the winners of matches were found to have an average probability of 0.71 while the losers averaged 0.62.)

According to Table 6, the probability of the server winning from:
*   0-30 is always greater than from 15-40
*   40-15 is always greater than from 30-0
*   15-30 is always greater than from 30-40
*   40-30 is always greater than from 30-15
*   30-30 (or deuce) is greater than from 15-15 if $p < 0.50$, or less than if $p > 0.50$.
*   30-15 is greater than from 15-0 if $p > 0.63$ and less than if $p < 0.63$

Table 6

The probability of a server winning a game from various scorelines

					Probability p of the server winning a point					
Current score	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80
0-0	.099	.170	.264	.377	.500	.623	.736	.830	.901	.949	.978
15-0	.211	.311	.424	.542	.656	.758	.842	.905	.949	.976	.990
30-0	.412	.523	.631	.742	.813	.879	.927	.960	.980	.991	.997
40-0	.710	.787	.850	.900	.938	.963	.980	.990	.996	.998	.999
0-15	.051	.095	.158	.242	.344	.458	.576	.689	.789	.870	.930
15-15	.125	.296	.286	.389	.500	.611	.714	.804	.875	.928	.964
30-15	.284	.381	.485	.589	.688	.775	.847	.903	.944	.970	.986
40-15	.586	.672	.751	.819	.875	.919	.951	.972	.986	.994	.998
0-30	.020	.040	.073	.121	.188	.271	.369	.477	.588	.696	.795
15-30	.056	.097	.153	.225	.313	.411	.515	.619	.716	.802	.873
30-30	.155	.225	.308	.401	.500	.599	.692	.775	.844	.900	.941
40-30	.409	.496	.585	.671	.750	.820	.877	.921	.953	.975	.988
0-40	.004	.009	.020	.037	.063	.100	.150	.213	.290	.380	.481
15-40	.013	.028	.049	.081	.125	.181	.249	.328	.414	.506	.602
30-40	.047	.079	.123	.180	.250	.329	.415	.504	.592	.675	.753
40-40	.155	.225	.308	.401	.500	.599	.692	.775	.844	.900	.941

Proper analysis of the mathematical theory of tennis would not be complete without a discussion of the importance of each point in a tennis match. A popular definition of *importance I*, (Morris [12]), is based on the following difference between two conditional probabilities.

I(point) = P(Server wins game | server wins point) – P(Server wins game | server loses point)   (2)

Croucher [11] gives a complete list of the importance of each point for values of p ranging between 0.30 and 0.80. Since the receiver's probabilities are the complement of those of the server, every point is equally important to both players. A summary of the relative importance of points is shown below.

- The first point (0-0) is always of only average importance. That is, it never ranks highly or near the bottom for any value of p.
- The point 30-40 has top ranking for $p \geq 0.50$ but has decreasing importance as p falls below 0.50.
- The point 40-30 has top ranking for $p \leq 0.50$ but has decreasing importance as p rises above 0.50.
- No point has a consistently high (or low) ranking for *all* values of p. However, the points 30-30 and deuce never rank below 6th out of 16.
- For $p \geq 0.50$, those points where the server is trailing rank highly, while for $p < 0.50$ those points where the server is ahead rank highly.
- The point 30-40 always ranks higher than the point 15-30.

Define I_{sr} = the importance of a point when the server has score s and the receiver score r

P_{sr} = the probability that the server will win the game when the score is s to r

In this context we use the values of s and r to be 0, 1, 2, 3, 4 where they represent the scores of 0, 15, 30, 40 and game, respectively. For the score of 15 – 30 we have:

$$I_{12} = P_{22} - P_{13} \tag{3}$$

And for the score 30 – 40 we have:

$$I_{23} = P_{33} - P_{24} \tag{4}$$

Since $P_{24} = 0$ as the server has lost, it follows from (4) that $I_{23} = P_{33}$. Also, since a game must be won by a margin of two points or more, $P_{22} = P_{33} = I_{23}$. Substituting into (3) yields:

$$I_{12} = I_{23} - P_{13} \tag{5}$$

Since $P_{13} > 0$, it follows from (5) that $I_{23} > I_{12}$, or that the point 30 – 40 is always more important than the point 15 – 30.

Using the more complex notion of 'time-importance' in which the importance of a point is weighted by the expected number of times the point is played, Morris develops several interesting conclusions. Firstly, he claims that, even though more points are served into the forehand court than into the backhand court, the same total time-importance is associated with both sides of the court. It follows that the higher *average* time importance is experienced in the backhand court and suggests the reason why doubles teams are advised to have the more experienced or stronger player on the left side. (This ignores whether the players might be a left-hand and right-hand combination which would presumably be another factor.)

Secondly, Morris claims that the total time-importance associated with even-numbered and odd-numbered service games are equal. Since there are more odd-numbered service games, the player who serves first generally will serve under less pressure. There is some logic in this statement, since professionals are always expected to win their own service. Winning the toss, a nervous starter may elect to receive, since there may be a high probability that service will be broken in the first game but, having settled down, a lower probability by the second when it will be his/her turn to serve first.

The winner of a 'classical' or 'advantage' set must get two games ahead once the game score reaches 6-6. This method often produces very long matches. The tennis tie-breaker, introduced to shorten the length of matches, has had its desired impact and its effect has been extensively examined in the research literature (Croucher [13] and Pollard [3]). The tie-breaker is played when the game score reaches 6-6 in a set. Some of the more interesting results from these papers surrounding tie-break sets are:

1. For two unequal players, the probability that the better player wins the set is greater for the classical version.
2. The expected value and variance of the duration of a match (in points played) are always smaller for the tie-break version.

3. If both players have a probability p of winning service points between 0.50 and 0.60, then the classical and tie-break sets have similar characteristics and so the tie-breaker rule ineffective. While this case applies for most women's matches where p does lie in this range. For most men, especially on grass surfaces, the value of p is almost always greater than 0.60.

4. DATA RECORDING

The field of notational analysis in all sports is one of growing importance as players and coaches try to gain a competitive edge. The idea is to find an effective way in which to accurately record appropriate information from a match in a way that is going to be of real benefit. In some circumstances immediate feedback may be required, (e.g., recording data on your next opponent), while on other occasions, it may not be necessary to analyse the data on the spot, (e.g., collecting information for a database).

A number of aspects of tennis require examination and proper analysis that can only be achieved through the collection of appropriate data. Details can be recorded that clearly show patterns of play along with the movement of players throughout the points. Of particular interest is the identification of the strengths and weaknesses of the players and the types of situations that bring these to the fore. Once these are known, a coach can recommend appropriate action to employ actions and tactics that will maximise the expected advantage to the player.

Recording information from tennis matches (or, indeed, any sport) can range from the simple use of pencil and paper to the employment of video cameras linked to computers. For many recreational tennis players, the former method may be quite sufficient as a friend could record basic information including, for example, the types of unforced errors made, serving statistics and data on particular strokes such as lobs and passing shots.

The professional level requires more elaborate analysis, not only on the player's own game but on that of the opponent. A player would be most unwise to enter a match with little or no information about the strengths or weaknesses of the opponent, who almost certainly has those facts on the player's game. In some of the more sophisticated techniques, data are recorded either directly into a computer or on data sheets that are placed into a specially designed database written for this purpose. These video/computer systems provide comprehensive sports analysis by transferring a videotaped sequence to computer memory. Once transferred, the sequence can be immediately retrieved and replayed for individualised coaching and instruction.

The data entry person makes one complete pass through the match in which relevant information or specific incidents are noted along with the time they occurred. Once this has been done, a coach can then later type in, for example, 'passing shots' and the video recorder will intelligently go to those incidents in the match where passing shots were in evidence.

One of the earliest and long-lasting techniques in the important area of data analysis in tennis illustrates the idea behind the modern data analytic techniques of

developing strategies. In 1982, a former South African amateur, Bill Jacobsen, began to record his son's tennis matches using hand-collected data that he fed into a microcomputer. He soon designed a four-pound portable computer with which a single observer could easily record ten times as much information. The system, now known as CompuTennis, was subsequently awarded a contract by the US Tennis Association to chart matches for all four national teams – Davis Cup, Wightman Cup, Federation Cup and Davis Cup under 18.

CompuTennis does not attempt to cover every shot in a match since players can rally for minutes at a time. Instead, the observer records only *key strokes* such as the serve, return of serve and the sequence of shots (up to 5 or 6) that lead to the end of the point.

The data can be fed into the computer by one of three ways:

1) At courtside while play is in progress,
2) from a video tape replay of the match,
3) by hand using CompuTennis hand charts.

Each key on the keyboard is programmed for a particular function to collect five types of data:

1. The shot type
2. The stroke description
3. The direction of the shot
4. The result of the shot
5. The location of each player at the end of each point

The system has the capacity to record shot descriptions (e.g. half volley, passing shot), whether the shot is forehand or backhand, the zones of the court travelled, and the result (i.e. forcing shot, error). There are console keys on which to record ten other statistics such as the number of times a player runs around his backhand.

Some of the fascinating statistics to be uncovered by CompuTennis are listed below. (See CompuTennis [14].)

- In 99% of matches, the player who wins the most points in the match wins the match.
- On 95.5% of occasions, the player who has the first match point will win the match.
- In matches played on grass, more than 80% of the points are over within three seconds.
- In the Wimbledon 1991 Men's Final between the two big servers Boris Becker and Michael Stich, the average length of a point was 2.6 seconds. This translated into actual playing time per hour of only three minutes 42 seconds. Only 9 minutes of the 2 hour 31 minute match was actually spent playing tennis.
- By contrast, the Monica Seles v Martina Navratilova U.S. Open final in 1991 saw an average point length of 5.3 seconds with an average playing time of 9 minutes 41 seconds per hour. This long playing time did not suit the older Navratilova who did not give herself enough time to recover between points.

Comparisons of several key statistics for three Grand Slam championship matches in 1991 are shown in Table 7. The reader will be able to make other interesting comparisons between not only male and female matches but between the type of surface on which the match was played. In particular, note the significantly longer playing time for each point on clay, along with the longer playing time per hour.

Table 7

Time of points for three Grand Slam championships in 1991

	French Open (clay)	US Open (Hard)	Wimbledon (grass)
Men			
No. of points	279	155	214
Avg. time per point	10.0 sec.	7.6 sec.	2.6 sec.
Actual play time	50 min	19 min.	9 min.
Avg. rest time between points	21.48 secs.	26.17 secs.	27.36 secs.
Total rest time	2 hrs 30 min.	1 hr 39 min.	2 hrs 22 min
Play time per hour	14 min 56 secs.	9 min 58 secs.	3 min 42 secs
Women			
No. of points	122	119	200
Avg. time per point	11.0 sec.	5.3 secs.	5.9 secs.
Actual play time	24 min	11 min.	20 min.
Avg. rest time between points	21.9 secs.	16.7 secs.	21.5 secs.
Total rest time	1 hr 7 min.	56 mins.	1 hr 48 min
Play time per hour	15 min 43 secs.	9 min 41 secs.	9 min 18 secs.

Source: CompuTennis

Using data that were taken from a selection of professional tennis players in various tournaments up to 1993, (Jacobsen [15]), the percentages of unforced errors are higher on all surfaces for women than men. Clay courts produce the highest values since the points are much longer producing more opportunity for a greater variety of strokes and tactics. As a result, tactical and fatigue errors tend to be higher on this surface. On the other hand, the tactical demands are less on grass; players are limited to fewer possible successful strategies or stroke selections, because of the speed of the court gives players little time to think *during* the points.

Many other aspects can be examined by match analysis, including whether players should serve differently to the forehand or backhand court, the importance of the ability to serve an ace or avoid double faults, the ability to return a clean winner off a serve and a measure of how aggressive a player is. These properties of a player's game are among those considered by CompuTennis and illustrate the power of proper data collection to effectively develop winning strategies. In fact, CompuTennis has developed the concept of an 'Aggressive Margin' that attempts to explain why a particular player has an edge in a crucial match and measures a player's improvement from match to match or against other players. (See Jacobsen [16].)

5. REMARKS

There is no doubt that the game of tennis, like many other sports, is undergoing revolutionary change with regard to the amount of information that can be gathered and analysed. While much research has centred on singles play, there is ample scope for proper investigation of doubles play in which extra variables involving the interaction between playing partners are clearly significant factors. One area for further investigation would be to characterise optimal combinations of players to maximise success in doubles and an examination of patterns of play. It would be particularly instructive, for example, to analyse video coverage of the Australian doubles combination of Mark Woodforde and Todd Woodbridge to discover what factors are relevant in making them the most successful doubles combination of all time.

Other areas for tennis research include the psychological profile of players, comparison of different types of match preparation and training techniques and the resultant effect of experiencing a perceived 'bad' line call. Different styles of play can also be compared in order to find optimal strategies for a particular situation, depending upon the opponent. Research by CompuTennis shows, for example, that less than 30% of the winners of Post-War of the Men's French Open (played on clay) were 'net rushers' while the last female net rusher to win was Martina Navratilova in 1984. This leads to the question of what tactics are most effective for different surfaces. There is even room for statistical analysis of the most effective way to rank professional players and whether the current ATP (Association of Tennis Professionals) method is the most reasonable. And if one wants to examine the underlying principles of the tennis scoring system itself, the groundwork for this has already been laid (Schutz [17] and Miles [18]). In particular, the latter claims that, largely because in top men's tennis the proportions of service points won by both players is so high, the efficiency of the traditional scoring system is unduly low. Indeed, he argues that the introduction of the tennis tie-breaker has further reduced efficiency and he proposes a simple model to make the scoring more efficient.

REFERENCES

[1] Kemeny J.G. and Snell J.L., *Finite Markov Chains*, (Van Nostrand. Princeton., 1960)

[2] Hsi B.P. and Burych D.M., Games of two players. *J. of the Royal Statistical Society – Series C*, 20, 86-92 (1971).

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

[4] George S.L., Optimal strategy in tennis. *J. of the Royal Statistical Society – Series C*, 22, 97-104 (1973).

[5] Gale D., Optimal strategy for serving in tennis. *Mathematics Magazine*, 5, 197-199 (1971).

[6]   King H.A. and Baker J.A., Statistical analysis of service and match-play strategies in tennis. *Can. J. of Appl. Sports Sc.*, 4, 4, 298-301 (1979).

[7]   Carter W.H., Jr. and Crews S.L., An analysis of the game of tennis. *The American Statistician*, 28, 130-134 (1974).

[8]   Weinberg R. *et al.*, Influence of cognitive strategies on tennis serves. *Perc. And Motor Skills*, 50, 663-666 (1980).

[9]   Croucher J.S., Replaying the 1994 Wimbledon men's singles final. *New Zealand Statistician*, 30, 1, 2–8 (1995).

[10]   Fischer G., Exercise in probability and statistics, or the probability of winning at tennis. *American J. of Physics*, 48, 1, 14-19 (1980).

[11]   Croucher J.S., The conditional probability of winning games of tennis. *Research Quarterly for Exercise and Sport*, 57, 1, 23-26 (1986).

[12]   Morris C. The most important point in tennis. In R. E. Machol (Ed.), Optimal Strategies in Sports, 131-140, (North Holland, New York, 1977)

[13]   Croucher J.S., The effect of the tennis tie-breaker. *Research Quarterly for Exercise and Sport*, 53, 336-339 (1982).

[14]   CompuTennis topics, Current Topics and Technical Briefs, CompuTennis, Los Altos, California (1996).

[15]   Jacobsen W., All unforced errors are not the same. *Sports Science for Tennis*, Summer Edition, 1-3 (1993).

[16]   Jacobsen W., Getting aggressive. *Tennis Match*, Sept./Oct., 17 (1994).

[17]   Schutz R.W., A mathematical model for evaluating scoring systems with specific reference to tennis. *Research Quarterly for Exercise and Sport*, 41, 552- 561 (1970).

[18]   Miles R., Symmetric sequential analysis: the efficiencies of sports scoring systems (with particular reference to those of tennis), *J. Royal Statist. Soc.- Series B*, 46, 1, 93-108 (1984).

A REFINED AERODYNAMIC MODEL FOR LOW TRAJECTORY FLIGHT

I.L. Collings[1] and N.J. de Mestre[2]

Abstract

A refined aerodynamic model of a golf ball is proposed where the lift and drag coefficients are based upon both velocity and spin rate which is also taken to be time dependent. That is, spin rate decay is built into the aerodynamic equations. The model is robust and permits the numerical calculation of low trajectory flight characteristic projections for a range of lift, drag and spin rate parameters.

Key words: Aerodynamics, Golf ball

1. INTRODUCTION

The simplest model to describe the trajectory of a driven golf ball is to assume that it lies in the vertical plane containing the initial velocity vector of the ball, and hence the motion may be considered to be two-dimensional. This is only possible if the ball spins about a horizontal axis perpendicular to the initial vertical plane of motion, since the lift force due to the Magnus effect remains in this plane.

The differential equations of motion are then given by

$$m\ddot{x} = -D\cos\psi - L\sin\psi$$

$$m\ddot{y} = L\cos\psi - D\sin\psi - mg \tag{1.1}$$

where
- x = horizontal co-ordinate
- y = vertical co-ordinate
- D = drag force
- L = lift force
- m = ball mass
- ψ = angle between velocity vector and x axis
- g = acceleration due to gravity.

Hence at any time $t, \dot{x} = v\cos\psi, \dot{y} = v\sin\psi$ where the speed $v = \sqrt{\dot{x}^2 + \dot{y}^2}$. Further, the initial conditions (when $t = 0$) are

$$x = 0, y = 0, v = v_0, \psi = \psi_0 \tag{1.2}$$

[1]   School of Computing and Mathematics, Deakin University, Geelong Vic 3217
[2]   School of Information Technology, Bond University, Gold Coast Qld 4229

Both Daish [1] and Bearman and Harvey [2] propose that the lift and drag forces be modelled by

$$L = \frac{1}{2}\rho S v^2 C_L \qquad\qquad (1.3)$$

and

$$D = \frac{1}{2}\rho S v^2 C_D \qquad\qquad (1.4)$$

where C_L and C_D denote the lift and drag coefficients respectively, ρ denotes the density of air (1.226kg/m^3 at sea level) and $S(\approx 0.00143\text{m}^2)$ denotes the projected or cross-sectional area of the ball.

The lift and drag coefficients can be shown by dimensional analysis to be functions of the spin parameter and the Reynolds number. Thus they change as the linear and rotational speeds of the ball change during flight. A number of experiments have been carried out to determine their values at different linear and rotational speeds.

Maccoll [3] considered a spinning smooth sphere supported by a spindle. Davies [4] spun a golf ball in a fixed position inside a wind tunnel until the desired spin rate was attained. The ball was then released, and the drift of the ball was used to calculate the lift coefficient. It has since been suggested that Davies' wind tunnel speed of just under 31.5ms^{-1} is too low to be representative of golf-ball drives.

Bearman and Harvey [2] utilised a $2\frac{1}{2}$ times scale model of a golf ball mounted on a thin wire. A small motor inside the ball drove the ball's rotation about the wire. The experiments were conducted in a wind-tunnel capable of ten different linear speeds, which simulated golf ball speeds from 14 to 88ms^{-1}.

Davies ball-drop method was updated by Aoyama [5] using current video and computer technology to automate the wind-tunnel experiments. Details of the experimental values were not given, but diagrams indicated that the variation of C_L and C_D with linear and rotational speed of the ball followed closely the results produced by Bearman and Harvey's graphs [2], with greater experimental accuracy being claimed.

More recently, Smits and Smith [6] obtained lift and drag measurements taking spin rate decay into account. They found similar results to Bearman and Harvey [2] for the lift coefficient, although for the drag coefficient their results indicated a stronger dependence on spin rate parameter over the entire spin rate regime than did Bearman and Harvey [2]. Their data also identified a decrease in drag coefficient at higher Reynolds numbers which had previously been unreported in the literature. They clearly identify the importance of spin rate (and hence spin rate decay) on both the lift and drag coefficients.

2. A REFINED MATHEMATICAL MODEL

Davies [4] assumed the lift force to be proportional to v and the drag force to v^2. That is, he took C_L to be proportional to v^{-1} and C_D to be a constant. Also, McPhee and Andrews [7] and MacDonald and Hanzely [8] have made the same or similar assumptions about C_L and C_D for analytical convenience.

As outlined in the introduction, Bearman and Harvey [2], Aoyama [5] and Smits and Smith [6] have all demonstrated the dependence of C_L and C_D on both the instantaneous velocity v <u>and</u> spin rate ω. Bearman and Harvey [2] integrated the momentum equations (1.1) using a numerical step by step process that introduced the appropriate C_L and C_D values from their wind-tunnel data at each step of the computation. One of the difficulties with this approach is that spin rate is not included in the formulation and an assumption is made about its initial value and how it changes during flight until impact.

Smits and Smith [6] propose $L = k_1 \omega^{0.4} v^{1.6}$ and $D = k_2 v^2 + k_3 \omega v$ (plus a sinusoidal nonlinear curve fit term). This paper assumes $L = k_1 \omega^{\gamma_1} v^{\gamma_2}$ and $D = k_2 \omega^{\alpha_1} v^{\alpha_2}$ where the k_i, γ_i and α_i are determined from the Bearman and Harvey [2] data by least squares analysis for a wide range of ω and v. In addition, following Bearman and Harvey [2] we take the spin rate decay to be proportional to $\omega^2 r^2 / v^2$ (where r = ball radius) and so

$$\omega = \frac{\omega_0}{1 + R_1 \omega_0 t} \tag{2.1}$$

where ω_0 is the initial spin rate and R_1 is determined from the data of Smits and Smith [6]. Spin rate decay is therefore built into the momentum equations.

Making the usual low trajectory approximation $|\ddot{y}| << \ddot{x}$ and applying the initial condition for $\dot{x}$ implied in (1.2), the horizontal momentum equation may be integrated once to obtain

$$\dot{x} = \left\{ \frac{(\alpha_2 - 1)k_2 \omega_0^{\alpha_1 - 1}}{mR_1(1 - \alpha_1)} \left[(1 + R_1 \omega_0 t)^{1 - \alpha_1} - 1 \right] + (v_0 \cos \psi_0)^{1 - \alpha_2} \right\}^{\frac{1}{1 - \alpha_2}}, \tag{2.2}$$

with $x(0) = 0$.

The vertical momentum equation is

$$\ddot{y} = \frac{-k_2 \left(\dfrac{\omega_0}{1 + R_1 \omega_0 t} \right)^{\alpha_1} \dot{x}^{\alpha_2 - 1} \dot{y}}{m} + \frac{k_1 \left(\dfrac{\omega_0}{1 + R_1 \omega_0 t} \right)^{\gamma_1} \dot{x}^{\gamma_2}}{m} - g, \tag{2.3}$$

with $\dot{y}(0) = v_0 \sin \psi_0$

and $y(0) = 0$.

Equations (2.2) and (2.3) are trivially coupled and may be easily integrated numerically using a symbolic manipulation package.

The constants are:

$$k_1 = 9.387 \times 10^{-6}, \gamma_1 = 0.6427, \gamma_2 = 1.388$$

$$k_2 = 2.264 \times 10^{-4}, \alpha_1 = 0.1587, \alpha_2 = 1.693$$

$$m = 0.0459, g = 9.81, R_1 = 2 \times 10^{-5}$$

where a 30% spin rate decay over a 6.5 sec. flight is taken.

The parameters are:

ω_0, the initial spin rate (rpm)

v_0, the initial velocity (ms^{-1})

ψ_0, the initial angle (radians).

3.   RESULTS

Equations (2.2) and (2.3) were solved using the Maple package for a range of initial spin rates, velocities and angles, with some results presented in the figures below.

The effect of initial velocity on distance is very close to linear for initial velocities in excess of $40\mathrm{ms}^{-1}$ (Fig 3.1), whereas the effect of initial velocity on both the horizontal component of landing velocity and maximum height is non-linear (Figs. 3.2, 3.3).
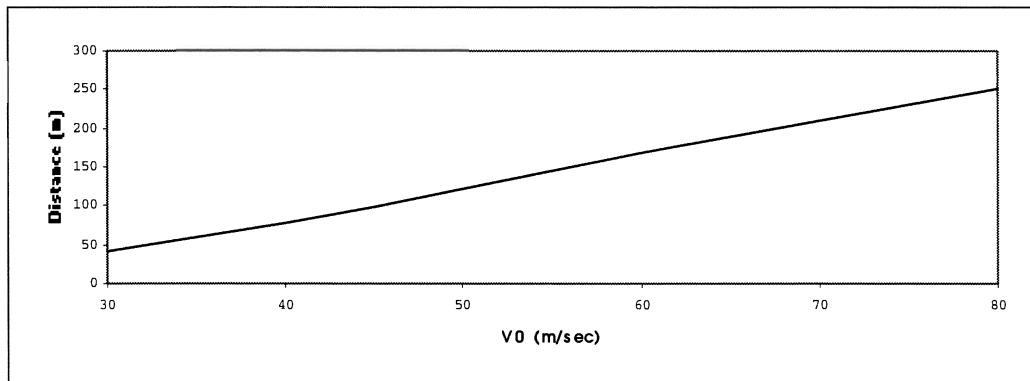
Figure 3.1: Plot of distance against initial velocity with $\omega_0 = 3500, \psi_0 = 10°$
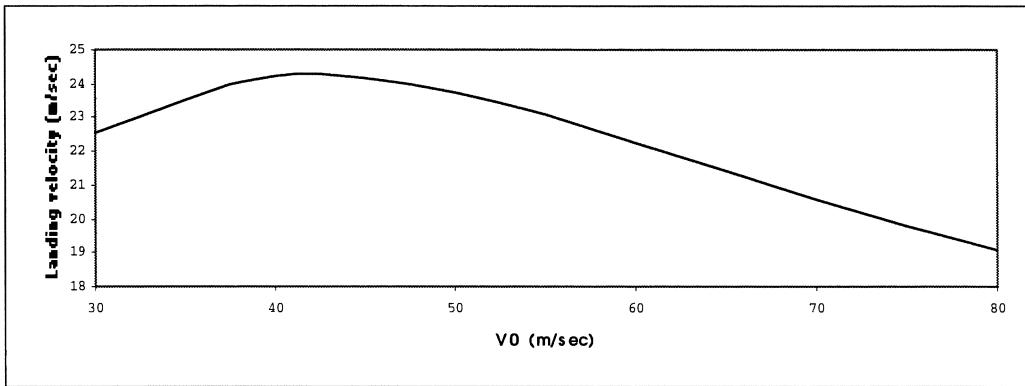
Figure 3.2: Plot of horizontal landing velocity against initial velocity with
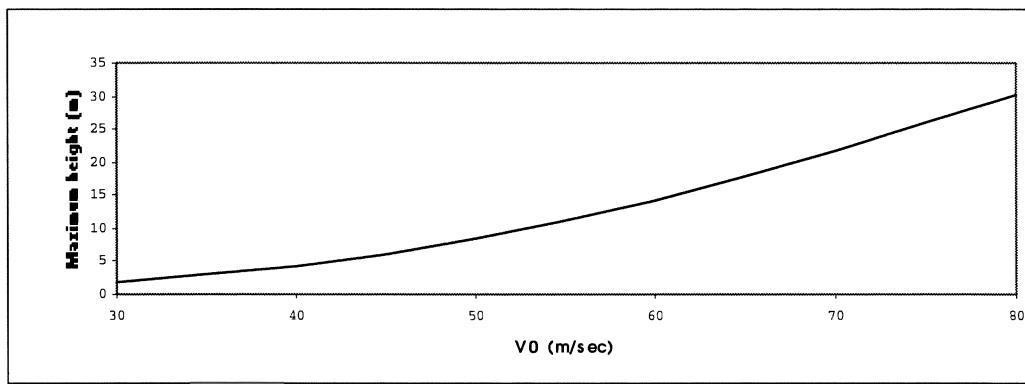
$$\omega_0 = 3500, \psi_0 = 10°$$

Figure 3.3: Plot of maximum height against initial velocity with $\omega_0 = 3500, \psi_0 = 10°$

As expected, with v_0 and ω_0 fixed, an increase in ψ_0 leads to an increase in carry, with of course a reduced horizontal velocity on impact. (Fig. 3.4). Finally, the effect of an increase in ω_0 with v_0 and ψ_0 fixed leads to an increase in maximum height and also an increase in distance of carry. (Fig. 3.5).
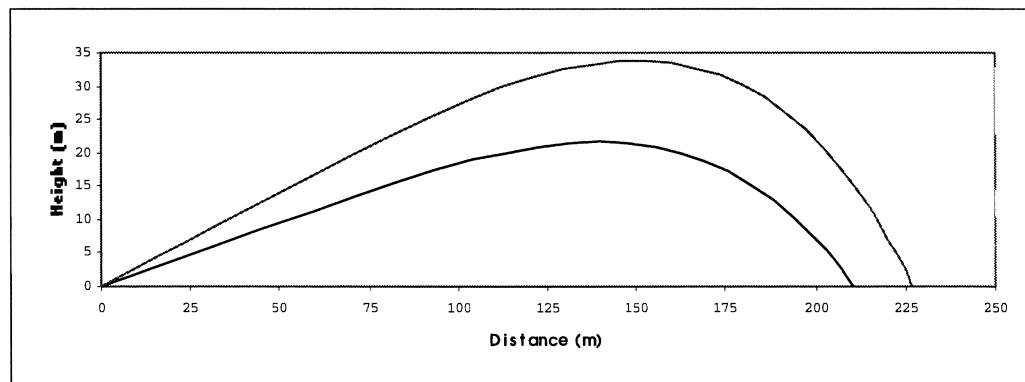
Figure 3.4: Trajectories for $\psi_0 = 10°$ and $\psi_0 = 15°$ with $v_0 = 70, \omega_0 = 3500$
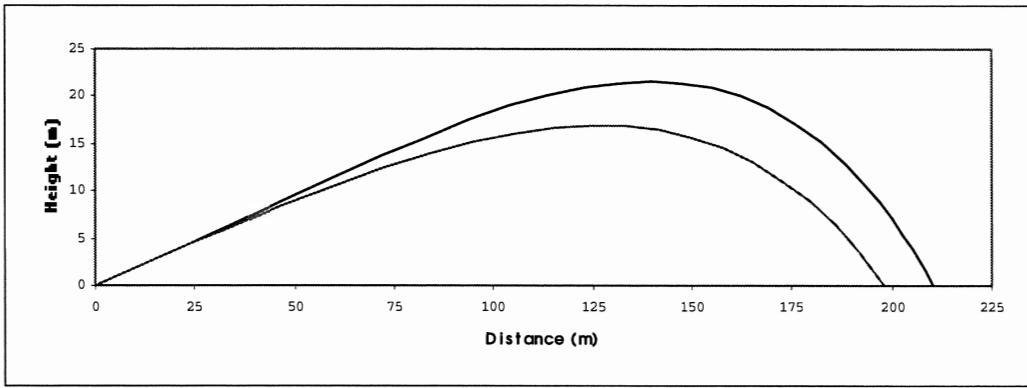
Figure 3.5: Trajectories for $\omega_0 = 2500$ and $\omega_0 = 3500$ with $v_0 = 70, \psi_0 = 10°$

4.   CONCLUSION

A model for low trajectory golf ball flight is proposed which incorporates spin rate decay in the momentum equations. The model permits a rapid calculation of the effects of initial velocity, angle and spin rate on low trajectory flight. It is robust, and changes in golf ball design leading to different lift and drag coefficients can easily be incorporated into the model and the flight characteristics determined. The model performs very well in matching the "real world" trajectory data as presented by Bearman and Harvey [2].

REFERENCES

[1]   Daish C.B., The physics of ball games, English University Press, London, (1972).

[2]   Bearman P.W. and Harvey J.K., Golf ball aerodynamics, *Aeronautical Quarterly*, 27, 112-122, (1976).

[3]   Maccoll J., Aerodynamics of a spinning sphere, *J. Royal Aeronautical Soc.*, 32, 777, (1928).

[4]   Davies J.M., The aerodynamics of golf balls, *J. Applied Physics*, 20, 821-828, (1949).

[5]   Aoyama S., A modern method for the measurement of aerodynamic lift and drag on golf balls, Science and Golf, (ed. A.J. Cochran), E. & F.N. Spon, London, 187-192, (1990).

[6]   Smits A.J. and Smith D.R., A new aerodynamic model of a golf ball in flight, Science and Golf II, (eds. A.J. Cochran and M.R. Farrelly), E. & F.N. Spon, London, 340-347 (1994).

[7]   McPhee J.J. and Andrews G.C., Effect of sidespin and wind on projectile trajectory, with particular application to golf, *American J. Physics*, 56 (10), 933-939, (1988).

[8]   MacDonald W.M. and Hanzely S., The physics of the drive in golf, *American J. Physics*, 59(3), 213-218, (1991).

THE EFFECT OF OAR FLEXNG ON ROWING PERFORMANCE

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

Abstract

Modern oars are made of fibre-glass, and bend to a significant extent under the loads occurring in normal use. This paper analyses the effect of such flexing on the performance of a coxless pair in a race over 2000 metres, but the conclusions reached are typical of all classes of racing shells. Some oar manufacturers offer oars classed as stiff, medium or soft, depending on their degree of flexibility. A quantitative comparison is made of the performances achievable with stiff, medium and soft oars, and (hypothetical) rigid oars. It is found that performance improves as the stiffness of the oars increases.

1.   INTRODUCTION

During the power stroke the loads on the oars of a racing shell are great enough to cause them to flex to some degree. The object of this investigation is to determine the amount of oar flexing, and its effect on the performance of a typical boat. A quantitative judgement is possible because of information available from an Australian Institute of Sport telemetry system for measuring oarlock forces.

Results will be obtained for one particular oarsman in a coxless pair, on the assumption that the oar involved is of a given stiffness. Similar analyses could be conducted for other rowers tested by the AIS, and for other types of boats, but would not be expected to yield results substantially different from those obtained here.

2.   THE FORCES ACTING ON A RIGID (NON-FLEXING) OAR

Figure 1 shows the important forces acting on a rigid oar during the power stroke, and some of the notation to be used. The concept of a rigid oar is theoretical, since no real material is completely rigid.

---

[1]   85 Dandarriga Drive, Clifton Springs   Vic   3222
[2]   School of Information Technology, Bond University, Gold Coast Qld 4229
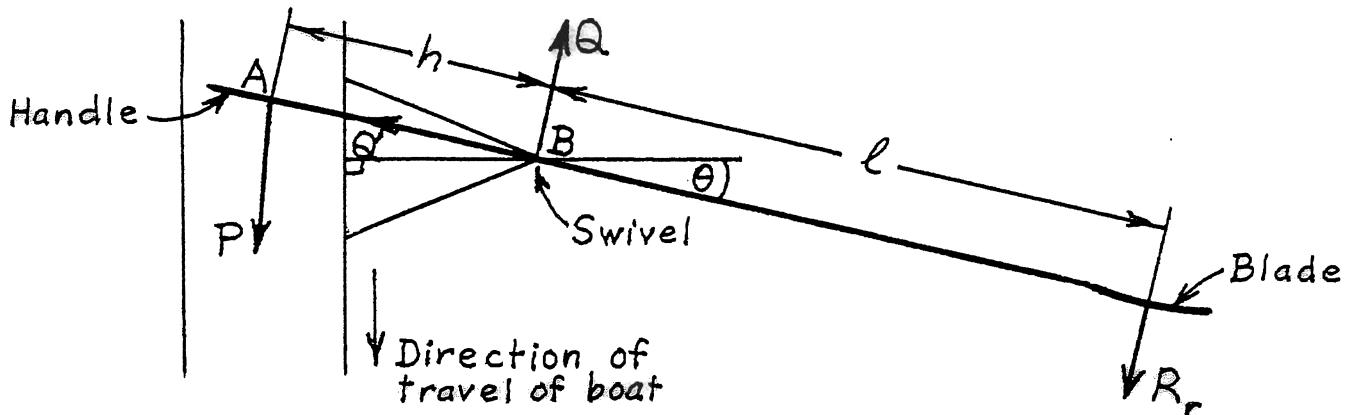
Figure 1: Plan view of the forces acting on a rigid oar during the power stroke

The notation shown includes the following:

P   = force exerted by the rower (taken to act between hands),

Q, Q^1 = components perpendicular to and parallel to the oar of the force exerted at the swivel,

R_r = force exerted by the water (taken to be at blade centre),

θ = angle between oar and the "square off" position.

The force R_r is assumed to be perpendicular to the blade, the lateral frictional force of the water being neglected. The suffix r on R is used to indicate that this force is for a rigid oar.

A modern oar is light enough for its mass to be neglected entirely. As a result, taking moments about the point A of the forces on the oar gives

$$R_r(\ell + h) - Qh = 0$$

that is,
$$R_r = h(\ell + h)^{-1}Q = 0.3003Q \qquad (1)$$

on using the typical values h = 1.00 m, ℓ = 2.33 m.

3.   THE MECHANICS OF OAR FLEXING

The forces on the oar cause it to bend, thus changing the angles which the blade and the force R_r make with the direction of motion of the boat. To determine the amount of this change, information provided by a manufacturer of oars can be used.

Oars are usually divided into three classes, depending on their flexibility. They are called stiff, medium or soft oars depending on the degree to which they bend under an applied load. Figure 2 shows the arrangement which one manufacturer [1] uses to check the flexibility of an oar.
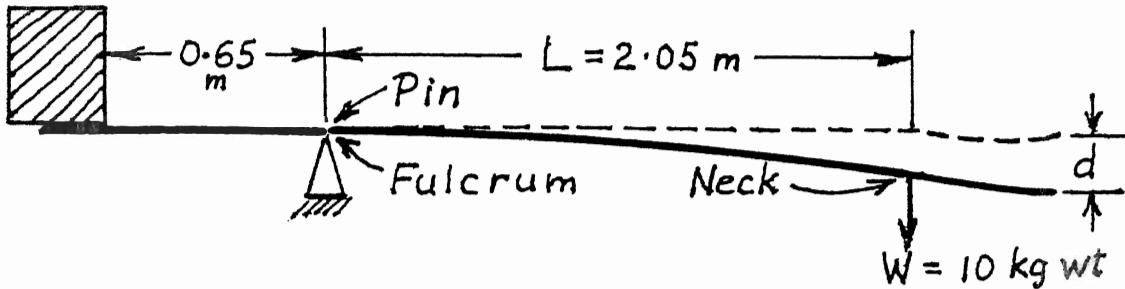
Figure 2: Deflection test of oar flexibility.

Let d = the end deflection of the oar caused by a weight W of 10 kg wt (=98N) at the neck.

Consideration will at first be restricted to the case of a medium oar, for which [1]

$$d = 3.85 \pm 0.25 \text{ cm} = 0.0385 \text{ m}$$

The theory of beam deflection [2] may be applied, taking the situation in Figure 2 to be that of a cantilevered beam of length L = 2.05 m under an end load W = 98 N. The relevant formula for the end deflection is

$$d = WL^3/(3EI), \tag{2}$$

where E = the modulus of elasticity of the oar material,

I = the moment of inertia of the oar cross-section.

The formula (2) assumes that I is constant along the oar shaft, which is not actually the case since the shaft tapers towards the blade. If taper is taken into account it is found that calculated values of blade deflection angle are increased by only about 10% above those calculated using (2). This difference can be shown to affect the final conclusions of this paper by less than 1%, so that shaft taper may reasonably be neglected. The fact that the deflection d in Figure 2 is not measured just where W is applied will not cause significant error because the blade length is small compared with L.

On substituting into (2) the values mentioned above for d, L and W it becomes

$$0.0385 = 98 \times (2.05)^3/(3EI),$$

that is, $\qquad EI = 7310 \text{Nm}^2 \tag{3}$

In the theory of beam deflection, the quantity EI is called the flexural rigidity of the beam.

Now consider how the shape of an oar changes when flexing occurs under the loads experienced during the power stroke. This situation is depicted in Figure 3.
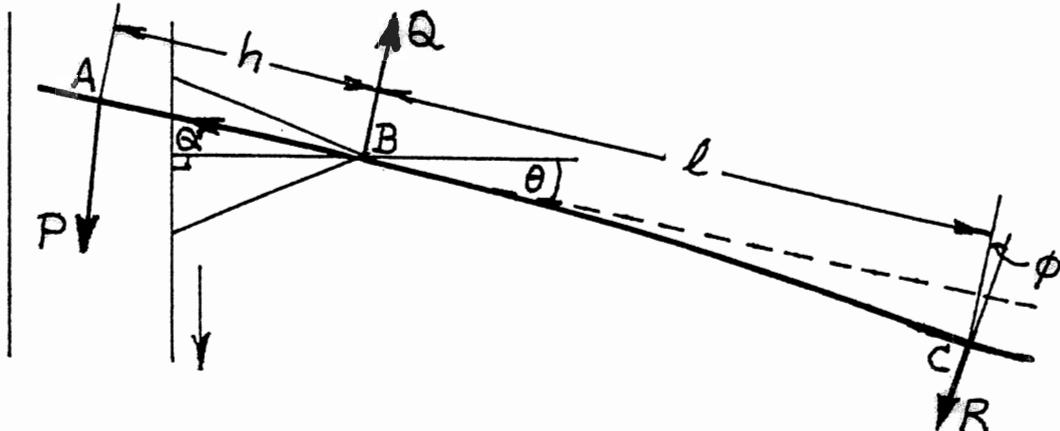
Figure 3: Plan view of oar bending during the power stroke

Let ϕ = change in angle at blade centre due to load.

The force R of the water on the blade then has the direction shown in the figure. The value of Q may reasonably be assumed to be the same as for a rigid oar, and this is the appropriate assumption for a comparison between the performances of rigid and flexible oars. Because the degree of flexing is small, the value of ℓ is virtually the same as for a rigid oar.

The oar is very stiff in the region AB because of its greater diameter there and because of the wooden handle which is inserted for some distance into the hollow fibre-glass shaft. This causes the oar to behave virtually as if cantilevered outboard of the point B.

Goodman [2] gives the slope at the free end of a cantilevered beam (in the notation of Figure 3) as

$$\tan \phi = (R \cos \phi)\, \ell^2 / (2EI).$$

For the present, attention will be restricted to the case of a medium oar. On using (3) and the typical value ℓ = 2.33m, the last equation gives

$$\tan \phi_m = 3.713 \times 10^{-4} R_m \cos \phi_m, \tag{4}$$

where the suffix m is used on ϕ and R to indicate that these quantities are for a medium oar.

From Figure 3 it can be seen that taking moments about the point A gives (on using R_m for R and ϕ_m for ϕ)

$$(R_m \cos \phi_m)(\ell + h) - Qh = 0.$$

Hence, with the aid of (1)

$$R_m \cos\phi_m = h(\ell + h)^{-1}Q = 0.3003Q. \tag{5}$$

When used in the right-hand side of (4), this yields

$$\tan\phi_m = 1.115 \times 10^{-4}Q.$$

The angle of flexing at the blade of a medium oar is therefore

$$\phi_m = \tan^{-1}(1.115 \times 10^{-4}Q). \tag{6}$$

When the oarlock force Q is known, (6) enables the value of ϕ_m to be calculated, after which the accompanying value of the force R_m on the blade can be found by using (5) in the form

$$R_m = 0.3003Q \sec\phi_m = R_r \sec\phi_m. \tag{7}$$

Equations (6) and (7) have been derived by according the blade of the oar the same degree of flexibility as the shaft, an assumption which will not greatly affect the accuracy of later conclusions because the blade length is small compared with ℓ.

4. A PARTICULAR ILLUSTRATIVE EXAMPLE

The Australian Institute of Sport has conducted experiments [3] which measure the oarlock forces Q shown in Figure 3 for various oarsmen using oars of medium flexibility in different classes of shells under race conditions. The values of Q were measured during the power stroke as functions of the time t and of the oar angle θ. From these data the blade flexing angle ϕ_m of a medium oar and the force R_m on its blade can be determined from equations (6) and (7), and the corresponding blade force R_r for a rigid oar from equation (1).

The process will be illustrated for the case of one particular oarsman in a coxless pair. The graphs of Q versus t and Q versus θ are shown in the Appendix. The values of Q at t = 0, 0.1, 0.2, …, 0.9 seconds were scaled off the second graph and transferred to the first graph to reveal the corresponding values of θ at each of these times. The data obtained are shown in the first three columns of Table 1. By convention, θ is taken as negative at the start of the power stroke.

The results of all these calculations are shown in Table 1 for each value of t. The values of ϕ_m are taken as negative to conform with the sign convention used for θ.

Table 1

Data for one particular oarsman using a medium oar at racing speed in a coxless pair

t (sec)	Q (N)	θ (deg)	R_r (N)	ϕ_m (deg)	R_m (N)	$\theta + \phi_m$ (deg)	F_r (N)	F_m (N)
0	72	−58.8	21.6	−0.5	21.6	−59.3	11.2	11.0
0.1	375	−52.6	112.6	−2.4	112.7	−55.0	68.4	64.6
0.2	810	−43.9	243.2	−5.2	244.2	−49.1	175.3	159.9
0.3	930	−34.0	279.3	−5.9	280.8	−39.9	231.5	215.1
0.4	1106	−23.3	332.1	−7.0	334.6	−30.3	305.0	288.7
0.5	1050	−10.2	315.3	−6.7	317.5	−16.9	310.3	303.8
0.6	897	4.5	269.4	−5.7	270.7	−1.2	268.5	270.7
0.7	623	17.3	187.1	−4.0	187.5	13.3	178.6	182.5
0.8	143	27.3	42.9	−0.9	42.9	26.4	38.2	38.5
0.9	0	34.2	0	0	0	34.2	0	0

From Table 1 it can be seen that, as expected, the values of R_m and ϕ_m follow those of the oarlock force Q; the magnitudes of all three variables peak at the same time $t = 0.4$ s after the start of the power stroke. The last two columns of the Table will be explained in the next Section.

5.   THE EFFECT OF OAR FLEXING ON RACE PERFORMANCE

From Figure 1 it can be seen that the forward component F_r of the force provided by one rigid oar is

$$F_r = R_r \cos \theta \qquad (8)$$

The suffix r on F is used to indicate that this forward force is for a rigid oar.

Similarly Figure 3 shows that the forward component F_m of the force provided by a medium flexible oar is

$$F_m = R_m \cos (\theta + \phi_m) \qquad (9)$$

The values of F_r and F_m in the case of the particular oarsman discussed in Section 4 are shown at each value of t in Table 1. These may be averaged over a complete stroke in the following way.

The experiment described in Section 4 was conducted at a pace of 28 strokes per minute, which corresponds to a duration of about 2.1 seconds for a complete stroke. The mean values $\bar{F}_r$ and $\bar{F}_m$ of F_r and F_m over a complete stroke are therefore given by

$$\overline{F}_r = \left(1/2.1\right)\int_0^{2.1} F_r dt, \quad \overline{F}_m = \left(1/2.1\right)\int_0^{2.1} F_m dt \qquad (10a, b)$$

The integrals here were evaluated numerically, using the values listed in Table 1 for $0 \le t \le 0.9$ together with zero values for F_r and F_m during the period $0.9 \le t \le 2.1$ occupied by the recovery phase. For this purpose, Simpson's 1/3 Rule was used in $0 \le t \le 0.8$, together with the trapezoidal rule over $0.8 \le t \le 0.9$. It was found that

$$\overline{F}_r = 75.55\,N, \quad \overline{F}_m = 73.17\,N$$

To save repeating all of the calculations for the second oarsman in the pair, it was assumed that the mean forward components of the force on the whole boat during a complete stroke could be found by doubling those found for the first oarsman. This gives, for the whole boat, for rigid and medium oars respectively,

$$2\overline{F}_r = 151.1\,N, \quad 2\overline{F}_m = 146.3\,N \qquad (11a,b)$$

As expected, the forward force component is greater for the rigid oar than for the medium flexible oar. To determine how this difference would affect the performance of the boat over a 2000 m race the following approximate method may be used.

At the speeds attained in the "steady state", the resistance D of the water is approximately proportional to the square of the boat speed, for all classes of racing shells; that is,

$$D \propto v^2$$

Hence in the "steady state" situation, the mean drag over a complete stroke is

$$\overline{D} = k\overline{v}^2 \qquad (12)$$

the constant k taking account of any small difference between $\overline{v^2}$ and $\overline{v}^2$.

In the "steady state", each mean forward force in (11a,b) is balanced by the drag $\overline{D}$ in (12), so that

$$k\overline{v}_r^2 = 151.1, \quad k\overline{v}_m^2 = 146.3 \qquad (13a,b)$$

where $\overline{v}_r$ and $\overline{v}_m$ are the mean boat speeds with rigid and medium flexible oars respectively. Hence

$$\overline{v}_r/\overline{v}_m = 1.016 \qquad (14)$$

There is therefore an increase in mean boat speed of 1.6% if rigid oars are used instead of medium oars. If this change applied over the whole of a 2000 m race, the improved boat position would be

$$2000 \times 0.016 = 32 \text{ m,}$$

or about 3 boat lengths for a pair. The improvement has not been shown to apply during the acceleration period near the start of a race, but it is reasonable to suppose that it would be comparable with that calculated for the "steady state" part of the race.

It is clear that the use of rigid oars would result in significantly better race times than are obtainable by flexible oars, in all classes of shells. Completely rigid oars cannot, of course, be made, but it is likely that more rigid oars than are now available will eventually be produced.

It would also be of interest to calculate how much superior to medium and soft oars are the stiff oars already available, and this will be done in the following Section.

6.   COMPARISON OF STIFF, MEDIUM AND SOFT OARS

For a stiff oar the deflection d in Figure 2 has the value 3.25 ± 0.25 cm = 0.0325 m [1]. In place of equation (3) for medium oars it is found that, for stiff oars,

$$EI = 8659 \text{ Nm}^2.$$

The force of the water on the blade, and the angle of flexing at the blade will be denoted by R_s and ϕ_s respectively, the suffix s being used to indicate a stiff oar. By the same argument as in Section 3, it is found that the analogues for stiff oars of equations (6) and (7) are

$$\phi_s = \tan^{-1}\left(9.414 \times 10^{-5} Q\right), \tag{15}$$

$$R_s = 0.3003 Q \sec \phi_s = R_r \sec \phi_s. \tag{16}$$

These equations enable the values of ϕ_s and R_s to be found for the times listed in Table 1. By the same method as described in Section 5, the mean forward component F_s of the force R_s for a stiff oar can then be found from the equation analogous to (9), namely

$$F_s = R_s \cos\left(\theta + \phi_s\right). \tag{17}$$

The mean value of F_s over a complete stroke is given by the equation analogous to (10b), namely

$$\overline{F}_s = \left(1/2.1\right)\int_0^{2.1} F_s \, dt. \tag{18}$$

The integral may be evaluated numerically as described in Section 5; it is found that, when doubled to give the forward force for a pair of oarsmen,

$$2\overline{F}_s = 147.1\text{N}.$$

To compare the performances of medium and stiff oars the process described in Section 5 may be used. In the "steady state", the relevant equations are

$$k\overline{v}_m^2 = 146.3, \quad k\overline{v}_s^2 = 147.1 \tag{13b,c}$$

where $\overline{v}_s$ is the mean boat speed with stiff oars. Hence

$$\overline{v}_s / \overline{v}_m = 1.00273. \tag{19}$$

The increase in mean boat speed if stiff oars are used instead of medium oars is 0.27%. Over a race of 2000 m this would improve the position of the boat by a distance of

$$2000 \times 0.0027 \text{m} = 5.4 \text{m},$$

or about half a boat length for a pair.

A similar analysis may be conducted for soft oars, for which the deflection depicted in Figure 2 is d = 4.45 ± 0.25 cm. It is found that, in place of (3),

$$EI = 6324 \text{ Nm}^2.$$

Using the suffix 0 to denote soft oar quantities, it is found that , in place of (6) and (7), one obtains

$$\phi_0 = \tan^{-1}\left(1.289 \times 10^{-4} Q\right),$$
$$R_0 = 0.3003 Q \sec \phi_0 = R_r \sec \phi_0.$$

These equations enable the soft oar equivalent of Table 1 to be produced, including a column showing the forward force component given by

$$F_0 = R_0 \cos\left(\theta + \phi_0\right).$$

The mean value of F_0 over a complete stroke is found by numerical integration to be

$$\overline{F_0} = 72.84 N, \quad \text{with} \quad 2\overline{F_0} = 145.7 N$$

The "steady state" equation governing the mean boat speed $\overline{v}_0$ with soft oars is

$$k\overline{v}_0^2 = 145.7 \tag{13d}$$

Then (13c) and (13d) yield

$$\overline{v}_s / \overline{v}_0 = 1.00479,$$

which shows that the performance of a pair over 2000 m would be improved by about 9.6 m (or about one boat length) if stiff oars were used instead of soft oars.

7.   SUMMARY AND CONCLUSIONS

An analysis of the amount of oar flexing that occurs during the power stroke under race conditions was made possible by applying beam deflection theory to an oar. The effect of oar flexing on rowing efficiency was also determined. Precise results were made possible by a knowledge of the oarlock forces occurring during the power stroke. These forces were measured during experiments performed by the Australian Institute of Sport under race conditions.

Attention was confined to the case of a particular coxless pair, but the conclusions reached will certainly be typical of all classes of boats. The main conclusion is that the stiffer an oar the greater its efficiency. For a pair in a race of length 2000 m it was calculated that its final position would be improved by the following distances:

4.1 m if medium oars used instead of soft oars;
5.4 m if stiff oars used instead of medium oars;
9.6 m if stiff oars used instead of soft oars.

If rigid oars existed they would improve the position of a pair over 2000 m by about 32 m (or 3 boat lengths) above that achievable by medium oars.

8.   ACKNOWLEDGEMENTS

REFERENCES
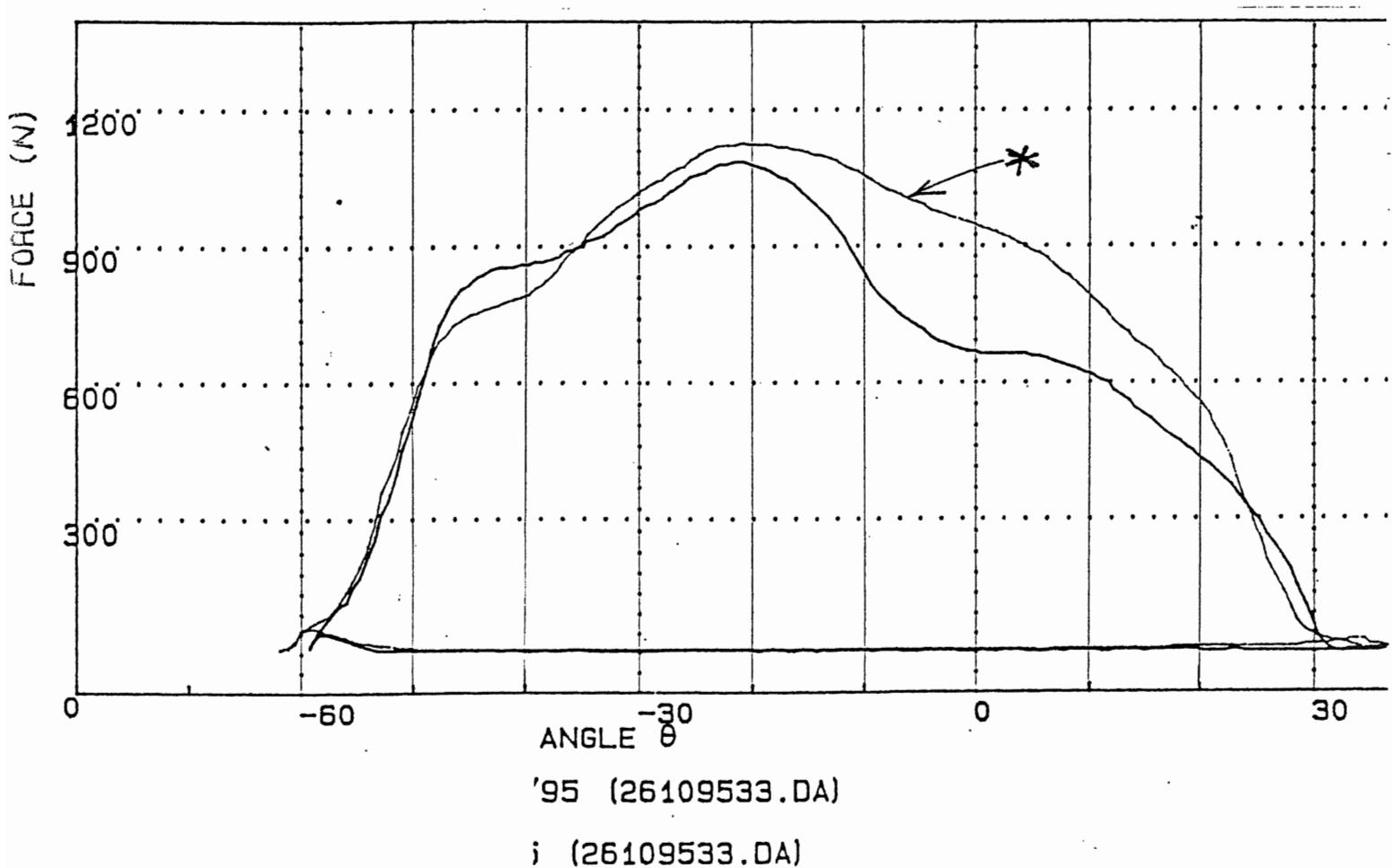
[1]   Advertising brochure, Concept II Oars Inc., Vermont, USA.

[2]   J. Goodman: Mechanics applied to engineering, Vol I, (Longmans, Green & Co, London, 1930)

[3]   Australian Sports Commission: AIS Rowing Telemetry System (Internal Report, 1995)

APPENDIX

Extract from AIS Rowing Telemetry System Report ([3]; reprinted with permission of the AIS)

ROWTEL V2.9   AIS ROWING TELEMETRY SYSTEM – Force/Angle Graph
HM2– Location: CANBERRA Date: 26/10/95
Start/end time (secs): 254 361 #strokes: 48   26/10/95

'95 (26109533.DA)

¡ (26109533.DA)
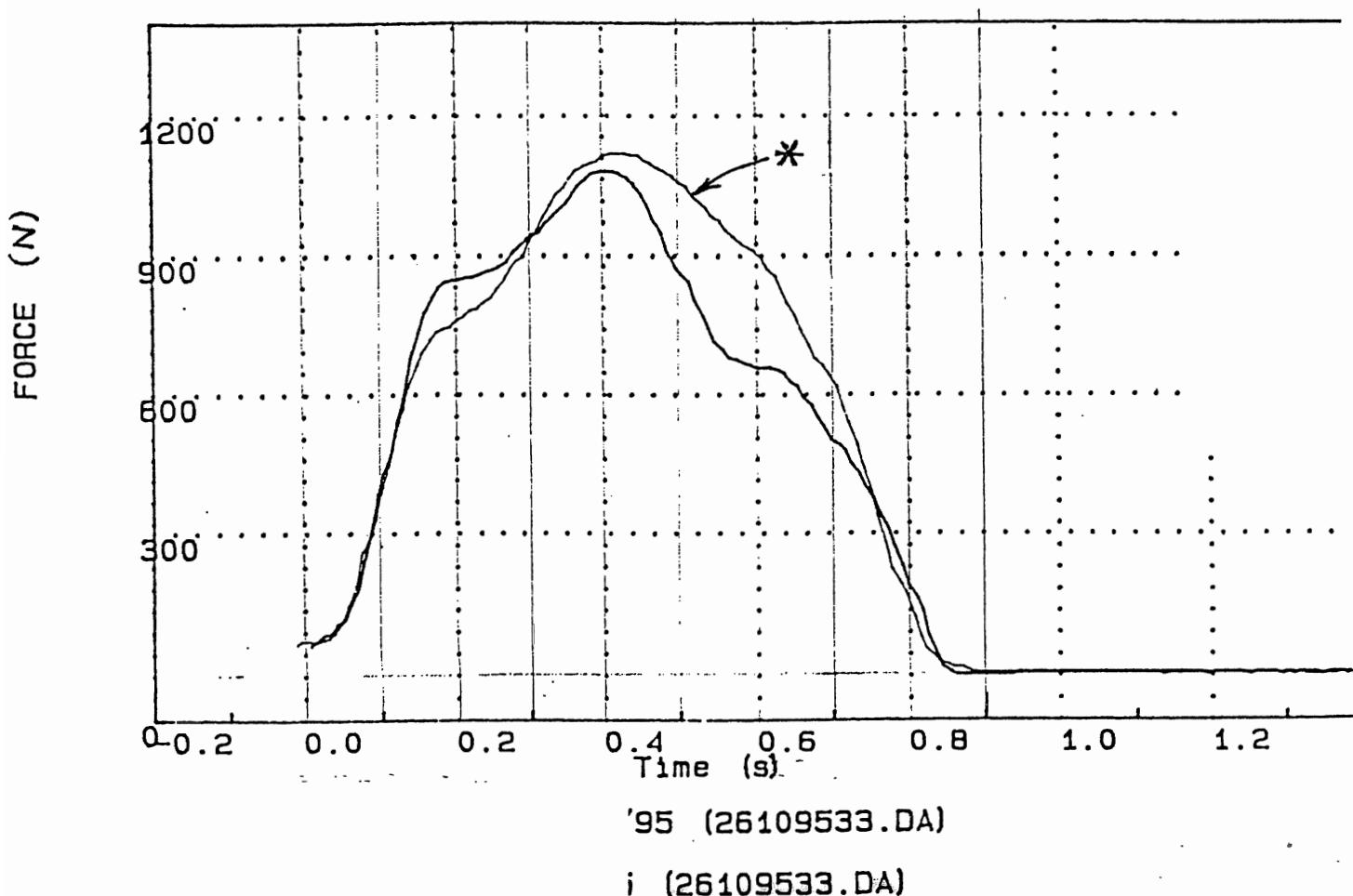
Measured oarlock force versus oar angle θ for oarsmen rowing in a coxless pair at 28 strokes per minute. The results are summarised in Table 1, Section 5, for one oarsman (*).

APPENDIX (CONTINUED)

ROWTEL V2.9   AIS ROWING TELEMETRY SYSTEM – Force/Time Graph
HM2– Location: CANBERRA Date: 26/10/95
Start/end time (secs): 254 361 #strokes: 48

'95 (26109533.DA)

i (26109533.DA)

Measured oarlock force versus time t for oarsmen rowing in a coxless pair at 28 strokes per minute. The results are summarised in Table 1, Section 5, for one oarsman (*).

A GEOMETRICAL ARGUMENT DEFINING AN ORDER RELATION FOR ALL-ROUND PERFORMANCES IN CRICKET

Iain Skinner[1]

Abstract

A simple geometrical argument shows that the geometric mean of runs scored and wickets taken is the appropriate number with which to measure the achievements of cricketing all-rounders.

1.   INTRODUCTION

Anyone who is involved with cricket knows that the compilation and comparison of statistics and records is an integral part of the game. Indeed, for many it is intrinsic to their enjoyment of the game. These numerical records, which purport to measure the quality of performances, pose endless questions for debate and discussion. Whereas, for example, the total runs scored or the batting average, as appropriate, are usually trusted to measure the best batting performances in various contexts, on some matters the popularly presented numbers are less than helpful, and require slightly more mathematical analysis before providing reliable indications of the relative merits of deeds. This paper addresses one such case by resolving the shortcoming of current methods' inability to assess all-rounders. In doing so, it also provides an instructive example of how mathematics can be used to clarify confusion.

Generally cricketers specialize at either batting or bowling, but a very talented few excel at both skills simultaneously. These are cricket's all-rounders. Contrasting with that of bowlers and batsmen/batswomen, though, the traditional consideration of all-rounders in cricket's records is, alas, cursory, confusing, inconsistent, and somewhat arbitrary. I believe this to be due to the lack of an accepted objective criterion for rating all-round performances, since people tend to be interested in records to make comparisons. There is, therefore, a clear need for cricket's statisticians to have a simple method to measure the records of all-rounders. If considering batting, for example, one can total runs, see who scores the most, and an order of merit naturally follows. Similarly, for bowlers there are natural orders: number of wickets taken, bowling average, strike-rate, etc. But how can one rank all-rounders and have a way of determining best-performed, because players both score runs and take wickets? Which performance, of those depicted in Figure 1 (or Figure 2), is the best all-round achievement? How can one appropriately combine runs and wickets into a single summary number that can then be ordered? Below, I show one

[1] School of Electrical Engineering, The University of New South Wales, Sydney NSW 2052
(email: I.Skinner@UNSW.edu.au)

way this can be done. A simple geometrical argument demonstrates that a suitable definition for a single measure of all-round merit is the geometric mean of runs scored r and wickets taken w. Arithmetic ordering then allows tabulation of orders of merit. To verify its utility, this method is used to determine rankings with three specific sets of data: test match series, Australian first-class seasons, and entire test careers. First, though, to motivate the need for applying some mathematics, the existing confusion in measuring the deeds and defining the achievements of cricketing all-rounders is illustrated.

2.   MOTIVATION

As a notational shorthand, define $\{r,w\}$ to be the compound achievement of scoring r runs and taking w wickets, e.g. $\{100,10\}$ means the combination of 100 runs and 10 wickets.

2.1   The Confusion

To see the absence of a widely accepted standard defining all-round merit and some of the confusion that ensues, let us compare several collections of records and note which all-round doubles are recognized and the methods employed to order them.
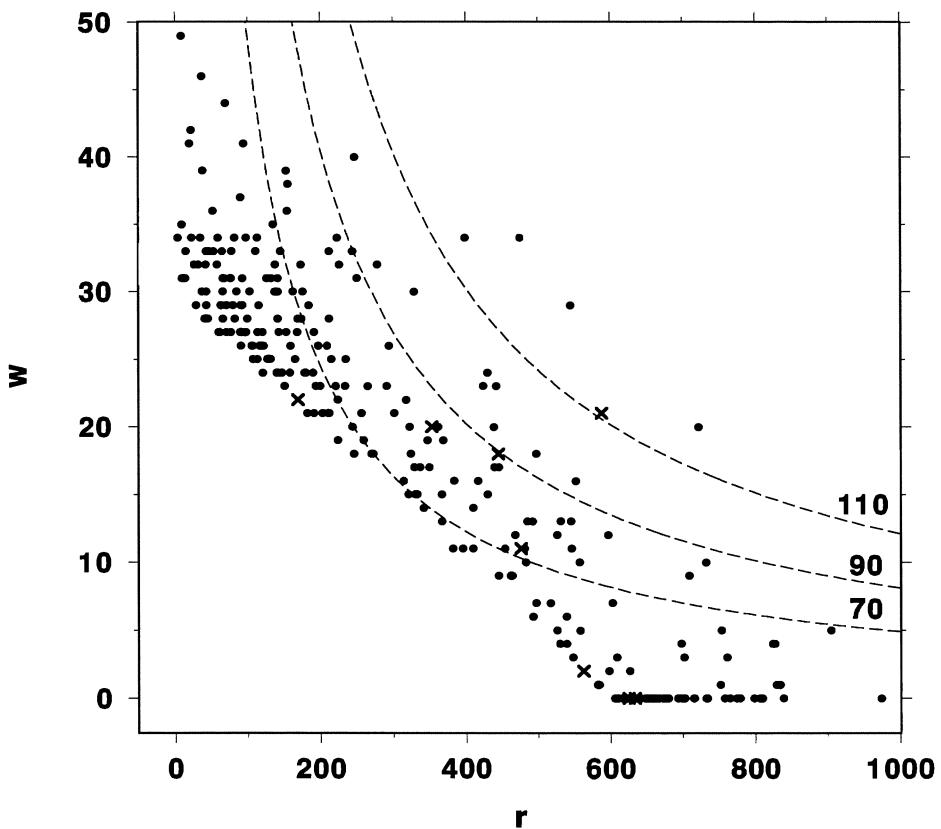
Figure 1: Outstanding achievements by players in test match series. These 266 dots show the most productive performances in terms of the combination of runs scored and wickets taken in a single test series for all men's tests up to May 1997. The 9 crosses denote performances not recognized as test series, though of equivalent standard and involving representative sides. Along the bottom row, corresponding to specialist batsmen who took no wickets, many dots merge together. The three hyperbolae (broken curves) correspond to all-round ratings (i.e., $\sqrt{rw}$) of 70, 90, and 110. Specific details pertaining to performances rating at least 90 are given in Table 1.

As a first example, consider performances in test match series. The best ones of these are illustrated in the scatter diagram of Figure 1. As worthy of special recognition, Frindall [1], Dawson and Wat [2] nominate{250,20} and list players chronologically; Gibb [3] has a chronological listing for {200,20}; Frith [4] includes both {200,20} and {300,15} (Why not also {220,19}, etc?) and lists alphabetically; Matthews [5] chooses {300,20} and, most peculiarly, orders by the number of wickets taken; but the definitive collection of cricketing records, namely the annual edition of *Wisden Cricketers' Almanack* [6], contains no corresponding table, and any recognition of all-rounders is completely absent from Dundas and Pollard [7]. Cricinfo [8] has two chronological listings: those who, in the same series, simultaneously either took most wickets and scored most runs or topped both the batting and bowling averages for their teams.

As a second example, consider cumulative career totals from test matches. The classic benchmark [1,2,3,6] is {1000,100}, associated with an alphabetic listing. Nemeruck and Meher-Homji [9], however, chooses to list players by the number of games required to attain that cumulative total. Frith [4] has two lists and two orders: those with {500,50} are ordered alphabetically, whereas those combining a batting average over 20 with 30 wickets are ordered by the ratio of batting to bowling average. In Matthews [5] are those with {2000,150}, ordered . . . apparently randomly!

In detailing the best all-round peformances from an English season, Matthews [5] this time ranks by the ratio of averages, and Wisden [6] details four different combinations: {3000,100}, {2000,200}, {2000,100}, and {1000,200}, all ordered chronologically.

2.2   The Problems

In the absence of a way of measuring runs and wickets simultaneously, neutral orders - chronological or alphabetical - are frequently favoured, but they do not provide any basis for comparisons. Furthermore, even these neutral orders must be accompanied by some nominated qualification (usually a minimum values of r and w) to restrict consideration to "genuine" all-rounders, and so, themselves, are not exempt from confusion and inconsistencies. Such minimum qualifications are, at best, sanctioned by tradition, but always present a fundamental problem because the precise choice reflects an individual assessment of what defines the all-rounder. For example, the classic {1000,100} means Merv Hughes is worth recognising as a test-match all-rounder, but Jack Gregory is not. Rankings that depend on an arbitrarily defined qualification are unsatisfactory and this is the first and universal problem of existing methods.

Ordering all-rounders by the number of games required to reach a specific, career milestone is appropriate for such milestones, but, despite its use for such in Nemeruck and Meher-Homji [9], provides no meaningful basis for comparing non-career (i.e., series, season or match) doubles. That Bruce Taylor performed the blue-ribbon match double (a century and five wickets in an innings) in his first test does not make it a better all-round performance than that of Ian Botham (his fifth) in his sixty-fourth match.

Likewise, the ratio of averages, i.e., *(batting ave)/(bowling ave)*, and other orderings based on summing functions of the averages [e.g., 10, 11] are flawed. They always allow one outstanding average to distort rankings. In particular, if you examine them, you will see that bowling averages are restricted to a much narrower band than are batting averages, so such methods invariably favour "batting" all-rounders. By almost any method based on averages, Michael Slater (one wicket) dominates all Australian test match all-rounders. It is to stop such nonsensical outcomes that minimum qualifications are specified.

In summary, we see that there are few widely recognized benchmarks for all-round achievements and no universally accepted order relations for comparison of the numerically defined achievements. Such orderings that exist are complemented by arbitrarily chosen minimum qualifications which, thereby, make the methods even more unsatisfactory.

3.   MODELLING

Before beginning, note that detailed analysis was limited to men's cricket, since its results are more readily available. There is no *a priori* reason why a similar argument and result does not hold in women's cricket. Also, analysis is dominated by records of test matches, which are more readily available than those of other first-class games.

3.1   The Runs-Wickets Equivalence

In wishing to combine runs and wickets into a single summary statistic, it is appealing to identify an equivalence between them. In this context, to see what an individual player can achieve, it is instructive to use a scatter diagram for $\{r,w\}$ in a specified collection of games. Figure 1 shows outstanding examples of the scoring of runs and taking of wickets in test cricket series. It is interesting to note the limit on an individual's achievements. It seems that the best one can ever do is score about $r=1000$ (Bradman's best was 974, though three times he exceeded 800) or $w=50$ (which is half the absolute maximum available in a five test series), but not both. Getting more of one is associated with less of the other. Understandably so; players tire doing one and so cannot perform as effectively at the other. In Figure 2 are $\{r,w\}$ for all players in the 1994-95, 1995-96 and 1996-97 Australian first class cricket seasons. Again there is a hint of an invisible line limiting the total productivity of any one player.

Both Figures 1 and 2 are consistent with the lines given by *constant* $= r+20w$, representing contours defining approximately equal densities of points on the scatter diagrams. Here, the constant indexes a specified level of achievement, with fewer players attaining beyond the higher values, and all points along the line representing an equivalent feat. Hence, from the scatter diagram, we may conclude that, as a first approximation, scoring 20 runs is equivalent to taking 1 wicket. This factor of 20 is consistent with, as examples, the recognized and tallied fundamental innings achievements of scoring 100 runs or taking 5 wickets, with the widespread acceptance [e.g. 1,2,3,4,7] of 25 wickets and 500 runs as the benchmarks for test series performances by bowlers and batsmen, respectively, and with the ratio of the chosen,

arbitrary qualifications for special recognition of run-making or wicket-taking during test careers [6,12]. That it is not unambiguously endorsed is shown by the 1997 *Wisden* having other factors elsewhere in matched tables with runs or wickets as qualifications (20 is used 5 times, but 25, 21, 17, and 13 also appear). I have only found one instance of this equivalence being specifically identified as such in a collection of cricket records, namely Miller [13], wherein the most prolific all-rounders in Sheffield Shield cricket were identified using 120<*(r+20w)/match*.

Patterns similar to Figures 1 and 2 resulted from plotting the series returns of all players who participated in various selections of post-War test match series. Further evidence of the balance between runs and wickets, this time over an extended period, was found by plotting, separately, the totals obtained by leading players over both first-class and test match careers. In the latter case there was one significnt, noticable difference: career batting totals tended to a higher maximum achievement. This is probably evidence that batting skill lasts longer than the more physically demanding bowling prowess. It is for this reason that the examples which are localised in time(namely, season or series) give a better idea of what an individual can accomplish as an all-rounder. All-round skill is about being able to perform with bat and ball at the same time.
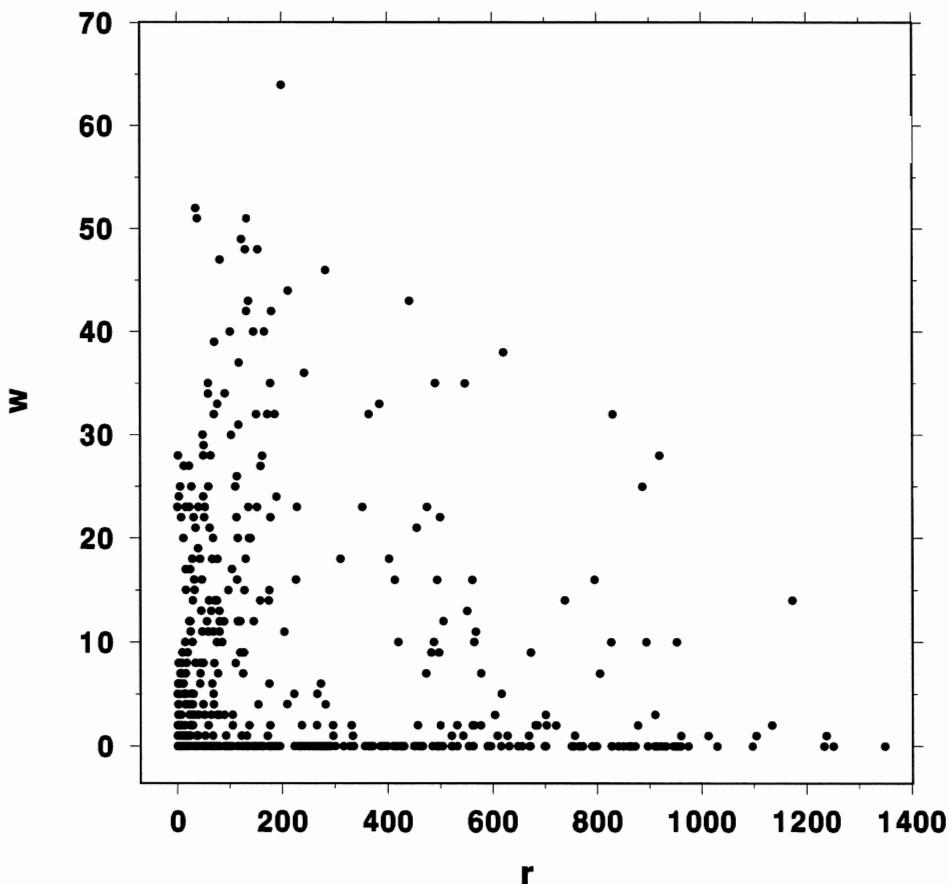
Figure 2: Players' achievements in recent Australian seasons. These 496 points show the wickets taken and runs scored by players in the 1994-95, 1995-96 and 1996-97 Australian first class seasons, with separate entries for each player for each season.

This apparent equivalence between 20 runs and 1 wicket means that equally meritorious deeds can be identified: one point per run, twenty per wicket, and add, as is done in Miller [13]. However, this equivalence between compound performances does not define all-round deeds of equal value. It is important to distinguish between these two ideas: scoring 600 runs (i.e., {600,0}) is assessed to be as good as taking 30 wickets (i.e. {0,30}), but, though both are equivalent to {300,15}, neither is as good an all-round effort. Indeed, neither is an all-round achievement at all. But {200,20} is; is it better than {300,15}? I think not. I claim that, of all equivalent achievements, the best all-round deed is that with an equal number of points each from batting and bowling, i.e. perfectly "balanced".

3.2   Modelling All-Round Merit

Having identified the best all-round feats from a set of eqivalent compound deeds, the next step is to compare between different levels of achievement.

To quantify all-rounders' performances, note the geometry illustrated in Figure 3. Pure batting and bowling deeds define the ends of lines of equivalent performances, i.e. points A and B, respectively. Let us assume there is an equivalence between M runs and 1 wicket, so that the line joining A and B is defined by $P = r + Mw$ with P constant. (Above I suggested that $M=20$, but in this analysis it remains general.) It follows that the equivalent, perfectly balanced all-round performance falls half-way between these batting and bowling achievements, namely at point C, with $r=M$ w, i.e., $\{P/2, P/2M\}$. To allow for variation away from this balance, it is appealing to divide the line AB into three equal segments, each defining a sector of skill: batting, all-round, and bowling. It is interesting to note that this sets the bowling boundary, point D, as the ratio $M/2$ runs to 1 wicket, which, if $M=20$ as argued above, has the value 10 which, in turn, is widely used to define benchmarks for all-round success (e.g., {1000,100} in a specified collection of matches). As a bowler scores more runs, his performance moves toward the boundary with all-rounders to become a player described by critics as "a bowler who makes useful runs". Likewise, there are those who bat and double as "useful change bowlers", and, with enough wickets, would be recognized as all-rounders. The geometry has provided a quantitative measure of qualitative terms.
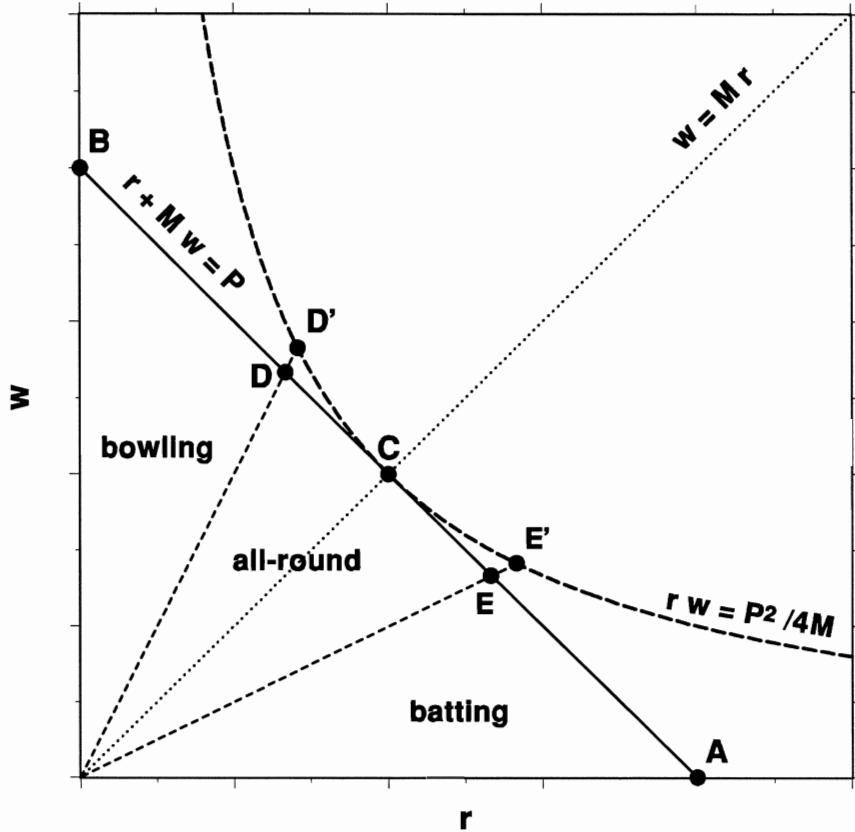
Figure 3: The geometry used to define all-round excellence.   Here r denotes runs and w wickets. The broken curve is the hyperbola defined by $w = P^2 / 4Mr$.

One might argue that anywhere on the relevant line is an equivalent all-round performance. However, those deeds at the ends (with either no wickets or no runs), though equivalent *achievements*, have no *all-round* value at all. Similarly, those at the mid-point of the line should rate higher as all-round performances than those off to either side. In Figure 3, the feat indicated by D is equivalent to those at A, B and C, but it is inferior to A and B, respectively, as a batting or bowling effort, and point C, being the middle, should outrank it as an all-round deed. But what about an off-centre point on a higher rating line? Return to the case examined above. What happens in comparing {220,20} with {300,15}? It rates higher as an achievement, but it is not clear whether it is a better all-round feat. Thus, this method of allocating points to measure all-round value has two severe limitations:

* The equivalence that sets $M=20$, while supported by persuasive evidence, is arbitrary. There is no intrinsic reason why it, or any other value for M, is valid. If it is not, rankings are invalid.

* There is no obvious way to compare the all-round merit of two differently rated compound achievements.

To resolve this second problem, we need to supplement the scatter diagrams with contours defining deeds of equivalent all-round skill. These curves need three properties.

* They should approach infinity along either axis so that no wickets or no runs means a zero rating as an all-round deed. This eliminates the need for arbitrary qualifications.

* They should be symmetric about the line $r = Mw$. This ensures that runs and wickets are treated equivalently, so that twice as many of one compensates for only half as many of the other.

* They should be tangential to the mid-point of the line $r+Mw = constant$. This means that the balanced feats are rated highest.

Given these considerations, a suitable, simple model for the merit of all-round performances is given by the hyperbolae defined by

$$r\,w = P^2 / 4M \qquad\qquad (1)$$

which is constant. All points on one of them represent equivalent levels of all-round achievement, with the precise level indexed by the constant. They show in Figure 3 that points D' and E', and not D and E, are all-round feats equivalent to C, preserving the ratios M/2 and 2M, respectively. The constant, in turn, can be determined for any particular combination of runs and wickets. That this model of all-round merit satisfies the first two requirements is trivial. That it satisfies the third, and, furthermore, does so for all values of M, requires a little algebra which I leave the reader to confirm.

Since doubling both r and w means that the all-round achievement becomes twice as meritorious, in using Equation (1), it follows that a rating of all-round performances can be defined by

$$rating = constant \sqrt{rw} \,. \qquad\qquad (2)$$

In other words, *the geometric mean of wickets and runs is a simple, natural measure for all-round merit*. The constant in Equation (2) is arbitrary; the following three examples show different ways it can be chosen usefully.

4.   EXAMPLES

The validity of using the geometric mean to combine the measures of batting and bowling skill into a single number which, in turn, measures all-round skill was determined by considering the sensibility of rankings generated in three different situations. There does not appear to be any systematic nonsense generated in any of the examples.

Consideration of the aggregate return from test match series gives the order of merit listed in Table 1. In this case, the constant was chosen to be 1. With the aid of the hyperbolae, the respective positions can be identified in Figure 1. Table 1 should be compared with those listing the most productive batting and bowling performances in test series (Wisden [6], pp 151, 162 respectively).

Table 1

Most prolific all-round performances in a test cricket series. The comparative merits of these performances are shown in Figure 1. The rating is $\sqrt{rw}$.

rating				M	runs	ave	wkts	ave
127	G.Giffen	Aus v Eng	1894-95	5	475	52.77	34	24.11
126	G.A.Faulkner	SA v Eng	1909-10	5	545	60.56	29	21.89
120	G.S.Sobers	WI in Eng	1966	5	722	103.14	20	27.25
116	I.T.Botham	Eng v Aus	1981	6	399	36.27	34	20.58
111 *	G.S.Sobers	RoW in Eng	1970	5	588	73.50	21	21.52
102	A.W.Greig	Eng in WI	1974	5	430	47.78	24	22.63
101	J.M.Gregory	Aus v Eng	1920-21	5	442	73.67	23	24.17
99	Imran Khan	Pak v Ind	1982-83	6	247	61.75	40	13.95
99	R.Benaud	Aus in SA	1957-58	5	329	54.88	30	21.93
99	G.S.Sobers	WI v Ind	1962	5	424	70.66	23	20.56
95	G.S.Sobers	WI in Aus	1968-69	5	497	49.70	18	40.72
94	Kapil Dev	Ind v Pak	1979-80	6	278	30.88	32	17.68
94	W.J.Edrich	Eng v SA	1947	5	552	110.40	16	23.12
94	K.R.Miller	Aus in WI	1955	5	439	73.16	20	32.00

* not recognized as an official test series

Similarly, an order was obtained for performances in an Australian season of first-class cricket and is given in Table 2. This complements lists detailing the most runs scored or wickets taken in an Australian season (Dundas and Pollard [7], pp 86, 112 respectively). This time the arbitrary constant was chosen as $\sqrt{5/3}$, meaning that {600,30} is rated as 1000 points. For comparison, the best bowling was C.T.B.Turner's 106 wickets and the most prolific batting Bradman's 1690 runs.

Table 2

Most prolific all-round performances in an Australian first-class season. The rating is $7.45\sqrt{rw}$.

rating			M	runs	ave	wkts	ave
2159	G.Giffen (SAus/Aus)	1894-95	11	902	50.11	93	22.54
2044	G.A.Faulkner (SAfr)	1910-11	14	1534	59.00	49	25.59
1804	R.B.Simpson (WAus/Aus)	1960-61	15	1541	64.21	38	31.79
1788	G.S.Sobers (SAus)	1962-63	9	1128	80.57	51	28.25
1736	M.H.Mankad (Ind)	1947-48	13	889	38.65	61	26.14
1688	G.S.Sobers (SAus)	1963-64	10	1006	62.87	51	26.56
1496	J.N.Crawford (MCC)	1907-08	16	610	26.52	66	25.19
1486	F.R.Foster (MCC)	1911-12	13	641	35.61	62	20.19
1475	L.C.Braund (MCC)	1907-08	16	783	35.59	50	32.88
1462	K.R.Miller (Vic/Aus)	1946-47	13	1202	75.13	32	23.06
1435	A.J.Richardson (SAus)	1925-26	11	904	50.22	41	19.92
1422	G.S.Sobers (WI)	1968-69	10	1011	67.40	36	31.11
1420	J.M.Gregory (NSW/Aus)	1920-21	12	844	60.28	43	22.37
1414	K.R.Miller (NSW/Aus)	1950-51	14	1332	78.35	27	28.22
1405	M.A.Noble (NSW/Aus)	1903-04	12	961	56.52	37	21.86

A similar method to the above, through Equation (2), could be used to evaluate who was the most prolific all-rounder in test cricket history. It is obvious, though, that more matches would mean more runs and wickets, so that there is need for a different quantity, one which fulfils the same role in comparison of all-rounders as averages do when batting achievements are compared. As a first suggestion, we could, of course, divide by the number of matches and the result would be a better guide to ability. This was done to produce Table 3: *rating = constant* $\sqrt{(r/match)(w/match)}$, where the constant was chosen to make the top rating 100. One could argue a need to increase the arbitrarily chosen qualification (i.e., the minimum number of games which only excluded three players, all of whom were bowlers), but even with this very weak condition, of those listed, only Godfrey Lawrence is not afforded all-rounder status by Jenkins [14], and he is in fifteenth position.

Table 3

Rankings of all-round performances for entire test careers (qualification: played at least 3 matches). The rating is $6.48\sqrt{rw}$ /matches. For comparison, the final column is batting ave / bowling ave.

rating			M	runs	ave	wkts	ave	ratio
100	H.Hordern (Aus)	1910-12	7	254	23.09	46	23.36	0.99
100	A.E.Trott (Aus/Eng)	1895-99	5	228	38.00	26	15.00	2.53
98	G.A.Faulkner (SA)	1906-24	25	1754	40.79	82	26.58	1.53
96	G.S.Sobers (WI)	1954-74	93	8032	57.78	235	34.43	1.68
90	I.T.Botham (Eng)	1977-92	102	5200	33.54	383	28.40	1.18
89	M.J.Procter (SA)	1967-70	7	226	25.11	41	15.02	1.67
88	T.L.Goddard (SA)	1955-70	41	2516	34.46	123	26.22	1.31
87	R.J.Hadlee (NZ)	1973-90	86	3124	27.16	431	22.29	1.22
86	Imran Khan (Pak)	1971-92	88	3807	37.69	362	22.81	1.65
86	M.H.Mankad (Ind)	1946-59	44	2109	31.47	162	32.33	0.97
86	G.J.Thompson (Eng)	1909-10	6	273	30.33	23	27.73	1.09
84	J.M.Gregory (Aus)	1920-28	24	1146	30.96	85	31.15	0.99
84	Amar Singh (Ind)	1932-36	7	292	22.46	28	30.64	0.73
83	K.R.Miller (Aus)	1946-56	55	2958	36.97	170	22.97	1.61
81	G.B.Lawrence (SA)	1961-62	5	141	17.62	28	18.28	0.96
80	A.W.Greig (Eng)	1972-77	58	3599	40.43	141	32.20	1.26

Note: In women's test cricket, B.(Betty) Wilson (Aus) earnt a rating of 143 with this system.

The method could be much improved. Rather than runs per match, the batting average, which gives the expected score in some sort of notional innings, is widely accepted as a reliable indication of batting ability. (For a discussion about improvements, see Kimber and Hansford [15].) To replace wickets per match (an artificially high value of which follows for those who do disproportionally more bowling), a more suitable definition of effectiveness for bowlers, giving some expected number of wickets in a notional performance, awaits development. A further refinement would be evaluation of the geometric mean on a match-by-match or season-by-season basis, and calculate a further mean from these values. This would remove from consideration players who excelled with either ball or bat, in turn, but not both simultaneously, but who, nevertheless, finished with the cumulative figures of a genuine all-rounder. I leave such closer examination of the careers of all-rounders to such others as compile the detailed cricketing records and statistics that I have used.

5.   CONCLUSION

Ultimately, the test of any application of a mathematical structure is whether it enhances understanding of the subject to which it is applied. Here, the question is whether the orders of merit generated are sensible ones. I believe they are. They have generated no systematic criticism from others. There are no performances recognized that are not noteworthy, and no players without pretence to all-round ability are ranked highly. As with all cricketing records, though, there will be endless debate about whether a mathematical method can ever generate records that correctly recognize noteworthy performances. Concerning the sport's statistics, the doyen of cricketing critics, Sir Neville Cardus [16] wrote, "We might as well add up the quavers and crotchets in Rossini's operas."

Like all modelling, that in cricket has shortcomings. The specific problem here is an intrinsic property of the statistics and records: very rarely do they consider the playing conditions, strength of opposition, amount of luck, etc. For example, runs are relatively easier to score now than when first-class cricket was first played in Australia. (Is this why tradition has recognized the ratio 10 as defining all-round merit?) A similar problem exists with the natural order relations defined for batting and bowling, but, nevertheless, ranking records are produced. The method described herein allows the process to be extended to afford some recognition, albeit imperfect, of all-round excellence.

REFERENCES

[1]   Frindall W.H., The Wisden Book of Cricket Records, 2nd edn (London, Queen Anne Press, 1986)

[2]   Dawson G. & Wat C., Test Cricket Lists (Balwyn (Vic), Five Mile Press, 1992)

[3]   Gibb J., Test Cricket Records from 1877 (London, Collins, 1979)

[4]   Frith D., England v Australia Test Match Records 1877-1985 (London, Willow Books, 1986)

[5]   Matthews P. (ed), The Guinness Encyclopaedia of International Sports Records and Results, 4th edn (Enfield (Eng), Guinness Publishing, 1995)

[6]   Wisden Cricketers' Almanack 1997 (Guildford (Eng): John Wisden & Co)

[7]   Dundas R. & Pollard J., Highest, Most and Best: Australian Cricket Statistics 1850-1990 (Sydney: Angus & Robertson, 1990)

[8]   "Cricinfo -- the Home of Cricket on the Internet", http://www-aus.cricinfo.org

[9]   Nemeruck M. & Meher-Homji K., Those mighty all-rounders, *World of Cricket Monthly*, 40 (July 1976)

[10]   "The Coopers & Lybrand Ratings of Test Cricketers", http://coopers.co.uk/coopers/cricketratings/index.html

[11] Derriman P., The Top 100 and the 1st XI (Sydney: Fairfax Library, 1987)

[12] Cashman R. & Weaver A., Wicket Women: Cricket & Women in Australia (Sydney: UNSW Press, 1991)

[13] Allan's Australian Cricket Annual (Busselton (Austr): Allan Miller)

[14] Jenkins C.M., World Cricketers: a Biographical Dictionary, (Oxford: Oxford University, 1996)

[15] Kimber A.C. & Hansford A.R., A statistical analysis of batting in cricket, *J R Statist Soc* A, vol 156, 443-455 (1993)

[16] Cardus N., Charles Macartney and George Gunn, Wisden Cricketers' Almanack 1959, 87 (1959)

ONE DAY CRICKET SCORING PROGRAM

Sean Innes[1] and Steve Sugden[2]

Abstract

This work has grown out of a final-semester project to create a cricket scoring program to run on a Windows-based laptop computer. A major design goal of the program was ease-of-use by a normal person who is a professional cricket scorer or just an avid cricket lover. It appears that there are or no similar programs available. The cricket scoring program has been designed for one day cricket matches and the project was created using Borland's 32-bit Delphi 3 toolset. The application is capable of running on any system running the Windows 95 or Windows NT environment, with sufficient speed and RAM. The cricket scoring program will allow users to enter information for a match, see batsman and bowler statistics, general statistics, as well as line and column graphs.

1.   INTRODUCTION

The basic aim of the project was to create a program which could be used to score a cricket match at a live game; thus, the design goal of ability to run on a laptop, with its relatively limited screen real estate, was constantly before us. From feedback from program testers, it has been noted that trying to use this cricket scoring program while watching a televised game is very difficult and sometimes impossible due to the fact that the television networks don't provide you with sufficient time to enter in player information. They also do not provide adequate details on where the ball was exactly hit on the field. This is because the camera work is sometimes focussed on the batsman running between wickets and it is therefore impossible to click on the appropriate field position.

As noted, the programming environment that was used in this project was Delphi 3 client/server. This language has very powerful components (Halogram [1]) and abilities to easily calculate any mathematical expressions involved in the application. Its graphing abilities also made drawing graphs very easy and less tedious than more traditional programming environments. All that was needed for the graphs was the appropriate information for each column or point, and after this information was provided to Delphi the appropriate graphs were automatically created by Delphi run-time system (Vivrette [2]).

---

[1]   School of Information Technology, Bond University, Gold Coast Qld 4229
        Email: sinnes@surf.its.bond.edu.au

[2]   School of Information Technology, Bond University, Gold Coast Qld 4229
        Email: stephen_sugden@bond.edu.au

2.   PROJECT OBJECTIVES

The program is designed for the cricket lover who would like to keep a personal record of information and statistics for a limited-over (one day) cricket game. This program can easily be taken to a cricket match and used on a laptop. The entire program will be *object-oriented* and almost entirely *ball-driven* that is, after every ball is bowled the user enters in the required information.

During the entire life cycle of this project the *clean room* approach was used. With this the *waterfall model* (Schach [3]) for software development was used and the following phases were completed during the product life-cycle. The phases include Specifications, Planning, Design, Implementation and Testing. The documentation also includes some of the test cases that were performed on the implemented code along with the expected results and the actual results. During the life-cycle, it was made sure that the product was robust, meaning that the program should not crash under non-standard operating conditions, would perform at a reasonable speed without taking up an excessive amount of resources, fulfils what the user wants, is reliable and satisfies the output specifications independent of its use of computing resources.

The information that is gathered from each ball in this application includes the following:

- The bowler

- The striker batsman and non-striker batsman

- The number of runs scored by the striker

- The direction of the shot based on a cricket field map

- The wickets or retiring details resulting from this ball

- The extras situations:

    ➢ No ball – requires extra ball to be bowled

    ➢ Wide – requires extra ball to be bowled

    ➢ Leg Bye

    ➢ Bye
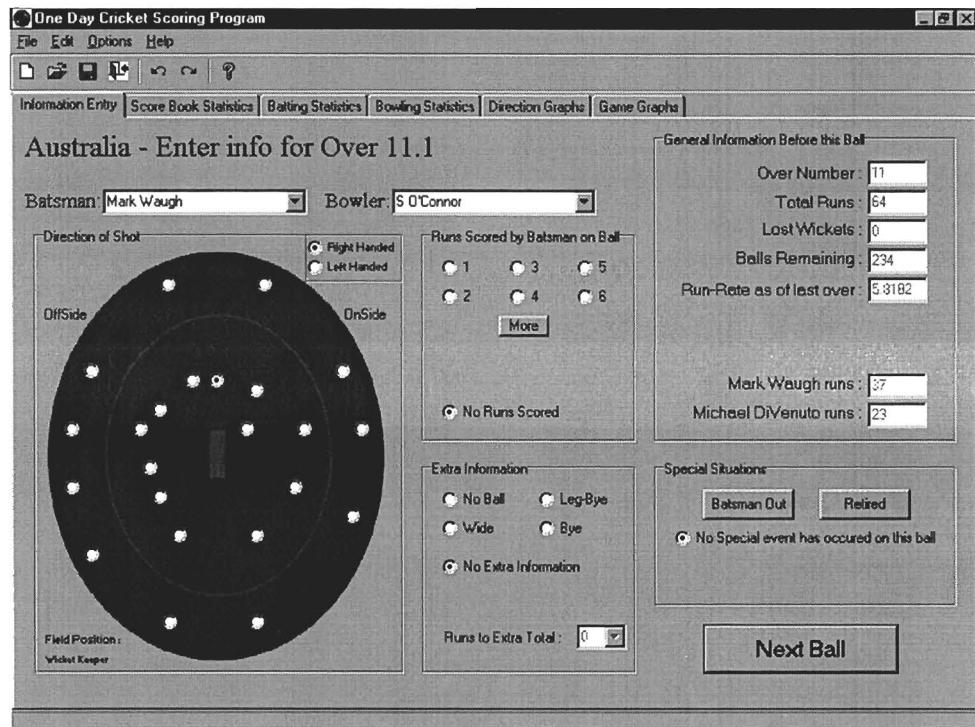
Figure 1: Screen Shot of Main Interface

A number of special situations may arise. These occur, for example, when the batsman may score runs from a no ball or wide or situations that require leg byes or byes to be added.

An *Undo* feature is also provided so the user can easily go back a step when entering in this information. This undo ability is multi-leveled and also allows users to perform a re-do later and move back to the current position in the match. The program also allows the user to easily save an entire match and later load this match back into the program.

Many batting and bowling statistics are available in this application. Together these statistics with the other general statistics make up the scorebook view that is very common to the person who is familiar with scoring on paper.

The program also produces a few different types of graphs. The first is a column graph showing the amount of shots hit to a particular direction for each cricket team. The second is a column chart for each innings showing the number of runs scored per over. Finally the third type of graph compares one team with the other in a line graph showing the run rate for each side. This line moves slowly in an upward direction as the number of overs increases.
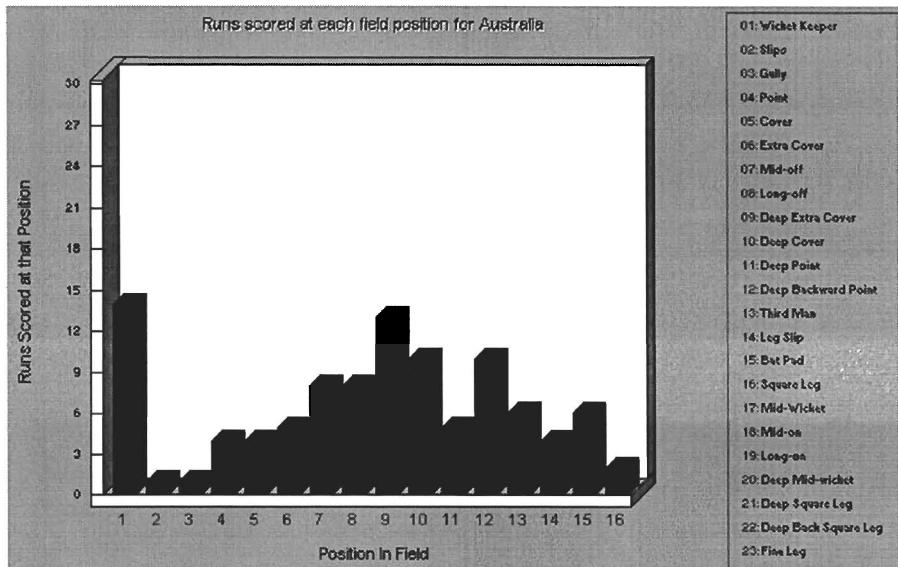
Figure 2: Fielding Positions Runs Scored Example
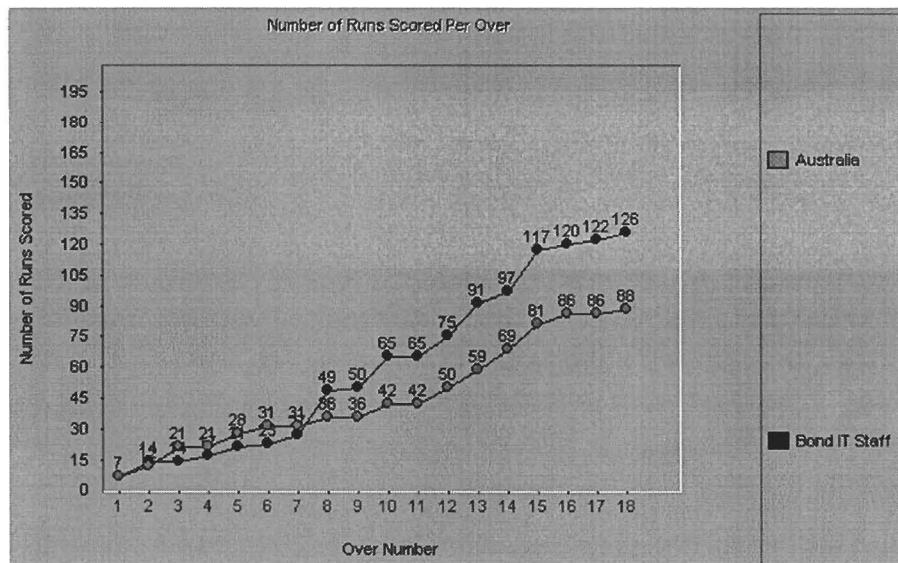
Figure 3: Line Graph Comparing Side A and Side B

3.   FUTURE CONSIDERATIONS

We discuss a number of items on our *wish list* of further enhancements. These are, of course, extra features or options that could be added to the product in another version. Foremost among these are making new statistics and new graphs available. It is expected that these will be easily done without any necessary modifications to the ball-driven structure of the program. In this manner, a user from Version 1 will be able to load a saved cricket match which was created in Version 1 yet still see the new statistics and new graphs which are created in Version 2 of the product. Apart from more statistics and more graphs the following options are high on the list for future versions.

3.1   Free Editing Option

This option was originally in the first specification document, but due to time considerations and difficulty of the task this option was removed from Version 1 of the product. It is still at present achievable through the use of the saved files. A cricket match can be saved and the user may decide to open this created text file and edit it with a simple text-editing program such as Windows Notepad or Wordpad. With this simple tool, users may easily change the values of the computer fields for each ball bowled. When the user later re-loads that cricket match, it will reflect the changes which were made during the manual editing of the saved file. However, saving a file and making manual modifications can be very time consuming and, due to problems which can arise in the middle of a cricket match, it would be very desirable to implement this in the next version.

3.2   Team Databases

Currently, with the commencement of a cricket match, the scorer must enter the complete team names and information every time. With a database of teams, one could simply have the team already defined in the database and when the team is loaded, simply select the appropriate team. When modifications to teams are required, these are simply saved to the database.

3.3   Rain-affected Cricket Matches

In future versions, we would like to implement the ability to still score and use the program in a rain-affected cricket match. We believe that implementing this feature for a rain-affected first innings is relatively straightforward, however implementing it for the second innings would be very difficult. This would still not be a very good option either at present as a complete overhaul of the rules for rain-affected cricket matches is currently in progress. It might be some time before a final decision has been made on how rain-affected matches are going to be played.

3.4   Web connectivity of players

Real-time updates of matches could be easily hooked into the Internet. This way someone at the game can in the future use the program that has a direct link to the Internet. Then ultimately the application can be used on any platform throughout the World Wide Web.

REFERENCES

[1]   Halogram Publishing, Third Party Add-on Tools for Delphi, http://www.halogram.com/ menus/Delphi.html

[2]   Vivrette Robert., *ChartFX Sample Code*, The Unofficial Newsletter of Delphi Users, Issue 7, August, 1996.

[3]   Schach Stephen R., *Classical and Object Oriented Software Engineering*, Times Mirror Higher Education Group: United States of America, 1996.

MATHEMATICAL MODELS THAT PREDICT PERFORMANCE DECLINE IN ELITE VETERAN ATHLETES 30-90 YEARS IN THE SPRINT, DISTANCE AND JUMP EVENTS

Ian Heazlewood[1] and Gavin Lackey[1]

Abstract

A number of researchers have measured strength, power, speed and aerobic power to plot the course of, as well as, to explain the decline in motor performance. Some previous research indicated these trends were linear and in other instances non-linear. However, a number of important research questions can be addressed. These are: Do the track and field events that are dependent upon the different energy-liberating mechanisms display similar declines and do both males and females follow a similar trend in the decline of motor performance with increasing age (30-90 years)? Do the performance declines fit specific mathematical models that reflect the aging process and which would allow accurate predictions of future performance with increasing age for veteran athletes? The method consisted of selecting state (NSW) and national (Australian) veteran championship performances and world records as the data sets. It was hypothesised that the short sprint (100m), high jump, long and triple jump events reflect the ATP-CP system; the 200 and 400m the anaerobic-glycolytic system; and the longer distances such as 1500m, 5000m and 10000m the aerobic system. It was also hypothesised that if a general aging process is expressed in the performances, then a similar mathematical function should display this trend across all the selected events and for both males and females. The mathematical function that best fitted the data (highest R2, lowest p-value and smallest residuals) was selected for the explanatory model. The results indicated for the majority of events and for both males and females that the best fit were quadratic and cubic mathematical functions with R2 values ranging in the mid .8's to mid .9's. These findings suggest that declines in motor performance are similar for the events and constructs studied. These findings are in contrast to some of the previous research and suggest a common aging process is expressed across all the events and energy systems.

"Time is seen as a prison that no one escapes; our bodies are biochemical machines that, like all machines, must run down," (Deepak Chopra [1]).

"Age can bring illness and dependency, a decline in the functioning of body and mind, of memory and of mobility, but age can also signify inner riches, experience that can be handed on and new creativity," (Professor Roman Herzog, President, Federal Republic of Germany [2]).

REVIEW OF LITERATURE

The effect of aging on human motor performance is currently a very popular research topic and reflects the current explosion in participation in veteran sports (Harridge [3]; Heazlewood [4]; [5]; Herzog [2]; Lexell [6]; Richter [7]; Suominen [8];

---

[1] Australian Catholic University, 40 Edwards Street, North Sydney   NSW   2060

Reaburn [9]; Reaburn, Logan and MacKinnon [10]). The research topics addressed include changes in the enzyme and energy substrate level, anatomy, functional anatomy, physiology, exercise physiology, biomechanics, competitive performance and sports psychology. Currently, entire international conferences (Huber [11]) are addressing and reporting research findings on healthy aging, activity and sports. A recurring theme within this research domain is, what is the rate of age-related decline beyond the third and fourth decade of life in the various biochemical function, physiological systems and anatomical structures, and are these declines reflected in gross motor performance or the total human organism motor responses?

Another important question is, do we evaluate the changes in a partistic manner, that is biochemistry, anatomy, physiology, exercise physiological, biomechanical and sports psychological constructs are researched in isolation and usually in a laboratory setting, or do we attempt to utilise an integrated approach (a holistic approach) to understand the changes in total motor behaviour of the human organism? That is, behaviour that is usually expressed in the field or the sports-competition setting? The partistic approach may or may not describe the decline in competitive performances that are thought to reflect the declines in the different factors, such as strength, power, speed and endurance.

The past research had addressed the declines in function from a partistic approach and in some cases attempted to relate decline in a specific function to decline in a competitive situation. The different factors that have been, and are identified to change with increasing age will be addressed under appropriate subheadings and will highlight the areas of research focus.

Aerobic Function

Reaburn ([9] and [12]) noted declines in maximal aerobic capacity, maximal heart rate and increases in adiposity (adiposity effects relative aerobic power) in males and females between 30 and 80 years. Linear models were fitted to each of these constructs, however it is clear that these changes based on the scatterplots provided by the author are non-linear in nature.

DeVries and Housh [13] plotted the loss of function of a number of physiological constructs from ages 25 to 65 years and noted that there were both increases, decreases and no changes in some of the constructs. In this review the researchers made no effort the fit the changes, and in most instances the declines, to any specific mathematical functions.

Specifically they noted:
- There were increases in body weight, heart volume, systolic blood pressure and diastolic blood pressure.
- Essentially no changes in heart rate and oxygen pulse at a constant power load (100W).
- There were declines in blood volume, total haemoglobin, vital capacity, forced expired volume, maximum oxygen uptake, maximum heart rate, maximum stroke volume, maximum cardiac output, maximum ventilation, maximum

respiratory frequency, maximum diffusion capacity, maximum blood lactate values and maximum strength.

Although the researchers did not fit specific mathematical functions to the data, the declines in function that were plotted appeared not to conform to a consistent function. The construct of strength has relevance to aerobic power as aerobic activities require repetitive force production by the striated voluntary muscles to complete the tasks of running, walking, swimming and cycling.

Astrand [14] examined the exercise physiology of the mature athlete and noted some functional and performance changes. Specifically, marathon times for men and women declined from age 35 to 80 plus years. Although Astrand [14] did not fit any mathematical functions to quantify the declines, the scatterplot appeared to conform to either quadratic or cubic functions.

Exercise physiological changes revealed declines in maximal oxygen uptake from years 55 to 70 for both males and females, significant reductions in muscle strength and muscle mass beyond 50 years of age. The reductions in muscle strength and muscle mass were attributed to the loss of motor neurones which resulted in muscle fibre degeneration. Although the loss of muscle strength occurred after 55 years there was no comment that this loss would translate to an accelerated decline in maximal aerobic power although the scatterplot of marathon performance (Astrand [14]) may actually reflect an accelerated and non-linear decline in maximal aerobic power past 50 years of age.

Thompson and Dorsey [15] examined the cardiac function of masters athletes and they believed that the decreases in endurance athletic performance reflected decreases in maximal aerobic power at a rate of approximately 1% per year past the third decade of life. They cite the decline in marathon performance with increasing age, 30 to 70 years, to support their argument. Although the declines in marathon performance are presented as though they are linear, an examination of the graph where increasing time of performance is plotted against increasing age, the line of best fit appears to be either a quadratic or a cubic mathematical function and not a linear fit.

Some of the decrease in maximal aerobic power was attributed to age-related decreases in maximal heart rate, maximal cardiac stroke volume and in maximal arterial-venous differences. Declines in maximal aerobic power in both trained and untrained people are attributed to the factors mentioned previously (Thompson & Dorsey [15]). Specifically, maximal heart rate decreases at a rate of 0.4 to 0.95 beats per minute per year and accounts for between 30% to 50% of the decreases in VO_2 max., decreases in maximal stroke volume were from 25% to 30% and declines in maximal arterial-venous differences were from 25% to 30%. These findings were based on average values from cross-sectional and longitudinal studies, and both Thompson and Dorsey ([15], p. 310) admit that "individual variation in exercise performance are generally greater than those produced by age alone." They believe that the declines are not reflective of mechanical efficiency (a biomechanical construct).

Saltin [16] examined the performance declines in physically-active endurance athletes, specifically orienteers. Some of the analyses were based on cross-sectional data and others on longitudinal data.

Saltin [16] discovered that:
- The decline in maximal oxygen uptake declined at approximately 0.73 ml kg^{-1} min^{-1} per year or 0.9% from 20 to 70 years (30%) and a figure consistent with other researchers.
- Cardiac output declines approximately 16%.
- Maximal heart rate declines approximately 11%.
- Stroke volume declines approximately 5%.
- Maximal oxygen pulse, a measure of stroke volume and arterial-venous oxygen differences, declined from approximately 0.42 to 0.32 ml kg^{-1} min^{-1} beat^{-1} from 26 to 66 years. The decline in arterial-venous oxygen difference (14%) is thought to be more important than the change in stroke volume in the decline of maximal oxygen pulse.

It is interesting to note that capillary density is similar in both young and older active men and does not appear to explain the decline in maximal oxygen uptake. In terms of actual performance declines of 35 to 40% have been observed in the sport of orienteering from the ages 20 to 25 through to 65 to 70 years. Similar declines have been observed in marathon running.

Saltin ([16], p.77) believes, " Thus the decline in top performance in the older athletes would seem to be age-related, something which cannot be overridden or compensated for by training" and reflects the age-related declines in maximal heart rate and increased stiffness (hardening of the arteries) of the arterial tree which are expressed as reduced maximal aerobic power.

Hagerman, Fielding, Fiatarone, Gault, Kirkendall, Ragg and Evans [17] examined high performance oarsmen over a 20 year period from 1972 to 1992 (1972 mean age = 23.8 years for the young adult males, 1992 mean age 44.2 years for the middle age males).

The major changes were:
- Decreased peak blood lactate (106%).
- Decrease in peak power, ventilation and aerobic power were similar (40%).
- Decrease in relative aerobic power (corrected for lean body mass) and absolute aerobic power (30%).
- Increases in body fat (12.3% to 15.6%).
- Significant but smaller changes in body weight, heart rate and oxygen pulse.

These findings by Hagerman et al. [17] indicated that the different constructs measured changed in their own unique manner with increasing age and may be reflected as different mathematical functions.

Drinkwater [18] examined an aging nexus in female athletes and indicated that the slope of the decline in maximal aerobic power was approximately the same for sedentary, active and highly-trained woman athletes, however the absolute values

for these groups are differently. Changes in isometric strength displayed a moderate decline up to age 50 years and then a rapid decline in strength after 50 years onwards. These findings indicate that loss of function with age in some constructs appears to be unique to that construct. This implies that different mathematical functions that may fit the data should be unique to specific constructs such as aerobic power, strength and range of motion.

Pulmonary Function

Makrides [19] when researching the effects of physical training in young and old people also noted the declines in maximum oxygen uptake with increasing age past the third decade. The declines were of the order of 0.5 to 1.0% per year which were attributed to reductions in maximum heart rate, cardiac output, declines in breathing capacity and pulmonary gas exchange function.

Hagberg, Yerg and Seals [20] examined the pulmonary function in young and older athletes and untrained men of similar age, and normalising the data for age and height, discovered declines in vital capacity, total lung capacity and forced expired volume (one second test) when older (mean = 66 years) sedentary people were compared to young (mean = 27 years) sedentary adults. No actual mathematical function describing the decline was generated for the data. It was interesting to note that age and height normalised data indicated that the masters' athletes (mean = 65 years) actually displayed higher values than young athletes (mean = 24 years) for vital capacity, total lung capacity and forced expired volume (one second test). The findings for the masters' athletes indicated that exercise might actually increase lung function in contrast to the majority physiological systems which display declines with increasing age past the fourth decade of life.

Flexibility

Drinkwater [18] observed that the loss of flexibility was most rapid in those arthroses (joints) that were least often used, whereas those arthroses most frequently used displayed minimal decline in range of motion (ROM). These findings indicate that the age-related declines in range of motion with increasing age might be more a reflection of lifestyle and physical activity patterns more than a true aging effect.

Muscle Function

Green [21] addressed the changes in aging human skeletal muscle and observed the following changes.

- Significant loss (25%) in muscle fibres from 30 to 72 years.
- No significant changes in muscle-fibre distribution.
- Type I fibres are relatively insensitive to age associated changes up 60-70 years, however Type II fibres may display more significant changes with age.
- The changes in capillarization are unclear and there might be sex differences where males displayed decreases whereas females did not.
- Energy metabolic changes were thought to be reasonably invariant as many enzymes, the aerobic substrate-end oxidation and anaerobic glycolysis did not appear to deteriorate with age.

- There appears to be a reduction in the volume fraction of mitochondria and mean mitochondrial volume from 16 to 76 years.
- In terms of muscle function it is suggested that declines in maximal voluntary contraction capabilities (MVC), both static and dynamic, begin in the thirties. Values between 26% to 38% decline in the quadriceps from 20 to 65 years have been reported with appreciable losses after 40 years of age.
- In terms of maximal static strength (dorsi and plantar flexors), 80-90% of the young adult capacity is retained into the seventh decade of life. Rapid losses in strength were only reported in the age group 80 to 100 years. These losses were explained in terms of loss of fibre number and reduction in muscle cross-sectional area. The magnitude of age-associated loss in performance functions appears to be dependent on the type of physical activity and the degree of muscle involvement.
- In gross motor activities the ability to produce energy aerobically does not appear to follow the trends in muscle function.

Meltzer [22] addressed the age dependence of Olympic weightlifting ability and noted the decline from 30 to 80 plus years was approximately 1-1.5% per year and was a similar decline in performance that was observed for masters' sprinters and jumpers. The declines were non-linear in nature and the second derivative of the curve (performance versus age) repeatedly changed sign. The best non-linear fit of the data was for a quadratic function (function $S = a + bx + cx^2$), however this fit may not reflect the changes in curvature (that expressed by the second derivative).

Anaerobic Metabolism

Reaburn and MacKinnon [23] found no significant difference between age groups for measures of maximal blood lactate concentration, time to reach maximal blood lactate concentration and the half recovery time to baseline concentrations of blood lactate after a maximal sprint swim. Reaburn and MacKinnon [23] concluded that the decline in sprint-swimming performance with age may be due to factors other than changes to the anaerobic-glycolytic capacity of the swimmer. The declines in anaerobic-power output were corroborated on a anaerobic bicycle ergometer test (maximal anaerobic capacity 30 second test) by Reaburn ([9] and [12]) and once again, a linear mathematical function was fitted to explain the trends in the data.

Martin [24] reviewed the effects of age and exercise on short term maximal performance based on the changes in the physiological systems. Both aerobic and anaerobic changes with age were addressed. Loss of muscle function can effect both anaerobic and aerobic processes as the muscles release the energy to complete the mechanical work, which is transduced by levers, pulleys and, wheel and axle arrangements within the human skeleton, and which results finally in movement.

The loss of maximal aerobic power may reflect the loss of muscle function in terms of loss of muscle fibres, reductions in mitochondrial enzymes (25 to 40% reductions in SDH, CS and b-HAD), reduced mitochondrial density, reductions in capillary density and capillary to fibre ratio.

The anaerobic processes indicated that no age-related changes in terms of glycolytic enzyme or high energy phosphates have been demonstrated. However, energy

substrates have shown some declines such as a 61% decrease in muscle glycogen when sedentary older men (65 years) are compared to younger sedentary men (24 years). As well, men and women, 52 to 79 years, have 5% less creatine phosphate in striated muscles when compared to younger adults. The neural activation, based on the current evidence, appears not to change markedly with increasing age, and some changes thought to be neural in nature may in fact represent changes in muscular and neuromuscular function (myoneural junction).

Although some of the enzymes and energy substrates display different age-related changes that may fit different mathematical functions, Martin [24] does not attempt to fit any specific mathematical functions to the age-related changes. However, Martin [24] did attempt to integrate the changes associated with aging into a schematic conceptual figure indicating the specific links, although the specific statistical weights do not indicate the relative importance of the different constructs.

Body Composition

Going, Williams, Lohman and Hewitt [25] in an extensive review of literature evaluated the relationship between aging, body composition and physical activity based on both cross-sectional and longitudinal studies. The cross-sectional studies indicated that when young male and female adults (20 years plus) were compared to elderly males and females (up to 80 years), the fat free mass (FFM) decreased by 15 to 30%. However, it must be noted that the rate and degree of FFM loss was influenced by gender, age, level of physical activity and individual variability. Longitudinal studies on the loss of FFM using the total body potassium method (TBK), although fewer longitudinal studies have been conducted than for cross-sectional studies, were approximately the same, that is 15 to 30% FFM loss.

The losses in FM were attributed to declines in muscle mass, total body water (TBW) and bone minerals. Specifically:

- The losses in FFM were attributed to loss of muscle mass which were 40% of peak values in the seventh decade when compared with values in the second decade of life.
- Total body water declines by 55% in men and 45% in women from early adulthood to 70 years and is thought to reflect an increase in the fat mass over this time. After age 70 the TBW may actually increase as body fat decreases in this age group.
- Bone mineral loss was at a rate of 0.3% per year for males after 50 to 60 years of age and 1.0% for females between 45 to 75 years although variability in mineral loss is related to the particular site in the human skeleton.

Determining the changes in fat mass are somewhat more problematical and Going et al. ([25], p.54) believe due to the methodological differences and measurement error related to different age groups, state that, "there are no exact descriptions of age related changes in total body fat in American and other populations."

Biomechanical Factors

Hamilton [26] investigated the changes in sprint-stride kinematics with age and found that stride length of athletes declined with age. The change is not linear with the greatest change occurring at age 60 years. No significant changes in stride period were found in the age groups below 90 years. The support time of runners was found to change significantly (p<.05) at 70 years, swing time showed no change and flight time was shown to decrease with age (p<.01). The changes in decreased stride length probably reflects the reduction in force production or strength which has been noted previously in this review of literature.

Mathematical Modeling

Mathematical modeling is a valid method of examining trends within data sets. The methodology has both theoretical and practical value as it permits more meaningful description of the data and prediction of trends in other the sets of data that are based on a specific mathematical model (Arya and Lander [27], Haefner [28]). According to Haefner [28] there are three primary uses for scientific and mathematical models and these are:

- Understanding - Of a various phenomena whether biological, physical, chemical or geological.
- Prediction - Of some future situation (extrapolation), past situation (extrapolation) or some other situation that is currently unknown (interpolation).
- Control - To constrain or manipulate a system to produce a desirable condition.

In the realm of sport and human movement, mathematical models have been utilised extensively to predict various movement constructs (de Mestre [29], de Mestre [30]). Specifically, the principles have been applied to predict triathlon times (Heazlewood & Burke [31]), bicycle motor cross times (Politi & Heazlewood [32]), future Olympic performances in track and field (Heazlewood & Lackey [33], Heazlewood & Lackey [34]), future Olympic performances in the different swim strokes (Heazlewood & Lackey [33]) and to integrate diverse findings generated from the different disciplines within the sports sciences (Heazlewood [35]).

The valuable insights provided by these models have identified what factors influence performance and how performance will be influenced if the factors are manipulated, what future level of performance will be required to reach Olympic finals in swimming and track and field, and provided a greater understanding of the complex interactions that are expressed as human movement.

Summary

A number of anatomical and physiological factors have been identified that are thought to decline with increasing age, such as cardiac output, stroke volume, maximal heart rate, the number of functioning muscle fibres, muscle-fibre size, bone-mineral density, mitochondrial number, mitochondrial density, muscle-glycogen stores, creatine-phosphate stores, FFM, body height, blood volume, total haemoglobin, pulmonary function, maximum diffusion capacity and maximum blood-lactate values.

These decline in the anatomical and physiological factors have been presented to explain the declines in the exercise physiological constructs of maximal aerobic power, maximal anaerobic power, maximal work output, endurance capacity, maximal strength and strength performance related to weightlifting.

Although some researchers have attempted to explain in mathematical terms the age-related changes in athletic performance, especially events based on aerobic power (marathon), the age-related declines were presented as essentially linear when the plots of the data were clearly non-linear functions in nature.

Other researchers, although noting the variability in the decline in a number of anatomical and physiological factors did not attempt to fit any mathematical functions to the data.

In some research, physiological factors, especially anaerobic enzymes did not change appreciably and yet sprint, jump and hurdles performances do decline appreciably with increasing age past 30 years of age as evidenced by state, national and world records for veteran-masters' athletes.

Few researchers have attempted to explain or understand human competitive gross motor performance as an expression of the changes that are occurring at the anatomical, physiological or biochemical levels as they relate to the process of aging. As well, few researchers have attempted to plot mathematically the declines in human competitive performance that may accurately reflect the aging process, although some researchers believe that the aging process is operating independently of the fitness level of the individual, especially for those motor activities that are dependent on cardiovascular-pulmonary fitness.

RESEARCH PROBLEMS

The research indicates that certain constructs such as strength and aerobic ability have been examined in some depth and specific trends have been delineated. However a number of research problems can be identified that to a degree can be answered.

1. The research conducted thus far indicates that certain anatomical structures, physiological systems, exercise physiological performance, biomechanical performance, biochemical enzyme concentrations and biochemical substrate utilisation changes to a degree in their own unique way that describes declines in human function with increasing age past the third decade of human life. An interesting but expected outcome of the declines in the various factors should be that the declines in gross motor performance of the different events in track and field should reflect different underlying constructs, such as sprints-100m, 200m (ATP-PC systems), longer sprints 400m, 800m (speed endurance-lactic acid system-anaerobic glycolysis), endurance events 1500, 5000, 10000 (aerobic power), throws shot, discus, javelin (ATP system) and the jumps (ATP-CP power systems). The performance declines, if they conform to a specific pattern, could be expressed as specific mathematical functions, such as linear, quadratic, cubic, exponential declines that represent how these underlying energy transductions or systems decline with age (30 years onwards).

2. If the performance decrements reflect homogenous changes generally, that is one mathematical model fits all changes then one or two mathematical models should adequately describe and predict the age-related changes.

3. If the performance decrements reflect heterogeneity of change, that is a number of mathematical models are required for each event, as the underlying constructs that predict performance change in their own unique manner.

4. The genders in some constructs displayed slightly different declines with increasing age, and if these differences are expressed in competitive performance, then the genders should display different trends related to age and performance. However, if the aging process is general to both genders then they should display identical trends in performance decrements across the different events.

5. Humans display different athletic abilities, however they all age and competitive performance ultimately declines. This concept is easily demonstrated when world records for the different veteran athletic age groups for the different events are examined. The issue is whether or not the athletes of different ability display similar or different trends in the decline in motor performance when expressed by competitive motor performance. That is, will they display similar or different mathematical functions that describe the performance declines?

RESEARCH QUESTIONS

1. Does one model fit all events?
2. Do the sexes (males-females) display different rates and trends in performance decline?
3. Do the models reflect the decline in the underlying energy-liberating systems; ATP-CP (sprints, jumps, throws), anaerobic glycolysis (400m, 800m) and aerobic power (1500m, 5000m, 10000m)?
4. Are there complex interactions related to specific events and specific sexes in terms of performance declines?
5. Do individual changes in strength and aerobic power when measured by laboratory assessment reflect what occurs in the field (competitive setting) for events that are thought to be dependent on these constructs?
6. Is the aging process independent of the athlete's ability?

RESEARCH HYPOTHESES

1. If the process of aging is independent of the training status of the athlete, the mathematical functions that represent the aging process should be common to all athletes and across all track and field events.

2. If the process of aging is independent of the ability level of the athlete, the mathematical functions that represent the aging process should be common to all athletes and across all track and field events.

3. If the process of aging is independent of the gender of the athletes, the mathematical functions that represent the aging process should be common to both genders and across all track and field events.

4. All events will fit a specific mathematical function irrespective of the events dependence on speed, power, strength or aerobic endurance and will reflect the general aging process.

RESEARCH METHODS

The data was collected from public domain data, that is, official competition results that result in state, national and world records. Specifically, national veteran and world veteran records (Niemi [36]) and data provided by the NSW Veterans' Athletics Club Inc. [37]. The events examined were the sprints (100m, 200m and 400m), middle distances (800m, 1500m), distances (3000m, 5000m and 10000m), and jumps (long, triple and high) were utilised. The problem with analysing the throwing events is the weights of the different implements decline with increasing age as well as the genders (males and females) have different weights, making comparisons between the different ages and genders difficult. A similar situation occurs with the hurdles where the heights are different based on gender and age group. Due to the problems with the throwing and hurdles events, these events were not included in the analysis.

All other events were performed under identical distances (sprint and running events) or rules (jumps). The variability in weather conditions were not considered as these factors (temperature, humidity and wind) are difficult to include in the mathematical models. The data for the best performance for each event and gender at the state (New South Wales), national (Australia) and world records served as the sets of data in this research. In the majority of cases the persons holding the respective records were different people.

The data were coded for event, age, gender, time, distance or height. The independent variable was age group and the dependent variable was the event time, distance or height achieved. The software program utilised was SPSSX 6.1.3 (Norusis [38]; SPSS Inc. [39]. The data was assessed against both linear and non-linear mathematical models. The potential mathematical functions (curve-estimation regression models) that could be fitted to the data were linear, logarithmic, inverse, quadratic, cubic, compound, power, sigmoidal (S), growth, exponential and logistic in nature.

General Method of Determining the Appropriate Regressions Models

A number of criteria were selected to assess which function best fitted the data. These were the R^2-value, the F-value, p-value and residuals (difference between the model and data points). The model that best conformed to these criteria was selected as the best fit. If two or more models were identical these were examined against the logical empirical trends displayed by the performance decrements. For example, it was unlikely that the older age groups 70 to 80 years would display performance increments with increasing age when compared to younger age groups.

The Coefficient of Determination

The coefficient of determination (R^2) is a measure of accuracy of the model used. A coefficient of determination of 1.00 indicates a perfectly fitting model where the

predicted values match the actual values for each independent variable (Norušis [38]; SPSS Inc. [39]). Where more than one model was able to be selected due to an equal R^2, the simplest model was used under the principle of parsimony, that is the avoidance of waste and following the simplest method (SPSS Inc. [39]).

Residuals

The residuals are the difference between the actual value and the predicted value for each case, using the regression equation (Norušis [38]; SPSS Inc. [39]). The smaller the residual, the better the fit of the model. For each model the residuals were generated by the SPSS program. A large number of positive residuals indicate that the prediction is an over-estimation or faster than the actual performance and a large number of negative residuals indicates an under-estimation or slower time than the actual performance.

Level of Significance

The level of significance, or p value, is a representation of the relationship between the model and the data. The smaller the p value, the higher the level of significance and the greater the relationship. A small p value indicates a small possibility of the closeness of the predicted values to the actual values.

Logical Acceptance Based on Extrapolations

The ability of the model to generate extrapolations that appear to be reasonable when compared to previous performances was also taken into consideration. When a model generated extrapolations that appear to be inconsistent with the actual results this model was discarded and the model with the next highest coefficient of determination was selected.

<div align="center">

RESULTS

</div>

In the majority of events the models (mathematical functions) that best fitted the data were the quadratic and cubic models. The models of best fit for each event, at state (NSW), national (Australian) and world record level are shown in Tables 1 and 2. The only event to be best described by a model other than cubic or quadratic was the Women's 800m at the national level. This event was best described by an inverse model (R^2 = 0.949, p=0.026). In this case the coefficients of determination for the cubic and quadratic models were 0.945 and so it is feasible that either of these models could be utilised to describe the trends of performance decline at the national level. The coefficients of determination ranged from 0.732 for the Women's 800m NSW record to 0.997 for the Men's 5000m NSW record.

In the majority of events the significance of the fit of each model, whether it was cubic, quadratic, linear, inverse, compound and so on, was high (in most instances p<0.001). The main exception to this was the Women's 800m at the state level where p = 0.0517 and was considered non-significant.

Table 1

Regression Models, Coefficients of Determination and Levels of Significance (p) for Men's New South Wales, Australian and World Records.

Event	R^2	State Model	p	R^2	National Model	p	R^2	World Model	p
100m	.994	Cubic	<.001	.989	Cubic	<.001	.951	Cubic	<.001
200m	.921	Cubic	<.001	.984	Cubic	<.001	.959	Cubic	<.001
400m	.994	Cubic	<.001	.953	Cubic	<.001	.970	Cubic	<.001
800m	.986	Cubic	<.001	.987	Cubic	<.001	.943	Cubic	<.001
1500m	.927	Cubic	<.001	.900	Cubic	.001	.958	Cubic	<.001
5000m	.963	Cubic	<.001	.984	Cubic	<.001	.945	Cubic	<.001
10000m	.980	Cubic	<.001	.976	Cubic	<.001	.953	Cubic	<.001
Long Jump	.972	Quadratic	<.001	.960	Quadratic	<.001			
High Jump	.965	Quadratic	<.001	.918	Cubic	<.001			
Pole Vault	.928	Quadratic	.005	.926	Cubic	.001	.989	Cubic	<.001
Triple Jump	.608	Cubic	.038	.919	Quadratic	<.001	.981	Cubic	<.001

Table 2

Regression Models, Coefficients of Determination and Levels of Significance (p) for Women's New South Wales, Australian and World Records.

Event	R^2	State Model	p	R^2	National Model	p	R^2	World Model	p
100m	.990	Cubic	<.001	.980	Cubic	<.001	.949	Cubic	<.001
200m	.961	Cubic	<.001	.966	Quadratic	.001	.962	Cubic	<.001
400m	.981	Quadratic	<.001	.997	Cubic	<.001	.901	Cubic	<.001
800m	.732	Cubic	.517	.949	Inverse	.026	.977	Cubic	<.001
1500m	.996	Quadratic	.004	.972	Quadratic	.028	.992	Cubic	<.001
5000m	.997	Cubic	.058	.996	Cubic	.060			
10000m							.902	Cubic	<.001
Long Jump	.974	Quadratic	<.001	.966	Quadratic	<.001	.949	Cubic	<.001
High Jump	.965	Cubic	.035	.999	Cubic	.001	.974	Cubic	<.001
Pole Vault							.976	Cubic	<.001
Triple Jump	.876	Quadratic	.002	.969	Quadratic	<.001	.968	Cubic	<.001

Tables 3 and 4 show the regression equations derived for selected events at state, national and world record level.

Table 3

Regression Models and Equations for Selected Men's Events.

Record	Model	Equation
Men's 100m NSW	Cubic	$y = 20.95 - .63age + .01age^2 - 6.0*10^{-5}age^3$
Men's 100m Australian	Cubic	$y = 12.65 - .06age - 5.0*10^{-5}age^2 + 1.7*10^{-5}age^3$
Men's 100m World	Cubic	$y = 12.31 - .002age^2 + 3.8*10^{-5}age^3$
Men's 400m NSW	Cubic	$y = -10.95 + 4.15age - .096age^2 + .0008age^3$
Men's 400m Australian	Cubic	$y = 53.64 - .01age^2 + .0002age^3$
Men's 400m World	Cubic	$y = 65.05 - .02age^2 + .0003age^3$
1500m NSW	Cubic	$y = 352.63 - .16age^2 + .002age^3$
1500m Australian	Cubic	$y = 348.82 - .16age^2 + .002age^3$
1500m World	Cubic	$y = 288.11 - .09age^2 + .001age^3$
Long Jump NSW	Quadratic	$y = 7.4 + .002age - .001age^2$
Long Jump Australian	Quadratic	$y = 5.99 - 7.0*10^{-5}.09age^2 - 6.0*10^{-6}age^3$
Long Jump		
High Jump NSW	Quadratic	$y = 2.49 - .017age - 1.0*10^{-5}age^2$
High Jump Australia	Cubic	$y = 2.51 - .02age - 3.7*10^{-7}age^3$
High Jump World		

Table 4

Regression Models and Equations for Selected Women's Events.

Record	Model	Equation
100m NSW	Cubic	$y = 12.31 - .08age - .002age^2$
100m Australian	Cubic	$y = 13.27 - .003age^2 + 4.8*10^{-5}age^3$
100m World	Cubic	$y = 10.8 - .0001age^2 + 1.7*10^{-5}age^3$
400m NSW	Quadratic	$y = 112.32 - 3.19age - .04age^2$
400m Australian	Cubic	$y = 70.93 - .03age^2 + .0005age^3$
400m World	Cubic	$y = 105.61 - .063age^2 + .001age^3$
1500m NSW	Quadratic	$y = 422.98 - 9.25age + .15age^2$
1500m Australian	Quadratic	$y = 494.45 - 12.45age + .17age^2$
1500m World	Cubic	$y = 210.18 - 1.33age + .06age^2$
High Jump NSW	Cubic	$y = 1.4 + .0006age^2 + 1.0*10^{-5}age^3$
High Jump Australian	Cubic	$y = 1.67 + .0001age - 3.0*10^{-6}age^3$
High Jump World		
Long Jump NSW	Quadratic	$y = 2.12 + .21age - .003age^2$
Long Jump Australia	Quadratic	$y = 1.17 + .26age - .003age^2$
Long Jump World	Quadratic	

Figures 1 to 4 show the predicted values for each gender according to the model of best fit. The decline in performance for each event is readily seen in the figures, as is the similarity, between the genders of the curves of best fit. Figure 3 which shows the models for the 1500m shows interpolated values for the Women's 30 World record to be faster than that of the Men at the same age. The actual records do not reflect this.
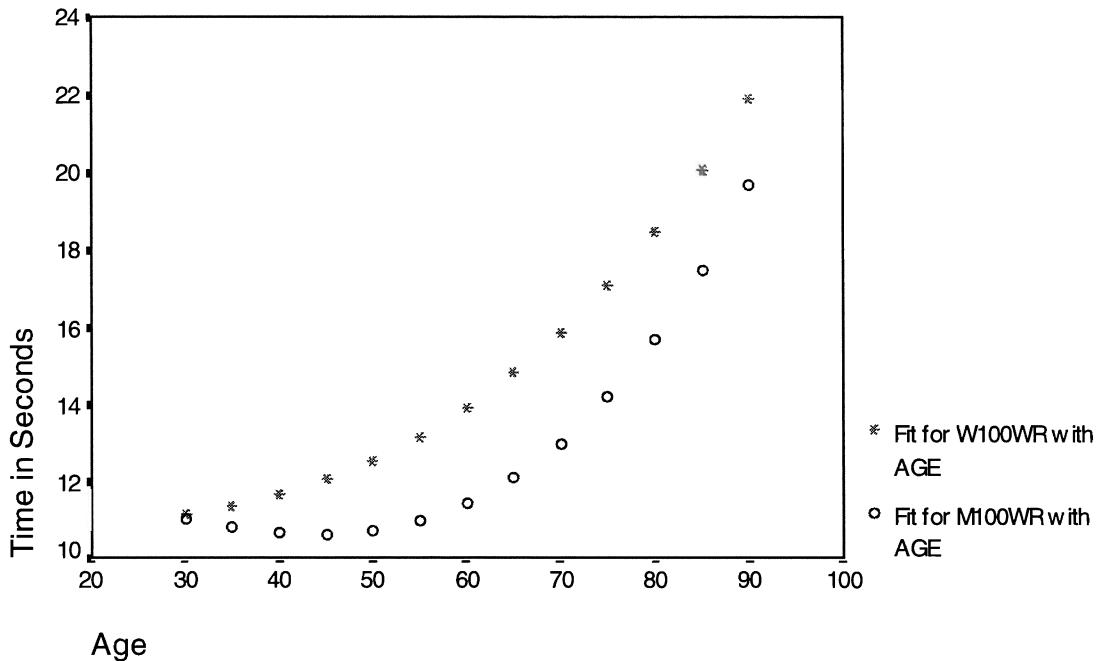
Figure 1: Cubic Models Representing the Men's and Women's 100m World Records.
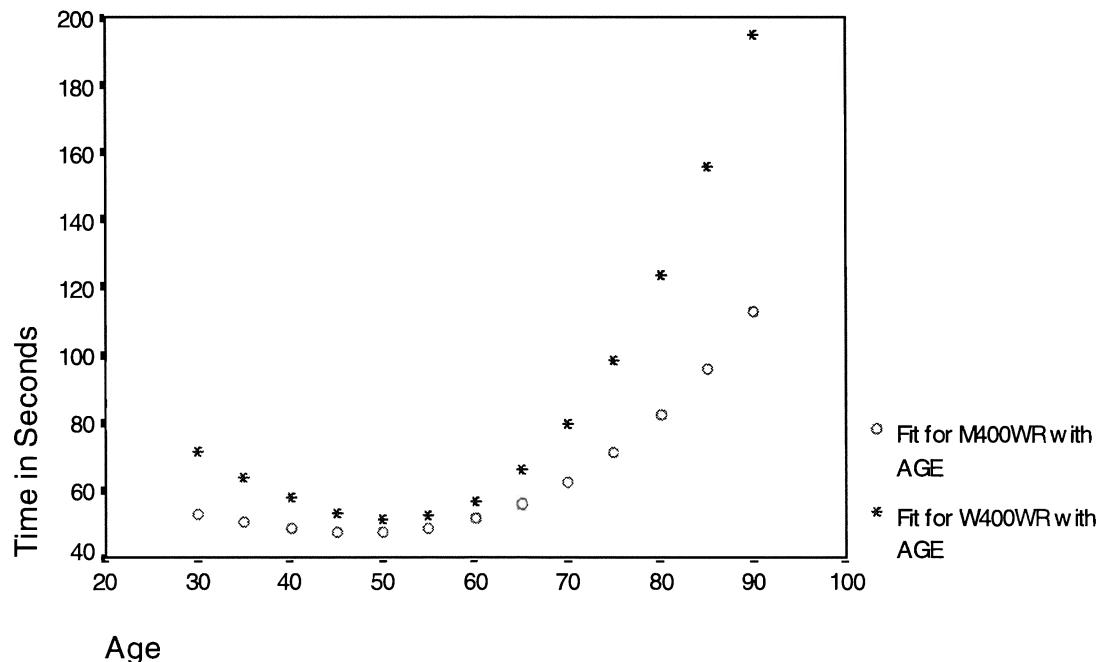
Figure 2: Cubic Models Representing the Men's and Women's 400m World Records.
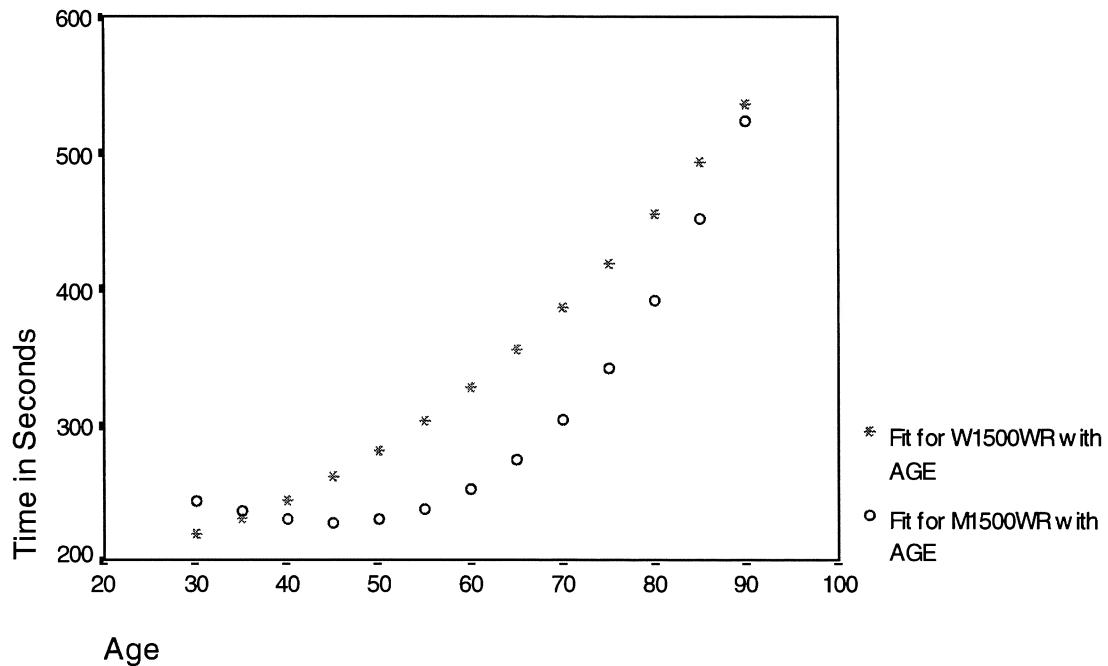
Figure 3: Cubic Models Representing the Men's and Women's 1500m World Records.
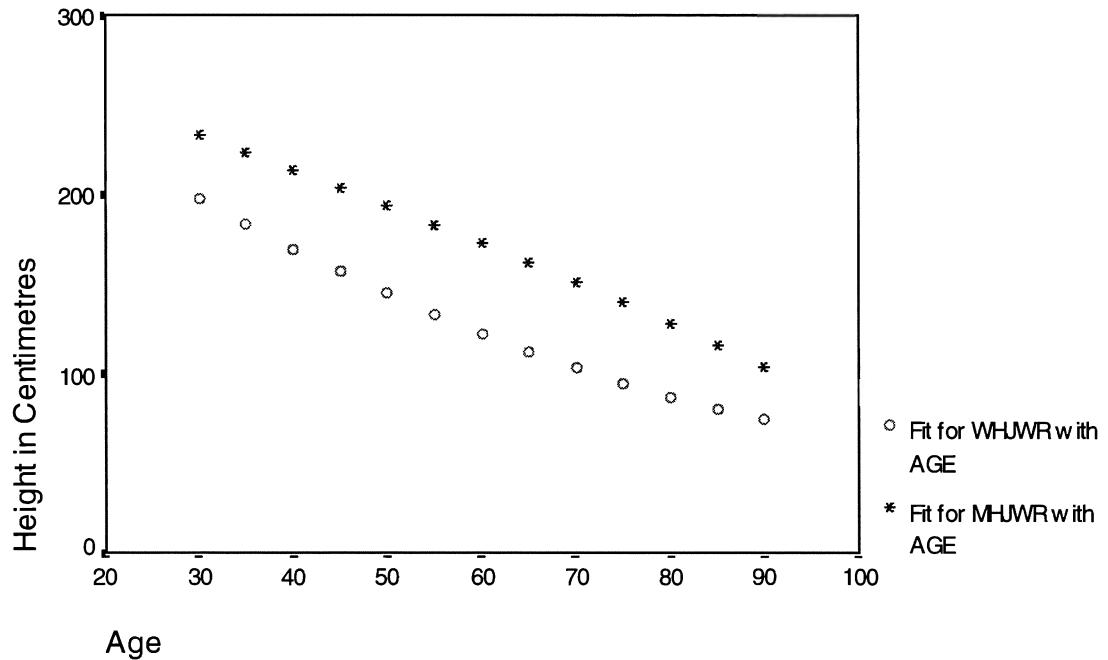
Figure 4: Cubic Models Representing the Men's and Women's High Jump World Records.

Figures 5 and 6 show the trends of decline, using cubic models, with age at each of the State, Australian and World record levels for the Men's and Women's 100m. The curve for the Men's NSW record shows an apparent anomaly at the 80 year age, where the interpolated time is faster than those for the national and world records. In the Women's 100m a similar occurrence appears at the 80 years group where the NSW record is predicted to be faster than the Australian record. These are not possible occurrences based on the current actual data.
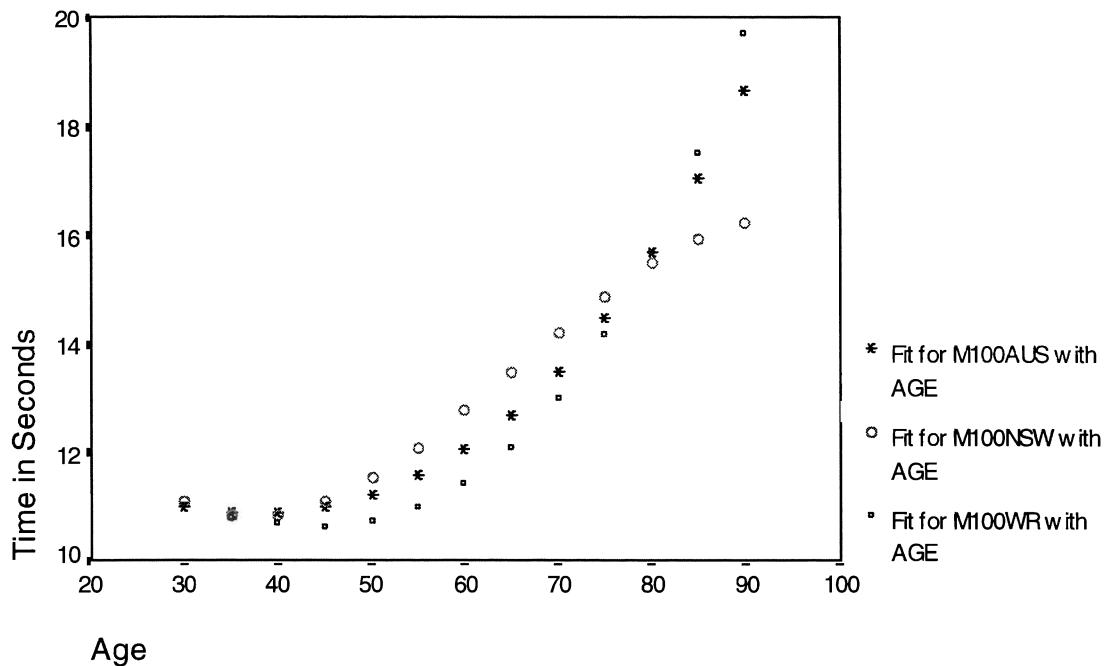
Figure 5: Cubic Models Representing the Men's 100 Records at State, National and World Level.
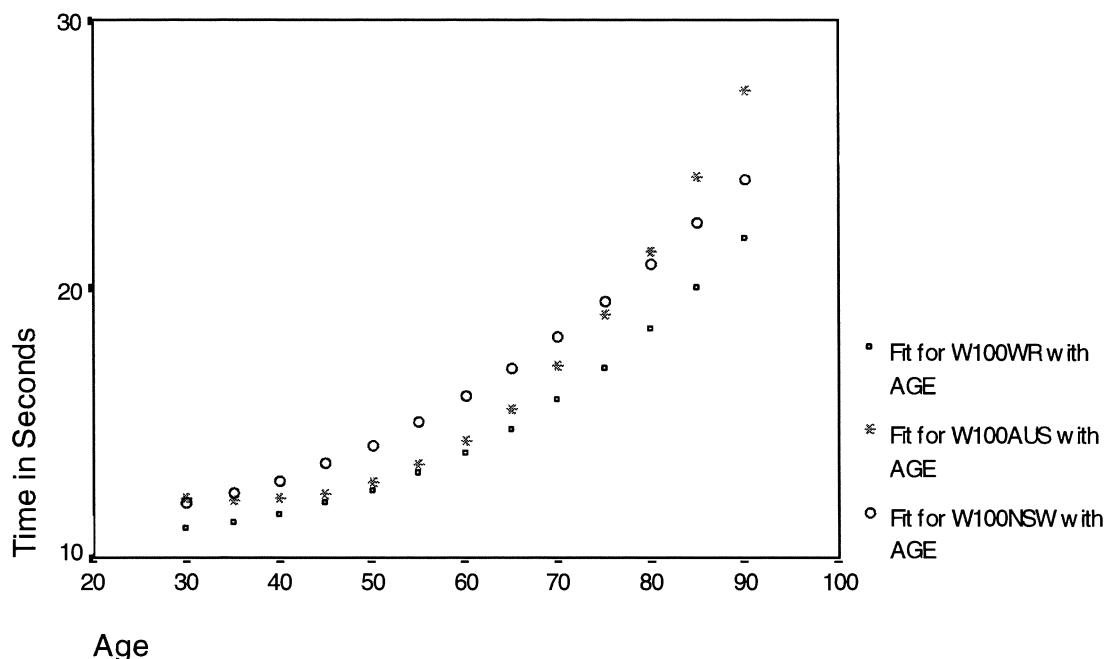
Figure 6: Cubic Models Representing the Women's 100m records at State, National and World Level.

Figures 7 and 8 show the relative fits at each level for the Men's and Women's 400m record. For both genders the predicted world record at the 30 year age are slower than those for the NSW and Australian records. At the 90 year age group the interpolation for the Australian record are faster than those for the world record.
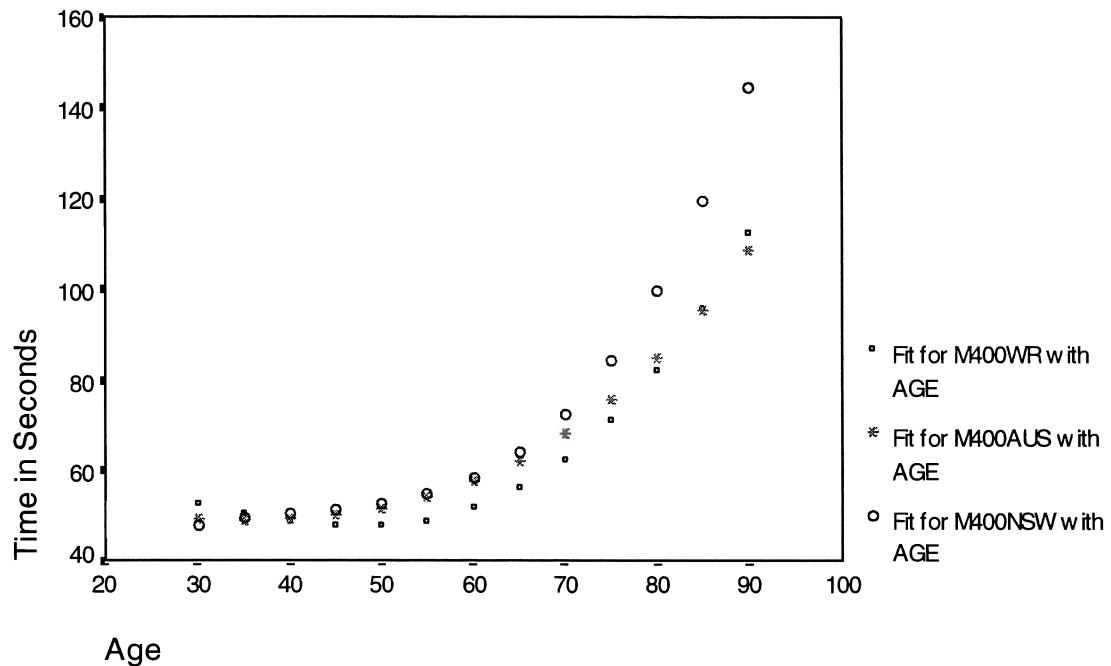
Figure 7: Cubic Models Representing Men's 400m Records at State, National and World Levels.
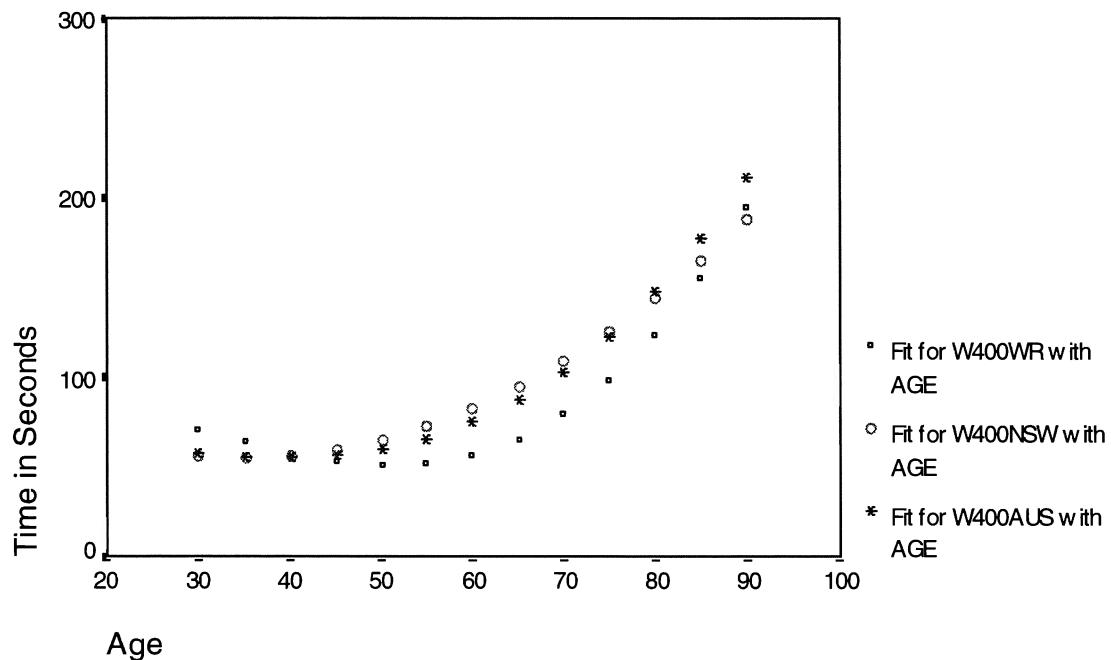
Figure 8: Cubic Models for Women's 400m records at National and World Level, and a Quadratic Model Representing the Women's NSW 400m Record.

The Men's and Women's 1500m records are represented in Figures 9 and 10. In these figures there does not appear to be any deviation from the expected relative records.
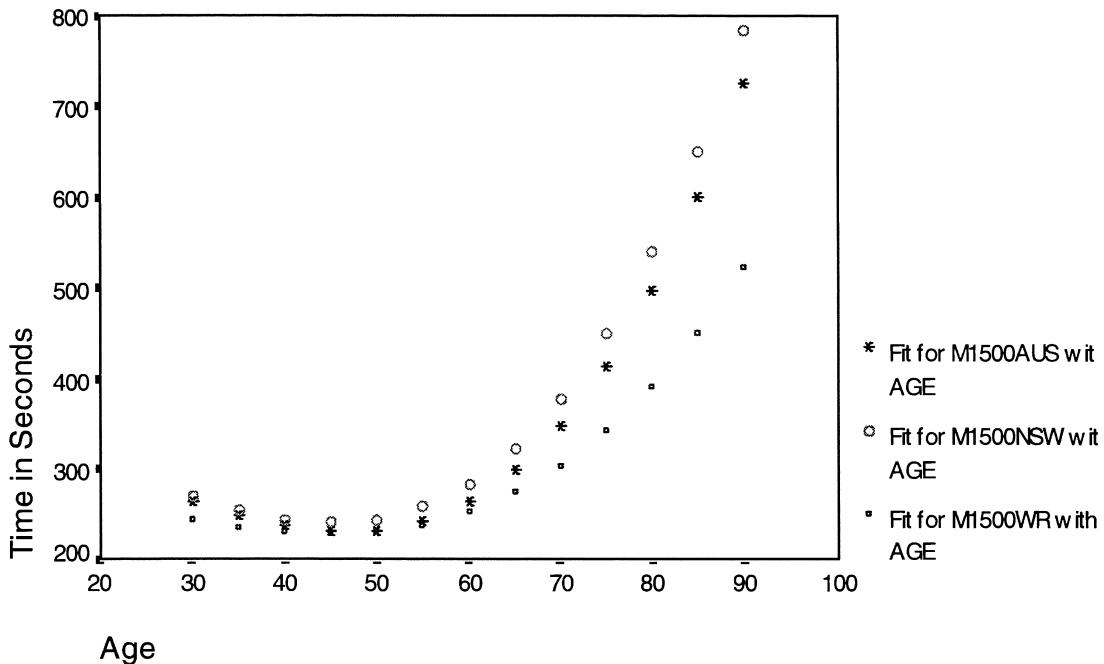
Figure 9: Cubic Models Representing the Men's 1500m records at State, National and World Level.
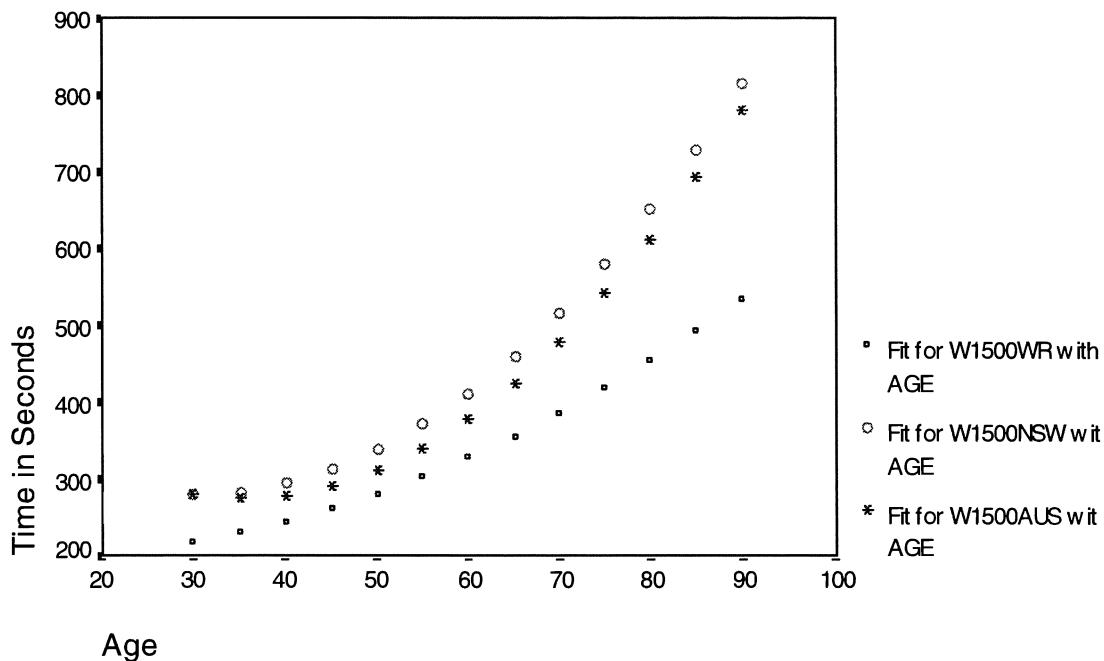
Figure 10: Quadratic Models Representing the Women's 1500m Records at State and National and a Cubic Model Representing the Women's 1500m World Record.

The models for the Men's and Women's High Jump records are shown in figures 11 and 12. In Figure 11 a deviation from the expected relative position of the records is seen at age 80, where the interpolated Australian record is higher than that for the world record. The negative heights shown for the NSW records may be due the fact that data only to the 50 year age group was collected for this record which produces predictions that do reflect realistic predictions, in fact they predict zero and minus heights.
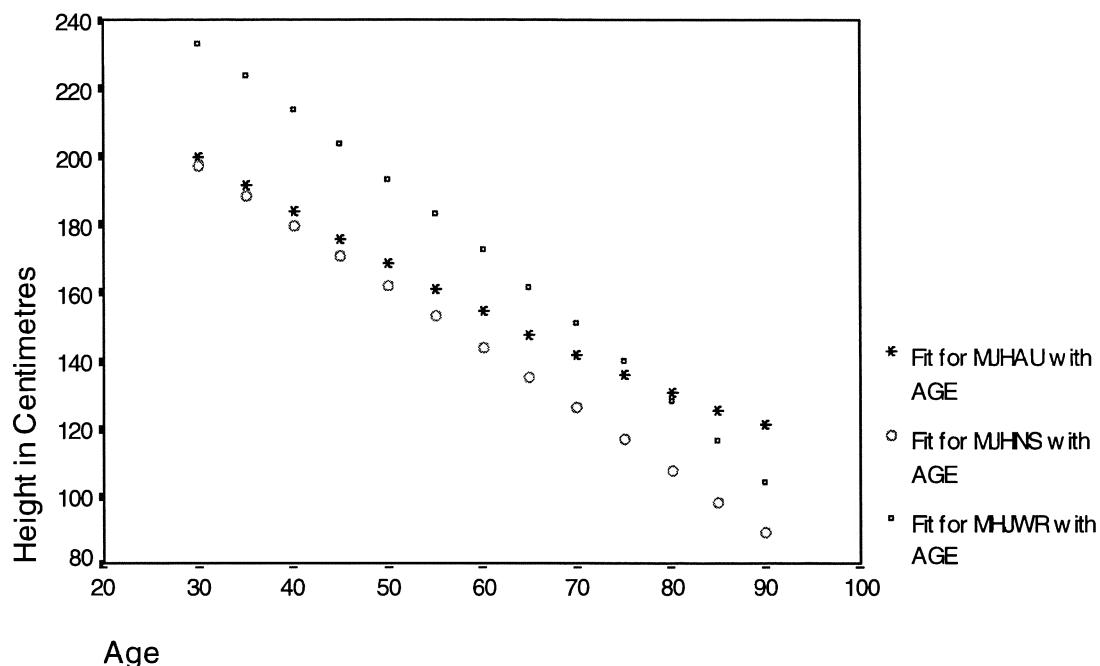
Figure 11: Cubic Models Representing the Men's High Jump records at State, National and World Level.
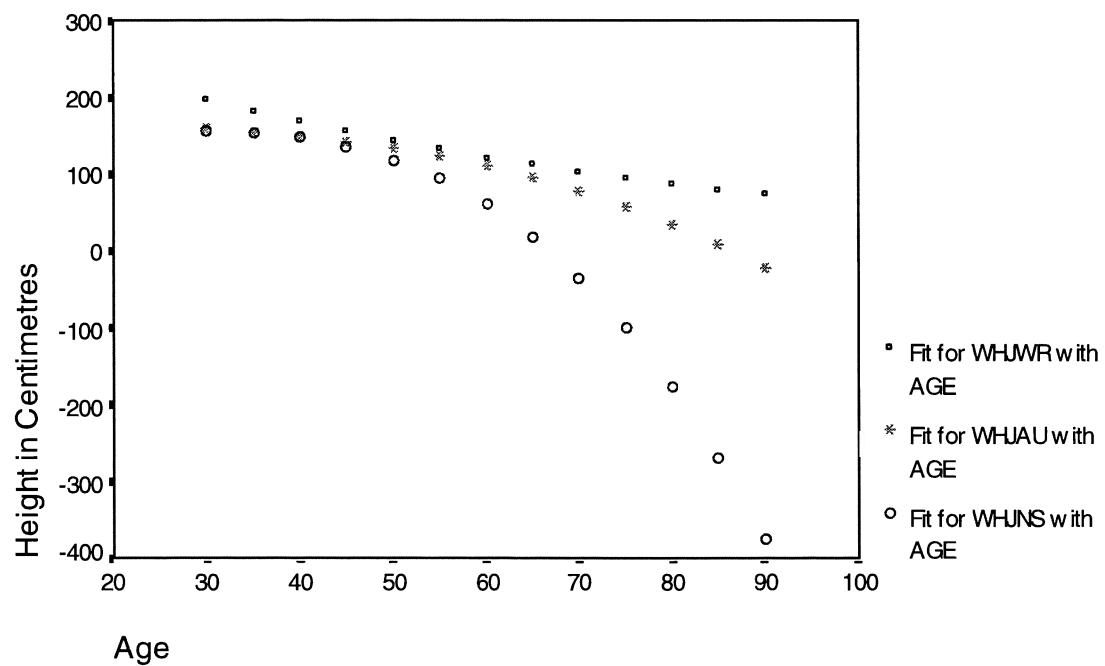
Figure 12: Cubic Models representing the Women's High Jump records at State, National and World Level.

DISCUSSION

General Aging Construct

The most important question to be addressed is, is aging a generalised aging process expressed throughout the body or do different systems and structures within the human age in their own unique manner and then will the aging process be expressed as performance decline in a specific age-related manner? Based upon the findings of

this research as the majority of the decline was expressed as a cubic mathematical function across all the different athletic events and for both genders especially for the data based on world records. This result suggests that the mathematical function describes an aging process that is general to the human organism and that we are at any time (assuming we are alive) on some segment of the curve represented by the cubic function. Although humans display different levels of athletic ability such as by state, national and world records these people are still conforming cubic decline trend in performance, its just that the curve has different constants and coefficients.

Energy Systems

In the men's events the decline in performance in the majority of cases conformed to a cubic model closely followed by the quadratic model. Specifically, the events related to the ATP-CP system, that is the 100m, long jump and high jump, fitted the data very closely with R^2 ranging from 0.918 to 0.990 for the state, national and world records. The Men's Triple Jump although having a lower R^2, that is $R^2 = 0.608$, the best fitting model was still the cubic model.

Events that are thought to depend on the anaerobic-glycolysis and some ATP-CP input, that is the 200m and 400m, also revealed a good fitting cubic model (R^2 0.921 to 0.994).

The events that were dependent predominantly on aerobic metabolism displayed cubic models for the men (1500m, 5000m and 10000m) and cubic-quadratic models for the women (1500m and 5000m). It is interesting to note that the 800m which is thought to be dependent on the anaerobic-glycolytic and aerobic systems also displayed a cubic model for the males and in the majority of cases a cubic model for the females.

When all the mathematical models describing the age-related declines across all the events and ability levels (state, national and world) are counted 26 were cubic and five were quadratic-cubic for the men and 19 were cubic, eight were quadratic-cubic and one was inverse for the females. When world records were examined alone, all models were cubic for both men and women with R^2 close to 1 for all events (R^2, from 0.901 to 0.992).

These data indicate that the decline appears to independent of the event representing the different energy-liberating mechanisms, although the review of literature indicated that different anatomical and physiological constructs displayed unique patterns of decline or loss of function. The review of literature would suggest that the declines of specific events that depend predominantly on the three different energy producing systems should display unique mathematical functions that reflect the changes in the underpinning systems that are expressed as performance. This was not the case based on the comprehensive data sets in this research.

Models Related to Gender

Gender differences in anatomical and physiological constructs in some cases were presented as gender specific, such as changes in body composition, whereas other constructs were not, such as strength, cardiovascular and pulmonary function. The

constructs of strength, cardiovascular and pulmonary function which are thought to underpin the declines in anaerobic and aerobic power with increasing age and should be reflected as changes in the sprints, jumps, middle distance and longer distances were not markedly different between the genders based on the findings of this research. These findings may indicate that age-related declines in strength, cardiovascular and pulmonary function are far more important in determining declines in competitive performance than changes in body composition as the mathematical function describing performance changes with age were almost identical, especially based on world record performances.

Some additional mathematical models were derived based on national records from Finland to assess if the trends based on New South Wales State Records and Australian National Records were replicated. These results are presented in Appendix I for the interest of the reader. It is important to note that the sprint, middle distance, long distance and the jump events produced the consistent cubic model followed by the quadratic model in the majority of events, and that described the declines in competitive performance with increasing age from 30 to 90 years.

CONCLUSIONS

1. The decline in human motor performance for the majority the data sets based on the veteran athletic events, that is the sprints, horizontal jumps, vertical jumps, middle distance and long distance events conformed closely fitted a cubic mathematical function followed by a quadratic mathematical function. In a small number of events multiple mathematical functions closely fitted the data, that is cubic, quadratic, compound, growth, exponential and logistic functions.

2. The model fits, based on the fit indices of R^2, p-values, residuals and the logical fits based on interpolation and extrapolation were exceptionally close to the data sets with the majority of R^2's greater than 0.95.

3. The model fits based on the cubic and quadratic functions were consistent across all events analysed in this research.

4. The model fits based on the cubic and quadratic functions analysed in this research were consistent across all events and for the participants' level of competitive ability.

5. The model fits based on the cubic and quadratic functions were consistent for gender across all events analysed in this research.

6. The general aging process as reflected by the consistent declines in performance appears to be independent of ability level, athletic event and gender.

7. The ability to predict the age-related decline in performance for an individual could be predicted with reasonable accuracy independent of ability level, athletic event and gender.

8. This research indicates that the general aging process and the aging process reflected in the consistent declines in human motor performance based on speed, power and endurance, and is the dominant factor that explains declines in gross

motor behaviour. The training process is then just an overlay to the general trends of human aging.

9. The partistic analysis of anatomical and physiological functions such as energy systems, muscle function, cardiovascular system, the pulmonary system, FFM and FM do not explain or represent the complex interactions that are expressed as competition human motor performance.

10. The partistic approaches must be integrated into a more holistic analysis to explain human motor behaviour both within the laboratory and field (competitive) settings.

REFERENCES

[1]   Chopra D., Ageless body, timeless mind, Sydney: Random House Australia (1993).

[2]   Herzog R., Welcoming address. Fourth International Congress Physical Activity, Aging and Sport. Heidelberg, Germany, August 27-31, 1996. Programme and Abstracts, 4 (1996).

[3]   Harridge S., Contractile proteins: Skeletal Muscle, Aging and Physical Exercise. Paper presented at the Fourth International Congress Physical Activity, Aging and Sport. Heidelberg, Germany, August 27-31, 1996.

[4]   Heazlewood I., Training programs of elite veteran track and field athletes. In Healthy Aging, Activity and Sport Proceedings. G. Huber (Ed). Fourth International Congress Physical Activity, Aging and Sport. Heidelberg, Germany, August 27-31, 1996, 446-449. Gamburg: Health Promotion Publications (1997).

[5]   Heazlewood I., Factors contributing to continuation and discontinuation in veteran athletics. In Healthy Aging, Activity and Sport Proceedings. G. Huber (Ed). Fourth International Congress Physical Activity, Aging and Sport. Heidelberg, Germany, August 27-31, 1996, 450-452. Hamburg: Health Promotion Publications (1997).

[6]   Lexell J., Adaptability and coupling function: Skeletal muscle, aging and physical exercise. Paper presented at the Fourth International Congress Physical Activity, Aging and Sport. Heidelberg, Germany, August 27-31, 1996.

[7]   Richter C., Mitochondria: Skeletal muscle, aging and physical exercise. Paper presented at the Fourth International Congress Physical Activity, Aging and Sport. Heidelberg, Germany, August 27-31, 1996.

[8]   Suominen H., Connective tissue: Skeletal muscle, aging and physical exercise. Paper presented at the Fourth International Congress Physical Activity, Aging and Sport. Heidelberg, Germany, August 27-31, 1996.

[9]   Reaburn P., Masters sports: Training for enhanced performance. Brisbane: ACHPER (Queensland) (1995).

[10] Reaburn P., Logan P. and MacKinnon L., The effect of resistance training on anaerobic work capacity in veteran sprint runners. Paper presented at the Fourth International Congress Physical Activity, Aging and Sport. Heidelberg, Germany, August 27-31, 1996.

[11] Huber G., (Ed.) Healthy Aging Activity and Sports - Proceedings. International Congress Physical Activity, Aging and Sports (PAAS IV). Heidelberg, Germany, August 27 - 31, 1996 (1997).

[12] Reaburn P., What does research tell the masters athlete. *Sports Coach*. Autumn, 1997, 22-23 (1997).

[13] deVries H. and Housh T., Physiology of exercise (4th ed.) Dubuque: Wm. C. Brown (1994).

[14] Astrand P-O., Exercise physiology of the mature athlete. *In Sports Medicine For The Mature Athlete.* J.R. Sutton and R.M. Brock (Eds.), 3-13 (1986).

[15] Thompson P.D. and Dorsey D.L., The heart of the masters' athlete. In Sports Medicine For The Mature Athlete. J.R. Sutton and R.M. Brock (Eds.), 309-318 (1986).

[16] Saltin B., The aging endurance athlete. In Sports Medicine For The Mature Athlete. J.R. Sutton & R.M. Brock (Eds.), 59-80 (1986).

[17] Hagerman F.C., Fielding R.A., Fiatarone M.A., Gault J.A., Kirkendall D.T., Ragg K.E., and Evans W.J., A 20-yr longitudinal study of Olympic oarsmen. *Medicine and Science in Sports and Exercise.* 28, 9, 1150-1156 (1996).

[18] Drinkwater B.L., Exercise and aging: The female masters athlete. In Sport Science Perspectives for Women. J.L. Puhl, C.H. Brown and R.O. Voy (Eds.). Champaign: Human Kinetics, 161-170 (1988).

[19] Makrides L., Physical training in young and older healthy subjects. In Sports Medicine For The Mature Athlete. J.R. Sutton and R.M. Brock (Eds.), 363-373 (1986).

[20] Hagberg J.M., Yerg H. and Seals D.R., Pulmonary function in young and older athletes and untrained men, *Journal of Applied Physiology*, 65(1), 101-105 (1988).

[21] Green H.J., Characteristics of aging human skeletal muscles. In Sports Medicine For The Mature Athlete, J.R. Sutton and R.M. Brock (Eds.) 17-26 (1986).

[22] Meltzer D.E., Age dependence of Olympic weightlifting ability, *Medicine and Science in Sports and Exercise.*,26, 8, 1053-1067 (1994).

[23] Reaburn P.R.J. and MacKinnon L.T., Blood lactate responses in older swimmers during active and passive recovery following maximal sprint swimming. *European Journal of Applied Physiology*, 61, 246-250 (1990).

[24] Martin J., The effects of age and exercise on short term maximal performance: A model based on physiological systems (1998). Http://www.krs.hia.no/-stephens/maxpower. htm

[25] Going S.B., Williams D.P., Lohman T.G. and Hewitt M.J., Aging, body composition, and physical activity: A review. *Journal of Aging and Physical Activity*, 2, 38-66 (1994).

[26] Hamilton N., Changes in sprint stride kinematics with age in masters' athletes. *Journal of Applied Biomechanics*, 9, 15-26 (1993).

[27] Arya J.C. and Lardner R.W., Mathematics for the biological sciences. Englewood Cliffs: Prentice-Hall, INC (1979).

[28] Haefner J.W., Modeling biological systems: Principles and applications. New York: Chapman & Hall (1996).

[29] de Mestre N., (Ed.) Second conference on Mathematics and Computers in Sport. Conference Proceedings. Bond University, Queensland, 11-13 July 1994.

[30] de Mestre N., (Ed.) Third conference on Mathematics and Computers in Sport. Conference Proceedings. Bond University, Queensland, 30 September - 2 October 1996.

[31] Heazlewood I. and Burke S. Exercise physiological, sport psychological and biomechanical predictors of the triathlon. Conference Proceedings in Active Connections, 20th Biennial/International ACHPER Conference. University of Melbourne, 14-19 January, 118-122 (1996).

[32] Politi S. and Heazlewood I., Exercise physiological, biomechanical and kinanthropometric predictors of Bicycle Motor Cross in young adults. Conference proceedings 3rd Conference on Mathematics and Computers in Sport. Bond University, Queeensland, 30 September - 2 October, 123-154 (1996).

[33] Heazlewood I. and Lackey G., The Use of Mathematical Models to Predict Elite Athletic Performance at the Olympic Games. Conference proceedings 3rd Conference on Mathematics and Computers in Sport. Bond University, Queensland, 30 September-2 October, 185-206 (1996).

[34] Heazlewood I. and Lackey G., The use of mathematical models to predict elite athletic performance at the Olympic Games. Conference proceedings The 21st Biennial National/International ACHPER Conference. Key into Life. Adelaide, January 12-16, 60-64 (1998).

[35] Heazlewood I., Multidisciplinary and multivariate approaches to problem
      solving in the movement sciences. Conference proceedings The 21st Biennial
      National/International ACHPER Conference. Key into Life. Adelaide, January
      12-16, 65-69 (1998).

[36] Niemi O., Veteran's track and field records.
      Website: http://koti.kolumbus.fi/~geodun/niemi/records.htm

[37] NSW Veterans Athletics Club Inc. 27th NSWVAC Inc. Track & Field
      Championships. March 22-23 1977. Programme (1997).

[38] Norušis M. J., SPSS for Windows: Base systems user's guide. Release 6.0. (SPSS,
      Chicago, 1993).

[39] SPSS Inc., SPSS for Windows Release 6.1 [Computer Software], 1994.

APPENDIX I

Table A1

Regression Models - Finnish National Records for Women's Events.

Event	Model	R^2	Level of Sig.
100m	Cubic	.978	<.001
	Quadratic	.975	<.001
200m	Cubic	.980	<.001
	Quadratic	.972	<.001
400m	Cubic	.987	<.001
	Quadratic	.985	<.001
1500m	Quadratic	.989	<.001
	Cubic	.988	<.001
	Compound	.962	<.001
3000m	Cubic	.961	<.001
	Quadratic	.949	<.001
5000m	Cubic	.992	<.001
	Quadratic	.986	<.001
10000m	Cubic	.984	<.001
	Quadratic	.980	<.001
Marathon	Cubic	.989	<.001
	Quadratic	.987	<.001
High Jump	Compound	.969	<.001
	Growth	.969	<.001
	Exponential	.969	<.001
	Logistic	.969	<.001
	Cubic	.966	<.001
	Quadratic	.966	<.001
Pole Vault	Cubic	.959	0.041
	Quadratic	.941	0.059
Long Jump	Cubic	.949	<.001
	Quadratic	.948	<.001
	Linear	.948	<.001
Triple Jump	Cubic	.936	<.001
	Quadratic	.936	<.001
	Linear	.930	<.001
Discus	Compound	.967	<.001
	Growth	.967	<.001
	Logistic	.967	<.001
	Exponential	.967	<.001
	Cubic	.962	<.001
	Quadratic	.962	<.001
	Logarithmic	.962	<.001
Shot (3kg)	Cubic	.979	<.001
	Quadratic	.979	<.001
	Compound	.977	<.001
	Growth	.977	<.001
	Logistic	.977	<.001
	Exponential	.977	<.001

Table A2

Regression Models - Finnish National Records - Men's Events.

Event	Model	R^2	Level of Sig.
100m	Cubic	.967	<.001
200m	Cubic	.871	<.001
400m	Cubic	.850	<.001
800m	Cubic	.990	<.001
	Quadratic	.986	<.001
1500m	Cubic	.843	<.001
3000m	Cubic	.988	<.001
	Quadratic	.986	<.001
5000m	Cubic	.970	<.001
10000m	Cubic	.990	<.001
	Quadratic	.987	<.001
Marathon	Cubic	.991	<.001
	Quadratic	.984	<.001
High Jump	Cubic	.954	<.001
	Quadratic	.947	<.001
Pole Vault	Cubic	.993	<.001
	Quadratic	.993	<.001
	Linear	.992	<.001
Long Jump	Cubic	.991	<.001
	Quadratic	.991	<.001
Triple Jump	Cubic	.975	<.001
	Quadratic	.972	<.001

THE USE OF MATHEMATICAL MODELS TO PREDICT ELITE SWIMMING PERFORMANCE

Gavin Lackey[1] and Ian Heazlewood[1]

Abstract

The knowledge of future levels of sporting performance has been identified as beneficial in talent identification, the setting of goals by competitors and team selection. The trends of Olympic swimming performance from 1924 to 1996 have been analysed using one linear and ten non-linear models with the intention of predicting future performance levels. The mean from each Olympiad of each Freestyle swimming event that was contested at the 1996 Olympic Games was used as the data set for the regression analyses. The regression models used were linear, inverse, logarithmic, quadratic, cubic, power, sigmoidal, compound, logistic, exponential and growth. The coefficient of determination, level of significance and an analysis of the residuals were used to select the most appropriate model. The logical acceptance of the model was based on extrapolations of future performance. The final model was used to predict, through the use of extrapolation, future performance means. Factors that may have had an effect on swimming performance at previous Olympic Games such as hand-timing, the use of drugs and the political boycotts of 1980 and 1984 were also incorporated into the models. The model used for extrapolation of performance in the year 2000 was for every event a non-linear model with coefficients of determination ranging from 0.543 for the Men's 50m Freestyle to 0.977 for the Men's 400m Freestyle. It may also be possible to incorporate other factors such as the total number of competitors, number of nations represented and temperature of the pool into future regression equations. The opportunity exists to investigate the trends of performance in other events where performance is measured by time or distance. Where performance is measured by subjective means this type of research would not be appropriate.

1.   INTRODUCTION

The knowledge of future levels of sporting performance has been identified by Banister and Calvert [1] as beneficial in the areas of talent identification, both long and short term goal setting, and training program development. In addition, expected levels of future performance are often used in the selection of representative teams. For example, the Australian Sports Commission [2] stated that to be eligible for selection in a national team a swimmer must record a time equal to or better than that of the 16th ranked swimmer in the world in the previous year. At this stage in the analysis of swimming data it appears that the data from the previous year is the most accurate indicator of current performances. This does not, however, account for any underlying trends that may have been occurring over a number of years. For example, a year of poor results in one event does not necessarily infer that poor results will continue in this event in the near future.

[1]   Australian Catholic University, 40 Edwards Street, North Sydney NSW   2060

The prediction of performances that might be expected of competitors in future competitions is a recurring theme in sports, especially during Olympic years. Often these predictions are purely speculative and are not based upon any substantial evidence, rather they are based on the belief that records are made to be broken and that performances must continue to improve over time.

The accessibility of data in the form of results from Olympic Games, world records and world best performances in a specific year allows the analysis of performances in any number of events. From such analysis changes in performance levels, that is, times, can be observed and predictions of future performance levels can be made via the process of mathematical extrapolation.

2.   RESEARCH PROBLEM

A review of the published literature revealed an apparent lack of analysis of the mathematical trends in the change of swimming performance at the Olympic Games. The fact that performances have improved is well documented, as are many of the possible reasons behind these improvements such as changes in training methods, changes in technique, improvements in the competition and training environments, the use of performance-enhancing drugs and increases in the participation rate. The manner of these improvements as a result of the genetic and environmental changes does not appear to have been investigated. Therefore, the problem was to use scientific methods in conjunction with the available data to determine if these improvements in swimming performance have occurred in a regular manner and whether the trends of change can be used to predict future performance levels based on the mathematical-statistical modeling.

3.   RESEARCH QUESTIONS

A number of research questions were generated by the reviewed literature which relate to the constructs of the changes in swimming performance. These questions were:

1. What is the best mathematical model for the available data in each event?

2. Can the effects of androgenic and anabolic steroids and other performance-enhancing substances be accounted for, and thus provide more accurate models?

3. Does countering the effects of hand-timing prior to 1964 improve the models?

4. Does each event have a specific model or does one global model apply?

5. Will the models be gender specific as well as event specific?

6. Will the mathematical model allow accurate predictions of future performance?

7. Will the models derived generate absurd predictions such as those of a zero or negative time?

8. How much variance will be explained in the different models for each event and gender?

4.   RESEARCH HYPOTHESES

According to Rothstein [3], research hypotheses are predictions that are tested through the use of research. Statistical hypotheses, which are normally stated as null hypotheses suggesting that there will be no difference, are used to evaluate research hypotheses. Statements that there will be a difference are classified as alternative hypotheses. The following hypotheses are stated as alternative hypotheses.

1. The modeling procedure will allow predictions of future swimming performance for the Freestyle events at the Olympic Games. These events are 50m Freestyle, 100m Freestyle, 200m Freestyle, 400m Freestyle, for men and women, the 800m Freestyle for women and the 1500m Freestyle for men.

2. More variance will be explained through the use of non-linear models, therefore reducing the amount of error variance (the unexplained component). This means more accurate models describing the relationship between event times and the specific Olympiad will be derived.

3. The effects of steroids and other performance enhancing substances will be seen to a greater degree in the Women's results and correction for these effects will allow better predictions based on the derived models predicting future performance.

4. Corrections for hand-timing prior to 1964 will not greatly improve the fit of the derived models.

5. Each event will be best described by a specific mathematical-statistical model.

6. The models in addition to being event specific will be gender specific. This means that each gender in each event will be described by a specific mathematical model.

7. The use of non-linear models will result in less absurd predictions such as those of a zero time or negative times for all swimming events analysed in this study.

5.   METHODS AND STATISTICAL ANALYSIS

The basic premise applied was one of plotting the mean result of the finalists for each event against the year of performance, then determining which of a number of regression models provided the straight line or curve of best fit. This model was then used to predict future performance in each event based on mathematical extrapolation.

As linear regression models such as those used by Edwards and Hopkins [4], despite high coefficients of determination, have major deficiencies, a number of non-linear regression models were applied to the data to determine the lines or curves of best fit for past performances. The non-linear regression models are based on the curve of best fit as compared to the line of best fit that is used with linear regression (Norušis [5]).

The sampling procedure in this study was, in effect, self selecting. Given the rewards and accolades of Olympic success it was assumed that each swimmer gave their best effort on a number of occasions. These occasions are:

1. National selection trials.

2. Qualifying heats at the Olympic Games.

3. The final of the Olympic Games.

Following these assumptions it is reasonable to expect that those swimmers who competed in each Olympic final were the best eight swimmers in the world at that time of competition.

The results for the finalists in each current Olympic Freestyle swimming event were collected. Times were recorded to one hundredth of a second which is the recording method used by Federation Internationale de Natation Amateur [6]. These times were then converted from a minutes and seconds format (mm:ss.th) to a seconds only format (ss.th) to facilitate calculations when applying the regression methods. The mean of the finalists in each event for each year in the study was then calculated. The mean was used as it is a measure that is representative of all scores in each group (Rothstein, [3]). The use of the mean of the finalists in this study may be more representative of the changes in human performance that have occurred as used by Jokl and Jokl [7], [8] and [9]) and Edwards and Hopkins [4]. A world record holder's performance may be far in advance of that of any other competitor and not be representative of overall performance. For example, the women's 400m Freestyle world record as set by Tracey Wickham in 1978 was not bettered until 1988 at the Seoul Olympic Games (Wallechinsky [10]).

The calculated means were then included as a data set for analysis by the Statistical Package for the Social Sciences (SPSS) program version 6.1 (Norušis [5]; SPSS Inc. [11]) in order to derive a number of regression equations for each event.

To determine if this method was actually viable the means from the 1996 Olympiad were initially excluded from the data set and the extrapolated means using the regression models were compared to the actual means from the 1996 Olympic Games.

Predictions Based Upon Years Where Androgenic and Anabolic Steroids are not Known to be Involved

The use of steroids and other performance enhancing substances in the 1970's, 1980's and early 1990's is well documented (Georges [12]; Helmsteadt [13], [14]; Muckenfuss [15]). In an attempt to determine the effects that the use of steroids and other performance enhancing substances may have had on elite swimming performances the results from the 1976 to 1996 Olympic Games were removed. It is well documented that many of the finalists at these Olympiads had used or were using some form of performance-enhancing drugs.

Corrections for Hand-Timing Prior to 1964

Hypothesis 4 states that, corrections for hand-timing prior to 1964 will not greatly improve the fit of the models. Prior to 1964 timing at the Olympic Games was performed by hand. This hand-timing is a source of what may be an error in the

actual time taken to complete the event. Whilst the difference between automatic timing and hand-timing is swimming does not appear to have been quantified, a number of researchers have investigated reaction time. Abernethy, Kippers, Mackinnon, Neal and Hanrahan [16] and Schmidt [17] have noted that reaction time to an anticipated signal for a simple motor task ranges between 0.1 seconds and 0.2 seconds. In athletic events that are timed by hand a 'correction' of 0.24 seconds is applied to counter the effect of the timekeepers' reaction times (Athletics Australia [18]). In acknowledging this, a correction of 0.24 seconds was added to the mean times of the results prior to 1964.

In relation to the duration of each event 0.24 seconds is quite small and thus it is expected that such a small factor would not greatly change the models. Any change would be expected to become smaller as the length of time taken to complete each event increases. For example, the effect should be greater in the 100m Freestyle than in the 1500m Freestyle.

All events that were contested prior to 1964 were included in this data set, with the exclusion of the 1996 results. A set of regressions were performed and the selection of the most appropriate model made according to the criteria previously stated. Predictions of 1996 performance in these events were then made and compared to the actual performances. For this hypothesis to be supported the models generated need to be similar to those without the correction factor included.

Regressions Incorporating the Boycott Years of 1980 and 1984

A number of swimmers were unable to compete at the Olympic Games of 1980 and 1984 due to political boycotts of these games by their countries. It is possible that the swimmers in the finals at these Olympics were not the best in the world at that time. In fact, Wallechinsky [10] noted a number of instances where competitors at the Goodwill and Friendship Games held, soon after the Olympic Games, performed better than the medallists of the Olympics.

Combination Sets of Regressions

Two further series of regressions that combined each of the above series were also performed. These series were 'Hand-timing and steroid free' and 'Boycott years and hand-timing'.

GENERAL METHOD OF DETERMINING THE APPROPRIATE REGRESSIONS MODELS

To investigate the hypotheses of model fit and prediction, the eleven regression models were individually applied to each of the swimming events. 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 from these eleven equations. The specific criteria to select the regression equation of best were the magnitude of R^2, the significance of the analysis of variance alpha or p-value and the residuals.

The Coefficient of Determination

The coefficient of determination (R^2) is a measure of accuracy of the model used. A coefficient of determination of 1.00 indicates a perfectly fitting model where the predicted values match the actual values for each independent variable (Norušis [5]; SPSS Inc. [11]). Where more than one model was able to be selected due to an equal R^2, the simplest model was used under the principle of parsimony, that is the avoidance of waste and following the simplest method (SPSS Inc. [11]).

Residuals

The residuals are the difference between the actual value and the predicted value for each case, using the regression equation (Norušis [5]; SPSS Inc. [11]). The smaller the residual, the better the fit of the model. For each model the residuals were generated by the SPSS program. A large number of positive residuals indicate that the prediction is an over estimation or faster than the actual performance and a large number of negative residuals indicates an underestimation or slower time than the actual performance.

Level of Significance

The level of significance, or p value, is a representation of the relationship between the model and the data. The smaller the p value, the higher the level of significance and the greater the relationship. A small p value indicates a small possibility that the possibility of the closeness of the predicted values to the actual values is due to chance is small.

Logical Acceptance Based on Extrapolations

The ability of the model to generate extrapolations that appear to be reasonable when compared to previous means was also taken into consideration. When a model generated extrapolations that appear to be inconsistent with the actual results this model was discarded and the model with the next highest coefficient of determination was selected.

APPLYING THE MODEL OF BEST FIT

After selection of the model to be used, according to the criteria previously stated, the equation of best fit was determined by applying the derived constants and coefficients to the generic formula for that model. Using this equation, a prediction of the mean result for the event at each Olympiad was calculated. At this stage, graphs representing the means of past and future performances for each event in each Olympiad were also generated in addition to predicted means using the appropriate regression equation.

FINAL PREDICTIONS FOR THE YEAR 2000 AND BEYOND

To predict the level of performance in the year 2000 and beyond, the data set that provided the greatest accuracy was chosen and the data from 1996 reincluded in the data set, where appropriate. A series of regressions were made using the best fitting

model and data set for each event. Using the constants and coefficients generated by regression models the future predictions were then calculated.

6.   RESULTS

The data set, model, R^2, level of significance of the best fitting model for each of the Freestyle swimming events are shown in Table 1. Also included in this table are the predicted means for each of these Freestyle swimming events as calculated using the appropriate regression model.

Table 1

Data Set, Model, Coefficient of Determination, Level of Significance and Predicted Mean for the Year 2000

EVENT		DATA SET	MODEL	R^2	LEVEL OF SIG.	MEAN	
Men's	50m Freestyle	All	Inverse	0.543	0.473	22.26	*
	100m Freestyle	Hand-timing	Compound	0.972	<0.001	48.17	
	200m Freestyle	All	Sigmoidal	0.862	0.001	105.57	
	400m Freestyle	All	Cubic	0.977	<0.001	219.03	
	1500m Freestyle	All	Cubic	0.963	<.001	854.72	
Women's	50m Freestyle	All	Inverse	0.732	0.347	25.11	
	100m Freestyle	Hand-timing	Cubic	0.959	<.001	54.81	
	200m Freestyle	Boycott	Sigmoidal	0.812	0.037	117.90	
	400m Freestyle	Steroid Free	Cubic	0.976	<0.001	217.97	
	800m Freestyle	Boycott	Sigmoidal	0.692	0.040	496.38	

* Time is recorded in seconds.

Comparisons of the actual mean from the 1996 Olympic Games with the predicted means for the 2000 Olympic Games are shown in Table 2. Also included in this table is the percentage improvement required to achieve the predicted means.

The predictions, with the exception of the Women's 400m Freestyle, seem reasonable in the light of the past performances with all other required improvements being equal to or less than 6.10%.

Table 2

Actual 1996 Means, Predicted Means and Percentage Improvement Required

Event		Actual 1996		Predicted 2000		Percentage Improvement
Men's	50m Freestyle	22.47	*	22.26	**	0.93
	100m Freestyle	49.31		48.17		2.31
	200m Freestyle	108.40		105.57		2.61
	400m Freestyle	230.57		219.03		5.00
	1500m Freestyle	910.29		854.72		6.10
Women's	50m Freestyle	25.37		25.11		1.02
	100m Freestyle	55.37		54.81		1.01
	200m Freestyle	119.95		117.90		1.71
	400m Freestyle	250.00		217.47		13.01
	800m Freestyle	514.89		496.38		3.59

* Time is recorded in seconds. ** Time is recorded in seconds.

The regression equations as generated by the SPSS program are shown in Table 3. In this table 'y' represents the predicted mean and 'year' is the prediction variable, the year of performance.

Table 3

Final Regression Equations

Event		Type	Equation
Men's	50 Freestyle	Inverse	$y = -44.77 + 134064.199 / year$
	100 Freestyle	Compound	$y = 42747.22 \times .99^{year}$
	200 Freestyle	Sigmoidal	$y = e^{-1.33+11971.98/year}$
	400 Freestyle	Cubic	$y = 12572.61 - 8.79 \times year + 6.53 \times 10^{-7} \times year^3$
	1500 Freestyle	Cubic	$y = 40457.13 - 27.31 \times year + 1.88 \times 10^{-6} \times year^3$
Women's	50 Freestyle	Inverse	$y = -71.68 + 193578 / year$
	100 Freestyle	Cubic	$y = 7417.11 - 5.48 \times year + 4.5 \times 10^{-7} \times year^3$
	200 Freestyle	Sigmoidal	$y = e^{-0.54+8451.02/year}$
	400 Freestyle	Cubic	$y = 1565 - 1.68 \times 10^{-7} \times year^3$
	800 Freestyle	Sigmoidal	$y = e^{-1.26+14938.8/year}$

The regression curves as generated by the SPSS program are shown in Figures 1 to 10. The change in performance level is apparent in the charts, as is the closeness of fit for the regression curves.
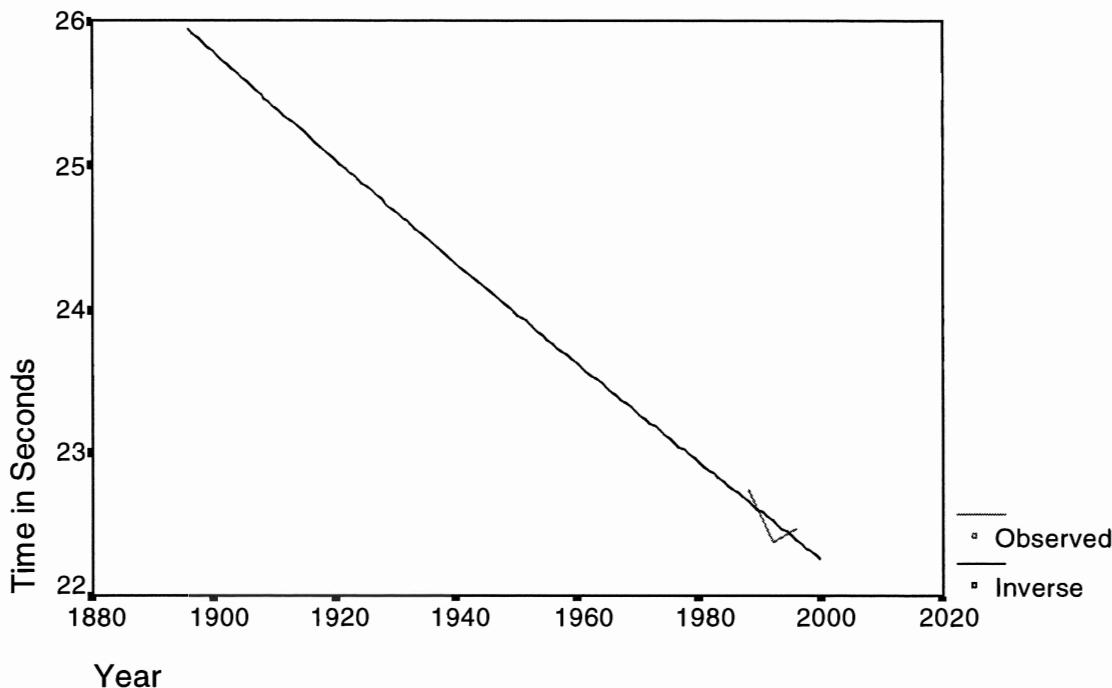
Figure 4.1:   Men's 50m Freestyle. $R^2 = 0.543$, p=.473, using all data.
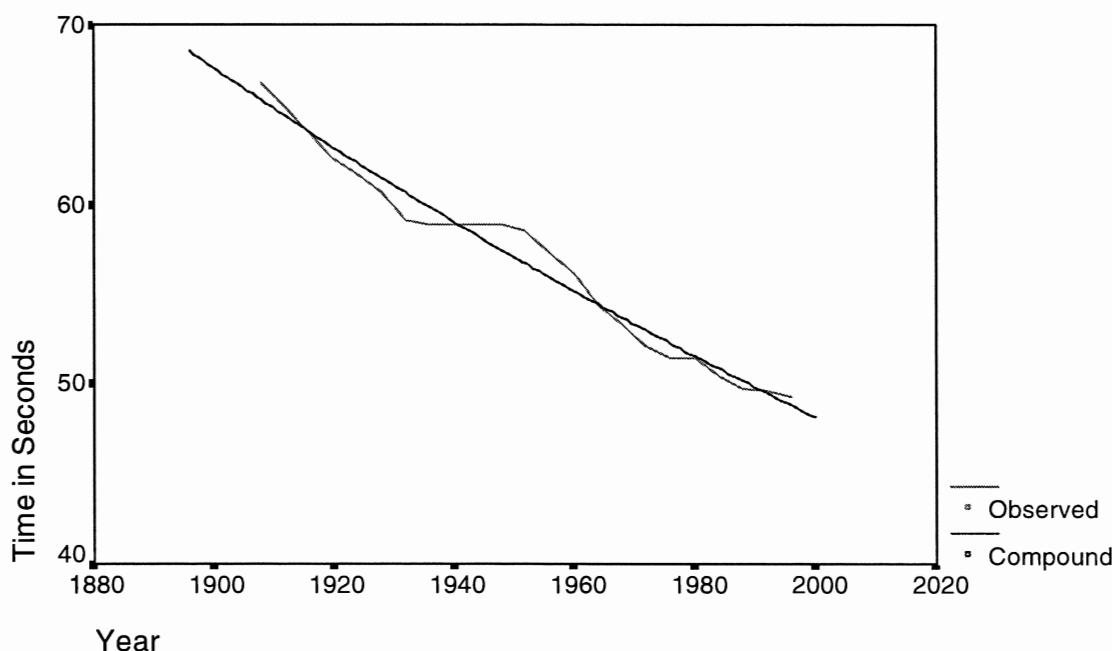
Figure 4.2: Men's 100m Freestyle. $R^2 =0.972$, p<.001, using the hand-timing data set.
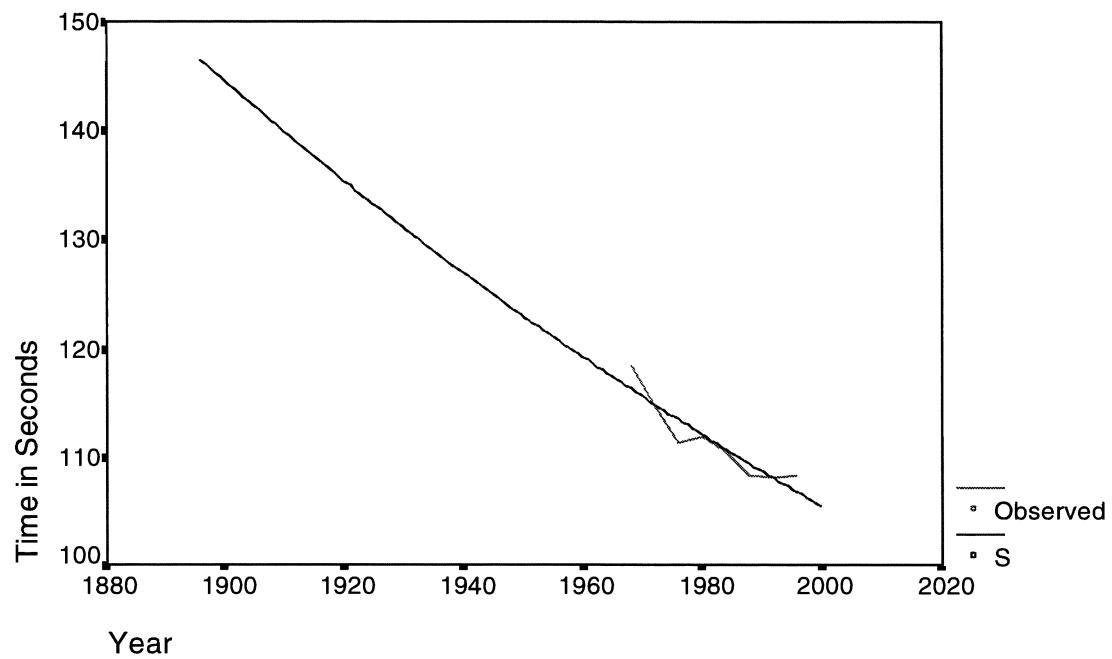
Figure 4.3:   Men's 200m Freestyle. $R^2 = 0.862$, p<.001, using all data.
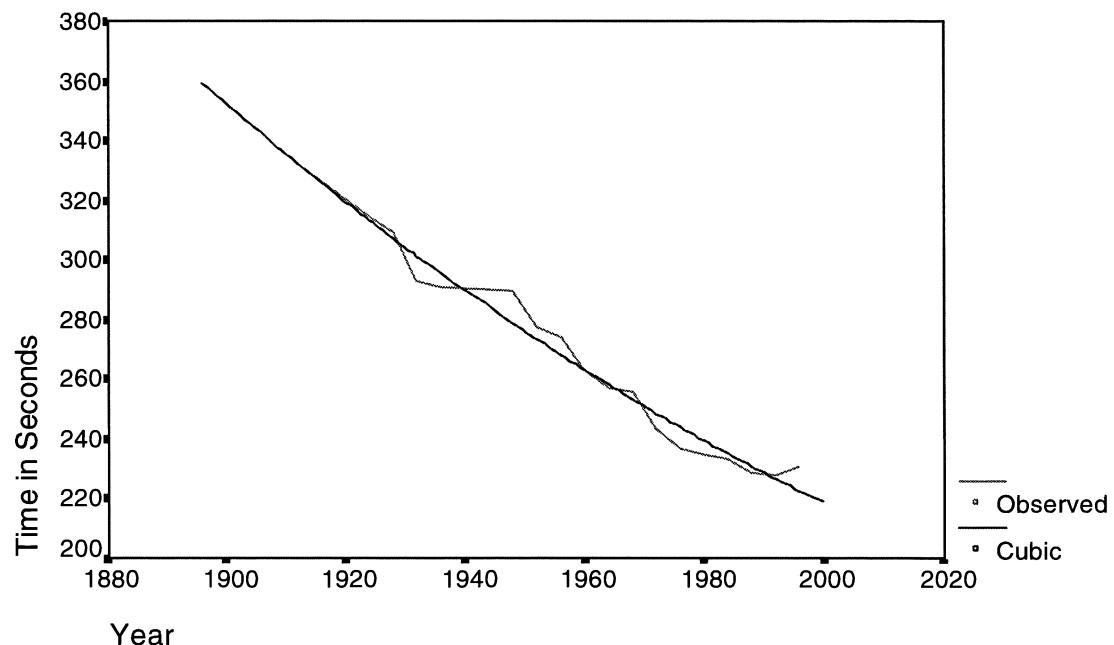
Figure 4.4:   Men's 400m Freestyle. $R^2 = 0.977$, p<.001, using all data.
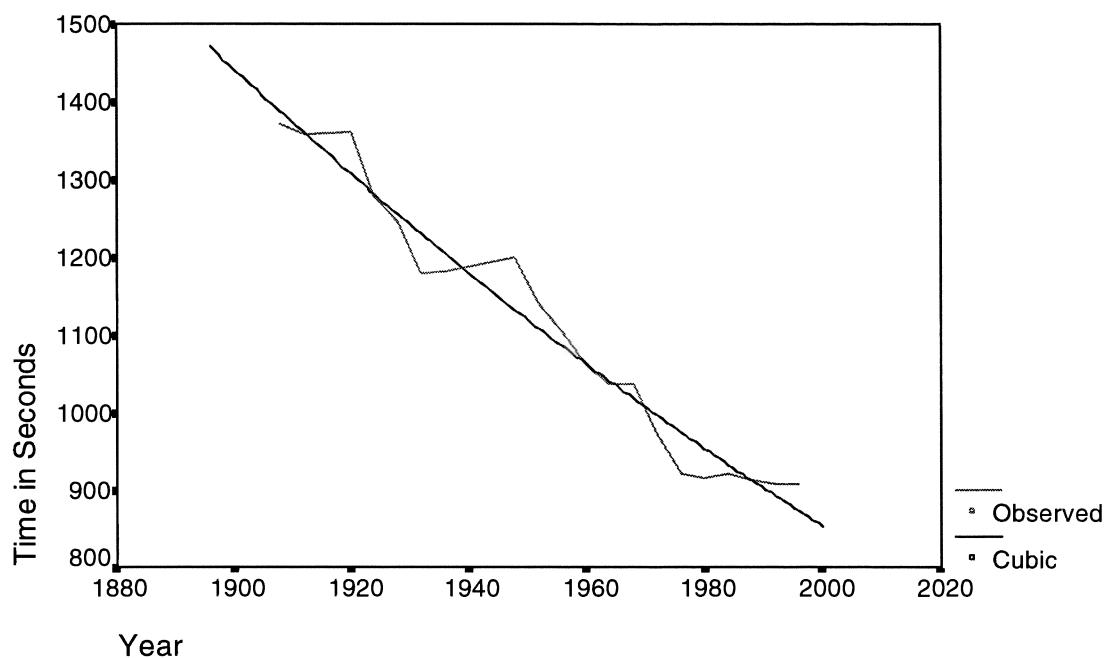
Figure 4.5: Men's 1500m Freestyle. $R^2 = 0.963$, p<.001, using all data.
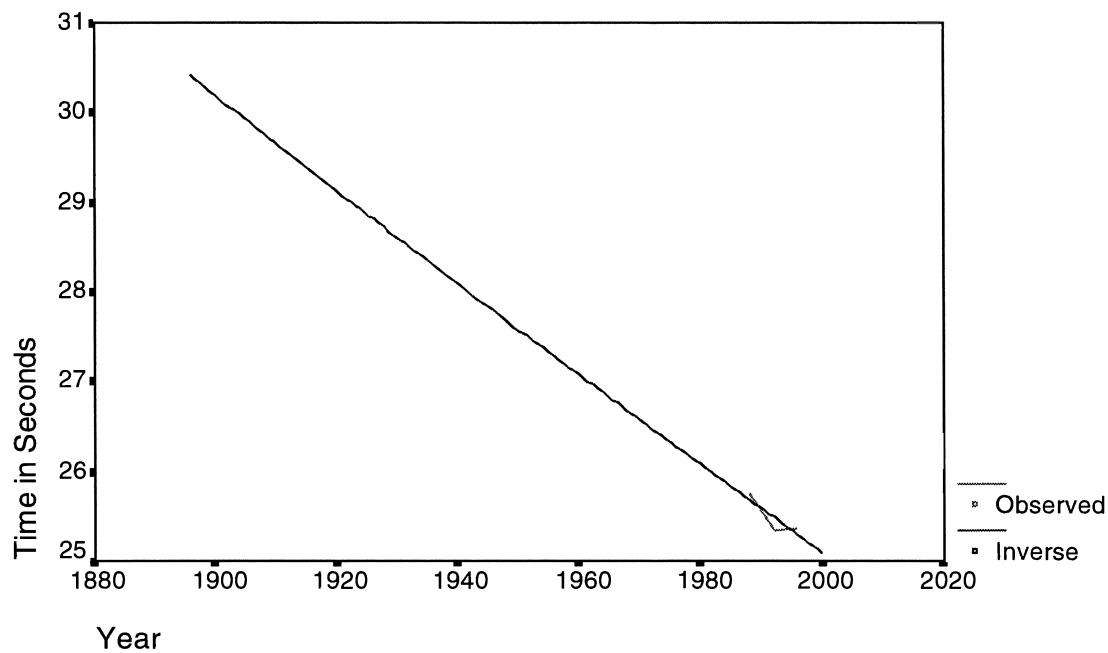
Figure 4.6: Women's 50m Freestyle. $R^2 = 0.732$, p=.347, using all data.
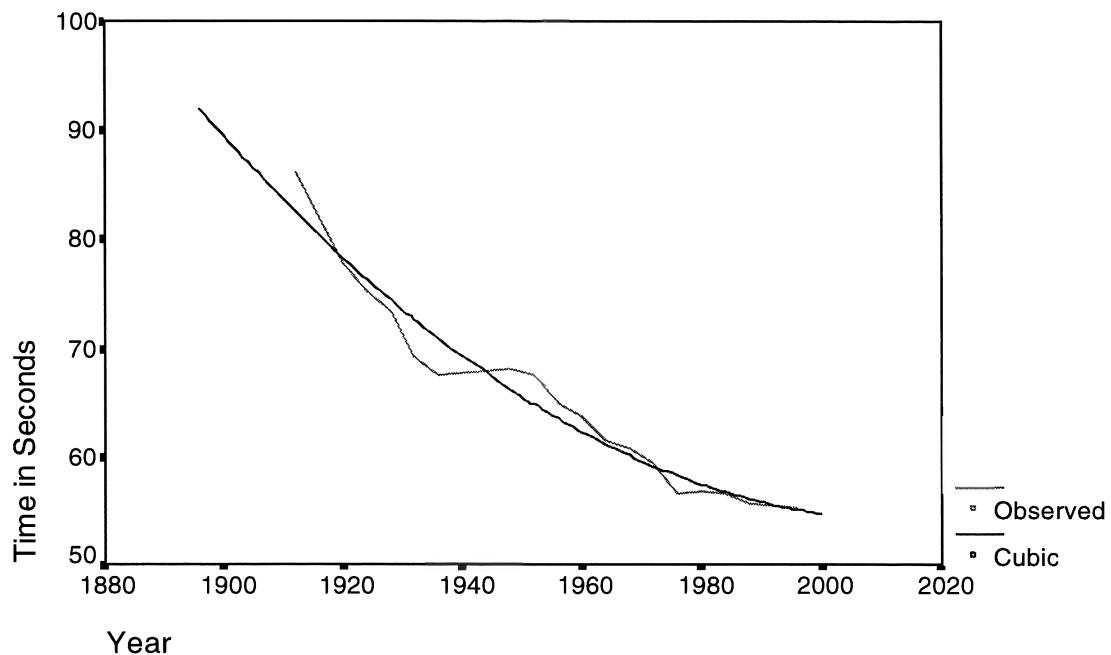
Figure 4.7: Women's 100m Freestyle. R^2 = 0.959, p<.001, using the hand-timing data set.
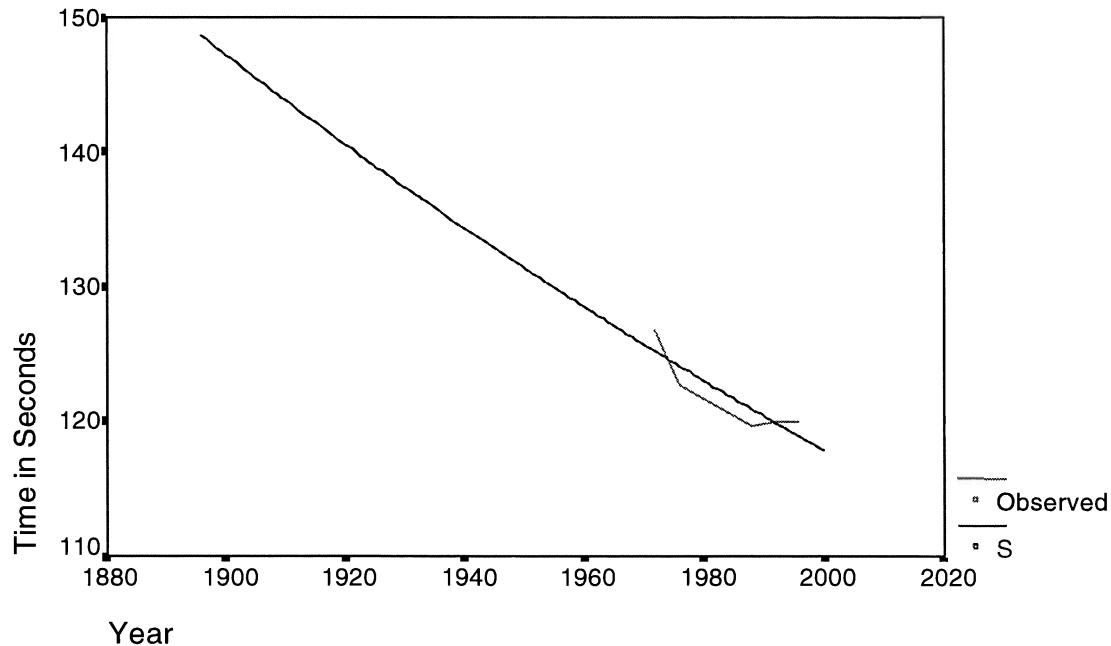
Figure 4.8: Women's 200m Freestyle. R^2 = 0.812, p<.037, using boycott data set.
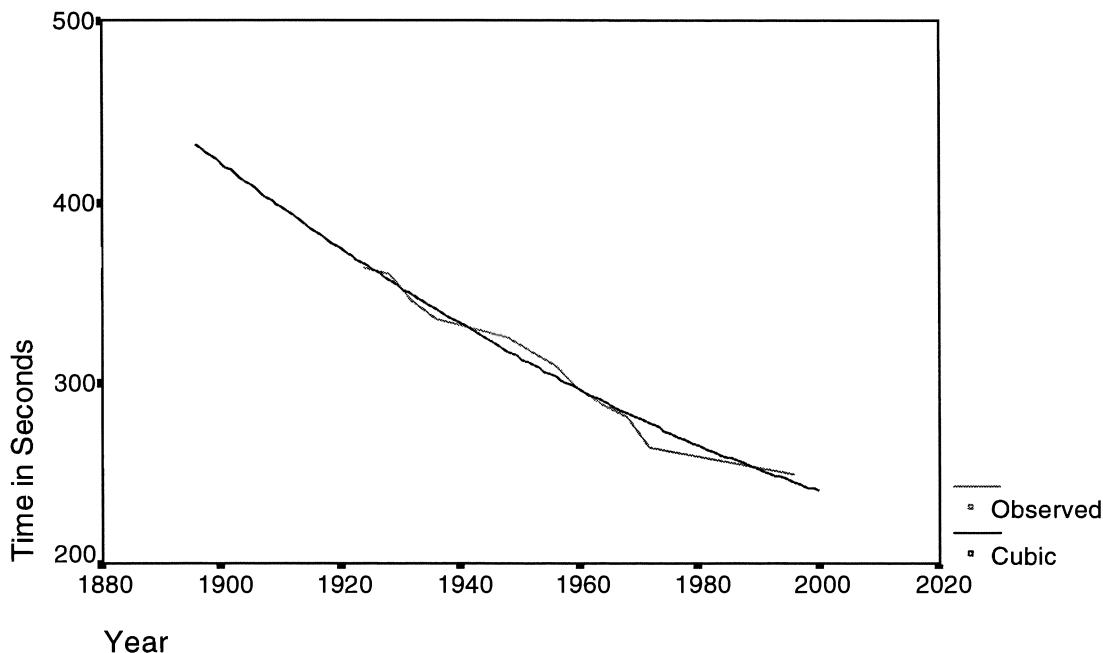
Figure 4.9: Women's 400m Freestyle. $R^2 = 0.968$, p<.001, using the 'steroid free' data set.
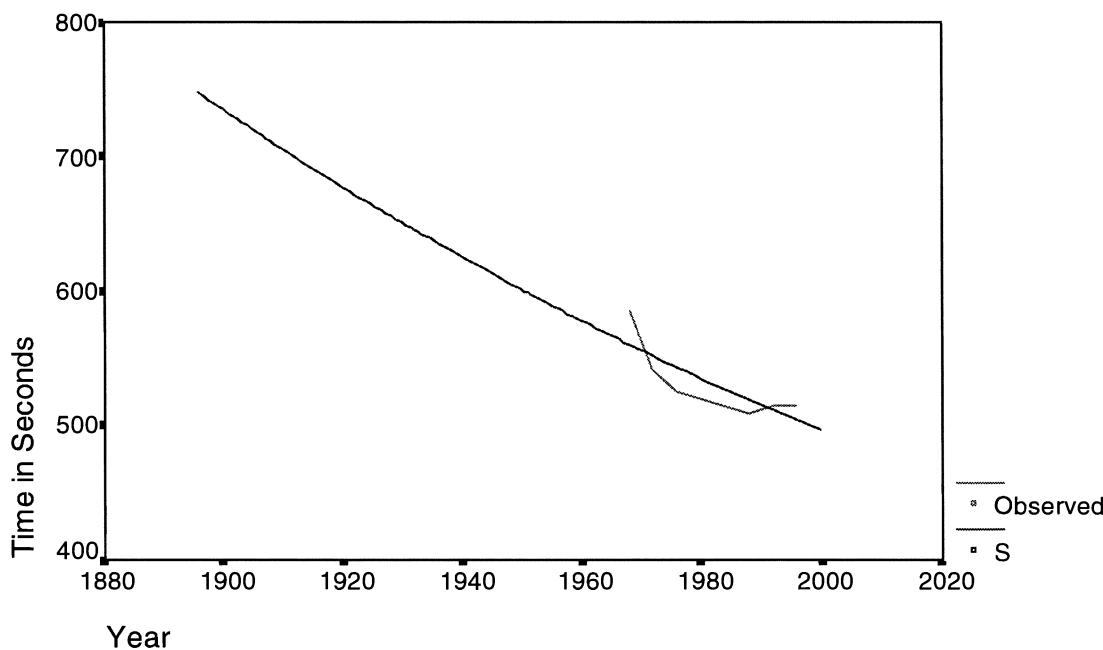
Figure 4.10: Women's 800m Freestyle. $R^2 = 0.692$, p=.04, using the boycott data set.

Checking for the Possibility of Absurd Predictions

Absurd predictions, such as those of a zero time, were able to be made by Edwards and Hopkins [4]. By using non-linear models it was projected that such predictions would be reduced or eliminated. An attempt to solve the regression models derived was made where a value of zero was used for 'y' (y = 0) to determine if it were impossible to predict a zero time using the model.

For those events were a zero time was able to be determined the year of this occurrence is shown in Table 4. The trend towards this zero time is illustrated in Figures 11 and 12 which represent the actual means to 1996 and the predicted means for the years past 2000.

Table 4

Year of Zero Time According to Regression Models

Event		Model	Year
Men's	50m Freestyle	Inverse	2994
Women's	50m Freestyle	Inverse	2700
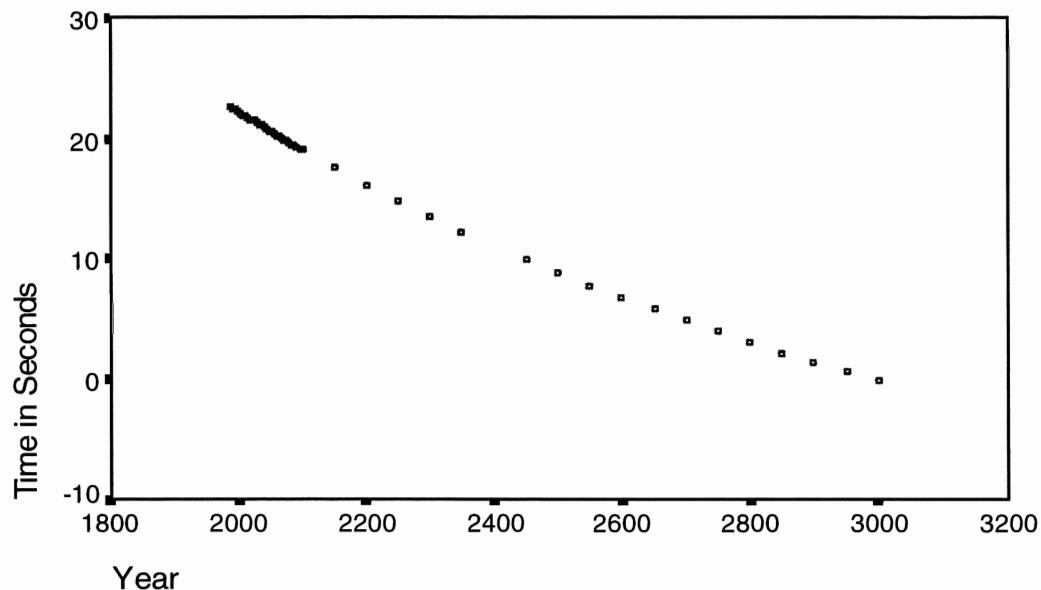
Figure 11:   Men's 50m Freestyle.
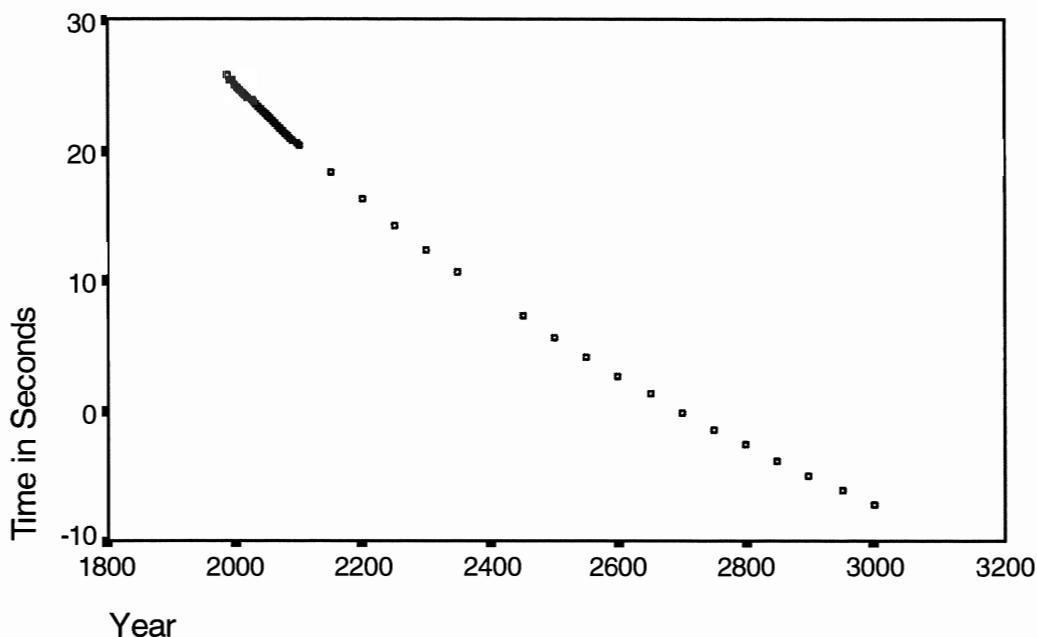
Figure 12:   Women's 50m Freestyle.

SUMMARY OF RESULTS

The major findings of the results of this study were:

1. Non-linear mathematical models provide better fitting curves of best fit than linear models and by using these models it is possible to predict swimming performance with more accuracy.

2. The removal of the data from the years 1976 to 1996, which are known to be confounded by the use of performance enhancing drugs such as androgenic and anabolic steroids, did not allow for more accurate predictions in most swimming events.

3. The hand-timing 'correction' of 0.24 seconds did not dramatically change the regression models of best fit.

4. There was not a single set of data, that is all data, 'steroid free', hand-timing, or the boycott free data set, or the combination sets, that was able to best describe all events.

5. The predictions of performance for the year 2000 made using the appropriate models were realistic when compared to the performances from 1996.

7.   DISCUSSION

The changes in performance in the Freestyle swimming events can readily be seen in Figures 1 to 10. These events have shown an improvement in performance (decrease in time) from inception to 1996. Performances are expected to improve over a period of time due to a number of factors, such as:

1. The use of more efficient swimming techniques, a biomechanical construct.

2. Improved training programs, an exercise physiological construct.

3. Enlarged population from which elite swimmers are drawn which results in an increased sample from the gene pool, a genetic construct.

4. Changes in human physiology, such as the recent trends in increasing height and weight (Dyer [19]).

5. The use of performance enhancing drugs, especially androgenic and anabolic steroids which have a masculinizing effect on women (Verroken [20]).

The overall manner of these improvements can only be seen after the appropriate regression model for each event has been determined. Banister and Calvert [1] have noted that changes in sporting performance appear to occur in a regular pattern. The task of determining this pattern for the Freestyle swimming events has been undertaken in this study. The results of this study are used as the base for the discussion which addresses the research issues that arise from the research questions and hypotheses generated previously.

Hypothesis 1

Hypothesis 1 stated that the models derived will allow predictions of future performance. This hypothesis was supported in entirety. The predictions for 1996, made using data up to 1992, were accurate within 5% in all but two cases. These events are the Men's 1500m Freestyle and the Women's 400m Freestyle. Attempts to further increase the accuracy of the models are made in later sections of the study.

It was also possible to determine, through the process of interpolation, a mean result for those years where a particular event was not conducted. Apart from 1916, 1940 and 1944 where the Olympic Games were cancelled due to war, the Men's 400m Freestyle was not held in 1920. The predicted means, according to the appropriate models using all data to 1996, were derived through interpolation and are shown in Table 1. Tables 4 and 5 shows the predicted means for events not contested either due to war or unspecified reasons. All of the means shown in Tables 4 and 5 are reasonable in that the actual results are not too different, in terms of performance time, from the predicted means when these means are compared to the means from the previous and subsequent Olympiads.

Table 4

Predicted Means (Using all data) for Men's Events not Contested at Various Occasions. Time is Recorded in Seconds.

Event			Year	Predicted Mean	Previous Mean	Subsequent Mean
Men's	100m Freestyle	*	1916	67.42	65.16	62.40
		*	1940	62.28	58.64	
			1944	61.48		58.65
	400m Freestyle		1916	324.00	332.24	314.88
			1920	319.71	332.24	314.88
			1940	287.81	291.31	
		*	1944	282.20		289.64
	1500m Freestyle	*	1916	1335.42	1358.13	1361.20
			1940	1182.37	1183.13	
		*	1944	1158.08		1200.24

* predicted mean outside of range of prior and succeeding means.

Table 5

Predicted Means (Using all data) for Women's Events not Contested at Various Occasions. Time is Recorded in Seconds.

Event			Year	Predicted Mean	Previous Mean	Subsequent Mean
Women's	100m Freestyle		1916	81.16	85.88	77.65
		*	1940	70.12	67.39	
			1944	68.57		67.90
	400m Freestyle		1940	330.71	335.55	
		*	1944	322.54		325.56

* Predicted mean outside of range of prior and succeeding means.

The method of substituting the geometric mean as used by Stefani [21], when omitting the data from 1968 due to the altitude of the host city, assumes that the times of the finalists will continue to improve from Olympiad to Olympiad. An inspection of the actual results indicates that this is not always the trend of improvement. The mean of the finalists in some events actually regresses before

improving. While Stefani [21] implies continuing improvement, the method used in this study allows for fluctuations in the actual mean times. This is reflected in the predicted means for the events marked with an asterix in Tables 4 and 5.

Jokl and Jokl [9] stated that in some athletic events the ultimate in performance appears to have been reached, but that in swimming events the prediction of ultimate performance is still not possible as the performance curves are still showing an improvement trend. This concurs with the trends shown in the current study.

In some cases the predictive models were able to be improved through the omission of, or adjustment to, certain data. The data that was omitted was that which is suspected to include the results of swimmers who had used performance-enhancing substances. This issue, and that of the adjustment of some data to address the issue of hand-timing is discussed in later sections.

Hypothesis 2

Hypothesis 2 stated that more variance will be explained through the use of non-linear models, therefore reducing the amount of error variance. The results of this study indicate that this hypothesis was supported in full.

In the present study, using all data, the linear models were the least accurate in all but four events (50m Freestyle for both men and women, Men's 1500m Freestyle and Women's 400m Freestyle). The highest coefficient of determination for these events was, in all cases, derived from a non-linear model.

The coefficients of determination for these events are, when looking at the least accurate model, in general, lower than those in Edwards and Hopkins [4] study, where the coefficients of determination where between 0.956 and 0.991. The study by Edwards and Hopkins [4] investigated the world record in only men's running events and as such did not include a large amount of data as used in the present study. Therefore, the trends of the Edwards and Hopkins study may not be totally representative of the changes in overall performance in running events and may be dissimilar to the trends of elite swimming performance as determined in the present study.

The concept of non-linear models and an asymptote of human performance was discussed by Jokl and Jokl [7], [8] and Telford [22]. Where a linear model is used it is expected that improvements occur at a steady rate and is entirely impossible to predict a point where the ultimate performance has been reached, as according to a linear model there is always the possibility of improvement. Telford ([22], p.2) noted that while "the asymptote of human performance is ever approached, it is never achieved". This is not to say that performances will improve at a steady rate, which is indicated by linear modeling, rather that performances will improve but according to the *Principle of Diminishing Returns* (Fowler [23]). This Principle indicates that improvements in performance will become smaller and more difficult to achieve as time progresses. This is apparent in the changes of world records which are being broken less regularly and by smaller amounts.

The *Principle of Diminishing Returns* is also discussed by Chatterjee and Laudato [24] who concluded in their study of World Record times that the rate of improvement for men is decreasing and this rate for women is increasing. This itself indicates a non-linear nature in performance curves. A linear trend would indicate that the rate of improvement remains constant over time. Such a rate of improvement would be represented on a time/date chart by the gradient of the line that shows the rate changes over time.

Hypothesis 3

The hypothesis that, the effects of steroids and other performance enhancing substances will be seen to a greater degree in the womens' results and correction for this will allow better predictions, was tested by removing the data from the years 1976 to 1992. These years were removed due to suspicions of drug use in swimming. Of the ten individual events contested in 1996 only five were able to be compared after the exclusion of the data from 1976 to 1992. Exclusion of events was due to those events being contested only twice in the time-frame considered as all of the models used are able to fit a line between two points with 100% accuracy, as indeed, is any model.

The hypothesis of greater accuracy was refuted for all events with the exception of the Men's 100m Freestyle. The predictions for all other events were considerably faster than the actual result when made using the data up to 1972. The comparisons of predictions and actual performances are shown in Table 6 Comparisons of Predictions Made Using All Data and 'Steroid Free' Data With Actual Performances in 1996.

Table 6

Comparisons of Predictions Made Using All Data and 'Steroid Free' Data With Actual Performances in 1996

Event		All data		Steroid Free	Actual
Men's	100m Freestyle	48.28	*	49.40	49.31
	400m Freestyle	218.54		213.81	230.57
	1500m Freestyle	854.92		851.66	910.29
Women's	100m Freestyle	54.62		52.62	55.37
	400m Freestyle	231.33		229.05	250.00

* All time recorded in seconds.

It was expected that removing the data from the years of steroid usage would have had the opposite effect to what is apparent in the data. That is, that the predictions of 1996 performances would have been more accurate with the years of known steroid use eliminated. Due to the greater effects of steroids when used by females it was

expected that the 1996 predictions for female events would have been greatly improved, given that it is acknowledged that the results from 1976 to 1988 were confounded by drug use by female East German swimmers (Helmsteadt [13], [14]; Verroken & Mottram [25]; Whitten [26], [27]). The 1992 results were similarly affected by drug use of Chinese female swimmers (Whitten [26], [27]). An assumption made when postulating this hypothesis was that the use of steroids and other pharmacological agents was not a factor at the 1996 Olympic Games. This assumption may have been false given that at least one finalist tested positive at the Atlanta Olympic Games and at least one other swimmer involved in a number of finals has been accused of drug use in the lead up to these Olympic Games (ESPNET [28]; Masters [29]). In the light of this, it is apparent that the use of steroids and other performance-enhancing substances has not been excluded from swimming events and may in fact be more widespread than is acknowledged. Given that former Eastern bloc countries tested their swimmers prior to competition and that a number of Chinese swimmers have been banned from elite swimming due to the use of performance-enhancing substances, it is entirely possible that swimming is still affected by the use of drugs. If this is the case then the models excluding the years 1976 to 1992 would not provide more accurate predictions as the trends during these years which include the use of drugs may become necessary parts of the models. The changes in performance may not only be due to the reasons stated by Dyer [19] such as improved technique, better facilities, larger pool of competitors and genetic changes, but also to the development of methods of chemical manipulation to enhance performance (The Parliament of the Government of Australia [30]).

By excluding the years 1976 to 1992 a large amount of data was ignored leaving a number of events with very few data points to use in the regression models. This in itself would affect the accuracy of the models. The inclusion of the greatest amount of data possible is necessary to generate the most suitable models.

Another factor to which the difference between the predictions with all data compared to the 'steroid free' data may be attributed to the effects of steroids in swimming. If steroids were to have no real effect on a swimmer's performance, such as in the distance events, then the inclusion of all data would be expected to provide the more accurate predictions as was the case for most of the events.

The blanket approach of ignoring all data from 1976 to 1992, in an attempt to determine the effects of steroid and other performance enhancing substance use in swimming may have been inappropriate. An approach such as this necessitates ignoring numerous other factors that may have been causal in the improvement of swimming times. These factors include, but are not limited to, improved training techniques, better coach education, greater rewards for elite performance and larger number of competitors involved in elite swimming (Australian Sports Commission [2]; Dyer [19]).

It is however impossible to remove the results of only those swimmers who are known to have used performance-enhancing substances. While it is acknowledged that a number of swimmers from East Germany were involved in a drug program, there remains some doubt as to whether swimmers from other countries also attained the status of Olympic finalist with the assistance of such substances (The

Parliament of the Government of Australia [30]). It was not within the scope of this study to discriminate between 'clean' swimmers and those who used drugs, hence the total removal of the suspect years in this aspect of the study.

There are four possible explanations for the non-confirmation of the hypothesis, that removing the years 1976 to 1992 would provide more accurate predictions of 1996 performance. These are explanations are:

1. Steroids may still be in use in elite swimming.

2. Steroids may not have positive effects in elite swimming.

3. Not enough data was able to be included in the calculation of the regression models.

4. The omission of the years 1976 to 1992 to counter the effect of steroids on elite swimming performance levels also removed the effect of a number of other factors.

Further investigation into the use of performance enhancing substances in swimming is therefore required. FINA, the International Olympic Committee and Australian Olympic Committee have acknowledged that steroids and other drugs are still in use in elite swimming and many other sports and are running an out-of-competition drug-testing program in order to combat the situation (The Parliament of the Commonwealth of Australia [31]; Verroken & Mottram [25]).

Ethical issues that arise in the area of testing for steroids in humans arise from explanation point two, that is the actual effects that steroids and other substances have upon human performance. This is due to the fact that whilst medicinal doses of androgenic-anabolic steroids are known to have adverse effects on the liver and other structures of the body, athletes have reported using up to 1000 times the normally prescribed dose of a steroid (The Parliament of the Commonwealth of Australia [31]; Wadler & Hainline [32]).

By using more data, possibly from world rankings lists, the problem of lack of data may be overcome. World Championships for swimming, which are contested every four years, began in 1973, and so, all but one World Championship fall into the period of time that is in question (1976 onwards). The use of world rankings from the years prior to 1976 is possible, however this data is not of the same source as the rest of the data. World rankings lists are composed of all data from competitions from each year. It is entirely possible that the best swimmers were not competing against each other at all meets. While the pools are standardised, the data for one event may come from any number of competitions with different rewards for successful performance and differing qualifying criteria and as a result the data is not collected in the same environment.

Hypothesis 4

Hand-timing in sporting competitions is acknowledged to be a source of error. When hand-timing is used the reaction time to the starting signal by each timekeeper may differ. In addition, the recorded time will be slightly faster than the actual time due to this reaction. Automatic timing, which is initiated by the starting signal and

stops as the swimmer touches the wall, may give a slightly slower time, as no human reaction for the timekeeper is involved in this process. The time given by an automatic timing system is the total time from the starting signal to completion of the event with no extraneous error. At the 1964 Olympic Games automatic timing was introduced and thus this source of error was removed.

Hypothesis 4 states that, corrections for hand-timing prior to 1964 will not greatly improve the fit of the models. The models derived, after incorporating a correction factor of 0.24 seconds, did not improve the fit of the models for many events. In fact there was no major change in either direction for any event. Only the model for the 100m Freestyle for both men and women was improved with the addition of this factor. The improvement in coefficient of determination, which was the largest change, for the Men's 100m Freestyle was from 0.858 to 0.976 and for the Women's 100m Freestyle 0.955 to 0.956. The improvement in time for the Men's 100m Freestyle was 0.37 seconds and for the Women's 100m Freestyle 0.45 seconds. These changes are much larger than for all other events.

Due to the outcome that the correction factor is relatively minimal compared to the duration of the events, for example, in 1996 49.31 seconds for Men's 100m Freestyle to 910.29 seconds for the 1500m Freestyle, it was expected that this correction would not have a large impact on the models. The correction factor is 0.49% of the Men's 100m Freestyle mean and 0.03% of the 1500m Freestyle mean in 1996.

For most of the events that were able to be included in the hand-timing corrected data the model of best fit did not change. Changes in the models were seen for the Men's 100m Freestyle. The 100m Freestyle model changed from a Sigmoidal model (R^2 = .858) using the data to 1992 to a Compound model (R^2 = .969) with the incorporation of the hand-timing correction. This appears to be the only major change in the models of best fit. It must however be noted that the Sigmoidal model in the corrected data set also has a coefficient of determination of .969. The difference is only seen at the fourth decimal place.

Based on these findings, Hypothesis 4 was therefore supported.

Hypothesis 5

The type of model for the Men's Freestyle events changed as the event duration increased. The 100m and 200m events were described by Sigmoid models and both the 400m and 1500m events were described by cubic models.

For the women's Freestyle events the model again changed as the distance swum increased. The 100m Freestyle was described by a cubic model, the 200m and 800m by s models and the 400m Freestyle by an inverse model.

An apparent anomaly, in this situation a point where the predicted times become progressively slower, occurs when a cubic model is used to describe events. According to the data up to 1992 this point does not occur in the men's 400m and 1500m Freestyle before the year 2100. Nevertheless, a turn-around point at 2106 is found for the women's 100m Freestyle. Such anomalies are due to the nature of the

cubic model which has an minimum value at the vertex of the curve. At this point the times are to predicted to in fact, decrease in the future.

The fact that the model changed as the duration of the Freestyle events increased, for both men and women, may be a reflection on the differing energy systems that dominate each event. For example, the 50m Freestyle events last less than 30 seconds and are performed utilising energy primarily derived from the anaerobic system, whereas the 400m, 800m and 1500m Freestyle events last between 3:40 and 15:30 and therefore rely predominantly on the aerobic energy system.

Most swimming events are longer than one minute in duration and a number of researchers (Åstrand & Rodahl [33]; Brookes & Fahey [34]; Brookes, Fahey & White [35]; Fox & Mathews [36]) have noted that aerobic glycolysis becomes a more important source of ATP for muscle contraction as the duration of the event increases. With this in mind, the fact that a number of events are described by similar models becomes reasonable.

In those cases where the same model described two or more events, the curve of best fit itself differed in the constants and coefficients that produce the curve. In this way no two events were described by exactly the same curve of best fit.

Hypothesis 6

Hypothesis 6, that the models will also be gender specific, was also partially supported. No single model was able to fit any or all events perfectly, that is $R^2 =$ 1.00. Again, where two or more events were represented by the same model, the coefficients and constants of the lines of best fit differed between events.

The final regression equations for a number of events were of the same type for both genders. The 50m Freestyle were both inverse functions, the 200m Freestyle events were best represented by Sigmoidal curves and the 400m Freestyle events by cubic curves. For the remaining events the model differed between genders.

Each regression equation was specific to the event and gender. Where one model described two or more events, the constants and coefficients that determine the lines of best fit cause the actual lines describing each event to differ.

The physiological differences between genders as noted by Wells and Plowman [37], O'Brien, Davies and Daggett [38], and Brukner and Khan [39] also helps to explain the differences in the lines derived for each gender in each event. These differences mean that each gender must perform each event in slightly different manners, utilising different muscle masses and distributions to move different body shapes through the water.

The similarities between the curves of best fit for each gender in a number of events can be seen in the Figures 1 to 10. There are many differences between the genders in terms of physiology (Carbon [40]; O'Brien et al. [38]; Rushall & Pyke [41]; Wells & Plowman [37]).

That the performance of both genders were able to be described by the same model for a number of events indicates that similar trends have occurred for both genders in these events. This may also indicate that there is a common construct or constructs underlying performance in both genders, as despite their physiological differences males and females are showing similar adaptations to training. The differences in swimming performance are, according to Kennedy, Brown, Chengalur and Nelson [42], and Arellano, Brown, Cappaert and Nelson [43], due to males having a greater stature and using this to develop longer stroke lengths.

A sigmoid model was by far the most common across both genders, in all events. This would seem to indicate that similar adaptations, by both genders in these events, to the new training techniques that have been implemented. This is in concurrence with the conclusions made by Chatterjee and Laudato [24] that overall improvements in performance, when represented graphically, depict an Sigmoidal-shaped curve. This shape of this curve is due to early variations in performance, rapid improvements and then followed by ever-decreasing improvements as the 'ultimate performance' is neared.

Table 7

Events Represented by the Same Model for Both Genders.

Event	Model
50m Freestyle	Inverse
200m Freestyle	Sigmoidal
400m Freestyle	Cubic

This does not however explain the differences, between the genders, in the form of line for the 100m Freestyle. A compound and a cubic model were used, respectively, to describe the 100m Freestyle for men and women. Given the similar energy demands for each gender in each event, these results appear to be an aberration. As the models this event are drawn from the same data set for each gender, in all cases the hand-timing set, this difference becomes hard to explain. Differences in the amount of data used may account for this. For example, the Men's 100m Freestyle has been contested since 1896 (data from 1908 onwards used) and the Women's 100m Freestyle was first contested in 1912. However, such differences in the amount of available data did not cause the same effect in other swimming events.

The models used to describe the trends in these two events were chosen after comparison of the coefficients of determination (R^2). For these events the coefficients of determination of a number of models were the same to the third decimal place and the particular model chosen from the fourth decimal place. At three decimal places the compound model for the Men's 100m Freestyle is no better than any other model with the exception of the linear and quadratic models. The cubic model for the Women's 100m Freestyle was the most superior in all respects. Therefore, it is

possible that the performances of both genders could be adequately described through the use of the same form of line.

If these changes to the models used were adopted both genders in each event would then be described by the same form of line and so the changes in performance within each event would be following similar trends.

Both 50m Freestyle events were described by inverse lines. The coefficients of determination for these events were both low (.543 for men and .732 for women). These events have only been contested at the past three Olympiads and perhaps as more data becomes available in the future these events may be able to be described by more accurate models.

Chatterjee and Laudato [24] concluded that the curve for the improvement within any one event would be Sigmoid-shaped. In three events analysed in this study, the curve of improvement is Sigmoid-shaped. For the two events represented by inverse curves, both 50m Freestyle events, there was not a large amount of data available for the analysis. With the collection of more data in these particular events, it may be possible that a Sigmoidal curve will be suited to the 50m Freestyle. The overall finding of non-linearity is in contrast with Edwards and Hopkins [4], who found a high degree of linearity for the progression of the World Record in a number of athletic events. The authors did acknowledge that it was expected that lines of improvement would be curvilinear in nature. Edwards and Hopkins [4] however, did not attempt to fit any non-linear models to the data used. It must also be noted that the Edwards and Hopkins [4] study involved World athletic records and the current study was used to investigate the changes in performance over time of the means of the Olympic finalists in swimming events.

Hypothesis 7

Hypothesis 7 stated that the use of non-linear models will result in less absurd predictions such as those of a zero time. An absurd prediction might also be one for the 2000 Olympic Games that is far in excess of current (1996) performances.

For all events, with the exception of the 50m Freestyle for both Men and Women, it was not possible to predict a zero time. The 50m Freestyle events were both described by an inverse line and a zero time was predicted for the year 2994 for men and in 2700 for women (refer to Table 4.33). These two events have only been contested at the past three Olympic Games and thus a large amount of data regarding the 50m Freestyle has not been collected at this stage. With such a small amount of data it is not possible to accurately analyse the long term trends in performance. In addition, the coefficients of determination for theses events (men 0.543; women 0.732) must be recognised. Low coefficients of determination result in less accurate predictive models. The mean performance of finalists for both men and women in 1996 in the 50m Freestyle were slightly slower than in 1992. Without further data to include in the models, it may be thought that ultimate performances in these events may have been reached. The predicted years of zero time are well in advance of the present year and with more data it is entirely possible that the models for each event will change and future models in addition to being more

representative of the changes in performance, that is the models will have higher coefficients of determination, may not allow such predictions.

It must be noted that the models derived are part of an evolving process and that as new data becomes available after each Olympiad this data should be included in the data set before new regressions equations are derived. In this manner the new models may be entirely different.

The predictions of zero time in athletic events by Edwards and Hopkins [4] using linear models ranged from 2228 to 2893. Predictions of zero time were made in all events. In the present study the predictions of zero time were much further into the future than for Edwards and Hopkins [4] study.

SUMMARY OF THE DISCUSSION

The changes in swimming performance at the Olympic Games are evident. Previous research of this type has primarily investigated changes in World Record performances. The mean of the finalists at the Olympic Games has been used in this study as it is believed that this would be more representative of overall changes in swimming performance. The best swimmers in the world compete against each other at the Olympic Games, under the same conditions, in an effort to be recognised as the best swimmer in the world.

Mathematical modeling has previously been used to predict sporting performance. However this is often based on linear regression. This study has shown that the use of non-linear models, when predicting future sporting performance, is more appropriate than the use of linear models. For every event considered, a non-linear model had a higher coefficient of determination, greater level of significance and smaller residuals than the linear model for the same event. The use of non-linear models also resulted in less absurd predictions being made. The 50m Freestyle events were the only events where a prediction of zero time was able to be made.

The use of steroids and other performance enhancing substances in swimming in previous years (1976 to 1992) is acknowledged. The removal of those years from the data set with the intention of deriving more accurate predictions was not successful. Only three events were predicted with greater accuracy with the years 1976 to 1992 removed from the data set. This may be a reflection of the lack efficacy of such drugs in swimming performance or perhaps be an indication that drugs are still being used by elite swimmers however the actual extent of this use of drugs is not detected by current drug testing protocols.

The correction for hand-timing, as expected, did not have a large impact on the fit of the models. The models for 100m Freestyle events for both men and women were improved only slightly with this correction. As these are the shortest events that could be analysed in this section this is an expected outcome.

It has been shown that a particular line was derived for each event. The models of each line however were remarkably similar with the majority of events predicted using a sigmoid line.

Finally, it was also shown that the mean performance of the finalists in each swimming event at the Olympic Games is expected to continue to improve in the near future.

CONCLUSIONS

While it is apparent that the times of Olympic finalists have, in general, been improving the manner of this improvement has not previously been investigated. Previous investigations have assumed that these improvements are linear in nature and that linear models sufficiently fit the available data.

The findings of this research has resulted in a number of conclusions and these are:

1. The results of this study indicate that the use of non-linear models in the analysis of phylogenetic trends in human performance is more appropriate than linear modeling.

2. In each event studied and in all aspects investigated a non-linear model generated the line of best fit.

3. From the regression equations generated it was possible to predict levels of future performance.

4. The predictions made as a result of this study may then be used by swim coaches, officials and swimmers to set training programs, training and competition goals and selection standards for future teams.

This research was not intended to reduce elite sporting performance to a series of mathematical functions, but rather to determine whether such functions can be used to extrapolate future levels of performance. It has been shown that mathematical models are able to be used for the purpose of extrapolating future results in addition to interpolation for those events which were not contested at one point in the past for any particular reason.

The regression equations may be useful for coaches and swimmers in determining the level of performance required at future competitions. Banister and Calvert [1] have identified this important issue that knowledge of future performance levels are essential in the development of training programs and the setting of goals for swimmers.

It has also been shown that the use of non-linear modeling techniques are more appropriate than the linear models as derived previously by Edwards and Hopkins [4]. By using the means of Olympic finalists it has also been possible to investigate the changes in performance at a greater depth. Edwards and Hopkins [4], Jokl and Jokl [9] and Prendergast [44] used world record performances in the predictions of future performance. As a World Record is a superlative performance, changes in this are not a true reflection on overall changes in performance.

Whether the trends at the Olympic Games for swimming over the past approximately 90 years (1908 to 1996) are representative of the phylogenetic changes in humans, such as physique of swimmers, or environmental changes, such as training, drugs in sport, sport, increased participation in swimming and physique of

swimmers (this construct can be changed by training as well as being genetic control), is at this point in time have been partially answered by this study. The development of each stroke into its current form, changes in training techniques and increased rewards and number of competitors have been identified as possible causal factors that improve swimming performance.

Carter [45], [46] has noted changes in the physical structure of Olympians. Competitors at the Olympic Games have been becoming taller and more muscular. Again, it is not known if these changes are entirely due to phylogenetic changes or to the auto-selection of athletes who have body types that are better suited to particular events. It is also possible that the structure of Olympic athletes is changing as a result of the development of new training techniques. Counsilman [47] has stated a belief that performances are improving as a result of new training methods, that is manipulation of environmental factors.

It has been shown in this study that performances have been improving and should continue to do so in the near future. Many events have displayed similar trends, in terms of the regression models that best fit the data, thus indicating that there may be an underlying construct of performance that is changing.

REFERENCES

[1]   Banister E.W. and Calvert T.W., Planning for future performance: Implications for long term training. *Canadian Journal of Applied Sports Science*, 5, 170-176 (1980).

[2]   Australian Sports Commission, Evaluation of the Australian Sports Commission's impact on sports performances and participation in Australia, (Canberra, 1993).

[3]   Rothstein A.L., Research design and statistics for physical education, (Prentice Hall, Englewood Cliffs, 1985).

[4]   Edwards D. and Hopkins W., Targets for the future, *Modern Athlete and Coach*, 17, (3), 34-36 (1979).

[5]   Norušis M.J., SPSS for Windows: Base systems user's guide. Release 6.0. (SPSS, Chicago, 1993).

[6]   Federation Internationale de Natation Amateur, Welcome to FINA. (http://www.fina.org., 1997).

[7]   Jokl E. and Jokl P., Running and swimming world records, *Olympic Review*, 107/108, 536 - 543, (1976).

[8]   Jokl E. and Jokl P., Running and swimming world records, *British Journal of Sports Medicine*, 10, 203 - 208, (1976).

[9]   Jokl E. and Jokl P., Running and swimming world records, *Journal of Sports Medicine and Physical Fitness*, 17, 213 - 229, (1977).

[10] Wallechinsky D., The complete book of the Olympics 1996 edition, (Aurum press, London, 1996).

[11] SPSS Inc., SPSS for Windows Release 6.1 [Computer Software], 1994.

[12] Georges C., Banned, but not forgotten: steroids won't just go away, but maybe their use can be made less pervasive in the Olympic sports. *Swimming Technique*, 20, (3), 15 - 19. (1983, November).

[13] Helmsteadt K., Otto talks ... finally. *Swim Canada*, 22, (2), 16, (1995).

[14] Helmsteadt K., Renate's story. How the little blue pills were the basis for success. Swim Canada, 22, (10), 5 - 7, (1995).

[15] Muckenfuss M., Spilling the Beans. After years of rumours, former East German athletes are talking about steroid use in their native land. *Swimming World*, 30, (11), 39 - 42, (1989).

[16] Abernethy B., Kippers V., Mackinnon L., Neal R., and Hanrahan S., The biophysical foundations of human movement. (Macmillan Education, South Melbourne, 1996).

[17] Schmidt R.A.,. Motor learning and performance: from principles to practice. (Human Kinetics Books, Champaign, 1991).

[18] Athletics Australia, National qualifying times.
http://www.ausport.gov.au/aths/natqual.html., (1997)

[19] Dyer K.F., Challenging the men: Women in sport, (University of Queensland Press, St. Lucia, 1982).

[20] Verroken M., Drug use and abuse in sport., In R. N. Mottram (Ed.)., Drugs in Sport (2nd Ed.), (pp. 235 - 263), (E. and F. N. Spon, London, 1996).

[21] Stefani R.T., Trends in Olympic winning performances, *The Athletic Journal*, 58, (4), 44 - 46,(1977).

[22] Telford R.D., The Olympics: "We did well but we must proceed", *Excel*, 5, (2), 2, (1988).

[23] Fowler B., Principles of training, *Modern Athlete and Coach*, 21, (1), 25 - 28, (1983).

[24] Chatterjee S. and Laudato M., An analysis of world record times of men and women in running, skating, and swimming, *Journal of Strength and Conditioning Research*, 10, 274 - 278, (1996).

[25] Verroken M., and Mottram R.N., Doping control in sport, In R. N. Mottram (Ed.), Drugs in Sport (2nd Ed.), (pp. 18 - 85), (E. and F. N. Spon, London, 1996).

[26] Whitten P., Red star over Atlanta, *Swimming World and Junior Swimmer*, 35, (12), 50 - 52, (1994).

[27]   Whitten P., China's short march to swimming dominance: Hard work or drugs?, *Swimming World and Junior Swimmer*, 35, (1), 34 - 39, (1994).

[28]   ESPNET, ESPN Sports Zone. Http://espnet.com., (1996).

[29]   Masters R., Critics cry foul over pluck of the Irish. *Sydney Morning Herald*, (July 24, 1996, p1).

[30]   The Parliament of the Commonwealth of Australia, Drugs in sport. An interim report of the Senate Standing Committee on the Environment, Recreation and the Arts, Commonwealth Government Printer, Canberra, 1989).

[31]   The Parliament of the Commonwealth of Australia, Drugs in sport. Second Report of the Senate Standing Committee on the Environment, Recreation and the Arts, (Commonwealth Government Printer, Canberra, 1990).

[32]   Wadler G.I. and Hainline B., Drugs and the athlete, (F. A. Davis Company, Philadelphia, 1989).

[33]   Åstrand P-O and Rodahl K., Textbook of Work Physiology: Physiological bases of exercise, (McGraw-Hill, New York, 1977).

[34]   Brookes G.A. and Fahey T.D., Exercise physiology: Human bioenergetics and its applications, (Macmillan Publishing, New York, 1985).

[35]   Brookes G.A., Fahey T.D. and White T.P., Exercise physiology: Human bioenergetics and its applications, (2nd ed.), (Mayfield Publishing, California, 1996).

[36]   Fox E.L. and Mathews D.K., The physiological basis of physical education and athletics (3rd ed.), (Saunders College, Philadelphia, 1981).

[37]   Wells C. and Plowman S., Sexual differences in athletic performance: Biological or behavioural?, *Physician and Sports Medicine*, 11, (8), 52-63, (1983).

[38]   O'Brien M., Davies B. and Daggett A., Women in sport, In B. Davies and G. Thomas (Eds.), Science and sporting performance: Management or manipulation?, (Clarendon Press, Oxford, 1982).

[39]   Brukner P. and Khan K., Clinical sports medicine, (McGraw-Hill Book Company, Sydney, 1995).

[40]   Carbon R.J., The female athlete, in J. Bloomfield, F. A. Fricker, & K. D. Fitch, Textbook of Science and Medicine in Sport, (Blackwell, Melbourne, 1992).

[41]   Rushall B. and Pyke F., Training for sports and fitness, (Macmillan Education, Melbourne, 1990).

[42]   Kennedy P., Brown P., Chengalur S.N. and Nelson R.C., Analysis of male and female Olympic swimmers in the 100 meter events, *International Journal of Sport Biomechanics*, 6, 187-197, (1990).

[43] Arellano R., Brown P., Cappaert J. and Nelson R.C., Analysis of 50-, 100-, and 200-m freestyle swimmers at the 1992 Olympic Games, *Journal of Applied Biomechanics*, 10, 189-199, (1994).

[44] Prendergast K., What do world running records tell us?, *Modern Athlete and Coach*, 28, (1), 33 -36, (1990).

[45] Carter J.E.L., Age and Body Size of Olympic Athletes, In J.E.L. Carter (Ed.), Physical structure of Olympic athletes, Part II, Kinanthropometry of Olympic athletes (pp. 53 - 79), (Switzerland: Karger, Switzerland, 1984).

[46] Carter J.E.L., Somatotypes of Olympic Athletes, In J.E.L. Carter (Ed.), Physical structure of Olympic athletes, Part II, Kinanthropometry of Olympic athletes (pp. 80 - 109), (Karger, Switzerland, 1984).

[47] Counsilman J.E., The science of swimming, (Prentice Hall, Englewood Cliffs, 1978).

THE AFL FINALS: IT'S MORE THAN A GAME

George A. Christos[1]

Abstract

The McIntyre system, which is currently used in the AFL finals, is *inconsistent, uncompetitive and offers the top two or three teams in the final-eight too much advantage* to win the premiership flag. If one takes into account home-ground advantage and the form of the teams over the season, one would conclude that the top two teams each have of the order of a 30% chance, or more, to win the premiership, while the teams finishing in seventh and eighth position only have around a 1% chance, or less, of winning the premiership. **We believe that this differential is far too great to generate public interest in the finals series, especially since the top teams play the bottom teams in round one in the McIntyre system.**

We have considered various other finals systems which *progressively* aim to correct the imbalance towards the top end teams, and make the matches *much closer* and more interesting. Some of these systems use a specific rule, while others involve an element of chance in that a draw is conducted to see which teams, under certain constraints, actually get to play each other in each of the finals. In all of these new systems, there are 10 finals matches, compared to 9 in the McIntyre system. This should offer the AFL an *additional incentive* to improving the finals system.

1. THE MCINTYRE SYSTEM

The AFL uses the McIntyre system for the finals matches. The top 8 teams from the home-and-away series of matches take part in the finals. In the first round in the McIntyre system, the team on top of the ladder plays the team in eighth position, the team in second position plays the team in seventh position, the third team plays the sixth team, and the fourth team plays the fifth team, as illustrated in Figure 1 below. The matches are played at the home-grounds of the top 4 teams.
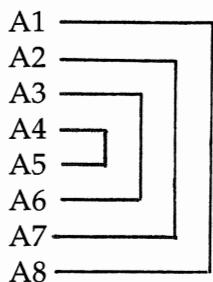
Figure 1: The first round matches in the McIntyre system, for the final-eight teams in specific order A1(top) to A8(bottom). Connected teams play each other at the home-ground of the higher ranked team.

---

[1]   School of Mathematics and Statistics, Curtin University of Technology, GPO Box U1987, Perth, Western Australia, 6845. Email: christos@cs.curtin.edu.au

The teams are then ordered with the 4 winners, keeping the same relative order as in the final 8, followed by the 4 losers also in the same relative order as in the final 8. The 2 lowest ranked losers are eliminated from the competition. As an example, let us suppose that the 4 winners from round one (r1) are teams A1, A3, A5, and A7, then the 6 teams that proceed to round two (r2) are ordered as follows:

```
A1 - bye
A3 - bye
A5 ⌐
A7 ┤ ┐
A2 ┘ │
A4 ───┘
```
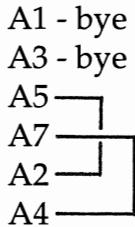
Figure 2: An example of the 6 remaining teams in the finals after r1, also showing how these teams are matched to play in r2 in the McIntyre system.

In this example, teams A6 and A8 were eliminated after r1, as they were the 2 lowest ranked losers. The top 2 teams, here A1 and A3, do not play in r2, while the other 4 teams play as indicated above, A5 plays A2, and A7 plays A4, at the home-grounds of A5 and A7 respectively. **The McIntyre system is flawed or inconsistent** here, since it is *unfair* that a lower ranked team A4 gets to play a weaker opponent (A7) than A2, who must play A5, even though A2 is ranked above A4. Notice also that, in this example, A4 did not win in r1 to earn this privilege. By the same token, A5 also undeservedly gets a much tougher assignment than A7. The reason that this particular choice of matches is made in the McIntyre system is clear since otherwise, if A5 plays A4, and A7 plays A2, one would have the same matches as in r1. The 2 losers from the r2 matches are eliminated, leaving 4 teams, the 2 winners and the 2 teams that had a bye in r2, in this example A1 and A3. For illustration purposes let us suppose that A5 and A4 win in r2. In this case the 4 remaining teams are then ranked in the order shown in Figure 3.

```
A1 ───┐
A3 ⌐  │
A5 ┘  │
A4 ───┘
```
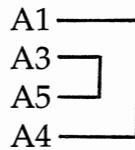
Figure 3: The r3 matches in the McIntyre system for the example considered.

In round three (r3), or what are also called the preliminary finals, the teams play each other as indicated above, in this example, A1 plays A4, and A3 plays A5. The 2 winners proceed to the Grand Final.

2.   THE RAW 'PREMIERSHIP PROBABILITIES' IN THE MCINTYRE SYSTEM:ASSUMING EVERY GAME IS A EVEN CHANCE FOR BOTH TEAMS

It is a simple matter to calculate the probability that any team in the final 8 will go on to win the premiership flag from their position in the final 8, under the assumption that every game is even, or 50:50 for each team to win. Consider for

example the team finishing in third position A3. Out of the 16=2x2x2x2 possible outcomes to the 4 matches in r1, A3 will finish in a top 2 position on 6 occasions, in a bottom 4 position on 8 occasions, and is eliminated on 2 occasions (when A1 and A2 also lose). From a top 2 position in r2, A3 needs to win 2 more games to win the premiership (probability 1/4) and from a bottom 4 position, A3 needs to win 3 games to win the premiership (probability 1/8). Therefore the probability that A3 wins the premiership is equal to $\frac{6}{16} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{8} = \frac{5}{32} = 0.15625 = 15.625\%$. In a similar way one can calculate the premiership probabilities for all of the other teams in the finals competition, which are given in Figure 4.

Final 8	raw premiership probability
A1	18.75%
A2	18.75%
A3	15.625%
A4	12.5%
A5	12.5%
A6	9.375%
A7	6.25%
A8	6.25%

Figure 4: The raw premiership probabilities that each of the teams in the final 8 will win the premiership flag assuming that every match is an even chance for either team to win.

Another fault of the McIntyre system is that the top 2 teams have the same premiership probabilities, so there is no incentive to finish on top of the ladder (minor premiers) after the home and away series. The McIntyre system also treats A7 and A8 equally and also makes no real distinction between 2 pairs of teams playing in r2, in our example between A5 and A7, and between A2 and A4.

3.   THE 'DRESSED' PREMIERSHIP PROBABILITIES IN THE MCINTYRE SYSTEM: TAKING FORM AND HOME-GROUND ADVANTAGE INTO ACCOUNT

If one disregards the inconsistencies in the McIntyre system mentioned above, the raw premiership probabilities in Figure 4 look quite reasonable, but they do not take into account

- the form of the teams that are playing each other.
- the fact that the top teams generally get to play the weakest teams, for example, in r1, A1 plays A8, and A2 plays A7.
- the home-ground advantage to the top teams.
- the fact that 2 of the top teams get a week off in r2. This gives them an advantage since their players have more time to recuperate from injuries.

We have performed calculations which take into account the first three of these factors, as detailed below. The probabilities calculated in this way are referred to as 'dressed' probabilities, in comparison to the 'raw' probabilities calculated in Section 2, where every match was considered to be an even chance for both teams playing. *These calculated dressed premiership probabilities are given in Figure 5.*

To illustrate our technique consider the first round match between the teams that finished first and eighth on the ladder last year (1997) after the home and away games, that is St Kilda and Brisbane respectively. St Kilda won 15 games out of 22, while Brisbane won 10 games and drew one game out of 22. Therefore the probability that St Kilda wins in general (or its 'win-match ratio') is $15/22 \approx 0.6818$, and the probability that it loses is $7/22 \approx 0.3182 = 1 - 0.6818$. Brisbane has a probability of $10.5/22 \approx 0.4773$ to win and a probability of 0.5227 to lose. The probability that St Kilda would win against Brisbane is therefore proportional to $(0.6818)(0.5227) \approx 0.3564$, and the probability that Brisbane would win against St Kilda is proportional to $(0.4773)(0.3182) \approx 0.1519$. These factors need to be normalized so that either one of these outcomes occurs with probability one, ignoring the possibility of a draw. This means that the probability that St Kilda wins is equal to $0.3564/(0.3564 + 0.1519) \approx 0.7012$, and the probability that Brisbane wins is equal to $0.1519/(0.3564 + 0.1519) \approx 0.2988$. In our calculations we have generally made an allowance for the home-ground advantage by shifting these probabilities by 0.1, or 10% from the team playing away to the team playing at home. In some cases where there was no clear home-ground advantage no adjustment was made, while in other cases smaller adjustments were made. We believe that our modelling of home-ground advantage is quite conservative. One could perform more detailed calculations in this regard by looking at each team's specific record in the home and away games, in much more detail. If one uses a simple shift in probabilities then the probability that St Kilda would win is equal to $0.8012 \approx 80\%$ and the probability that Brisbane would win is equal to $0.1988 \approx 20\%$.

Using the win-match ratios for the 8 finalists in 1997 (SK 15/22, G 15/22, WB 14/22, A 13/22, WC 13/22, S 12/22, NM 12/22, B 10.5/22) one can calculate the probabilities for the outcomes to each of the other r1 matches and from this deduce the probability for each of the 16 possible outcomes to the 4 r1 matches, which now do not occur with equal probability 1/16, as in the undressed situation.

For each of the 16 possible outcomes to the r1 matches there are 4=2x2 possible outcomes to the two r2 matches. One can calculate the probabilities for each of these r2 outcomes to occur, using the updated win-match ratios for the competing teams. Each r2 outcome leads to a different set of r3 matches which in turn each have a further 4 possible outcomes with an associated probability. Finally one can calculate the probability that either one of the two remaining teams wins the Grand Final. In essence, there are $512 = 16 \times 4 \times 4 \times 2 = 2^9$ (9 finals matches in all) different possible scenarios that can arise, which we have examined in deriving the results given in Figure 5. Each of these 512 possible outcomes has an associated probability of occurrence, which is given by the product of all of the associated r1, r2, r3, and Grand Final probabilities that lead to this result. The probability that some team, say Adelaide, goes on to win the flag is given by summing all of those associated probabilities, out of the 512 possible cases, where Adelaide wins the Grand Final.

Final 8 (1997)	dressed premiership probabilities
SK	~30%
Geel	~18%
WB	~17%
Adel	~14%
WC	~7%
Syd	~5%
NM	~8%
Bris	~1%

Figure 5: The dressed premiership probabilities, in the McIntyre system, for each of the teams in the 1997 AFL final 8 in the McIntyre system. Results from a detailed calculation sifting through all 512 possible match scenarios by hand.

It is clear from these results that the top few teams have a much higher premiership probability, and the bottom few teams have a much lower premiership probability, than calculated previously, where form and home-ground advantage were completely ignored. We regard this as **the most serious fault of the McIntyre system**. Note that, in these calculations Geelong's premiership probability was adversely affected by the fact that they had to play North Melbourne at the MCG, North Melbourne's home-ground, in r1. If for example, a non-Victorian team had finished in seventh position, then Geelong's premiership probability would have also been around 30% like St Kilda, and that team in seventh position would have had a premiership probability of around 1% like Brisbane. We believe that the McIntyre system offers far too much advantage to the top few teams in the final 8 to win the premiership flag. The problem arises because the top teams generally get to play the weakest teams, and generally get to play at their home ground. In addition, as we noted earlier, we have not taken into account the fact that two of the top teams do not have to play in r2.

It is also important to note that last year (1997) the AFL competition was regarded as one of the closest ever, so our results in Figure 5 would be further exaggerated if the competition was dominated by one or two teams, as is usually the case. In these circumstances the premiership probabilities for the top teams may be closer to 40% and the premiership probabilities of the bottom two teams may be much less than 1%.

We have investigated a number of other finals systems, which progressively aim to improve the prospects for the teams at the lower end of the final eight. Some of the systems that we have considered use a specific rule (or deterministic) algorithm, while others involve an element of chance, or randomness, to see who plays who under certain constraints.

4.   THE OUTER PAIRING SYSTEM $O_1 O_2 O_3$

In a deterministic system, that is, a system where there is a specific rule about who plays who, it is important that when the teams are paired to play each other,

that none of the lines, which join together these teams, cross each other, or otherwise a lower ranked team is necessarily given an easier task than a team which is ranked above it. This was observed in r2 of the McIntyre system, see Figure 2. To ensure 'fairness' in the fixtures, it is essential that these lines are in a concentric, or nested formation. The fairest arrangement of matches in r1 is as shown in Figure 1, as used in the McIntyre system. In the $O_1O_2O_3$ outer pairing system this principle of nested outer pairing is applied in r1, r2, and r3 (hence the notation), as depicted in Figure 6. The r1 matches are the same as in the McIntyre system. The 2 lowest losers from r1 are eliminated, leaving 6 teams in relative order to compete in r2. The difference with the McIntyre system comes about in r2, where now all 6 teams participate. The top 3 teams play the bottom 3 teams at their home grounds, and the 2 lowest losers are eliminated. In r3, the top 2 teams play the bottom 2 teams, and the 2 winners proceed to the Grand Final.
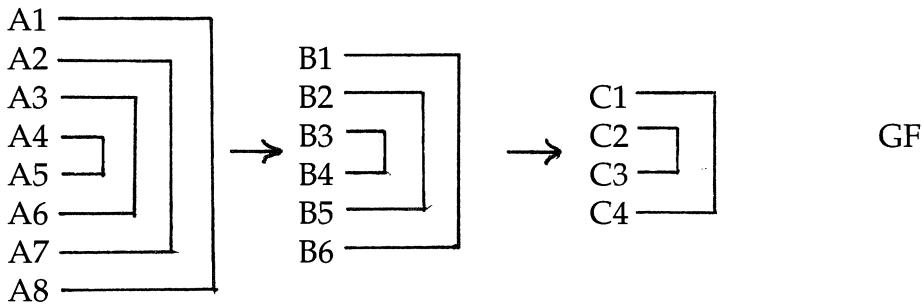
Figure 6: The r1 r2 and r3 matches in the $O_1O_2O_3$ system.

Using similar techniques as before, the raw premiership probabilities for the teams in the final 8 can be calculated. The results are given in Figure 7. As an example consider the team which finished in second position. A2 has a probability of $4/16=1/4$ that it will finish top of the 6 after r1, a probability of $4/16=1/4$ that it will finish in second position, a probability of $4/16=1/4$ that it will finish in fifth position, and a probability of $4/16=1/4$ that it will finish in sixth position. An analysis of the 8 possible outcomes to the r2 matches reveals that from these positions the probability that A2 will proceed to r3 is equal to 1, $6/8=3/4$, $4/8=1/2$, and $4/8=1/2$ respectively. Multiplying the corresponding probabilities together and summing over all of the possible outcomes gives the probability that A2 will make it to r3. This probability is given in curly brackets below. From r3 there is a further probability of 1/4 that A2 will win the premiership, 2 more games, so the overall premiership probability for A2 is equal to $\left\{ \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} \right\} \cdot \frac{1}{4} = \frac{11}{64} \approx 17.19\%$. Similarly one can calculate the premiership probabilities for each of the other teams in the final eight (Christos [1]), which are given in Figure 7.

Final 8	raw premiership probability
A1	18.75%
A2	17.19%
A3	14.45%
A4	12.11%
A5	12.11%
A6	9.77%
A7	7.81%
A8	7.81%

Figure 7: The raw premiership probabilities for each team in the final 8 in the $O_1O_2O_3$ outer pairing system.

One can see from these results, that in the $O_1O_2O_3$ system, A1 and A2 do not have the same raw premiership probabilities, as in the McIntyre system. The teams that finish seventh and eighth on the ladder also have a slightly higher raw probability to win the premiership than in the McIntyre system, but this increase is really quite small. Also in the $O_1O_2O_3$ system there are no logical flaws, as were found in r2 of the McIntyre system, where the team in sixth position had an easier match than the team above it in fifth position. Another good feature of the $O_1O_2O_3$ system is that the top 2 teams after r1 must also play in r2, and the bottom 4 teams in r2 do not have the same probabilities to proceed to the next round, as in the McIntyre system. This also means that there is an extra game played in the finals in the $O_1O_2O_3$ system, here 10, compared to 9 in the McIntyre system.

The $O_1O_2O_3$ system has certain advantages over the McIntyre system as we have noted above, but one may still argue that it does not go far enough in improving the prospects for the bottom end teams to win the premiership flag since in the $O_1O_2O_3$ system the top teams still play the weakest teams, in r1 and also in r2, and also enjoy a home-ground advantage.

A complete listing of all possible outcomes for the $O_1O_2O_3$ system reveals that there is a 1 in 8 chance that 2 teams that played each other in r1 will have to play each other again in r3. A similar listing for the McIntyre system reveals another interesting fact, which is, that it is impossible in the McIntyre system to have a Grand Final between A1 and A7 or between A2 and A8. In the context of last years finalists this means that a Grand Final between St. Kilda and North Melbourne, or between Geelong and Brisbane was impossible in the McIntyre system. This may be considered as yet **another fault of the McIntyre system**. There are no such unallowed possibilities in the $O_1O_2O_3$ system.

5.   THE RANDOM PAIRING SYSTEM $R_1R_2R_3$

Another 'clever' way around the problem of fairness in match fixtures is to introduce an element of chance into the system. We have considered a number of such systems. In the first of these random pairing systems, called the $R_1R_2R_3$ system, each team in the top half is randomly paired to play with a team from the

bottom half, for all matches in r1, r2, and r3. In r1 there are 24 different combinations of such matches that can be arranged in principle. Since these draws are conducted randomly, if some team is unlucky to draw a relatively more difficult opponent than a team ranked below it, then this is deemed to be the luck of the draw and the system cannot be blamed per se, as it is not the result of any specific rule. The top 4 teams still enjoy a home-ground advantage in r1, just as in the McIntyre system and the $O_1O_2O_3$ system. As before the 2 lowest ranked losers are eliminated in going to r2, and the winners and losers that proceed to r2 are ordered with respect to their relative order in the final 8. In r2, each of the top 3 teams are randomly paired to play a team in the bottom 3, with all matches played at the home ground of the top 3 teams. Once again the 2 lowest ranked losers are eliminated, leaving 4 teams to play in r3, where each team in the top 2 is randomly paired to play a team in the bottom 2, at the home-ground of the top 2 teams. A possible arrangement of matches in the $R_1R_2R_3$ system is shown in Figure 8.
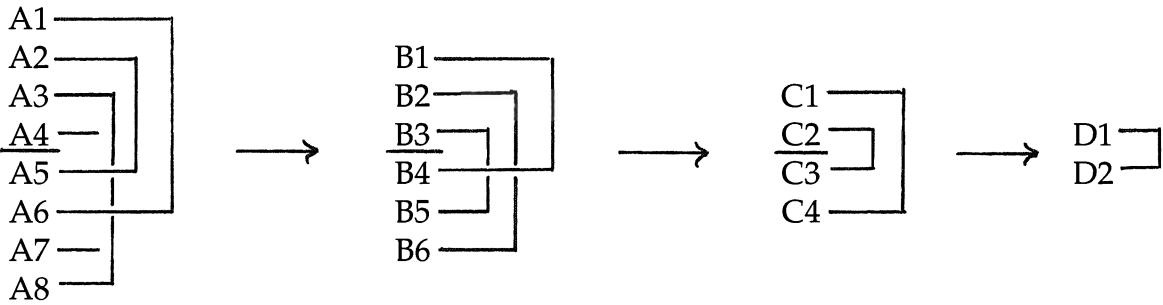
Figure 8: A possible arrangement of matches in the random pairing system, $R_1R_2R_3$, where in each round, teams from the top half are randomly paired to play teams in the bottom half. In each transition to the next round the 2 lowest ranked losers are eliminated.

On average one might expect that the matches in this system are closer than in the $O_1O_2O_3$ system, but this is actually not the case. If one was to count the differences in ladder position between the teams playing each other in r1 in the $O_1O_2O_3$ system (or the McIntyre system) the sum is 16. The difference in position between A1 and A8 is 7, the difference between A2 and A7 is 5, the difference between A3 and A6 is 3 and the difference between A4 and A5 is 1. This difference index in r1 in the $R_1R_2R_3$ system is actually also equal to 16, for any combination of matches (Christos [1]). At least here, two of the games (referring in particular to the matches A1 vs A8 and A2 vs A7, in the McIntyre system and the $O_1O_2O_3$ system) are not practically decided before the ball is even bounced. The difference index in r2 in the $R_1R_2R_3$ system is also the same as in the $O_1O_2O_3$ system, both equal to 9. Consequently the matches in the $R_1R_2R_3$ system are not really any closer than in the $O_1O_2O_3$ system, but at least the weakest teams do not necessarily get to play the best teams. In this respect the $R_1R_2R_3$. system is a progressive improvement over the $O_1O_2O_3$ system.

It turns out that the raw premiership probabilities for each team in the final 8 in the $R_1R_2R_3$ system are the same as in the $O_1O_2O_3$ system (Christos [1]). The

$R_1R_2R_3$ system does however give the bottom end teams a slight advantage over the $O_1O_2O_3$ system because, as noted above, they may now not have to play the very best teams.

In the $R_1R_2R_3$ system, there is potential problem because there is quite a high probability that two teams that played each other in r1 may be selected to play each other again in r2 or in r3, and two teams that played each other in r2 may be chosen to play each other again in r3. It is appropriate then to ask whether it is possible to rearrange matters so that rematches are excluded. In other words, is it possible to find combinations of matches in later rounds where there are no rematches between teams that played in earlier rounds. It turns out that there is a 1 in 8 chance that a r1 rematch in r3 cannot be avoided, even with a redraw option (Christos [1]).

6.   THE GENERAL RANDOM PAIRING SYSTEM $G_1G_2G_3$

As note above, the $R_1R_2R_3$ system is only marginally better than the $O_1O_2O_3$ system in that the bottom 2 teams will on average play less formidable opponents. One can actually do better if, in the random pairing process, one does not insist that a team in the top 4 needs to play a team in the bottom 4, but instead allows the teams to be paired together from any position on the ladder. In this system, which we have called the general random pairing $G_1G_2G_3$ system, A1 can play any one of the other teams. There are actually 105=7x5x3 different combinations of matches possible in r1. All games are played at the home-ground of the team in the highest relative ladder position.
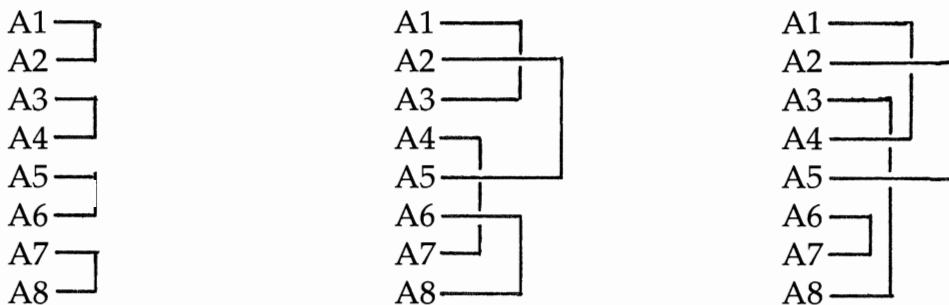
Figure 9: Some possible arrangement of matches in r1 in the $G_1G_2G_3$ system where teams are randomly paired from any ladder position to play each other. The matches are played at the home-ground of the team with the highest ranking for each chosen pair. The matches in this system are clearly much closer in general than in the $R_1R_2R_3$ system.

The key advantage of the $G_1G_2G_3$ system over the $R_1R_2R_3$ system (and the previous $O_1O_2O_3$ system) is that the matches are now in general much closer, see Figure 9. It also turns out (Christos [1]), that rematches can be totally avoided in the $G_1G_2G_3$ system, because there are many more combinations of matches possible in this system. In the $G_1G_2G_3$ system, some of the bottom 4 teams also have a small probability that they might host a home-final. Clearly A1 will always host a home-final in r1, while A8 can never host a home-final in r1. A2

will host a home-final unless it is drawn to play A1, which occurs with probability 1/7. Therefore A2 plays a home-final in r1 with probability 6/7. A7 will only get to host a home final in r1 if it plays A8, that is, with probability 1/7. These probabilities are directly related to the number of teams above and below each given team. From top position down to the eighth position these probabilities are equal to 1, 6/7, 5/7, 4/7, 3/7, 2/7, 1/7, and 0 respectively. In r2 the corresponding probabilities, from top position to sixth position, are equal to 1, 4/5, 3/5, 2/5, 1/5 and 0 respectively. This situation should be contrasted to the $O_1O_2O_3$ and $R_1R_2R_3$ systems, where only the top 4 teams in r1 get a home-final and the top 3 teams in r2 get a home-final. In this respect, the $G_1G_2G_3$ system further equalises the premiership probabilities for the teams in the final 8.

The raw premiership probabilities in the $G_1G_2G_3$ system can be calculated by tracing all possible paths with a positive outcome for each particular team to win the premiership, under all possible random draws of matches (Christos [1]). The raw premiership probability for A1 in the $G_1G_2G_3$ system is equal to 18.75%, which is exactly the same as in the McIntyre, $O_1O_2O_3$ and $R_1R_2R_3$ systems, but here, in the $G_1G_2G_3$ system, A1 (and A2) can expect a much tougher assignment than in any of the other systems. In fact A1 may even get to play A2. The premiership probability for A8 is equal to 7.5%, which is more than in the McIntyre system (6.25%), but is a little less than in the $O_1O_2O_3$ and $R_1R_2R_3$ systems (7.81%). We believe that the greater evenness in the competition in the $G_1G_2G_3$ system far outweighs this slight decrease in the raw premiership probability for A8, relative to the $O_1O_2O_3$ and $R_1R_2R_3$ systems. Recall that form and home-ground advantage had a devastating effect on the premiership probability for A8 in the McIntyre system, where it went from 6.25% to approximately 1% or less. In the $G_1G_2G_3$ system, the impact of these factors on A8 are substantially reduced. A7 has the same premiership probability as A8, that is 7.5%. A7 and A8 are however differentiated in the $G_1G_2G_3$ system because A7 has a small probability that it may host a home-final, against A8 in fact. The premiership probability for A2 is $\approx 16.79\%$, which is slightly less than in the $O_1O_2O_3$ and $R_1R_2R_3$ systems. Similarly the other raw premiership probabilities in this system can be shown to be comparable to corresponding probabilities in the other systems that we have considered, but what is more important here is that, in the $G_1G_2G_3$ system the competition is much more even, after form and home-ground advantage are taken into account, the matches are generally much closer (see below), no teams get a week off from the finals, and even the home-ground advantage is partly averaged, with small probabilities that some of the bottom 4 teams may also host a home-final. **We emphasis however that it is still of great advantage to finish closer to the top of the ladder, in this system, because the top teams have a higher probability to secure a home-final, their probabilities to survive each round are generally higher, and they generally play teams weaker than themselves.**

In the $G_1G_2G_3$ system, the difference index can be much smaller than in the previous systems considered, for instance for the first combination of matches shown in Figure 9 the difference index is only 4. The average difference index in the $G_1G_2G_3$ system in r1 is equal to 12 (Christos [1]), compared to 16 in the

$R_1R_2R_3$ and $O_1O_2O_3$ systems. In r2 the $G_1G_2G_3$ system has an average difference index of 7 compared to 9 in the $R_1R_2R_3$ and $O_1O_2O_3$ systems.

7. SPLIT PAIRING SYSTEMS

There is another group of deterministic systems where the outer pairing principle is not strictly adhered to, but any violations of fairness are compensated by other factors. Consider for example the situation in r1 where the top 4 teams play each other in a nested formation, while the bottom 4 teams also play each other in another nested formation, as shown in Figure 10.
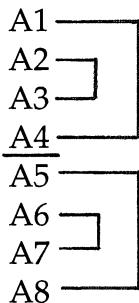
Figure 10: A deterministic system with a split competition in r1 where the top 4 teams play each other as indicated, and the bottom 4 teams play each other as indicated. The two losers in the bottom half are eliminated.

The top 4 teams are then reordered with the two winners, preserving the previous relative ordering of these teams, followed by the two losers, also preserving their previous ordering in the final 8. The final 6 for r2 is completed with the 2 winners from the bottom half matches. The 2 losers from the bottom half matches are eliminated. The nice feature about this system is that these matches are **much closer** than the in the McIntyre system, the $O_1O_2O_3$ system, or the two random pairing systems, $R_1R_2R_3$ and $G_1G_2G_3$, on average. In the deterministic split pairing system, illustrated above, the difference index is equal to (3+1)+(3+1) = 8 in r1, which is generally much less than that for the other systems considered so far.

Although one might argue that these split systems seem to be unfair in r1, in that A4, which is ranked above A5, gets a much more difficult opponent, namely A1, compared to A5, who gets to play A8 at home, it should be noted that A5 can be eliminated in r1, whereas A4 cannot, and also A4 will remain above A5 in r2 even if it loses and A5 wins. A5 and A6 can now be eliminated in this system, but this is compensated by the fact that A5 and A6 get a home-game and also get to play a weaker team than themselves, unlike the situation in the other deterministic systems.

Consider the system $S_1O_2O_3$, where split pairing is used in r1, and outer pairing is used in r2 and r3. The premiership probability for each of the bottom 2 teams is equal to $(1/2)^4$=0.0625, since these teams must win 4 games in succession in order to win the premiership. This probability is less than in the $O_1O_2O_3$ system, and the same as in the McIntyre system, but this is counterbalanced by the fact

that the bottom two teams now get to play less formidable opposition in r1, namely A5 and A6, instead of A1 and A2. The premiership probability for the top team A1 is equal to = 13/64 = 20.3125%, which is a little higher than in the McIntyre and the $O_1O_2O_3$ systems, but once again this is counterbalanced in the split pairing system by the fact that A1 must now play A4, a much better opponent than A8, as in the McIntyre system and the $O_1O_2O_3$ system. The team in second position in the final 8, that is A2, has a premiership probability equal to 12/64 = 18.75%, which is slightly less than the premiership probability for A1 in this split system, but slightly more than the corresponding probability in the $O_1O_2O_3$ system, and the same as in the McIntyre system. Although this probability is slightly higher in this split system, than the $O_1O_2O_3$ system, A2 must now play A3 instead of A7 in r1, as in the other deterministic systems. When the form of each team, and the home-ground advantage are taken properly into account, we expect that the dressed probabilities will be much more even in the split competition than in the other deterministic systems. We have not carried out these calculations here.

The $S_1O_2O_3$ system does however have a problem with regard to rematches. There are no rematches in r2 from r1, and no rematches in r3 of matches played in r2, but there is the possibility rematches in r3, that took place in r1. It turns out that in 1 in 8 of the outcomes to the r2 matches (for each outcome to the r1 matches), there is one rematch (between A4 and A1), and in another 1 in 8 outcomes to the r2 matches there are two rematches simultaneously (A1 vs A4 and A2 vs A3). The former rematch takes place when the outcome to the r2 matches is 'all bottom 3 teams win', and the latter double rematch occurs when the outcome to the r2 matches is 'all top 3 teams win'. These rematches can be avoided if the usual outer pairing combination in r3 (Figure 6) is changed so that C1 plays C3, and C2 plays C4. One can also consider systems such as $S_1S_2O_3$,where a (2+4) split pairing arrangement is used in r2 (Christos [1]) but this system is however plagued with a very serious and incurrable rematch problem.

Another interesting possibility is combine the splitting idea with random pairing. As an example, in r1 the ladder could be split into a top 4 and a bottom 4 and two pairs could be chosen at random in each of these groups. We might symbolize this system in r1 by the notation R^S_1. This system is interesting because it has an average difference index in r1 of $20/3 \approx 6.67$, which is even less than the difference index for the deterministic split pairing arrangement S_1. In r2 and r3 one could have a general random pairing arrangement, so a possible overall system might be $R^S_1G_2G_3$.

One could also consider systems where there are deterministic pairing rules in some rounds and random pairing rules in other rounds. For obvious reasons, if one wants to avoid rematches, it is necessary to have all of the deterministic sub-systems in the earlier rounds and the random pairing sub-systems in later rounds. As noted earlier, in the general random pairing arrangement G_3, there is always enough freedom in r3 to ensure that there are no rematches from previous rounds, but this may not be true for R_3. A particularly interesting example of such a system that utilize both determinism and randomness is $S_1G_2G_3$.

8. HANGING THE NUMBER OF TEAMS IN THE FINALS

Another variable in the consideration of finals systems is to vary the number of teams that take part in the finals. In the case of the AFL there are 16 teams in the competition, and the top 8 teams participate in the finals. One could either extend this number up to 10, 12, or even 16, or one could reduce it to 6 or even 4. The problem with increasing the number of teams in the finals is that there are too many games in the finals series, unless more than two teams are eliminated in some rounds. The problem with having too few number of teams in the finals is that there are very few finals matches, and for most teams the season may have already ended well before the home and away games are completed, a criticism that incidentally also effectively applies to the McIntyre system.

9. SUMMARY

We started this investigation with a review of the McIntyre system that is currently used by the AFL. When we took season form and home-ground advantage into account we found that this system heavily favoured the top few teams, which made it almost impossible for the bottom few teams in the finals to win the premiership. We estimated that the top 2 teams each had approximately a 30 % chance of winning the premiership while the bottom 2 teams only had around a 1 % chance, or less, of winning the premiership. In the McIntyre system, the top teams generally also get to play the weakest teams in the finals, generally get to play at their home-ground, and the 2 top teams after r1 get a bye in r2. The bottom 2 teams on the other hand must win 4 games straight, must play the best teams and furthermore generally must play away. In the 1997 season the competition in the AFL was regarded as being one of the closest on record, and in view of this, our estimates of the differential between the top teams and the bottom teams is probably quite conservative. The situation would be even worse if the competition was not so close. The exaggerated prospects for the top end teams in the McIntyre system also detracts interest from the finals series, particularly since in the first round the top teams play the bottom teams. We also found some other inconsistencies or flaws in the McIntyre system.

We then proceeded to progressively improve the McIntyre system, first by making sure that all teams played in round 2, then by introducing random pairings of teams, top 4 to bottom 4, and then any random pairing at all. We calculated the raw premiership probabilities in these systems for each team in the final 8. We were forced to examine in some detail the question of rematches in later rounds between teams that may have already played each other in earlier rounds. We then considered systems where the teams are split into two groups that separately play amongst themselves, and finally we considered systems where a combination of these various strategies is used in different rounds. In the systems considered in this paper, the matches are generally much closer than in the McIntyre system, the dressed premiership probabilities are much more evenly distributed from the top team to the bottom team, and as a consequence the games are much more interesting. Finally we note that our general strategies, such as 'outer pairing', 'random pairing' and 'split pairing' can be applied to other sporting competitions with more than or less than 8 teams in the finals.

REFERENCE

[1]   Christos G., More Competitive Final Systems for the AFL and other Team Sport
       Competitions, Curtin University, School of Mathematics and Statistics,
       Technical Report 22/97 (1997).

SCHEDULING

Jim Cross[1]

Abstract

This paper discusses some of the problems associated with writing a computer programme in scheduling draws for football and basketball leagues.

1.   INTRODUCTION

Miller, Mangelsdorf and Cross have been scheduling football and basketball leagues for some time. Currently we advise half-a-dozen competitions on their draws. Originally David Wilson and Jim Cross approached Jack Hamilton after the VFL unsuccessfully sought help from the computer programming companies. This is not an uncommon occurrence: computer programmers write small programs to do programming things, to run computers. Writing a program to do something in the external world requires the ability to think out an algorithm, and then the writing skill to turn the algorithm into a program. One of my colleagues was hired last month precisely for the algorithmic skill: the (small) company had tried several computer programmers, and after trial settled for a mathematician who could program.

David Wilson is a genius. His program came after six months dedicated and obsessive work. It was suited to the language (FORTRAN) and machine (VAX/PDP11) of the time, to the style of draw required. His draws were for 1983-1985 seasons. He left for more congenial work.

Chuck Miller rewrote the program in two stages. The first took a draw as given and assigned venues. The second stage was to apply the same engine to make the draw. Both allow one to fix certain games in certain rounds and at certain venues, and complete the draw around these.

When the load became too great a couple of years ago, Christine Mangelsdorf joined us, and successfully ran the two draws for this year.

The elements of a program of matches depends on what kind of competition is being run. If there are 8 games each week with every team playing, and at least 8 different venues, then almost any program will allow you to get a program. An integer programming package will do a lot of it as an allocation problem. But there is a catch. Solutions are sometimes hard to find. For the TFL in 1994 the successful run took 90 minutes of CPU time even with favourable initial conditions; we have run problems for 24 hours without a solution being found.

---

[1]   Mathematics and Statistics Department, University of Melbourne, Parkville Vic 3052

2.   ELEMENTS OF A FOOTBALL DRAW

There are two parts to a draw:

1. who plays whom **when,** which **round**, e.g., Essendon v St. Kilda in rounds 1 and 8 of the actual draw for 1907; we call this the **listing**.

2. who pays whom **where,** which ground or **venue**, e.g. Essendon v St. Kilda at the East Melbourne Cricket Ground in round 1 of the accompanying 1907 draw; we call this the **scheduling**. Usually these can be done separately; in fact, it is normal to do them in the above order, i.e., to decide who plays whom in all the rounds and then to decide where they play.   If the schedule of "wheres" or venues is impossible or unsuitable, then we change the listing of the "whens" or rounds.

3.   THE FORMULA BEHIND THE DRAW

A formula which does the listing of the games in each round is about 1000 words long or three tight pages. The formula for the scheduling is about 8000 words long or nearly 25 pages. Actually the formula is not a single algebraic expression like those one might learn at school but is more like a recipe, a list of instructions of how to go about working through all the possibilities so that if a list can be scheduled then you can eventually find the schedule or list of grounds which fits the requirements clubs and supporters and the AFL and other people impose. The program keeps track of several thousand pieces of information while it is assigning grounds to matches, and in fact there are 12 different types of information it keeps using and reusing, and it does this for each pair of clubs. These requirements cover the number of home, away and neutral games, the sequences of home/away/neutral games, the balance over the two halves of the season, the special requirements of shared grounds, and other features. There are three formulae for football draws: one by Ken McIntyre, perfect for an odd number of teams and easily modified for an even number. Then there is David Wilson's stroke of genius, first really seen in the draw for 1987.

4.   PREPARATION

When we begin a draw we like to have some information quite fixed: all consultants like to have solid information to begin with!

Apart from the number of teams and grounds, and when the grounds are available, there are at least five factors which make life difficult or easy:

(1) How many times will team A play team B? Once, twice, three times? Wrong choices for a whole competition will make the draw infeasible.
(2) How many games must teams play at home? away? neutral? This is usually not too restrictive.
(3) Grounds: do teams share ground? What are the conditions imposed on ground usage? This can be very savage.
(4) How many days between games? Five clear days means that if you play on the Sunday you can't be scheduled for the next Friday, for example.
(5) Sequences: many competitions like alternate home/away. With an odd number of teams and a bye, this is always possible (you have HH or AA mid season), but

the draw is inflexible. With an even number of teams it is impossible for all teams.

(6) First round: are the home teams assigned or not? This causes grief.

5.   LISTING OF ROUNDS

Over the years 1930--1980, the VFL used a cyclic permutation based on work by Mr. Ken MacIntyre. The MacIntyre draw gave rise to the giant wheel Jack Hamilton used for many years; on it each draw or listing took about an hour or so. When teams had to be interchanged in order to accommodate the scheduling or because the scheduling was unsuitable, the whole process took another hour or so. The scheduling was somewhat inflexible, and the listing was often unsatisfactory: this MacIntyre listing was used in 1986, where teams played Essendon (on top) one week and Sydney (second) the next, and the complaints were loud. The two rounds mid season where you played two teams once only were hard to choose from season to season to make things "equal", and insistence on home first round one year and away first round the next built-in even more inflexibility.

The Wilson draw gives a listing where teams which played on the Monday of a split round could be listed to play among themselves the following week (see rounds 4 and 5 of 1987). This listing was used in 1987 to accommodate the scheduling of venues, particularly because of the two trios who shared Princes Park and the MCG that year (ha, ca, fi at PP, me nm, ri at the G).

With Miller's new method of scheduling where games are played, it is became possible to interchange rounds much more freely than was done in 1985, 1986, and 1987. Even parts of rounds can be interchanged to avoid many of the undesirable features of the MacIntyre listing and to diagnose problems with the scheduling of individual games -- in some cases the first game scheduled for (or assigned to, say, Princes Park) might be assigned to the ground which makes a scheduling impossible, i.e., the game has to be assigned to some other ground for scheduling feasible.

6.   EVEN DISTRIBUTION OF HOME GAMES THROUGH THE SEASON

Compare the two draws for the 1907 season. They satisfy the 1 in 3 at home rule, and they give each team 7 at home and 7 away. But Essendon has 5 of the first 7 at home and then 2 of the last 7 at home in one draw, while the balance is better in the other draw, with 4 of 7 and then 3 of 7 at home: you can do no better than that. This constraint is imposed in the formula by a rule: a maximum of 4 at home in the first 7, a minimum of 3 at home in the first 7; naturally, the second seven then looks after itself! For 1908, the East Melbourne Cricket Ground could be regarded as always "home" for Essendon and University, whomever they played. This makes scheduling easier but life more difficult for the gatekeepers who have to let home supporters into one side of the ground and the opposition's into the other where they have to face into the sun. We give examples of two such schedules, one with the home and away labelled, one without. In 1998 people noted that every team played four or five of its first nine games at home, and every team but two played three or four of their first seven games at home. This was deliberate.

7.   ALTERNATING HOME AND AWAY

Tom Hafey had a dream, so he said at VFL Park long ago and Warwick Capper repeated on air last month, that the Swans would play every second Sunday afternoon at the SCG - and win too, I suppose! If a team has its ground available every second weekend, then its schedule can always be fitted to either of the patterns (H for a home game, A for an away game).

HAHAHAHAHAH ... or AHAHAHAHAHAHA

See the 1983 draw for Sydney. See one of the 1907 draws for Essendon. In fact **two** teams can be so scheduled providing the above left hand pattern is used for one and the right hand pattern is used for the other: they have to play each other, so they can't both be scheduled at home on that day! See one of the 1907 draws for Essendon and St. Kilda. For three teams this can be done in full for two and in part for the third, and we give examples for Essendon-St. Kilda-Geelong for 1907 and Essendon/St.Kilda/University for 1908. The third team must have a pattern matching one of the first two teams, and when it plays that team then its (the third team's) schedule must be varied with home and away reversed for those two games. Such scheduling sometimes affects other teams badly, depending of the other restrictions.

8.   SEQUENCES OF HOME GAMES

Up to the introduction of VFL Park teams almost always had one or two games at home and then one or two games away. If we use H for a home game and A for an away game again, the sequence of home and away games looked like

HAHHAAHAHHAHAA ...

In general at least one game in every 3 was at home and at least one game in every three was away. If you look carefully at the draws for 1907 and 1908 given elsewhere you will see that this feature is built in. With the introduction of VFL Park this sequence was extended: now over half the games were not at home but at one's opponent's ground or the neutral venue of VFL Park Waverley. Since over half the games were not at home -- eventually there were 9 at home and 13 not at home - it is almost impossible to force the sequence of one in 3 at home for everyone, so the sequence became a rule of at least one in 4 at home and not more than three-in-a-row not at home. Shifting of games to the MCG and elsewhere was reintroduced when Jack Hamilton was commissioner in 1986. Teams ended up with fewer home games in 1986, and Essendon finished 1986 with a fairly even distribution of home, away and neutral games, with about 7 home games, 8 away, and 7 neutral games. When the number of home games is 7 in a total of 22, it is no longer possible to guarantee for every team that it will have at least one in 4 at home. Due to Easter Sydney had to play North in Melbourne at the beginning of 1986 and so North ended up with 4-in-a-row away mid-season. In 1987 a careful reading of the footy fixture will show that Carlton, Hawthorn and Footscray have been "blessed" with a sequence of four games not at home, corresponding to a rule of at least one in 5 at home. To get the current rule of no more than two in a row away, other choices have to be given up: Brisbane would like alternate home/away because of the travel, and it almost has this for

1998, and Sydney might still like it, for habit and travel. But not everything is possible at once!

9.   SHARING OF GROUNDS

With two teams at the same ground as happened when Richmond and University entered the competition in 1908, it is easy to program the schedule. The difference between teams programs in 1907 and 1908 are minimal, with Essendon and University sharing the East Melbourne Cricket Ground near Jolimont. With three teams sharing the one ground the scheduling is not too bad, but there are certain facts of life which then determine the number of home games. Look at Princes Park in 1987. There are 23 Saturdays in the football season before the finals. With three teams to share these Saturdays, and a couple of Mondays from the split rounds, we have to divide 25 into three parts, say 8,8,9. These are the maximum numbers of home games possible. But there are other teams in the competition, and their demands for home games makes it almost certain that these maximum numbers of home games for the three cotenants are not reached. In 1987 Hawthorn had 7 home games, Carlton 8, and Fitzroy 8. With the three cotenants at the MCG the situation is different: the Victorian Government approved several Friday night games, and in 1987 there are about half a dozen of these. Throw in three or four Sunday games and we have about 33 possible times for using the MCG. Thus the number of home games for Melbourne, Richmond and North are near nine. In fact, since they have to play each other, it is possible for them to each have nearly 11 games at the MCG, and you might like to check the 1987 and 1986 draws to see exactly how many games each played there!

10.   NEUTRAL GROUNDS

In 1907 teams had 7 games at home and 7 away - away meaning at their opponents' grounds. In 1908, Essendon and University shared the East Melbourne Cricket Ground; teams had 9 home games and 9 away games, except University and Essendon played 10 at Jolimont and 8 on their other opponents' grounds. In 1987 Melbourne has 12 games at the MCG since it plays its cotenants Richmond and North Melbourne there. Today we follow a pattern of HOME, on your own ground; AWAY, on your opponent's ground, and NEUTRAL, on a ground not your's nor your opponent's, for example AFL Park, the MCG, or elsewhere: in 1952 a full round of matches was played on neutral grounds, in Sydney, Brisbane, Hobart, Albury, Yallourn and Euroa. Other such games were played in 1903, 1904, 1979, 1980, and 1981, and today in the SANFL. Also Fitzroy played St. Kilda at Princes Park in 1986 when their home grounds were Victoria Park and Moorabbin.

In the "far" past the pattern was 11 home games and 11 away games, and with the introduction of AFL Park around 1970 the pattern became 9 or 10 at home, 10 or 9 away, and 3 at AFL Park, and then with the Swans in Sydney 9 home, 9 away, 4 at AFL Park. With increasing pressure for the "big" games to be played where crowds could and would attend, the pattern of home, away, neutral became quite varied in 1986 and 1987 and up until today, with the following patterns being observed from either what happened in 1986 and what was in the footy fixtures for 1987: a range of from 7 to 10 at home, a range of from 7 to 10 away, and a range of from 3 to 7 at

neutral grounds. For scheduling, the inflexible pattern given by the MacIntyre listing and scheduling meant that the big games were immovable without enormous dislocation and trouble. With the Wilson program the allocation of games to grounds could be better controlled, and with the new Miller program there is even greater flexibility.

Other restrictions force games to "neutral venues": Optus Oval as it is now called can have two games in the one weekend only three times a season, a condition imposed by the residents and the council. When Carlton and the Western Bulldogs must have home games on the one weekend and they have used up the three permitted, then they must go to the MCG or Waverley. When the five teams at the MCG have problems, caused by the five-clear-days rule, and they must play on the Sunday, then one team must take its game elsewhere, as Collingwood did this year.

11.   HOW FAST ARE THESE PROGRAMS?

The speed with which a program produces a completed schedule with the games assigned to grounds for each round with a good balance of home, away, and neutral venues depends on

(1)   what you ask the program to do,
(2)   what machine you run the program on, and
(3)   the difficulties caused by the requirements imposed by clubs, supporters, newspapers, and say, the AFL.

David Wilson's program used to take three minutes on one machine and three seconds on a bigger machine to produce a draft. Miller's program takes about 42 seconds on a small machine and about 28 on a big machine to produce an answer. However there is a qualitative difference between the programs which cannot be measured in time but only in the better scheduling it produces. "Wrong" allocation of a single game can cause the programs to make long time-consuming searches, while rescheduling of that game to another venue can let the program complete the job in a few seconds! In 1993 we did the TFL draw, where the constraints were exceptionally heavy for any draw, and this one had byes as well as three groups of teams required to play alternately. This took 90 minutes to find a draw, and of course they modified it.

The original display of the associated materials was generously sponsored by the VFL under the late Jack Hamilton and the late Alan Schwab whose unfailing assistance and patience is again acknowledged most gratefully.

DEVELOPMENTS IN THE DUCKWORTH-LEWIS (D/L) METHOD OF TARGET - RESETTING IN ONE-DAY CRICKET MATCHES

Tony Lewis[1] and Frank Duckworth[2]

Abstract

In this paper we summarise the contents of the paper presented at the 3rd Conference on Mathematics and Computers in Sport (Duckworth and Lewis [1]).

We then relate the developments that have occurred both to the take-up of the Duckworth/Lewis (D/L) method by the cricketing authorities and also to the modelling and modifications on the implementation of D/L in the light of experiences.

1.   INTRODUCTION

Developments up to 1996

At the 3rd Conference on Mathematics and Computers in Sport we presented our method for resetting targets in interrupted one-day cricket matches (Duckworth and Lewis [1]). In that paper we developed a two factor model of the average runs scored $Z(u,w)$ from the u overs remaining when w wickets have already been lost. We then converted this function of average runs into one for percentages $P(u,w)$ of the average runs scored in 50 over innings, $Z(50,0)$, which is the standard length of innings for one-day international matches (ODIs). We then showed how, using the percentage of innings lost due to stoppages, to adjust the target score for teams batting second (designated as Team 2). We further showed how the method gave sensible targets in all known situations of stoppages whereas other methods in general use at the time gave sensible targets in few of these situations.

We summarised the results of approaches and presentations that had been made to two of the major cricketing authorities, the Test and County Cricket Board and the International Cricket Council, both of which are based in London.

Developments in 1996

Since the conference in September/October 1996 several developments occurred in rapid succession. Several cricketing authorities almost simultaneously began to show interest in the method. As a result our attention was quickly focused on the writing of regulations for the method's use in the field.

[1]   Department of Mathematical Science, University of the West of England, Bristol, BS16 1QY, UK
[2]   Moonrakers, Traits Hill, Stinchcombe, Gloucestershire, GL11 6PS, UK

Zimbabwe Cricket Union (ZCU)

The very first organisation to adopt D/L was the Zimbabwe Cricket Union. They wished to trial the method for England's tour in December 1996 and January 1997. The method was called into use for the very first time on New Years Day. Zimbabwe scored exactly 200 from their 50 overs and then rain between the innings led to 8 overs being deducted from England's quota. From the D/L table (Appendix 2), 42 overs left with no wicket lost represents a resource percentage remaining of 92.5%. Consequently England's target was to exceed 92.5% of 200, that is 185. In the event England only scored 179 and lost whereas they had exceeded the old average run rate target of 168. There was much comment and correspondence in the British press - not all of it complimentary to Duckworth and Lewis! Johnson [2] claimed that the method 'is so indeciferable (*sic*) that the admiralty might be interested in using it for a new secret code'. Wilde [3] warned his readers that 'anyone rash enough to try to understand it while nursing a new year hangover will soon be reaching out again for the aspirin'. However he did admit that being 'set a stiffer target than previous sides batting second only seems right as batting second in such circumstances used virtually to guarantee victory' (Wilde [4]).

England and Wales Cricket Board (ECB)

We were invited to make a presentation on 5th November to the Cricket Committee of the Test and County Cricket Board (TCCB) which was shortly to be reconstituted as the England and Wales Cricket Board (ECB). The emphasis of the talk focused on the 40-over per side matches of the Sunday League, the UK competition in most need of a suitable 'rain-rule'. For this we had prepared the table of percentages based on $Z(40,0)$. Although our presentation to the ICC had been focused on the 50-overs per side one-day internationals (ODIs) we found that the percentages were not inconsistent with percentages obtained from TCCB domestic data and so we had no hesitation in recommending a table using the average runs scored in 40 over innings based on the original D/L model as outlined in Duckworth and Lewis [1]. The virtue of this was that the percentages between the two different competitions were then internally consistent.

Following our presentation and subsequent discussion, this committee recommended that D/L should be trialed for 1997 in all three domestic one-day competitions and for the Texaco ODIs against Australia, subject to there being no problems from the use of the method in Zimbabwe. This recommendation to the TCCB was formally accepted at its final meeting in December 1996.

International Cricket Council (ICC)

Subsequent to our presentation to the chief executives of the full member countries of the ICC in July 1996 their management have favoured D/L over other methods and have adopted a policy that recommends countries try out the method for ODIs and perhaps domestic one-day competitions. The ZCU and the ECB were the first to do so, as outlined, and the ICC itself wished to try out the method for the ICC Trophy competition in Kuala Lumpur, March/April 1997. This is a limited-overs competition between the associate member countries of the ICC which is used as a qualifying competition for the World Cup, the top three countries gaining entry.

D/L was called into use nine times in all. The three most notable games were a preliminary round match in which Ireland beat Holland but would have lost using average run rate (Appendix 1 - Case 1.1), the third place play-off where Ireland were set a revised target higher than Scotland's first innings total following an interruption during Scotland's innings (Appendix 11 - Case 1.2), and the final itself in which Bangladesh scored the winning run from the last ball of their final over. Kenya had scored 241 in their 50 overs and then rain between the innings reduced Bangladesh to 25 overs and D/L gave a revised target of 166 to win (Appendix 1-Case 1.3).

2.   THE IMPLEMENTATION OF D/L

In early November 1996 following Zimbabwe's election to use D/L for England's one-day matches starting 30th November, we began to write formal regulations to cover all eventualities. It soon became clear that this was not going to be a simple task. Issues which we had to think through, in many cases, had not even been thought of in some of the other methods that were in use at the time. (See Duckworth and Lewis [1] or [5] for a description of these methods.)

The Extrapolation Problem

The minimum number of overs for a viable match in ECB one-day competitions is 10 per side. If, after 10 overs, Team 1's innings is prematurely terminated and then Team 2's innings is also restricted to 10 overs, the method of handling Team 1 interruptions, as outlined in Duckworth and Lewis [1], could lead to unrealistic targets for Team 2.

An example of a Team 1 interruption described in Duckworth and Lewis [1] was the match between England and New Zealand in Perth, Western Australia, January 1983. England had scored 45 for the loss of 3 wickets in 17.3 overs (17 overs and 3 balls) when extended rain led to 27 overs being lost from each team, allowing time for just 23 overs each. On the resumption, therefore, England only had 5.3 overs left. They reached 88 (for the loss of 7 wickets).   Linear interpolation within the table (Appendix 2) shows that England lost 45.3% of their innings leaving 54.7% available to them. With New Zealand knowing that they had only 23 overs from the start they had 65.0% of their innings available to them. In essence the calculation explained in the paper for New Zealand's target is to scale up England's 88 in the ratio of their percentages of innings available, 88 x 65.0/54.7 = 104.57 or 105 runs to win.

Using the ratios of proportions of innings available to the two teams, however, is liable to lead to grossly distorted targets. For example, suppose Team 1 have scored 80 runs without loss in 10 overs out of 50. From the table, Appendix 2, one sees that with 40 overs left and 0 wicket lost they have 90.3% of their innings remaining. They have had 9.7% available. If rain now causes their innings to be terminated and Team 2's innings is also restricted to 10 overs then Team 2 have 34.1% available. Direct scaling of Team 1's score in the ratio of percentages available gives a target of 80 x 34.1/9.7 = 281.24 or 282 runs to win in 10 overs; an unrealistic task!

The problem is the unlikely sustainability of the well-above-average run rate in those initial 10 overs, perhaps caused by a dry outfield, fielding restrictions but no doubt

not a little luck which cannot be expected to hold out for a full 50 overs. The extrapolation of the performance in the first 10 overs to the next 40 overs cannot be assumed.

Similarly suppose Team 1 have made a slow but solid start having made 15 runs without loss and the same situation occurs. Now Team 2's target is 15 x 34.1/9.7 = 52.73 or 53 to win which is on the low side and caused by the assumed, but unlikely, continuation of this low scoring rate. An acceleration would be likely to take place at some stage in the later parts of the innings.

We saw potential problems of credibility in the method if such situations occurred in practice.

Projected Score

What we felt was the more likely in these situations, however, and history generally confirms this, was a regression towards the mean for the runs obtained in the subsequent overs of Team 1's innings. Consequently we introduced the concept of the projected score based on the average performance in matches of the appropriate length. This average was denoted by G_N, or G50 when referring to the average in 50 overs-per-side matches. For ODIs the average from match records of *first* innings scores over several years was calculated to give G50 as 225 runs.

The projected score was then used as the basis for setting Team 2's target. Symbolically, if R_1 and R_2 represent the percentages of innings available to the two teams respectively and if Team 1 scored S runs then Team 1's projected score, P, is $INT[S + (1-R_1) G_N]$ and Team 2's target, T, is $R_2 P$.

In the match between England and New Zealand in 1983, described earlier, P is 189 and T is 122.85, that is 123 runs to win in 23 overs which, it may be argued, is rather harder than New Zealand deserved because of the fairly weak position England had been played into at the stoppage. And in the two hypothetical cases above, again using G50 as 225, the targets would be 97 and 75 runs to win respectively.

The introduction of G50 meant that each standard of cricket using D/L is required to calculate its own value. The instructions on this which we provided were to use only first innings totals from matches in which Team 1 had the opportunity of receiving all of their overs. For the ICC Trophy competition we calculated G50 as 190 based on the 50-over per side matches of the 1990 and 1994 competitions.

Match lengths other than 50 overs per side

We created tables of percentages for 50 overs per side matches because these are by far the most common lengths of innings in one-day matches around the world and was consistent with the target resetting method used for the 1996 World Cup and subsequently adopted as the standard ICC method (see Duckworth and Lewis [1] or [5]).

This causes difficulties for match lengths other than 50 overs per side. It was no trouble to provide separate tables, however, for the 40 and 60 overs per side matches

of the ECB domestic competitions simply by dividing $Z(u,w)$ by $Z(40,0)$ and $Z(60,0)$ respectively and to calculate appropriate values for G40 and G60 from match records. But if matches are shortened before they commence and are then subsequently reduced due to bad weather then either some readjustment is needed to the basic tables or separate tables are needed for every length of viable match.

The ICC method, according to the excerpt printed on the Cricinfo [6] database on the internet, had not thought this through clearly. This led to a confused situation in January 1998 between Kenya and England A. The match was shortened to 35 overs per side before it commenced. Kenya scored 177 and England had reached 146 for the loss of 3 wickets in 30.3 overs when rain caused the match to be abandoned. D/L was not in use for this match (but it would rightly have given England A the win by 3 runs [Appendix 1 - Case 1.4]). The ICC method was incorrectly used. The percentages had not been adjusted to a 35 over innings. England were declared the winners initially but the logical adaptation would have made Kenya the winners, albeit unfairly so because of the way the ICC method favours Team 1 for interruptions occurring at the end of the match. Kenya appealed to Lord's and eventually the match was declared 'no result'.

To avoid this problem we chose to provide a separate table for all match lengths from 60 down to 10 overs per side. In so doing we hoped to make it easier for scorers to do the calculations although we had also provided a purpose written computer program for speed and accuracy on match days. We needed also to provide values for G_N for all these match lengths. This was done by a scaling of the G50, ie 225, using the first column of our table for 60 overs so that the published values for G60 and G40 overs were reasonably consistent with results from past ECB matches. G60 became 241 and G40 became 203 runs by this process.

Stoppages in mid-over

The ICC has a policy of ignoring fractions of overs when resetting target scores which could lead to injustices in marginal situations. For the ICC therefore an over-by-over table is sufficient. The ECB has a policy of using *all* balls bowled in calculating target scores. Although it is quite feasible to produce ball-by-ball percentages for all values of w this would have increased the number of tables by a factor of six and there were already 51 tables! And so we chose to use just the set of over-by-over tables and use linear interpolation in *some* circumstances. In order to keep calculations for scorers as simple as possible we avoided interpolation for a stoppage and restart mid-over by working with the next higher complete number of overs left. However, when an innings was terminated then there was no alternative but to use linear interpolation. For example, if 14.2 overs were left and then 5 overs were lost, leaving 9.2 overs left, we would find appropriate percentages with 15 and 10 overs left respectively but would interpolate between 14 and 15 overs left if the innings was terminated with 14.2 overs remaining.

Penalties

If an innings overruns its allotted time then umpires can impose a penalty on Team 2 for bowling their overs too slowly. If rain has already occurred or occurs subsequently then the target adjustment is no longer the simple matter that other

methods imply by ignoring the issue. Although the situations do not often occur, when they do, the process needs to be clearly defined. A significant proportion of the regulations was given over to this rare occurrence, helping to create, perhaps, that impression of the 'unfathomable' nature of the method!

Margin of Victory

It has been traditional for results in rain affected matches to be awarded to one side 'by faster run rate'. This is now obsolete and in many cases inaccurate. For example in the Zimbabwe-England game on 1 January 1997, Zimbabwe scored at exactly 4 runs per over whereas England scored 179 in 42 overs, a run-rate of 4.26. They lost having failed to exceed the D/L target of 185 but actually had the faster run rate!

In the ECB regulation, teams exceeding a revised target win by the number of wickets, as is traditional, but teams failing to reach a revised target lose by the number of runs below the exact target and rounded up. In both cases 'D/L method' is appended to the result.

In abandoned matches teams win by runs in both cases with 'D/L method' appended, the number of runs above or below the exact target being rounded up as above. (See Example 1 below and Appendix 1- Case 1.4)

3.   DEVELOPMENTS IN 1997

1997 was a very interesting year for D/L. It was used a total of 33 times involving 1 ODI, 9 ICC Trophy matches and 23 ECB matches including one in which Australia played. Several issues came to light particularly during the early part of the ECB season

Communication of revised target/par score

Prior to the start of the 1997 season the ECB issued an instruction to counties that a section of the scoreboard should show the par score (the minimum runs required to be in a winning position) or Team 2's score relative to par (+/-) whether or not rain had interrupted play and should also show the revised target following rain. In the early part of the season this had not been implemented at many grounds and so the first occasion when D/L was used, a termination of the match involving Warwickshire and Glamorgan, the spectators left the ground not knowing who had won. The press, therefore but rather inappropriately, blamed us and the D/L method for the uncertainty. That Glamorgan won by 17 runs (batting second!) was not fully known to many spectators until the next day.

Similar situations occurred from time to time at other grounds around the country - and also occasionally the problem of the slowness in communicating the revised target to the captains and spectators at the resumption in play. These led to the provision of the facility to calculate revised target scores for all the possible number of overs that could be lost. It came in the form of output from the computer software that we had developed for the convenience of scorers and administrators and could be produced in anticipation of a restart and the revised target could then be communicated instantly.

Press and player reaction

On Sunday 8 June 1997 no fewer than seven of the eight matches were rain affected. Some sections of the media, but not all, were very critical of the method - the word 'unfathomable' was frequently in evidence in press reports the following day. The targets themselves, however, were not seriously disputed except in one case.

In this match (see Appendix 1 - Case 1.5) Durham scored 216 in their 40 overs. Sussex had begun their reply and had lost an early wicket. Rain was threatening and so they attempted to ensure that their score was at par by the 10 over point of their innings, the minimum number of overs per side required to make the match viable. With one wicket lost the par was 42. In attempting to achieve this however they lost more wickets and so par increased since it also represents the minimum return the team should have gained from the expenditure of resources to date in pursuit of the target. Hitting out even more they lost a wicket right at the end of the 10th over and had reached 39 for the loss of 4 wickets. The umpires took the players off the field at that point. Par was 85 (84.67) at this juncture and so Sussex were losing by 46 runs which would have been the margin of their defeat if the match were abandoned at that point and no further comment would have been likely.

After a lengthy stoppage, however, the rain cleared and the umpires decided that there was time for four more overs. The revised target was 118.37 or 119 runs to win (Appendix 1 - Case 1.5). And so Sussex required 80 more runs to win in 4 overs. They didn't even try, and ended by blocking out the final over in some sort of protest. The chief executive of Sussex was quoted as saying that this proves that D/L does not work. He put in a complaint to Lord's which the media interpreted as a complaint against D/L but subsequently we heard that the thrust of the complaint was not against D/L but against the umpires for keeping the players on the field whilst rain was falling in order to complete the minimum of 10 overs to make the match viable. The complaint was rejected by the ECB.

Nevertheless the match did focus debate on whether or not our method was fair since it did not retain the *probability* across the stoppage that Team 2 could win. We have argued strongly that our method retains any advantage in terms of *runs* that a team may have established. If the match were evenly balanced at the stoppage it remains evenly balanced upon restart. To retain the same probability would involve *changing* the advantage Team 1 had established and making the amount of the change dependent on the number of runs Team 2 had actually scored. We think this to be undesirable and also unworkable.

We drew an analogy with golf. Suppose after two rounds of a professional tournament one player is eight strokes ahead of the field. With two rounds left there is a good chance that someone would emerge to challenge and perhaps overtake the leader. If one round is lost to rain, however, there is now only one round left for the field to catch up and the probability that the leader goes on to win is increased. To retain the same probability of someone catching the leader, but now in only one round, would require the lead to be reduced. This would be totally unacceptable. Similarly in one-day cricket, any advantage in terms of runs established before an interruption must be maintained across the stoppage. Following a stoppage many cricketing scenarios are less likely or even impossible. We see no special reason that

just the probability of the win to Team 2 should be preserved. To do so would result in discontinuities in margins of defeat/victory.

In the Sussex situation had play not resumed then Sussex would have lost by 46 runs but if 1 ball (in theory!) were possible then providing the same probability of victory would change the requirement from the last ball to perhaps 6 and failure to achieve this would make the margin of defeat in single figures. We feel that continuity is an important feature of the modelling and in the cricketing logic.

We also feel that the Sussex players were using tactics appropriate to the old average run rate method - only the total mattering and not the wickets lost in the process. Having gambled everything and failed to reach the par score after 10 overs, on the assumption that the match would then be abandoned, we feel that Sussex had no reason to claim a second 'equally likely' opportunity to win the match when it happened to restart. Had they batted more conservatively, and reached say 35 runs but not lost any more wickets other than the one lost early on, then allowing for the possibility of a restart would have left them with a more 'gettable' target of 77, a further 42 in the last 4 overs [Appendix 1 - Case 1.5]. Even with only one over of play possible their task would have been a suitably challenging 16 runs.

A further counter to the Sussex 'complaint' is in the situation that would have occurred had they lost no wickets at all and scored, say, 68 runs in 10 overs. In this case there would have been no restart to the game at all!. With 4 overs left the target to win would have been 68 which they had already achieved, being 36 runs ahead of par. Would *Durham* have then complained? We doubt it. Having scored at well above the asking rate no-one believes it unusual or unreasonable to have already exceeded the revised total in an interrupted match. Yet Durham would have been denied the opportunity of getting back into the match by the loss of those 26 overs. *Their* probability of winning the match would have been reduced to zero.

In summary, these arguments confirm that the only sensible criterion is to maintain any run-differential across a stoppage and not the probability.

4.   MODIFICATIONS TO D/L

In the light of experiences during the 1997 season several modifications have been introduced to make the method simpler to use and to model more accurately the practical situation in certain circumstances.

Abolition of special terminology

In 1997, in order to tightly define the procedure for correctly setting the target using the set of 51 tables, several terms were defined which tended, again, to give an aura of complexity. The need for these terms such as *initial overs allocation, initial target score* and *revised target score* has been abolished with our revisions for 1998.

Reduction in the number of tables

A major (apparent) simplification has been to reduce the number of tables down from 51 to just one and this one table applies to all lengths of innings from 60 overs

downwards. This has been achieved by representing percentages of innings remaining as being relative to the resources available for a 50 over innings. We chose this as the standard for 100% because 50 overs per side matches are by far the most common around the world.

Consequently, innings which start with more than 50 overs commence with more than 100% of the resources compared with a 50 over innings. The table in Appendix 2 shows that 60 over innings start with 107.1% of the resources compared with a 50 over innings and 40 over innings start with 90.3% of resources compared to a 50 over innings. The figures are denoted as 'resource percentages remaining' for the overs that are left from 60 down to 0 and the wickets that have fallen from 0 to 9.

In reducing the tables by such an extent we could then afford the luxury and convenience of ball-by-ball tables in order to avoid the need for special rules or interpolation when stoppages occurred between overs. In our examples, for calculations of targets involving stoppages in mid-over we shall state the necessary ball-by-ball percentages but for checking the calculations readers will need to interpolate. Note that the use of linear interpolation yields no difference in the actual targets set or decisions on the winners of abandoned matches in our examples, but there are slight discrepancies in the decimal places between the exact ball-by-ball targets and the interpolated ones.

Modification of targets in Team 1 interruptions

We are indebted to a scorer for the Marylebone Cricket Club (MCC), for the identification of an inconsistency in the D/L procedure which would occur only in cases where Team 1's innings had been interrupted.

To illustrate this, consider the situation where Team 1 have successfully negotiated their first ball in a 50 over innings and then there is a long interruption. Play resumes with just enough time for Team 1 to complete a 10 over innings and for Team 2 also to receive 10 overs. Common cricket sense dictates that the two teams have been treated virtually identically except for one ball and so Team 2's target should be little different, if at all, to Team 1's score S in their total of 10 overs. Consider four cases. Suppose S = (i) 80 runs (ii) 60 (iii) 100 (iv) 120 runs.

In the process of projecting Team 1's total score, outlined above, R_1 would be 33.7% and R_2 would be 34.1%. Assuming G50=225 then in (i) Team 1's projected score, P, would be INT[80 + 0.663x225] = 229 and Team 2's target T would be 34.1% of 229 which is 79 to win and not dissimilar to Team 1's score of 80. But in (ii) T=71 for S=60, in (iii) T=85 for S=100 and in (iv) T = 92 for S=120. In other words in most cases the target score is unacceptably and illogically different from the score achieved by Team 1.

Revised method of application of D/L

The identification of this problem, part way through the ECB domestic 1997 season, led to a further rethink on how to handle the problems of Team 1 interruptions - there was no concern when just Team 2's innings was interrupted. The review of the process led to the following revised way of applying the D/L method. It has been

included in an overall review of usage of D/L in 1997 and the changes have been adopted for use in 1998.

Team 1 score S runs in their innings with resource percentage R_1 available for their innings of 50 overs. Team 2 expect to have resource percentage available of R_2.
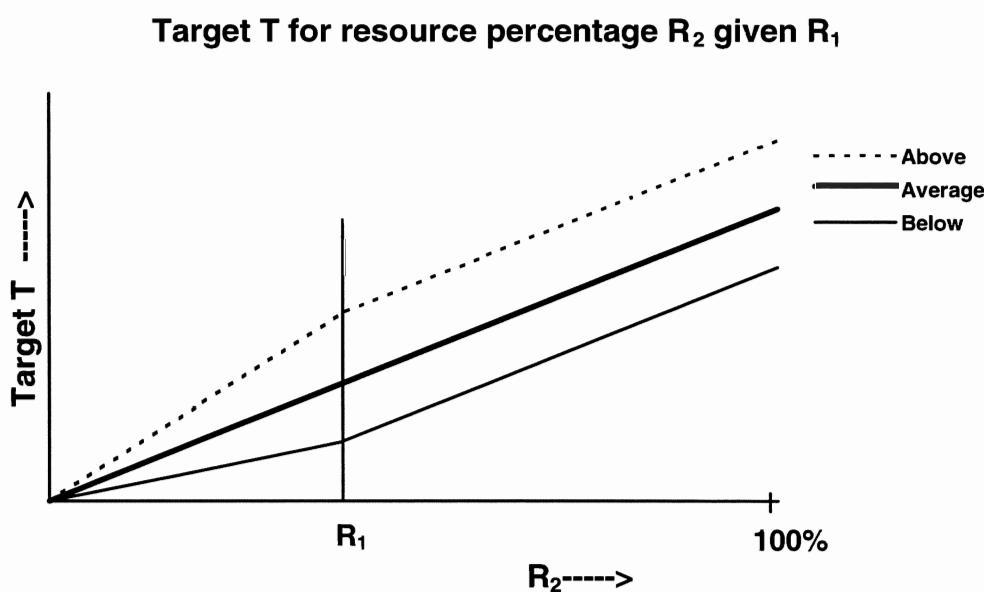
- If $R_2 \leq R_1$ then the target is set in proportion to the *resources* available to the two teams.
- If $R_2 \geq R_1$ then to Team 1's score S we add the amount of the excess percentage of G50, the average total in 50 overs,.
- If $R_2 = R_1$ then the target T is Team 1's score S and there is no discontinuity in the target
- For matches with overs per side different to 50, the *same* process is used and resource percentages from the table Appendix 2 are used. Further, G50 for the appropriate standard of competition is used in all cases since all percentages are based on the 50 overs innings being the standard 100%.

The process of setting the target T is summarised as follows

$$\text{If } R_2 \leq R_1 \quad T = S\, R_2 / R_1 \tag{1a}$$

$$\text{If } R_2 \geq R_1 \quad T = S + (R_2 - R_1)\, G50 / 100 \tag{1b}$$

The following diagram depicts the target T for all values of R_2 for arbitrary R_1

Target T for resource percentage R₂ given R₁

The heavy middle line represents average performance per unit of resource for Team 1. The other two lines represent Team 2's target for performances of Team 1 above and below average. The aspects of regression towards the mean lead to more realistic targets when initial performance has been unsustainably above or below average.

Worked examples

There follow several examples of the application of the currently adopted D/L target resetting procedure which illustrate its simplicity and fairness.

Example 1:   Premature curtailment of Team 2's innings

7 Feb 1998, Melbourne, Mercantile Mutual Cup. Victoria scored 223 runs from their 50 overs and WA had lost one wicket in scoring 188 runs in 43.2 overs. Play was then stopped by the weather, and the match was abandoned. A decision on the winner was required.

Victoria:
Their innings was uninterrupted: Resource percentage available: $R_1 = 100.0\%$

Western Australia:
At start of their 50 over innings: Resource % available   100.0%
At termination: 6.4 overs left, 1 wkt lost: Resource % left/lost   23.6%
                    Resource % available for WA = 100 - 23.6 $R_2 = 76.4\%$

Because $R_2 < R_1$ the target is set in *proportion* to the resource percentages available. From equation (1a):

$$T = 223 \times 76.4 / 100 = 170.37 \text{ ie 171 runs to win}$$

Since WA had scored 188 runs then D/L would have declared them the winners, 'by 18 runs (D/L method)'. (As discussed above, the runs above/below the exact par, here +17.63, are rounded up in giving the margin of victory)

Notes
(i)   The above result would have been quite fair as WA were clearly in a very strong position when play was stopped and would most probably have gone on to win the match if it hadn't rained.

(ii) Most other methods of target revision in use would, unfairly, have made Victoria the winners. The average run rate method gives 194 to win and the current ICC rain-rule would require 94.2% of the target ie 211 to win by the end of the 43rd over.

(iii) The method of Discounted Total Runs, the actual Australian rain-rule in use for the match, also required WA to have scored 211 by the end of the 43rd over to be declared winners. This method discounts, by 0.5% per over lost, the total of runs from Team 1's equivalent number of highest scoring overs.

(iv) Victoria were declared the winners much to the annoyance of Western Australians (Casellas [7]).

Par Score

This example illustrates further the idea of the D/L Par Score. In chasing their target Team 2 need to be aware of the minimum score necessary in order to be declared the winners if the match were abandoned at that stage. It is also a useful gauge of a

team's progress towards its target even if it doesn't rain. The par score depends not only on the number of overs bowled but also the number of wickets that have fallen.

In Example 1 if WA had lost several more wickets then the balance of the match would not have been quite so clear cut. The par scores to win, given below, show that WA, with 188 runs on the board after 43.2 overs chasing 223, would be ahead of par with up to 7 wickets lost. Had they lost 8 or 9 wickets at that stage then the balance of the match would have been with Victoria who would then have been fairly declared the winners at the abandonment.

w	0	**1**	2	3	4	5	6	7	8	9
Par to win	170	**171**	172	173	175	178	181	186	194	207
Winners	WA	**WA**	WA	WA	WA	WA	WA	WA	Vic	Vic

Example 2:   Interruption to Team 2's innings

23 Feb 1997, Eden Park, Auckland. In a ODI New Zealand scored 253 runs from their 50 available overs. England had not lost any wickets in scoring 47 runs in 6 overs. Play was then suspended and 24 overs were lost.

New Zealand:
Their 50 over innings was uninterrupted: Resource % available: R_1 = 100.0%

England:
At the start of their 50 over innings:	Resource % available:	100.0%
At stoppage, 44 overs left, 0 wkt lost:	Res % left: 94.6%	
At restart, 20 overs left, 0 wkt lost:	Res % left: 58.9%	
	Res % lost:	35.7%

Resource % *available* for England's innings = 100.0 - 35.7 R_2 = 64.3%

Since $R_2 < R_1$ England's target would have been set in proportion to the resource percentages available. From equation (1a):

$$T = 253 \times 64.3 / 100.0 = 162.68.$$

England would have needed 163 runs to win, a further 116 in 20 overs.

Note: Average run rate was in operation for this match which gave a total of 132 to win. Requiring only a further 85 more runs in 20 overs England won with more than 6 overs to spare.

Example 3: Interruption to Team 1's innings

8 June 1997. Derbyshire v Hampshire In an ECB Axa Life (Sunday) League match, Hampshire had lost 5 wickets in scoring 114 runs in 27 of an expected 40 overs when rain interrupted play and led to the match being shortened to 33 overs per side. Hampshire resumed to finish on 170 in their 33 overs.

Because of the different stages of the teams' innings that their 7 overs are lost, they represent different losses of resource.

Hampshire:
At the start of their 40 over innings, Resource percentage available: 90.3%
At stoppage, 13 overs left, 5 wkts lost: Res % left: 32.3%
At restart,   6 overs left, 5 wkts lost: Res % left   19.0%
                                                                  Res % lost   13.3%
Resource percentage available to Hampshire = 90.3 - 13.3 R_1 = 77.0%

Derbyshire:
At the start of their 33 over innings, Resource percentage available: R_2 = 81.5%

Since R_2 exceeds R_1, by 4.5%, (=81.5 - 77.0) we *add* to Hampshire's score 4.5% of G50. G50 is 225 for ECB matches. From equation (1b) the target is:

T = 170 + 4.5% of 225 = 180.13.

Derbyshire needed 181 runs to win in 33 overs.

Notes:
(i)   Even though this match began as a 40 over-per side we still use the average performance in 50 overs, G50, since the resource percentages used from the single table are all relative to resources available in 50 overs per side innings.

(ii) Derbyshire's target is higher than Hampshire's total in the same number of overs. This neutralises the advantage that Derbyshire would have had from knowing in advance of the reduction in their overs whereas Hampshire, who were pacing their innings to last 40 overs, suffered from an unexpected and untimely shortening of their innings.

(iii) All other target resetting methods in use would make no allowance for this interruption. They would set the target of 171 to win because both teams were to receive the same number of overs.

(iv) Derbyshire reached the D/L target of 181 in the last over.

Penalties

When umpires decide to impose a penalty on Team 2 for slow bowling and rain interrupts play before or after the application of the penalty then consideration is needed of the effect of the penalty on teams' resource availabilities. Nowadays, the general principle applied by the rules of most one-day competitions is that Team 1 receive the scheduled number of overs that weather permits even if time overruns. Umpires then make a decision on how many overs penalty should be imposed on Team 2 for the time over-run.

Let us suppose that Team 1 have had the opportunity to receive N_1 overs and the umpires impose a penalty of N_p overs. The D/L method converts the overs penalty into a corresponding resource penalty, R_p. It is calculated by

$R_p = P(N_1,0) - P(N_1-N_p,0).$

The Team 2 penalty is imposed, satisfying certain boundary conditions, by attributing Team 1's score S to R_p less resource than was actually available. Hence equations (1) are modified such that

$$T = S R_2 / (R_1 - R_p) \qquad\qquad \text{if } R_2 \leq R_1 - R_p \qquad\qquad (2a)$$

$$T = S + (R_2 - [R_1 - R_p])G50 \qquad \text{if } R_2 \geq R_1 - R_p \qquad\qquad (2b)$$

Separate equations have not been presented to the cricketing authorities. Instead we have given instructions on how to calculate the penalty R_p as above and then to revise *downwards* R_1 by this percentage to give an updated R_1 prior to applying equations(1).

Example 4: *Interruption to Team 1 and a bowling penalty for Team 2*

Scenario: Team 1, in an ODI, reach 230/7 in 48 out of their 50 overs when rain interrupts play so that Team 1's innings is terminated and there is then only time for Team 2 to receive 45 overs. Even allowing for the termination of Team 1's innings, the umpires decide that Team 2 should be penalised 2 overs due to slow bowling which has been incorporated into the reduced overs Team 2 will receive.

Team 1:

At the start of their 50 over inningsResource percentage available   100.0%
At termination: 2 overs left: 7 wkts lost: Res % left/lost   6.9%
     Team 1's Resource percentage available =100.0 - 6.9   **R_1 = 93.1%**

Team 2 penalty: Team 1 had 48 overs (= N_1) available to them. The resource penalty to Team 2 corresponding to the 2 overs penalty(=N_p) is:-
     Resource percentage for 48 overs left, 0 wickets lost   98.3%
     Resource percentage for 46 overs left, 0 wickets lost   96.5%
                                        Resource penalty   1.8%
     Updated resource % available to Team 1: 93.1 - 1.8 =   **R_1 = 91.3%**

Team 2:

At the start of their 45 over innings, Resource percentage available   **R_2 = 95.5%**

R_2 exceeds R_1 by 4.2% (= 95.5% - 91.3%). From equation (1b), using G50=225, Team 2's target is:
        T =   230 + 4.2% of 225 = 239.45
Team 2 would need to score 240 runs to win in 45 overs, 10 more runs than Team 1 actually made in 48 overs.

5.   DEVELOPMENTS IN 1998

Within innings data

Towards the end of 1997 we received, in electronic form, the over-by-over scores from all the ECB matches of the 1997 season. This has enabled us to begin to test out the accuracy of some of our estimates of the parameters in our model, $Z(u,w)$, of the average total runs obtained with u overs left and w wickets already lost. (See Duckworth and Lewis [1]).

The analysis of this data will be the subject of a future paper. Although more work still needs to be done, early indications are that the estimates of the parameters from the 1997 data are not sufficiently different from our initial estimates of those parameters currently in use and which are reflected in the percentages of Appendix 2. In the immediate future, therefore, Appendix 2, and its corresponding ball-by-ball tables are being used for all applications of the D/L method.

We are anxious to obtain as much data of scores from within innings as possible from around the world. Prior to the receipt of the 1997 ECB data much of our database with this detail has been from personal observation and manual recording from matches watched either at the grounds or on television. Data on total scores has been obtained from the Cricinfo [8] database and from Wisden [9] (previous year's editions). We need data from other sources which we would welcome either in printed format or, preferably, in electronic format.

Fielding restrictions and other innovations

In ODIs and several other limited-overs competitions there are restrictions on field placings. Within the first 15 overs, referred to as the '15 over rule', only two fielders are permitted to be beyond a certain distance from the bat at the moment of delivery of the ball. The purpose of the rule is to encourage more attacking batting in the early part of a team's innings when, usually, it is a period of foundation building for a later onslaught on the bowling. Although Clarke [10] has shown that this is not optimum, it has taken the introduction of the 15-over rule to change batting sides' tactics. There are possible implications for our percentages of innings as tabulated in Appendix 2 although our immediate response is that we do not attempt to model the *rate* of scoring of runs, but only the average runs obtainable in overs available. Nevertheless early indications from our analysis of not only our ODI database but also the ECB Benson & Hedges competition which also uses the 15 over rule, are that there is no significant difference of the average percentage of the total of runs scored, for the wickets lost in 15 overs, between actual matches and that predicted from the D/L model. In other words, on average there is an increase in the rate of wickets falling corresponding to any increase in the rate of scoring runs in the first 15 overs. There is no evidence, therefore, that the 15 over rule has any material effect on the validity of the D/L targets.

In the 1997/98 season the Australian Cricket Board (ACB) introduced a 30 over rule which retains some less severe restriction on field placings until the end of the 30th over. Although we do not have any data from the Australian season just past we are

confident that our findings in respect of the 15 over rule will similarly show that a 30 over rule would not have any material effect on the validity of D/L targets.

A further innovation in the Australian 1997/8 season was the introduction of the concept of a 12-player squad for its domestic one-day competition - the Mercantile Mutual Cup. A team may select a different set of 11 players from the squad for batting and fielding. The effects of this innovation are not yet known but are only likely to be on the parameters of the D/L model and not on the model itself. The effects on the parameters could well be to bring the curves (see Duckworth and Lewis [1] or [5]) more in proportion to each other since there is likely to be an increase in the number of specialist batsmen in the 'batting' squad. And the 'fielding' squad of their opponents will have more specialist bowlers so reducing the variability in the quality of both the batting and the bowling.

England and Wales Cricket Board (ECB) take-up of D/L

In October 1997 we were invited to Lord's to present a review of the application of D/L to the Cricket Advisory Committee. In this review as well as giving an overview of the use of D/L in the 1997 season we presented our recommendations, outlined above, on how the application of the method could be simplified. There was unanimous agreement to recommend adoption of the method again for the 1998 season, which included the recommendation also for the ODIs with South Africa in May and a Tri-nation series with South Africa and Sri Lanka in August. The ECB have also recommended to the ICC that D/L becomes the standard 'rain-rule' for all ODIs.

Subsequent to this meeting we have been exploring the possibility with the ECB of introducing the method to the Minor Counties level of cricket in the UK. As mentioned earlier, county scorers have a software program on a laptop computer which they use on match days to ensure speed and accuracy of the target resetting process. Minor counties scorers do not all have access to a laptop computer and so there would need to be a training programme of these scorers so that they could operate the D/L system entirely manually.

Early in 1998 we delivered a workshop to several ECB staff who would be responsible for the training of the scorers if the system were introduced at this level. The reaction was generally supportive of its use but doubts were expressed on the possibility of undertaking the necessary training in time for the beginning of the season. The expectation as we write, therefore, is that the method will be introduced into the later stages of the one-day competition for the MCC Trophy.

This step into the lower levels of cricket is confirmation of our strong conviction that the method is now simple enough to be used in limited overs matches at club cricket level. Indeed, at the end of the 1997 season we were asked to use D/L in arbitration of a club competition in Nottinghamshire by deciding on the winning team in their cup final which couldn't be finished by the end of their season.

International adoption of D/L

The situation described in Example 2 above undoubtedly was instrumental in persuading New Zealand to adopt D/L for their Shell Cup competition and for the ODIs with Zimbabwe and Australia in early 1998. In the event, due partly to a hot and dry El Nino induced summer, it was not invoked. Nevertheless New Zealand were the third of the ICC full member countries to adopt the method. - we only have six to go! And we are hopeful that the situation of the Victoria-Western Australia match in Example 4 may help persuade the ACB to adopt D/L.

As mentioned earlier, the management of the ICC are strongly supportive of our method and are making every effort to encourage the six other countries to at least try out our method in competition in the near future. The ICC have jurisdiction over the re-introduction of cricket into the 1998 Commonwealth Games in Malaysia in September. D/L will be the rain rule in use for this tournament. It will also be used in the ICC Knock-Out tournament to be held between all nine full-member countries, in Dhaka, Bangladesh in Oct/Nov 1998.

In May 1998 we addressed the ICC Cricket Committee (Playing) in an ICC-supported attempt to persuade those six countries to adopt D/L with the medium view to D/L being adopted as the official ICC rain-rule for ODIs and, of course, the 1999 World Cup which is to be held in the UK. At the time of writing the results of this presentation are unknown.

6.   DEVELOPMENTS FOR THE FUTURE

There is still considerable work to be done in the statistical analysis of the increasing data now becoming available to us. Although we are comfortable with the parameters at present and hence the percentages obtained in Appendix 2 there is little doubt that the character of the one-day game is changing both in terms of players' skills at the game and the tactics that are adopted.

Consequently, we intend to keep our database up-to-date, to review the parameters and, at the appropriate time, make a change to these which will affect the tables of percentages, to reflect these changing patterns of the game.

7.   CONCLUSION

We have given a summary of our experiences of the use of the Duckworth/Lewis method in 1997 both at the technical and practical levels. In the light of these experiences we have introduced some modifications to the way the method operates which have resulted in substantial simplification in its implementation.

The method has been used both domestically and internationally and, although there has been some scepticism in some sections of the media, it is gradually gaining acceptance.

Several countries have now adopted the method and we are hopeful that the method will become the accepted international standard and be used for the 1999 World Cup.

We recognise that there is still work to be done on the statistical analysis of current and future data in order to keep up-to-date the percentages used in the target resetting process.

REFERENCES

[1]   Duckworth F.C. and Lewis A.J., A fair method for resetting the target in interrupted one-day cricket matches Proceedings of the 3rd Conference, Mathematics and Computers in Sport, (ed. Neville de Mestre), 51 - 68 (1996).

[2]   Johnson and Martin, Daily Telegraph, 2 Jan 1997.

[3]   Wilde and Simon, The Times, 1 Jan 1997.

[4]   Wilde and Simon, The Times, 2 Jan 1997.

[5]   Duckworth F.C. and Lewis A.J., A fair method of resetting the target in interrupted one-day cricket matches, *J Opl Res Soc*, Vol 49 Pt 3, 220-227 (1998).

[6]   Cricinfo,
       http://www.cricket.org/link_to_database/ABOUT_CRICKET/RAIN_RULES/

[7]   Casellas Ken, West Australian, 8 Feb 1998.

[8]   Cricinfo, http://www.cricket.org/link_to_database/ARCHIVE/

[9]   Wisden, The Cricketers' Almanack, Guildford, Surrey, UK, 134th Edition (1997).

[10]   Clarke S.R., Dynamic Programming in One-Day Cricket - Optimal Scoring Rates, *J Opl Res Soc*, Vol 39, Pt 4, 331-337 (1988).

COPYRIGHT

APPENDIX 1

Case 1.1   ICC Trophy, Kuala Lumpur, March/April 1997

Holland scored 211 in their 50 overs. Ireland had reached 91 runs for the loss of 3 wickets in 23 overs when rain caused the match to be abandoned. [A minimum of 20 overs per side was required to make a match viable.]

At the stoppage: 27 overs left, 3 wkts lost: From Appendix 2, Percent lost 59.3%
Reduction in target: 59.3% of 211 = 125.12
Revised Target (Par Score) 211 - 125.12 = 85.88
At 91 (with 3 wkts lost) Ireland were declared the winners - by 6 runs

Case 1.2   ICC Trophy, Kuala Lumpur, March/April 1997

Scotland had reached 56 for the loss of 1 wicket in 19 of their 50 overs when rain led to the deduction of 5 overs from each innings. Scotland resumed to score 187 in their 45 overs.

G50 was 190 for the ICC Trophy matches
      With 31 overs left, 1 wkt lost, percentage of innings remaining: 74.4%
      With 26 overs left, 1 wkt lost, percentage of innings remaining: <u>67.2%</u>
      Percentage of innings lost due to the stoppage   7.2%
Using the method as applied in 1997, Scotland's projected total in 50 overs was INT[187 + 7.2% of 190] = 200
For Ireland R_2 = 95.5%, hence T = 95.5% of 200 = 191.00, 192 to win Ireland were all out for 141 in 39 of their 45 overs and lost

Case 1.3   ICC Trophy final, Kuala Lumpur, March/April 1997

Kenya scored 241 in 50 overs. Rain between the innings reduced Bangladesh to 25 overs.

Percentage of innings remaining for 25 overs left and 0 wkt lost: 68.7%
Revised target is 68.7% of 241 = 165.57 and 166 runs to win.
Bangladesh won on the last ball of the match.

*Case 1.4   **Tour match** Nairobi, Kenya, 3 Jan 1998, Kenya v England 'A'*

In a match shortened before the start to 35 overs, Kenya scored 177 and England 'A' were 146 for 3 wickets when further rain caused the abandonment of the match. The ICC method was in use for this match.

ICC Method - Logical Interpretation:
      For a 35 over match - percentage factor 84.0%
      For a 31 over match - percentage factor 77.8%
      For a 30 over match - percentage factor 76.0%
      Interpolation of factor for 30.3 overs   76.9%
      Target score calculation: 177 x 76.9 / 84.0 = 162.04 ie 163 to win
      Under the logical use of the ICC method at 146 England 'A' lose

ICC Method - As used omitting the correction for a 35 over match
      For a 31 over match - percentage factor 77.8%
      For a 30 over match - percentage factor 76.0%

Interpolation of factor for 30.3 overs   76.9%
Target score calculation: 177 x 0.769 = 136.11 ie 137 to win
Under this *misuse* of the ICC method at 146 England 'A' win

D/L Method - As revised for 1998 using the table Appendix 2:
Kenya: Completed their 35 over innings -resource percentage available   = **84.2%**
England 'A': For start of a 35 over innings, 0 wkt lost, res. percent avail   = **84.2%**
For 30.3 overs gone, 4.3 overs left, 3 wkts lost: percentage left/lost   = <u>16.0%</u>
Resource percentage available to England 'A'   **68.2%**
Target score calculation: 177 x 68.2 / 84.2 = ie 143.37 to win
Under the D/L method at 146 England 'A' win, by 3 runs (2.63 rounded up)

Case 1.5   *ECB Axa Life League, 8 Jun 1997, Chester-le-Street, Durham*
Durham scored 216 in their 40 overs. Sussex had lost 4 wickets and scored 39 runs in 10 overs. Rain interrupted play and 26 overs were lost. Play resumed with 4 overs left for Sussex to bat.

Durham:
Completed their 40 overs innings: Resource percentage available R_1 = **90.3%**
Sussex:
At start of their 40 overs innings percent of innings available   90.3%
         At stoppage, 30 overs left, 4 wkts lost: Percent left 54.9%
         At restart,   4 overs left, 4 wkts lost: Percent left <u>14.1%</u>
                                               Percent lost   <u>40.8%</u>
         Resource percentage available to Sussex = 90.3 - 40.8 R_2= **49.5%**
Sussex target is 216 x 49.5/90.3 = 118.41, ie 119 to win
Sussex required a further 80 runs in 4 overs

If only one wicket lost and 35 runs on the board
At stoppage, 30 overs left, 1 wkts lost: Percent left 73.1%
At restart,   4 overs left, 1 wkts lost: Percent left <u>14.8%</u>
                                               Percent lost   <u>58.3%</u>
         Resource percentage available to Sussex = 90.3 - 58.3 R_2 = **32.0%**
Sussex target is 216 x 32.0/90.3 = 76.54, ie 77 to win
With 35 runs on the board Sussex would require a further 42 runs in 4 overs.

Suppose Sussex had lost no wicket and had 68 runs on the board
At stoppage, 30 overs left, 0 wkt lost: Percent left   77.1%
At restart,   4 overs left, 0 wkt lost: Percent left   <u>14.9%</u>
                                               Percent lost   <u>62.2%</u>
         Resource percentage available to Sussex = 90.3 - 62.2 R_2 = **28.1%**
Sussex target is 31.1% of 216 x 28.1/90.3 = 67.22, ie 68 to win
With 68 runs already on the board Sussex would have won without any need for a resumption in play

Note: The calculation would have been undertaken by the scorers according to the way that D/L was implemented in 1997 using the table appropriate to a 40 over match. The actual targets differed only in the decimal places compared with those calculated above.

APPENDIX 2

The D/L (Duckworth/Lewis) method of adjusting target scores in interrupted one-day cricket matches

Table of resource percentages remaining - over by over

	Overs left	
60	to	0

Resource percentages remaining

overs left	0	1	2	3	4	5	6	7	8	9	overs left
					wickets lost						
60	107.1	98.0	87.9	76.5	63.8	50.0	37.8	26.5	16.4	7.6	60
59	106.4	97.5	87.5	76.3	63.7	50.0	37.8	26.5	16.4	7.6	59
58	105.8	97.1	87.2	76.1	63.6	49.9	37.7	26.5	16.4	7.6	58
57	105.2	96.5	86.8	75.8	63.5	49.9	37.7	26.5	16.4	7.6	57
56	104.5	96.0	86.4	75.6	63.4	49.8	37.7	26.5	16.4	7.6	56
55	103.8	95.5	86.0	75.3	63.2	49.8	37.7	26.5	16.4	7.6	55
54	103.1	94.9	85.6	75.0	63.1	49.7	37.7	26.5	16.4	7.6	54
53	102.4	94.3	85.2	74.8	62.9	49.7	37.7	26.5	16.4	7.6	53
52	101.6	93.7	84.7	74.5	62.8	49.6	37.7	26.5	16.4	7.6	52
51	100.8	93.1	84.3	74.1	62.6	49.5	37.7	26.5	16.4	7.6	51
50	100.0	92.4	83.8	73.8	62.4	49.5	37.6	26.5	16.4	7.6	50
49	99.2	91.8	83.3	73.5	62.2	49.4	37.6	26.5	16.4	7.6	49
48	98.3	91.1	82.7	73.1	62.0	49.3	37.6	26.5	16.4	7.6	48
47	97.4	90.3	82.2	72.7	61.8	49.2	37.6	26.5	16.4	7.6	47
46	96.5	89.6	81.6	72.3	61.5	49.1	37.5	26.5	16.4	7.6	46
45	95.5	88.8	81.0	71.9	61.3	49.0	37.5	26.4	16.4	7.6	45
44	94.6	88.0	80.4	71.5	61.0	48.9	37.5	26.4	16.4	7.6	44
43	93.6	87.2	79.7	71.0	60.7	48.7	37.4	26.4	16.4	7.6	43
42	92.5	86.3	79.0	70.5	60.4	48.6	37.4	26.4	16.4	7.6	42
41	91.4	85.4	78.3	70.0	60.1	48.4	37.3	26.4	16.4	7.6	41
40	90.3	84.5	77.6	69.4	59.8	48.3	37.3	26.4	16.4	7.6	40
39	89.2	83.5	76.8	68.9	59.4	48.1	37.2	26.4	16.4	7.6	39
38	88.0	82.5	76.0	68.3	59.0	47.9	37.1	26.4	16.4	7.6	38
37	86.8	81.5	75.2	67.6	58.6	47.7	37.1	26.4	16.4	7.6	37
36	85.5	80.4	74.3	67.0	58.2	47.5	37.0	26.4	16.4	7.6	36
35	84.2	79.3	73.4	66.3	57.7	47.2	36.9	26.3	16.4	7.6	35
34	82.9	78.1	72.4	65.6	57.2	47.0	36.8	26.3	16.4	7.6	34
33	81.5	76.9	71.4	64.8	56.7	46.7	36.6	26.3	16.4	7.6	33
32	80.1	75.7	70.4	64.0	56.1	46.4	36.5	26.3	16.4	7.6	32
31	78.6	74.4	69.3	63.2	55.5	46.0	36.4	26.2	16.4	7.6	31
30	77.1	73.1	68.2	62.3	54.9	45.7	36.2	26.2	16.4	7.6	30
29	75.5	71.7	67.0	61.3	54.3	45.3	36.0	26.1	16.4	7.6	29
28	73.9	70.2	65.8	60.4	53.5	44.9	35.8	26.1	16.4	7.6	28
27	72.2	68.8	64.5	59.3	52.8	44.4	35.6	26.0	16.4	7.6	27
26	70.5	67.2	63.2	58.3	52.0	43.9	35.4	25.9	16.4	7.6	26
25	68.7	65.6	61.8	57.1	51.2	43.4	35.1	25.9	16.4	7.6	25
24	66.9	64.0	60.4	55.9	50.3	42.8	34.8	25.8	16.3	7.6	24
23	65.0	62.3	58.9	54.7	49.3	42.2	34.4	25.6	16.3	7.6	23
22	63.0	60.5	57.3	53.4	48.3	41.5	34.1	25.5	16.3	7.6	22
21	61.0	58.6	55.7	52.0	47.2	40.8	33.7	25.3	16.3	7.6	21
20	58.9	56.7	54.0	50.6	46.1	40.0	33.2	25.2	16.3	7.6	20
19	56.8	54.8	52.2	49.0	44.8	39.1	32.7	24.9	16.2	7.6	19
18	54.6	52.7	50.4	47.4	43.5	38.2	32.1	24.7	16.2	7.6	18
17	52.3	50.6	48.5	45.8	42.2	37.2	31.5	24.4	16.1	7.6	17
16	49.9	48.4	46.5	44.0	40.7	36.1	30.8	24.1	16.1	7.6	16
15	47.5	46.1	44.4	42.1	39.1	35.0	30.0	23.7	16.0	7.6	15
14	45.0	43.7	42.2	40.2	37.5	33.7	29.1	23.2	15.8	7.6	14
13	42.4	41.3	39.9	38.1	35.7	32.3	28.2	22.7	15.7	7.6	13
12	39.7	38.8	37.6	36.0	33.9	30.8	27.1	22.1	15.5	7.6	12
11	36.9	36.1	35.1	33.7	31.9	29.2	25.9	21.4	15.3	7.5	11
10	34.1	33.4	32.5	31.4	29.8	27.5	24.6	20.6	14.9	7.5	10
9	31.1	30.6	29.8	28.9	27.6	25.6	23.1	19.6	14.5	7.5	9
8	28.1	27.6	27.0	26.3	25.2	23.6	21.5	18.5	14.0	7.5	8
7	25.0	24.6	24.1	23.5	22.7	21.4	19.7	17.2	13.4	7.4	7
6	21.7	21.4	21.1	20.6	20.0	19.0	17.7	15.7	12.6	7.2	6
5	18.4	18.2	17.9	17.6	17.1	16.4	15.5	14.0	11.5	7.0	5
4	14.9	14.8	14.6	14.4	14.1	13.6	13.0	11.9	10.2	6.6	4
3	11.4	11.3	11.2	11.1	10.9	10.6	10.2	9.6	8.5	6.0	3
2	7.7	7.7	7.6	7.6	7.5	7.4	7.2	6.9	6.3	4.9	2
1	3.9	3.9	3.9	3.9	3.9	3.8	3.8	3.7	3.5	3.1	1
0	0	0	0	0	0	0	0	0	0	0	0
overs left	0	1	2	3	4	5	6	7	8	9	overs left
					wickets lost						

CONSTRUCTING A PLAUSIBLE TEST CRICKET SIMULATION USING AVAILABLE REAL WORLD DATA

David Dyte[1]

Abstract

A basis is created for converting the averages of a batsman and a bowler into tables of probabilities of different events (runs, dismissals, sundries and so on) for individual deliveries where they play against each other. As a consequence of this method, a simple way of comparing players across different eras in cricket is also provided. A full cricket simulator created using this method is available for use on the world wide web.

1. INTRODUCTION

Previous investigations by authors such as Clarke and Norman [1], Johnston [2], and Chedzoy [3] have used simulated cricket matches to investigate the effects of such things as tactical changes or umpiring decisions in real cricket matches. These simulations usually involve constructing a table of probabilities of certain events for each ball in the match. The table is generally altered according to whether or not a specialist batsman is facing, and possibly by the run rate at which he is attempting to score (for one day cricket simulation), but a greater level of detail is usually not sought.

The Sim1 cricket simulator gains a degree of detail by taking into account the career statistics of both the bowler and the batsman involved in each ball. The first step in this process is to calculate the long term expected outcome for a given batsman-bowler combination. Each ball is then treated as an independent trial with the same expected outcome.

2. HOW IS IT DONE?

Averages are the principal statistic available in the cricket data, and are used here as the main basis for constructing a table of probabilities for the outcome of a given delivery. Cricket statisticians have a distinct advantage here over those constructing statistics for other sports. An average is a simple ratio of postitive events for the batting team (runs) to positive events for the bowling team (wickets). Such simple and intuitive measures of player performance are not so readily available in, say, baseball. Lindsey [4], James [5], and Thorn and Palmer [6] have all made increasingly complex attempts to quantify the contribution of players to their baseball teams.

---

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

Before we attempt to combine the averages of batsman and bowler for simulation, we require knowledge or an estimate of the overall averages in the contests where the two competitors acquired their averages. Once this is known, a relative player quality may be calculated for batsman and bowler. The adjustment from simple average to relative quality is made because of the sometimes severe differences between conditions in players' careers. Suppose we wish to compare the test careers of Sachin Tendulkar and Victor Trumper. Tendulkar's average of 54.8 and Trumper's 39.0 scarcely seem comparable unless we consider the conditions under which each batsman played. Similar adjustments could be made to compare, say, Syd Barnes' and Curtly Ambrose's achievements as bowlers.

Simply,

$$Q_{bat} = \frac{A_{bat}}{E_{bat}}$$

$$Q_{bowl} = \frac{A_{bowl}}{E_{bowl}}$$

where Q denotes relative quality, A denotes player's average, and E denotes the average over the era (or for the competition) in which the individual played. Note that for bowlers, a lower relative quality is better. Also note that $E_{bat} \neq E_{bowl}$ for reasons mentioned below.

So, for Trumper and Tendulkar's batting qualities, we have

$$Q_{bat,Tendulkar} = \frac{54.84}{32.38} = 1.69$$

$$Q_{bat,Trumper} = \frac{39.04}{26.79} = 1.46$$

So we give Tendulkar the edge. Don Bradman's relative quality of 3.02 appears somewhat impassable, however.

The relative quality of the player may then be multiplied by any suitable figure desired, to simulate cricket in particular conditions. To estimate a player's average had they played in the era of Trumper, multiply by 26.8. For Tendulkar's era, mutliply by 32.4. This method makes sense in a neat, intuitive way.

Gould [7] has found that players have tended closer to mean performance in baseball with time and increasing overall excellence in the game, which may indicate a benefit for discounting the relative quality of players from early eras somewhat. Such an analysis is yet to be performed on cricket data, however.

The picture is a little more complex than this, however. In viewing a cricket match as a series of head to head confrontations between batsman and bowler, some adjustment to the batsman's average does become necessary. Certain dismissal types such as run out or handled ball are not credited to the bowler, regarded instead as being solely a product of the batsman's error. Also, these types of dismissal may occur at either end of the pitch, not only to the striker. Hence, true average for this head to head calculation requires their removal from the batsman's statistics at this

point. They will be included at a later stage in the algorithm. This procedure may either be performed directly from the record of the particular batsman (preferable) or by using a generic rate of approximately one dismissal in thirty being of this type.

Similarly, since the 1983-84 season wides and no balls have accrued to bowlers' analyses. For the purposes of calculating head to head averages against a batsman, these figures must be removed also for modern players and reintroduced later. When sundries are introduced later, generic or reputation based figures must be substituted for older players where real data on wides and no balls is not available.

So,

$$Q'_{bat} = \frac{Q_{bat}(1 - R_{era})}{1 - R}$$

$$Q'_{bowl} = \frac{Q_{bowl}(1 - S)}{1 - S_{era}}$$

where Q' denotes the true quality in the head to head confrontation, R is the estimated proportion of non-bowler dismissals, and S is the estimated proportion of runs accrued to the bowler in sundries. Note, $S = 0$ for players prior to 1983. Now we have statistics which relate solely to the direct confrontation of batsman and bowler.

The heart of the matter is in combining batsman and bowler quality into a single figure. The philosophy behind the Sim1 simulation is this: a batsman twice as good as all comers will perform twice as well as all comers against a particular bowler; and a bowler with an average 0.7 times that of all comers will dismiss a particular batsman for 0.7 times the runs all comers would. This is based on the same kind of intuition as the relative quality ratings.

Clearly, then,

$$Q'_{comb} = Q'_{bat} \cdot Q'_{bowl}$$

gives us the relative quality of the head to head confrontation of the two players. We need only to multiply by the batting average for the era in which we wish the match to take place, suitably adjusted, and we have:

$$A'_{comb} = \frac{Q'_{comb} \cdot E_{bat}}{1 - R_{era}}$$

which is the long term expectation we require. Here, R_{era} is the proportion of non-bowler dismissals in the era the game is played. Insufficient data investigation has been performed to test whether the proportion of these types of dismissal has altered significantly with time. An eqivalent calculation would be:

$$A'_{comb} = Q'_{comb} \cdot E_{bowl}(1 - S_{era})$$

which is easier for earlier eras, as S_{era} will be 0.

Run rates or strike rates may also be taken into account if available, otherwise plausible results may be obtained by using a generic run/strike rate or simply estimating figures based on impressions of the players involved. Just as a combined average may be obtained first by comparing against the average for an era, then by a

multiplicative method, the same may be achieved for scoring rates. For want of better data at the present data source (the CricInfo web site at www.cricket.org) the Sim1 simulator presently operates with a fixed run rate, which is user specified. The average over all test cricket since World War I is approximately 2.4 runs per over, or 0.4 runs per ball.

Using a calculated head to head run rate will give better results than a simple generic run rate, since it provides for bowlers of varying economy (Brian Statham versus Devon Malcolm), or batsmen of varying speed (Chris Tavare versus Viv Richards).

It is apparent that if a team has one outstanding bowler, and others of lesser ability, the opposition's expected score will be reduced if the lesser bowlers concede runs at a slower rate. This leaves a greater share of all events (runs and wickets) to the better bowler, who also has the highest wickets / runs ratio. This is part of the basis for New Zealand fans' fond memories of Ewen Chatfield's efforts as an economical foil to star bowler Richard Hadlee.

However we arrive at the number, given some estimate of run rate per ball, r, we may then proceed to partition the event space for a single ball into probability of dismissal by bowler and some (arbitary) distribution of scores giving the calculated run rate.

$$\Pr(wicket) = p_w = \frac{r}{A'_{comb}}$$
$$\Pr(1\,run) = p_1 = 0.264r$$
$$\Pr(2\,runs) = p_2 = 0.1r$$
$$\Pr(3\,runs) = p_3 = 0.024r$$
$$\Pr(4\,runs) = p_4 = 0.1r$$
$$\Pr(5\,runs) = p_5 = 0.0008r$$
$$\Pr(6\,runs) = p_6 = 0.01r$$

Partitioning the dismissal event space into different types of dismissal can be done in a number of ways. For simplicity (and for want of good data) the simulator presently uses one of two fixed distributions according to the speed of the bowler. Analyses such as that performed by Croucher [8] may provide more accurate results if required.

Additional calculations must now be made in order to include run outs and sundries. Sundries other than no-balls may be placed into the "no runs or wickets" partition of the event space without damaging the balance already achieved. Run outs (and other, more exotic dismissals such as handled ball or obstructed field) must, however, be included as a separate event (occurring after the initial event) with probability adjusted to reflect the chance of the non striker being run out and the fact that certain events (boundaries) preclude any run out occurring. No balls complicate matters still further.

For the sake of brevity, we shall now assume no balls are exactly the same as other sundries and no other run scoring events occur with them. If no balls are to be treated correctly, the calculations become considerably more tedious.

Firstly, we define

$$p_0 = 1 - p_w - \sum_{runs} p_{runs}$$

then, if p_{s1} is the probability of one sundry given that no runs were scored from the bat, and so on, we can say:

$$p_{s1} = \frac{0.36rs}{p_0(1-s)}$$

$$p_{s2} = \frac{0.13rs}{p_0(1-s)}$$

$$p_{s3} = \frac{0.02rs}{p_0(1-s)}$$

$$p_{s4} = \frac{0.08rs}{p_0(1-s)}$$

Where s is the proportion of runs we assign to be scored in sundries. Again, this will vary by era. Note that $s \neq S$, the proportion of bowler's runs accounted for by wides and no balls. A typical value for s is around 0.05. The relative proportion of numbers here is quite arbitary, due to a lack of data. Type of sundries is, like dismissal type, chosen from two fixed distributions depending on bowler speed.

The formula for run out probability is a complex one. We assume the bowler's quality has no part in the probability of a run out (or other non-bowler dismissal), and allocate a chance to the non-striker also. Defining p_{nw} as being p_w if the non-striker were facing, we have the probabilities of run out given no other dismissal type, and no boundary having occurred:

$$p_{ro,strk} = \frac{p_w R_{strk}}{2Q'_{bowl}(1-R_{strk})(p_0 + p_1 + p_2 + p_3 - p_0 p_{s4})}$$

$$p_{ro,nstrk} = \frac{p_{nw} R_{nstrk}}{2Q'_{bowl}(1-R_{nstrk})(p_0 + p_1 + p_2 + p_3 - p_0 p_{s4})}$$

Note, we have assumed no-one will be foolish enough to be run out after a 5 is scored, assuming it is all run.

Apart from the construction of a table of outcomes for an individual delivery, some outside factors come into play. Clearly, in cricket, a captain cannot bowl his two lowest average bowlers all day. Some method of fatiguing and recovering players is required in order to guarantee a fair distribution of overs amongst the bowlers, in that not every bowler will be available to bowl all the time. In the Sim1 program, a simplistic method is employed revolving around stamina which largely depends on bowler speed. A more sophisticated system, which properly takes into account breaks in play, is in planning.

3.   WHAT IS MISSING FROM SIM1?

The effect of differing standards of fielding has not been included in this simulation. There are two important reasons for this. Firstly, one might argue that if we were to match the present day Zimbabwean team against the 1984 West Indians, that the Zimbabwean bowlers should suffer some sort of relative disadvantage because of their team's poorer fielding standard. The important point is, however, that this effect is already reflected in their averages. Unlike in baseball, bowlers do not receive any credit for dropped catches where they may have taken a wicket. So Heath Streak's average, good though it is, already includes the effect of Zimbabwe's poor fielding record. Secondly, the data for such events as dropped catches is sparse at best, and will do little to improve the simulation as it currently stands, whilst adding a whole raft of new calculations in order to be correctly implemented.

There are many commonly accepted facts of cricket life as yet unaccounted for in this simulation. These include such phenomena as ground effects, new ball effects, pitch wear effects, attacking or defensive tactics from either team, and so on. Again, where data is available such things may be estimated directly, otherwise generic effects or reputation based effects may be used. Some of these effects will be included in a future update.

Kimber and Hansford [9] have criticised the use of simple averages as a defining statistic for batsmen, and have constructed a more complex formula relating to the hazard of dismissal as a function of the batsman's present score, pointing out that at very low scores batsmen appear more likely to be dismissed than later in their innings. This would appear to point to two serious flaws in this simulation: each ball is not in fact a trial independent of the batsman's current score; and simple averages are not appropriate for predicting a batsman's true outcome. However, the effects noted by Kimber and Hansford are slight, and in any case require a great deal more data than the amount required to operate the simulation in its present form.

The Sim1 cricket simulator has been rendered in the Java programming language and may be played via the world wide web. It can be found http://www.swin.edu.au/sport/ .

ACKNOWLEDGEMENTS

REFERENCES

[1] Clarke S.R. and Norman J.M., To run or not? Some dynamic programming models in cricket, In R.L. Jensen & I.R. Johnson (Eds.), Proceedings of the Twenty-Sixth Annual Meeting of the Western Decision Sciences Institute, 744-746, (Hawaii, The Decision Sciences Institute, 1997).

[2] Johnston M.I., Assessing team and player performance in one day cricket using dynamic programming, Unpublished M.App.Sc., Swinburne University.

[3] Chedzoy O.B., The effect of umpiring errors in cricket, *The Statistician*, Vol 46, No 4, 529-540, (1997).

[4] Lindsey G., An investigation of strategies in baseball, *Operations Research*, Vol 11, 447-501, (1963).

[5] James B., The Bill James baseball abstract, (Ballantine, New York, 1982).

[6] Thorn J. and Palmer P., Total baseball, (Harper Collins, New York, 1993).

[7] Gould S.J., Life's grandeur: the spread of excellence from Plato to Darwin, (Random House, London, 1997).

[8] Croucher J.S., Anglo-Australian cricket dismissals 1946-1980, *Bulletin in Applied Statistics*, Vol 9, No 2, 179-193, (1982).

[9] Kimber A.C. and Hansford A.R., A statistical analysis of batting in cricket, *Journal of the Royal Statistical Society A*, Vol 156, No 3, 443-445, (1993).

HOW FAIR IS THE AFL DRAW?

Stephen R. Clarke[1]

Abstract

This paper discusses the fairness of the AFL home and away draw. The use of a linear model to estimate team ability and home advantage is shown. The degree of difficulty inherent in the draw is estimated by calculating the average standard of the opposition. However an extra term is necessary to allow for better teams playing weaker than average opponents. The technique is demonstrated using final ladder position as a measure of team ability. The draw difficulty is shown to be different for different teams, and the bias does not even out over the years. Another method compares the expected final ladder position under the actual draw with the fair position. The efficiency of the draw to reward better teams with higher final ladder position is estimated via a simulation. This shows the huge amount of variability due to randomness inherent in the League draw.

1. INTRODUCTION

All sports are affected by the overall rules of the competition. The Australian Football League along with individual clubs makes many decisions affecting the running of the competition. These are often based on financial aspects such as to maximise crowds or television exposure, but they also affect teams' chances of success in the competition. They range from relatively minor changes as in moving the venue of a single match or moving the home ground of a club for an entire season, through to decisions having major ramifications such as organising an unbalanced draw or alternative play-off structures. What effect do these decisions have on a team's chances? In the past these have not been quantified. Clarke [1] quantified the effects of the various final systems on teams' chances. Here we look at various methods for measuring the fairness of the home and away draw.

A major drawback of the AFL competition is that the draw is not balanced. In AFL football, 16 teams play 22 rounds, so teams do not play all other teams twice. The draw is unbalanced with respect to strength of opposition (each team plays a different set of opponents twice) and with respect to grounds (teams play a different number of matches on their home grounds). The introduction of interstate teams has seen an increase in the home advantage, so the difficulty of a match against an interstate opponent depends very much on where it is played. While the general public recognise this is inequitable, again it has never been quantified in a proper manner. At the very most a football writer may tabulate the number of times each team plays a weak team, or a finalist from the previous season, but never is it done at the end of a season when the true strengths of the teams are better able to be

---

1   School of Mathematical Sciences, Swinburne University, Hawthorn Vic 3122

estimated. This unfairness will not necessarily even out over the years. For example at one time the draw was made on the basis that the top teams in one year played each other twice the following year. Thus there was an ongoing bias in the draw.

2. MEASURING TEAM ABILITY

The end of year ladder is usually used as a measure of a team's success. This is often equated with their ability. However this presupposes that all teams had an equally difficult task. To correctly measure a team's playing level, strength of opposition must be allowed for, and this is often difficult to assess because the draw is not balanced.

Mathematical models can be fitted to estimate team ability and home advantage. Harville [2, 3], Stefani [4, 5, 6] do this for American football, Stefani and Clarke [7] for Australian Rules, Clarke [8] , Clarke and Norman [9], Kuk [10] for English soccer and Harville and Smith [11] for American basketball. Clarke [12] has shown that in Australian rules football, models allowing different home advantages for different teams are justified, so we model the winning margin w_{ij} when home team i plays away team j as

$$w_{ij} = u_i + h_i - u_j + e_{ij} \tag{1}$$

where u_i is a measure of a team's ability, h_i is team i's home advantage (HA) and e_{ij} is random error. Since the u_i are relative, we can require they sum to zero. The u_i, h_i can be found using least squares, and the results for 1995 are given in Table 1 along with the final ladder.
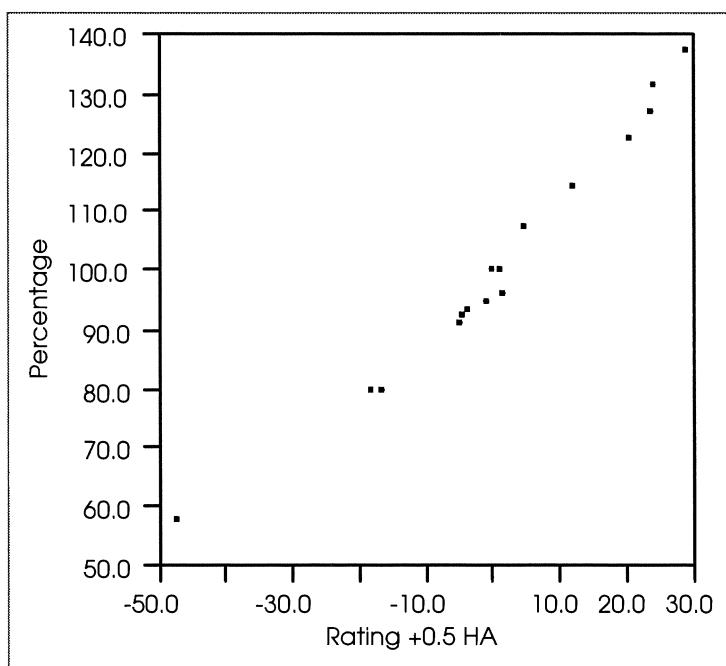
Since teams have traditionally played half their matches at home we might use $u_i +$ 0.5 h_i as a measure of a team's success through the year. This is in line with Harville and Smith [11] who suggest an equivalent measure for a team's overall performance level in relation to the average performance level. In Table 1 the rank order of the teams on this measure is also given. Note that some teams (Richmond, Footscray) have done better than they deserved, others (Collingwood, Sydney) have done worse. This may be due to draw difficulty, or just an effect of the large random variation due to teams winning or losing close matches.

For 1995, the measure $u_i +$ 0.5 h_i has a correlation of 0.90 with premiership points and 0.98 with percentage. Figure 1 shows a scatter plot of percentage against $u_i +$ 0.5 h_i and demonstrates the extremely close fit. Thus the u_i and h_i together give a good measure of a team's overall success through the year, but separately give a measure of how much contribution the effects of team ability and HA made. It also suggests that percentage is a better measure of a team's average performance level than premiership points, and may be a good surrogate to use for average strength of opposition rather than go to the trouble of fitting u_i and h_i by least squares.

Table 1

Actual final ladder for 1995, with team ratings and has shown

Team	Prem Points	Percent.	Rank by ladder	u	h	$u + .5h$	Rank by $u + .5h$
Carl	80	137.8	1	26.2	4.9	28.7	1
Geel	64	131.9	2	26.2	-4.9	23.7	2
Rich	62	107.9	3	0.9	7.0	4.4	6
Ess	60	127.6	4	26.9	-7.5	23.2	3
WC	56	122.9	5	6.6	26.9	20.1	4
NthM	56	114.8	6	22.5	-22.1	11.5	5
Foot	46	91.5	7	-2.3	-6.0	-5.3	13
Bris	40	95.3	8	-11.1	19.9	-1.2	10
Melb	36	100.7	9	3.8	-6.6	0.5	8
Coll	36	96.8	10	-1.6	5.6	1.2	7
Adel	36	80.1	11	-31.4	28.7	-17.1	14
Syd	32	100.7	12	-4.4	8.0	-0.4	9
Frem	32	92.8	13	2.3	-14.5	-4.9	12
StK	32	80.3	14	-15.3	-6.8	-18.7	15
Haw	28	94.0	15	-8.3	8.7	-4.0	11
Fitz	8	58.2	16	-41.0	-13.4	-47.7	16

Figure 1: Percentage versus ui + 0.5 hi for 1995

3.   FAIRNESS OF THE DRAW - AVERAGE STRENGTH OF OPPONENTS

Once ratings for teams are established, it is relatively simple in principle to quantify the unfairness of the draw after the season. Since the ratings of an opponent are a measure of the difficulty of that match, summing the ratings of the opponents of each team gives a measure of the difficulty of the draw for that team. (The home ground advantage of opponents could also be included as this contributes to the draw difficulty). This is equivalent to the approach of Leake [13] who suggested the average rating of opponents as a measure of schedule difficulty. However there are problems with this approach. Since the good teams do not play themselves, they will appear to have an easier draw than the others. Thus even in a balanced competition this method would give a measure of unbalance. For this reason we need to subtract the average strength of all possible opponents. Thus we are measuring the excess strength of the actual opponents over the average strength of all possible opposition. We show this is equivalent to adding a proportion of a team's own rating to account for the above bias.

If the measure of team ability in an N team competition is u_i, $i = 1$ to N, where $\Sigma u_i = 0$, then opponent j will exceed the average strength of all possible opponents of team i by

$$u_j - \frac{\sum\limits_{j \neq i} u_j}{N-1} = u_j - \frac{-u_i}{N-1} = u_j + \frac{u_i}{N-1} \qquad (2)$$

Summing this for all opponents is a measure of the total strength of opposition to team i.

While we could use the us as derived earlier, or better still $u_i + 0.5\,h_i$, there are advantages in using a measure that the general football follower would understand. Since percentage is highly correlated with $u_i + 0.5\,h_i$, this may be a good choice. Here we use a popular measure of a team's ability, final ladder position. This incorporates both team ability and some measure of home advantage. Unfortunately it also includes a component due to the factor we are measuring - draw difficulty, but we bear with this in the interests of having a simple measure. Table 2 was obtained by applying (2) using the ladder ranking above the mean at the end of the year. Because a low number indicates a high ranking and strong opposition, a negative total indicates the draw was more difficult than average, a positive number easier than average. Note that during the year 1986 all teams had a balanced draw, and this is true for several years prior to that. In other years, the difference between highest and lowest is generally about 35. This is clearly a significant amount, particularly for two teams in a similar position on the ladder, where the difference cannot be attributed to the different rankings of the two teams. For example in 1988, Geelong, one position on the ladder ahead of Richmond, had a more difficult draw by 36 ranking points. That is the equivalent of playing the top three teams instead of the bottom three teams. In the same year West Coast finished one spot above Melbourne with the same number of wins. However Melbourne's draw was 29 ranking points harder than West Coast. A similar draw could have given Melbourne three extra wins and put them second on the ladder. (They actually did win their way through to the

grand final). In 1995, the two teams with the hardest draw, Melbourne and Collingwood, both missed the final eight by one game, even though they had better percentages than Footscray who finished seventh and Brisbane who finished eighth. Again the difference in their draw difficulty could easily account for the difference. It is clear that the degree of imbalance that exists in the draw is enough to have a significant effect on the final ladder outcomes. Individual clubs should also look at their draw difficulty in assessing the measure of success they have achieved through the year.

Table 2

Measure of draw difficulty for AFL teams, 1980-1995

Team	Year										Sum
	86	87	88	89	90	91	92	93	94	95	Sum
Adel						-9	-3	4	-7	-1	-18
Bris		13	-9	-14	8	-6	-10	2	18	-5	-1
Car	0	-10	12	5	-9	-5	-4	0	-3	-7	-21
Coll	0	0	6	3	-7	0	14	-12	-6	-17	-19
Ess	0	-2	2	17	-4	6	-14	-4	-1	9	9
Fitz	0	-12	-12	12	2	-3	4	7	10	-4	5
Foot	0	-9	12	-7	0	5	8	-9	-11	6	-4
Frem										-8	-8
Geel	0	-11	-20	13	3	1	6	-14	-6	19	-8
Haw	0	0	13	14	-4	9	-5	2	-6	-6	16
Melb	0	16	-19	-12	5	4	-2	9	-2	-16	-17
NthM	0	5	-12	-16	1	1	1	20	-9	-1	-9
Rich	0	3	16	-7	3	5	6	4	3	15	47
StK	0	14	-9	-14	-9	9	-9	3	16	4	4
Syd	0	-1	10	12	4	-3	6	-12	12	9	38
WC		-6	10	-6	7	-14	2	0	-8	3	-12
No. of Teams	12	14	14	14	14	15	15	15	15	16	

Clearly the draw difficulty does not even out through the years. Richmond and Sydney appear to have had a long run of good draws, while Carlton has had a long run of more difficult draws. Many AFL clubs have criticised the level of financial support given to Sydney. They have also, it appears, received support from the schedule.

4.   USING A COMPUTER PREDICTION TO ASSESS FAIRNESS OF HOME AND AWAY DRAW

Clarke [14, 15] describes a computer prediction model that updates ratings and home advantages using exponential smoothing. The program includes a component that estimates a final ladder, and a simulation that estimates the chances of teams finishing in any position. This can be used to obtain alternative estimates of the effects of the draw on the success of clubs. This gives a measure of the effect in ladder positions. Any sporting competition is designed to produce a winner, and the rules should ensure the expected final positions reflect the abilities of the participants. The ladder prediction model provides a tool to investigate this. Given the ratings of each team and the draw it provides the expected finishing position of all teams. This takes into account not only the opponent, but the ground on which the matches are played. While this is usually used at the start of the season for forecasting, it can be run at the end of the year using the average ratings the teams actually achieved. The expected ladder position can then be compared with the team ratings. In a competition balanced for quality of opposition and HA, the expected final ladder would be roughly in the same order as $u + 0.5h$. Variations from this reflect unfairness in the draw. For 1995 this resulted in a predicted order in good agreement with the fair order. At most there was one game difference between expected and fair ladder position.

We might also wish to investigate the extent to which the final ladder position is affected by random variation. A season of football has a large random element, and most supporters recognise that luck plays some part in the success of their club. Also club success is not a linear function of ladder position. For example, obviously two seconds would not be equivalent to a first and third. For both these reasons it is appropriate to look at the probabilities of teams achieving certain goals. In racquet sports, for instance, this has resulted in the concept of efficiency of scoring systems, where the length of matches is traded off against the probability of the better player winning (Miles [16]).

While it is outside the scope of this paper to investigate alternatives, we do want to give an idea of the effects of random variation on the final ladder. The simulation model can be used to give an indication of its extent. Table 3 is the result of simulating the 1995 season 1000 times using the average rating for the teams as initial ratings. The table shows the number of seasons the team finished in the given position, and has been sorted in order of $u + 0.5h$, i.e. in the order that a fair draw should produce. While the most likely position was generally close to the fair ranking, the probability of this was often quite low. The table shows the huge variation possible in a season of football and demonstrates the dangers in putting too much emphasis on the final ladder position as a measure of the team's performance. It is possible for almost any team to finish anywhere from last to first due to the random effects. The range within which a team was an 80% chance to fall was about four positions for the very best and worst teams, up to about 10 positions for some of the middle teams. This dependence on chance can be demonstrated by looking at individual matches. In round 9, Adelaide beat Hawthorn 9.06 to 7.16 by two points. Had just one of Hawthorn's 16 behinds been a goal, Hawthorn's final ladder position would have been three places higher and Adelaide four places lower. In contrast, the

u_i and h_i for those teams as developed by linear models or exponential smoothing would have hardly altered at all. For this reason, measures obtained by model fitting are a more accurate reflection of a team's performance through the season.

Table 3

Chances in 1000 of ending in any position after home and away matches

Team	Final Ladder Position															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Car	302	218	136	113	83	63	35	26	11	6	3	4	0	0	0	0
WC	184	189	156	139	91	77	63	32	25	18	14	5	5	1	1	0
Geel	236	185	170	121	93	59	52	31	22	11	7	5	6	2	0	0
NthM	103	114	124	114	125	104	88	66	41	43	33	21	9	9	5	1
Ess	107	145	163	151	123	81	76	51	40	11	18	13	10	9	2	0
Melb	18	38	53	68	82	106	98	98	95	91	66	59	59	49	18	2
Rich	26	53	73	90	117	114	95	110	79	91	37	46	30	24	14	1
Coll	10	15	29	36	55	60	87	73	109	87	105	93	90	95	48	8
Foot	3	5	21	39	36	66	93	86	102	99	113	107	95	71	53	11
Haw	2	13	27	44	50	70	78	110	78	107	91	103	76	76	61	14
Adel	2	10	13	31	50	54	68	83	107	93	112	90	110	101	63	13
Syd	4	6	15	22	36	59	59	85	99	87	114	110	104	104	76	20
Bris	1	4	10	16	24	36	48	56	68	100	102	102	147	126	137	23
Frem	1	3	6	9	23	33	39	62	59	93	94	131	129	137	136	45
St. K	1	2	4	7	12	18	21	30	63	62	88	102	116	152	241	81
Fitz	0	0	0	0	0	0	0	1	2	1	3	9	14	44	145	781

5. CONCLUSION

The AFL draw is clearly not balanced for strength of opposition nor home advantage. The difference in the degree of difficulty faced by some teams is the equivalent of playing the three top teams rather than the three bottom teams. This bias does not even out over the years, and some teams appear to be consistently advantaged or disadvantaged. However there is some evidence that the bias appears to make at most a difference of one place in the expected and fair ladder position. The inherent variability in the draw due to randomness far outweighs this. An average team's actual finishing position can be anywhere from top to bottom due to the inherent randomness of football.

References

[1]   Clarke S.R., Calculating premiership odds by computer - an analysis of the AFL final eight play-off system, *Asia Pacific Journal of Operational Research*, Vol 13(1), 89-104 (1996).

[2]   Harville D., The use of linear model methodology to rate high school or college football teams, *Journal of the American Statistical Society*, Vol 73(358), 278-289 (1977).

[3]   Harville D.A., Predictions for National Football League games via linear-model methodology, *Journal of the American Statistical Association*, Vol 75, 516-524 (1980).

[4]   Stefani R.T., Football and basketball predictions using least squares, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol 7, 117-121 (1977).

[5]   Stefani R.T., Improved least squares football, basketball and soccer predictions. *IEEE Transactions on Systems, Man and Cybernetics*, Vol 10(2), 116-123 (1980).

[6]   Stefani R.T., Applications of statistical methods to American football, *Journal of Applied Statistics*, Vol 14, 61-73 (1987).

[7]   Stefani R.T. and Clarke S.R., Australian rules football during the 1980's, *ASOR Bulletin*, Vol 10(3), 11-15 (1991).

[8]   Clarke S.R., Home advantages in balanced competitions - English soccer 1990-1996, in *Mathematics and Computers in Sport*, N. de Mestre, Editor, Bond University: Gold Coast, Qld. 111-116 (1996).

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

[10] Kuk A.Y.C., Modelling paired comparison data with large numbers of draws and large variability of draw percentages among players, *The Statistician*, Vol 44(4), 523-528 (1995).

[11] 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*, Vol 48(1), 22-28 (1994).

[12] Clarke S.R., Performance modelling in sport, Unpublished PhD Thesis. Swinburne University of Technology. (1997).

[13] Leake R.J., A method for ranking teams with an application to college football, in Management Science in sports, R.E. Machol, S.P. Ladany, and D.G. Morrison, Editors. North Holland. 27-46 (1976).

[14] Clarke S.R., Computer forecasting of Australian Rules football for a daily newspaper, *Journal Operational Research Society*, Vol 44(8), 753-759 (1993).

[15] Clarke S.R., Computer and human tipping of AFL football - a comparison of 1991 results., in *Mathematics and Computers in Sport*, N. de Mestre, Editor, Bond University: Gold Coast, Qld. 81-93 (1992).

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

PLAYER ASSIGNMENTS IN AUSTRALIAN RULES FOOTBALL

Nikoleta Tomecko[1] and Jerzy A. Filar[1]

Abstract

This paper deals with work in progress on an application of operations research and statistical techniques to Australian Rules Football.

The Analytic Hierarchy Process is discussed as one of the ways for extracting expert knowledge from the coach and the selectors about the importance of game skills in each position. It is further proposed that this knowledge is used to construct "performance functions" which represent the expected performance of each player in each position and that these are exploited as inputs to the Assignment Problem. Finally, future research directions are presented.

1.   INTRODUCTION

Australian Rules Football is the national winter sport of many Australians. An "Aussie rules" match consists of four quarters, 30 minutes each. Each side has 18 players on the field at any one time. Scoring consists of goals, worth 6 points each, (when the ball passes between the two major goal posts) and "behinds", worth 1 point each (when the ball passes between a major and a minor post or it hits one of the major posts). The team with the higher score at the end of the match wins.

All of the major AFL teams have whole teams of "statisticians" who score all the team's matches, as well as the opposition's. This suggests that the clubs themselves recognise the need to be selective about both the statistics that they are going to collect, and about the methods of analysing the data so collected, to get the maximum possible benefit. However, little mathematical modelling is currently being done.

This project was run jointly with the Traralgon Football Club. The aim of the work is to assist an Australian Rules Football coach in his decisions. This involves assisting the coach in the selection of players, monitoring player performance and allocation of players to positions. It is not the aim of this work to devise an automated system which might, at some future stage, aspire to replace the coach. The aim is merely to capture some of the coaches' intuition and knowledge to produce a decision-support system that would simplify their decision-making tasks.

It is envisaged that simple statistical methods will be utilised to assist the coach with compiling the statistics and monitoring the performance of players and the team. When allocating players to positions, it is proposed that a classical operations

---

[1]   School of Mathematics, University of South Australia, Mawson Lakes SA 5095

research method, called the "assignment problem" be employed, while a technique called Analytic Hierarchy Process be used in conjunction with statistical results to evaluate the players' suitability for a position and in the process to supply the coefficients in the assignment problem.

If the system performs as well as might be expected it may be possible for the coach to enter the data of an opposition's "reshuffle" into the system and for an appropriate response to be calculated in "real time".

If this proves to be a successful method of assisting the coach in assigning the players to positions, it is possible to modify the method to adapt the system to any team sport where the positioning and allocation of individual players and the interactions between them is important.

2.   PERFORMANCE EVALUATION

If there were a known "performance function" that accurately predicted each player's performance in every possible position on a football field, then the problem of optimal assignment of players to positions would be the classical (and well-solved) operations research problem known in the literature as the "assignment problem" (Murty [1]).

However, such a performance function does not exist. Hence, the focus of the project is the evaluation of each player's suitability for each position on the field. Of course, this gives rise to a large number of possibilities (about 6,402 trillion in fact). In practice, each player is likely to be considered only for a few of the positions, nominated by the coach and/or the selectors.

The method which is going to be used to construct the performance evaluation function has to satisfy several criteria.

1. It has to be easy to use as the coach and the selectors cannot afford to devote a large proportion of their time interacting with the system.

2. The method has to be readily understood, so that they are clear on how it can be used in their decision making process and are not reluctant to take advantage of its capabilities.

3. There has to be the potential to use physical characteristics and training scores of a player, or his past performance data, or both.

4. The method has to be able to accurately represent the situation in that for every position there are a number of possible criteria that play a role. For example, a ruck has to be tall, have good ball skills and be able to run fast. A full forward has to have all of those abilities and in addition he must be good at marking and scoring goals.

Ball skills of a player could include his kicking abilities, which includes both accuracy and distance, marking and handpassing the ball to team-mates. Also, not all of the

mentioned abilities are of equal importance, in fact their importance varies with the position that is being considered.

This points to multiple-criteria decision-analysis techniques, and the number of "nested" criteria to be considered in the function suggests a possible hierarchical structure. To illustrate this on the full forward position, suppose that the broad attributes are as follows: ball skills, form, physical characteristics. Each of these attributes may consist of more sub-attributes. Figure 1 schematically illustrates this structure for the full forward position.

For instance, in this hierarchical model the sub-attributes are:

games	the number of games the player has played
AFL	the level of AFL experience of the player
position	the experience of the player in the position
distance	the player's kicking distance
accuracy	the player's kicking accuracy
physical	the player's physical form
psychological	the effect that the player has on the other players
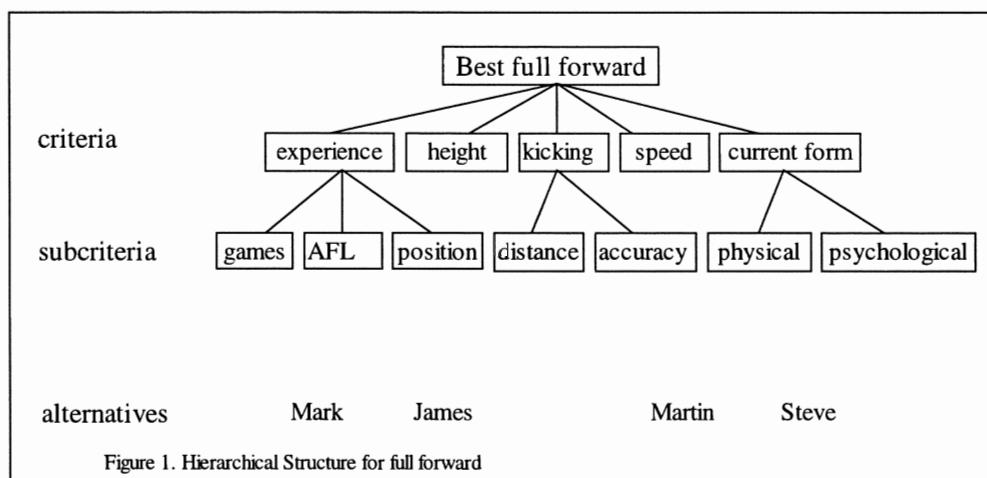
Figure 1. Hierarchical Structure for full forward

This is consistent with the way the hierarchy of attributes are constructed in the Analytical Hierarchy Process, see (Saaty [2], [3]). AHP also satisfies our other criteria in that the pair-wise comparisons being performed are reasonable since it is desirable to compare each player to every other candidate for that position. It is also very easily interpreted in terms of the football problem, because the priority vector is normalised, which means that the weights associated with attributes and sub-attributes can be interpreted as percentages.

The AHP also allows for the option of entering data, so training scores or past match statistics can be included in the process of determining the weights. Another bonus of the AHP approach is that the system is easily "personalised", in that the weights for the attributes and hence the final scores for the player's performance are all dependent on the initial pair-wise comparisons. This means that even given the same initial hierarchy, two different coaches would typically assign different weights to

attributes, hence it would be possible for different players to receive the top performance rating.

For example, suppose that for the full forward position, the top level priority matrix, (based on the pair-wise comparisons of the attributes) is as follows

$$\mathbf{A} = \begin{matrix} \text{speed} \\ \text{height} \\ \text{experience} \\ \text{form} \\ \text{kicking} \end{matrix} \begin{pmatrix} 1 & 1/4 & 1/5 & 1/4 & 1/5 \\ 4 & 1 & 2 & 1 & 1 \\ 5 & 1/2 & 1 & 2 & 1/3 \\ 4 & 1 & 1/2 & 1 & 1/4 \\ 5 & 1 & 3 & 4 & 1 \end{pmatrix}.$$

The priority weights assigned to these attributes, based on $\mathbf{A}$, are given by the vector

$$\mathbf{w} = (0.049 \quad 0.250 \quad 0.181 \quad 0.164 \quad 0.375).$$

Next, the process is repeated with pair-wise comparisons performed for the branches at lower levels of the hierarchy, for the full forward. The whole hierarchy including the weights for all levels is illustrated in Figure 2.
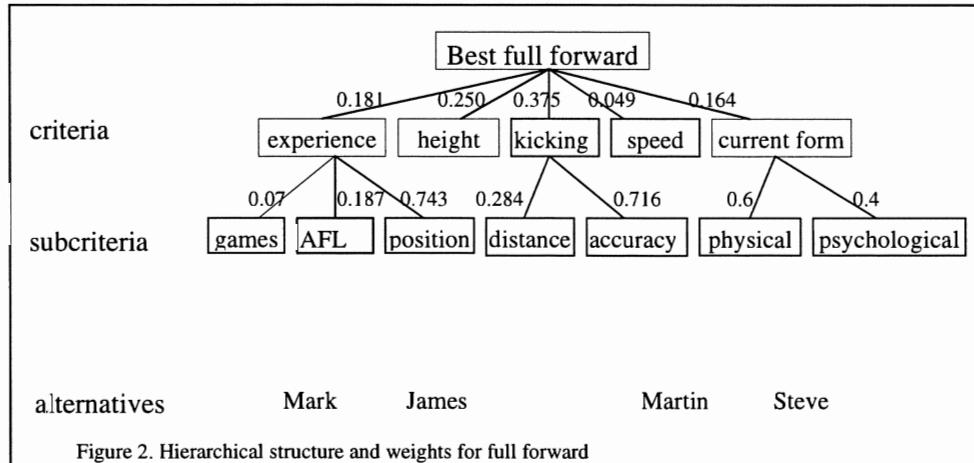
Figure 2. Hierarchical structure and weights for full forward

Similar hierarchical models can be formulated for all the positions on the field; each one consisting of attributes important in that position.

Once a hierarchy has been designed for each position and all the appropriate weights assigned, players are simply compared on each of the attributes. An overall score is assigned to each player according to the hierarchy for each of the positions that they are thought to be suitable for. This results in a single performance score for each player for each position that they are being considered for. Hence the performance indicator for player i in position j will be denoted as *per(i, j)*.

2.1   Scales

Even though each player is not considered for every position, this method could still involve a lot of comparisons. In addition to this, these scores have to be re-evaluated after every match to account for the latest performance figures as well as whenever any one player's situation changes (say, when a player becomes injured).

To overcome this, it is proposed that the "ratings" feature of AHP is utilised. This involves designing "scales" for each of the attributes. Then pair-wise comparisons of the levels of scales are performed and each level of the scale is assigned a score according to these comparisons. Then each player is simply given a rating or a score on each attribute according to the corresponding scale.

This greatly reduces the work involved since there is no need to perform all pair-wise comparisons after every match to update the data. Instead, any ratings that have changed are re-entered. Similarly, if any individual's performance score changed previously, every AHP hierarchy which included that player had to be re-evaluated. In the case of ratings, only the rating which has changed needs to be entered again, and the re-evaluation of the hierarchy is done automatically.

3. ALLOCATION OF PLAYERS

Once the performance functions have been designed, a linear programming formulation of the assignment problem can be invoked, to actually assign players to their positions. Recall that the assignment problem is of the form:

$$\max \sum_i \sum_j per(i,j) x_{ij},$$

subject to:

$$\sum_i x_{ij} = 1, \text{ where } j = 1, 2, \ldots n$$

$$\sum_j x_{ij} = 1, \text{ where } i = 1, 2, \ldots m$$

$$x_{ij} \geq 0, \quad i = 1, 2, \ldots m \text{ and } j = 1, 2, \ldots n$$

where the functions *per(i,j)* are the performance functions constructed using the AHP. An important feature of this linear program is that all basic feasible solutions are integer valued, thus there is no need to impose integrality constraints. Namely, every x_{ij} will be 0 or 1, with 1 corresponding to player i being assigned to the j-th position. Efficient special purpose algorithms are known for this problem (Murty [1]).

The constraints ensure that only one player is assigned to each position and that each position is only given to one player. Other convenient constraints could be added so that the problem takes full advantage of the structure of the matrix, for example no player with a score of 0 for any position will be assigned that position, etc. With the extra constraints we have to return to a 0-1 integer formulation, which should still be tractable numerically.

3.1 Opposition's Assignment of Players

It is common practice in most team sports to observe how the opposition is playing, watch for and try to pinpoint their strategies, and respond in a manner that lets the team capitalise on the situation (or at least to minimise the team's disadvantage). Australian Rules Football is not an exception in this case. A great deal of effort is being devoted by the top teams to monitor and score not only their own players' efforts, but those of the opposition's as well.

This means that the coach has to constantly observe both sets of players and re-evaluate his options based on how they are performing. Then he has to respond by changing the strategy or trying to assign his players to make the most of the situation. A lot of this effort is dedicated to the allocation of opposition's players and to finding an appropriate response.

This section presents and discusses some methods that can be employed to exploit the information on opposition's players collected during matches. This information can be either in quantitative form (data) or qualitative form (coach's impressions). It is proposed that this information is used to estimate the effect each player has on the performance of the opposing player from our team.

In real life the coach will not know the opposition's assignment for certain until the team is out on the field before the first bounce (sometimes not even then). However, most of the time there are a lot of very good "educated guesses" made by either the opposing team's officials or by sports commentators about how the teams are going to assign their players. Some sports commentators and reporters have, in fact, made an art-form of guessing and/or predicting the player allocations for a big match that is coming up. Any of these could potentially be used as the initial "best guess", to be altered once the selection becomes known.

This gives rise to an additional complication: ideally, the system would operate fast enough to be useable in "real time". That means that on the side-lines, or in the coach's box, when the coach/selectors see a reshuffle of the opposition's players or a new player coming on the field, they should be able to simply enter this information into the system and obtain an optimal response.

The assignment problem discussed previously is easily modified to include the opposition's players and their respective positions. The new linear program is of the form:

$$\max \sum_i \sum_j per(i, j, h_j) x_{ij},$$

subject to:

$$\sum_i x_{ij} = 1, \text{ where } j = 1, 2, \ldots n$$

$$\sum_j x_{ij} = 1, \text{ where } i = 1, 2, \ldots m$$

$$x_{ij} \geq 0, \quad i = 1, 2, \ldots m \quad \text{and} \quad j = 1, 2, \ldots n$$

where $per(i, j, h_j)$ is the estimated performance resulting from the matching up of our player i with opposition's player h_j in position j. It is still possible to develop the performance functions for our players by adjusting $per(i, j)$ to account for a "reduction" that will depend on the identity of the opposition's player. One simple construction would reflect a linear relationship between the performance of our player and the influence of the opposition's player in the corresponding position. For instance, this relationship might be represented by the equation

$$per(i, j, h_j) = per(i, j) - per(i, h_j),$$

where *per(i, h_j)* is the amount by which opposition player h_j reduces the performance of our player *i* in position *j*. Recall that the term *per(i, j)* is the already-defined performance of our player *i* in position *j*.

There are several ways that the opposition's contribution could be evaluated. It is likely, however, that all the information which is available on our players (such as the physical fitness data or training scores) might not be available for the opposition's players. This lack of information has to be taken into account when modelling the effect that the opposition's players have on the performance of our players.

One way of overcoming this could be by basing the performance functions of the opposition players on the performance model built for our own players. The latter might need to be modified to omit those variables/measurements which were monitored for our players but not for the opposition players. Most AFL teams collect at least some statistics on their players and some of the match statistics might be available from sources such as newspaper reports, etc.

Another way that may be considered is to rely on the coach and/or the selectors to give an estimate of the performance reduction based on the opposition players' past (recent?) games against our team, as well as against other teams. The coach may be able to provide a percentage estimate based on his intuition, for example, *"If 'h' is in full forward, he will reduce the efforts of our full back by a half,"* or, *"If 'h' is in full forward, he will significantly/moderately/insignificantly reduce the efforts of our full back"*. The latter qualitative statements by an "expert" such as the coach can then be converted to a numerical reduction term or a factor. For instance, we might obtain the expression

$$per(i,j,h) = per(i,j)red(i,h_j) \; ; \; 0 \leq red(i,h_j) \leq 1$$

where *red(i, h_j)* is now a multiplicative performance reduction factor.

Hence the player of the opposition is reducing our player's effectiveness by a certain percentage, where 0 value means that our player's efforts are completely annulled and a value of 1 means that the opposition player has no effect on the performance of our player. It is expected that *per(i, h_j)* = 0 or 1 would occur very rarely. However the case (where *per* = 0) is interesting from linear programming point of view, since this might have an effect on the interaction terms mentioned in the next section.

4.   INTERACTIONS

So far the discussion regarded each player i as an independent individual on the field. That is, player i contributes to his team total performance only via his individual performance in the assigned position. In a real game, players interact on the field and form good, indifferent, and bad partnerships of two or more players. A more sensitive system would account for these interactions.

From the mathematical modelling perspective, the issue of accounting for interactions appears deceptively simple. For instance, the interactions between pairs of players could be included in the objective function as terms of the form:

$$int\ (i, j, k, l)\ x_{ij}\ x_{kl}$$

where the interaction term *int(i, j, k, l)* represents the contribution of the pair of players *i* and *k* to the team if they are assigned positions *j* and *l*, respectively.

However, to consider the above extension fully would create almost insurmountable difficulties both from mathematical and practical perspectives. Firstly, the mathematical program to be solved now becomes an indefinite quadratic, which, algorithmically, can be prohibitively difficult. Secondly, the burden of estimating/modelling all the interaction coefficients *int(i ,j, k, l)* would be equally prohibitive.

Perhaps, the only practical way to include a (very) limited number of these interactions, is to identify a small number of 4-tuples *(i ,j ,k ,l)* which – according to the coach's judgement – are especially important to the performance of the team and to model the corresponding interaction coefficients for these 4-tuples only. The resulting, indefinite quadratic program with 0-1 variables might still be tractable via integer programming heuristics for non-linear programs (Pardalos and Rosen [4]).

A possible method for calculating the above interaction coefficients could involve observing and evaluating the success of such a combination in the past. For instance, for a combination of two particular players in the full forward and forward pocket positions, one might score the number of successful passes, and/or the percentage of times that the combination has scored after gaining possession, and compare this to a match (or a period during the match) when the full forward was unsupported.

Of course, in principle, interactions among three or more players could be incorporated along the same lines. Once again, this capability is limited by both the modelling/estimation and the algorithmic considerations.

5.   CONCLUSION AND FUTURE RESEARCH

A methodology has been discussed that exploits Analytic Hierarchy Process to construct performance functions for individual players in Australian Rules Football. Once these functions are constructed/estimated they can be utilised, as coefficients, in a mathematical program designed to maximise the overall performance of the team. The latter is of the form of "Assignment Problem", a classical Operations Research problem that calculates the optimal assignment of the players on the field. Methods were discussed that could be employed to extend this work to include interactions between players of the same team.

Once the performance indicators have been decided upon, the effect of opposition players may also be considered. To achieve this, two special forms of "performance reduction" functions were proposed.

Future research efforts will be concentrated on fine tuning of the player performance indicator functions and on the incorporation of the interactions between team-mates in the linear program.

REFERENCES

[1]   Murty K., Linear Programming, George Wiley and Sons, 1983.

[2]   Saaty T.L., The Analytic Hierarchy Process, McGraw Hill, 1980.

[3]   Saaty T.L., Fundamentals of Decision Making and Priority Theory with the AHP, RWS Publications, 1994.

[4]   Pardalos P.M., and Rosen J.B., Constrained Global Optimization: Algorithms and Applications, Springer-Verlag, 1987.

IT'S JUST NOT CRICKET

George A. Christos[1]

Abstract

"Rain, rain, go away, come again another day"

Cricket authorities have been wrestling with the problem of ensuring a fair contest when a one-day limited-overs cricket match is affected by rain. The original run-rate system was so biased towards the team batting second that whenever rain looked threatening the team that won the toss invariably chose to bat last. A new system was introduced in Australia and New Zealand for the World Cup of Cricket in 1992, but this was a failure since it gave the team batting first an unfair advantage. A modified version of this system is currently used in Australia. A completely new system was introduced by the International Cricket Council (ICC) in 1995. The problem with these next generation systems is that they are artificial and unnecessarily complicated. The public and most players do not understand how target scores are calculated, and it is difficult to foresee what might happen if there are further interruptions due to rain. We will review these various systems that have been tried, before we propose our own simple solution to this vexing problem. Our basic idea is that both teams compete on run-rate (who scores the most number of runs per balls faced), as in the old system, but when the number of overs is reduced the number of available batsmen or wickets is also reduced, possibly by a random deselection process. We propose that one wicket is made available for every 5 overs.

1.   INTRODUCTION

The first one-day international cricket match was played on 5 January 1971 at the MCG in Melbourne between Australia and England. This match was hastily arranged on the final scheduled day of the third test match, which had been abandoned because of rain. Since then over 1000 one-day international cricket matches have been played, and one-day cricket is now arguably more popular than test cricket. The basic idea behind one-day, or limited-overs, cricket is to see which team can score the most number of runs from a fixed number of balls faced (usually 50 overs), under similar playing conditions. In effect, the key to the game is to see which team can score at the highest average run-rate.

A problem soon emerged with this philosophy when the weather interrupted play, and time was lost, resulting in a shortening of one or both innings to less than 50 overs. Under what we will refer to as the **old system**, which was used in Australia up to 1991, and in most other cricketing nations until 1995/96, the team batting second would chase a target score obtained by multiplying the run-rate obtained by

---

[1]   School of Mathematics and Statistics, Curtin University of Technology, GPO Box U1987, Perth, Western Australia, 6845. Email: christos@cs.curtin.edu.au

the team that batted first by the number of overs that could be bowled in the second innings. This system clearly gives the team batting second a distinct advantage since they would generally be chasing a much lower total with all ten wickets available. This advantage was so great that if the weather looked threatening the team that won the toss would invariably choose to bat second. This system was known to be flawed since its inception, but was tolerated for want of a better system. The turning point really came in 1989 when the third and deciding final in the World Series Cup between Australia and the West Indies was ruined by the bias in these rules. Details of this match are given below. A new system was introduced in Australia in 1991/92 and was used in the World Cup of Cricket in 1992 but this system, to be discussed below, together with its variations, does not resolve the problem, as it now gives the team batting first a distinct advantage. In 1995 the International Cricket Council introduced yet another system, which we will also discuss below. We will argue that these systems are either inconsistent, unnecessarily complicated, or both. We will use explicit examples of one-day games to illustrate our arguments. We will then propose our solution to this problem by averaging the number of wickets over an innings.

As far as we know the ICC system, ratified in July 1995, has not been uniformly accepted. Certainly the system used in Australia is different, and up until recently the old system was still in use on the sub-continent. There is no universal rain-rule, a rule is usually decided at the start of a competition. The problem with this is that it is difficult to make proper comparisons between players and teams if more than one rule is in use. We should note that, in some county games in England, the game has been extended into the next day to ensure that both sides get the same number of overs, but this really defeats the purpose of having a one-day game as an entertainment spectacle, and the quality of the pitch can change over an extended period.

2.   FLAWS IN OLD SYSTEM

I will illustrate the flaws in the old system with three examples, which we will refer to later as old(1), old(2), and old(3) respectively.

1. In the World Series Cup match played in Brisbane on 16 January 1982 between Pakistan and the West Indies, Pakistan batted first and made 177 runs all out from their 50 overs. Rain interrupted play and the West Indies were set a target of 107 in 30 overs. The West Indies made 9 for 107 off 28 overs and 5 balls and subsequently won the match. One wonders if the West Indies would have been able to sustain this run-rate for 50 overs given that they had lost 9 wickets after 29 overs.

2. In the World Series Cup match between Australia and Pakistan played in Melbourne on 10 January 1989, Pakistan won the toss, Australia was put in to bat first and scored 4 for 258 after 43 overs, a run-rate of 6 runs an over. Rain interrupted play and Pakistan where set the comparatively easy target of 114 runs in 19 overs. Pakistan made 7 for 108 off 19 overs and Australia deservedly won that match, although Pakistan was clearly advantaged by these rules.

3. Our third example is the third and deciding final between Australia and the West Indies played at the SCG on 18 January 1989, which was mentioned earlier. In that match Australia batted first and were about 3 for 130 after about 28 overs when the match was interrupted by rain, and reduced to 38 overs. In the last 10 overs of that innings the Australian batsmen, lead by Dean Jones, scored at almost 10 runs an over to take Australia to the very respectable total of 4 for 226 off 38 overs, a run-rate of close to 6 runs an over. Before the West Indies came out to bat, rain interrupted play again and the West Indies were set a target of 149 off 25 overs. Rain interrupted play yet again, during the West Indies innings, and the target was then further reduced to 108 runs off 18 overs. The West Indies scored 2 for 111 off 13 overs and 2 balls, and won the match on run-rate. The West Indies had an enormous advantage in this match because they had all ten wickets available and only needed to sustain the run-rate obtained by Australia for less than half the number of overs. The Australian team was also disadvantaged when rain interrupted their innings after 28 overs.

3.   NEW SYSTEM

A **new system** was introduced in Australia in the summer of 1991/92 and was used during the World Cup of Cricket played in Australia and New Zealand in 1992. In this system, if play is reduced because of rain, the target of the team batting second is reduced by subtracting those runs from the total that were scored in the corresponding number of lost overs, in the first innings, that had the least number of runs scored from them. In this situation the team that bats second will need to chase a higher run-rate to compensate the fact that they are required to bat for less overs. The problem with this system is that if a there is a small reduction in the number of overs, the team batting second generally find themselves chasing almost the same total as the team that batted first but with less overs to achieve this total. In this case this system tends to favour the team that bats first, which reverses the situation with the old system. This fact was well known to teams, who generally elected to bat first when they won the toss. In the World Cup of Cricket held in Australia and New Zealand in 1992 there were 39 games in all, 2 were abandoned because of rain and 6 games were affected by rain. We believe that most of these shortened games were adversely affected by the new rules used in this competition. We will give 4 examples from this competition below, which we will refer to later as new(1), new(2), new(3), and new(4) respectively.

1. In the match between Australia and India played in Brisbane on 1 March 1992, Australia won the toss, batted first and scored 9 for 237 off 50 overs. Rained intervened and India was set the target of 236 runs off 47 overs, which is almost as much as what Australia scored but with 3 overs less. India were all out for 234 and subsequently lost the match. The irony is that India was effectively punished for the overs that they bowled where not many runs were scored, including maidens.

2. In the match between India and Zimbabwe, played in Hamilton on 7 March 1992, India won the toss, batted first and scored 7 for 203 in 32 overs. The number of overs for Zimbabwe were reduced after further rain and Zimbabwe were set a

target of 159 in 19 overs. Zimbabwe scored 1 for 104 off 19 overs and lost the match.

3. In the match between South Africa and Pakistan, played in Brisbane on 8 March 1992, South Africa batted first and made 7 for 211 off 50 overs. Pakistan were 2 for 74 after 22 overs when rain interrupted play and Pakistan's target was revised to 194 in 36 overs. Pakistan made 8 for 173 after their allocated 36 overs and lost the match. It is ridiculous that Pakistan had to chase a total only 17 runs less than South Africa with 14 less overs. Once again, Pakistan were effectively penalised for bowling so many good overs. One of the other difficulties with this new system is that it is difficult to foresee what the target will become if further time is lost to bad weather. Pakistan was also disadvantaged in this match when play was interrupted after 22 overs. From that point Pakistan required another 120 runs off the remaining 14 overs, a massive jump in the required run-rate.

4. Our last example concerns the semi-final in the World Cup of Cricket between England and South Africa played in Sydney on 22 March 1992. In that match rain delayed the start by 10 minutes. England batted first and made 6 for 252 after 45 overs. South Africa were unable to bowl the full 50 overs in the allotted time. South Africa were 6 for 231 after 41 overs and 5 balls, when rain interrupted play again. At that stage South Africa needed 22 runs off 19 balls to win the match and play in the final. When play resumed, 3 overs were lost, but South Africa's target did not change since they had bowled 3 maiden overs. South Africa now required an impossible 22 runs off the last ball to win the match.

Another problem with these biased systems is that when they give one side a huge advantage, the players and viewers generally lose interest in the match and the game in general. It is important to ensure that the rules are not only fair but are also clearly perceived to be fair.

4.   MODIFIED NEW SYSTEM

The new system used in Australia was modified in the summer of 1995/96, and is currently in use in Australia. In order to re-compensate the team batting second, which was clearly disadvantaged by the original new rules, it was decided to further reduce their target by 0.5% for each over lost. We assume that this reduction of 0.5% is applied to the preliminary target as calculated previously. For a target score of around 200 runs, this reduction corresponds to about 1 run for each over lost. Note that since the lowest scoring overs are still deducted to arrive at the preliminary target, that this system still penalises a team for bowling well in the first innings. Under these modified rules, in the game new(1), India would have had a target of 233 and would have possibly won that match. In the game new(2), Zimbabwe would have been set a target of about 149 runs, instead of 159, whereas, in the game new(3) Pakistan's target would have been 181 off 36 overs, instead of 194, and in the game new(4) South Africa would have still needed 19 runs off the last ball to win the match. Although this reduction in the target is in the right direction, this system probably still gives the team batting first a slight edge. More importantly this system, with or without it modification, is inconsistent, since it effectively penalises the team batting second if they bowl well in the first innings. We also believe that

the system used in Australia is unnecessarily complicated, as most players and public do not understand how targets are calculated, and how the situation might change if there is a further reduction in the number of overs due to rain.

A completely new method for calculating the target score of the team batting second was ratified by the International Cricket Council (ICC) in July 1995. In the **ICC system** : *"If the innings of the team batting second is delayed or interrupted and it is not able to receive its full quota of overs, the target score shall be calculated as follows: the score of the team batting first shall be multiplied by the percentage factor for the number of overs to be bowled to the team batting second, as set out on the Target Score Calculation Chart (see below). Fractions shall be rounded to the higher whole number. The percentage factors have been derived from a detailed mathematical analysis of a database of one day matches with an object to establish a "normal" performance."*

TARGET SCORE CALCULATION CHART								
Overs	% Factor		Overs	% Factor		Overs	% Factor
25	66.7		34	82.2		43	94.2
26	68.4		35	84		44	95.1
27	70.2		36	85.3		45	96
28	72.4		37	86.7		46	96.7
29	74.2		38	88		47	97.8
30	76		39	89.3		48	98.7
31	77.8		40	90.7		49	99.6
32	79.1		41	92		50	100
33	80.9		42	92.9			

As an example if the team batting first scored 250 runs in their 50 overs and the innings of the team batting second is reduced to 35 overs, then the target will be $250 \times \dfrac{84.0}{100} = 210$ runs. If the target was set on run-rates alone, as in the original rules, it would have been 175 runs, so clearly the ICC system tries to take into account that the team batting second has more wickets available for fewer overs. The ICC system is in many respects very similar to the system currently used in Australia, except that, it cleverly avoids the contradiction where the team batting second is effectively penalised for bowling well in the first innings. The ICC rule asserts that the target score calculation chart was established from a detailed mathematical analysis of a database of one-day matches with the object of establishing a "normal" performance. We suggest that this really means that these figures have devised so as to try and ensure a close game. The ICC rule also suffers from the criticism that it is too complicated for the players and the public to be able to calculate the target score themselves, and to determine the consequences of further interruptions due to rain. As the number of overs in the chart starts from 25, this may suggest that the minimum number of overs for a match to be declared valid is 25 overs per side, otherwise the match is declared to be a draw. The ICC rain-rule also does not explain what happens if the team batting first does not receive its full 50 overs. Clearly the team batting first is disadvantaged if its innings is suddenly reduced, especially in view of the fact that their run-rate generally accelerates in the last part of the innings. There is no compensation made for this in these rules. We presume that if the first innings is reduced, and the second innings is subsequently reduced further, that the way to use to chart is to divide the score of the first innings by its

appropriate percentage factor, corresponding to the number of overs in the first innings, before multiplying by the percentage factor for the second innings. Although the ICC system is probably the best system that has been devised by the cricket authorities so far, it is artificial, and cannot properly take into account the length, timing, and frequency of interruptions. In any case, we believe that the ICC system is much too complicated for the players and public to use and understand. We note that the ICC system would not have resolved the fiasco in the 1992 World Cup match between England and South Africa, example new(4). After the last interruption South Africa's target would have been recalculated to be $252 \times \dfrac{92.9}{96.0} = 244$ runs, which would have meant that South Africa would have required an equally impossible 13 runs off the last ball to win that match.

5.   PROPOSED SYSTEM

We believe that the only way to properly resolve the problem of fairness in a one-day game where one or both innings have been shortened because of rain, is to effectively average the 10 available wickets over 50 overs, with both teams competing on run-rate, as in the original rules. My proposal is that a team that bats its full 50 overs is deemed to have used all of its 10 wickets. If rain interrupts play and the number of overs in an innings is reduced, the number of available wickets is also reduced by one wicket for every 5 overs lost. More precisely, one wicket is available for every 5 overs, so if an innings is 35 overs for example there will be 7 wickets available to that team. It is obviously prudent to restrict the length of an innings to multiples of 5 overs, starting from a minimum of 15 overs say. The rule about the number of available wickets would also apply if there is a reduction in the number of overs in the first innings. If the team batting has already used up its appropriate number of wickets when rain interrupts play and the innings is shortened then that team will be deemed to be all out. Before we explain how the number of available wicket can be fairly reduced, it would be useful to go through the examples given above to see how the situation would have changed under this new **proposed system**.

• In the example old(1), the West Indies would have had to chase the target of 107 runs with only 6 wickets available. The West Indies would have probably lost that match since they were 9 out when they achieved that target.

• In the example old(2), Pakistan would have been chasing a target of 114 runs off 19 overs with 3 wickets in hand, or more appropriately the target may have been 120 runs off 20 overs with 4 wickets available, or 90 runs off 15 overs with 3 wickets available. Although Pakistan would have still lost this match, it would have been a much fairer contest under one of these scenarios.

• In the example old(3), when rain interrupted play after 28 overs in the Australian innings, and shortened the match to 38 overs, Australia would have lost 2 or 3 of its available wickets, depending on whether the number of overs was reduced to 40 or 35 overs respectively. When rain interrupted play again the West Indies would have been set a target of 149 off 25 overs with 5 wickets available, and when more overs where lost the target would have been either 119 runs off 20 overs with 4 wickets, or 90 runs off 15 overs with 3 wickets. Based on what did happen, the West Indies may

well have won that match but the situation may have been different if the West Indies knew that they only had 3 or 4 wickets available in their innings, especially since the West Indies were at one stage 2 out for 4 runs. This game would have certainly been much more entertaining and fairer in the proposed system.

- In the example new(1), India may have been set a target of 214 runs off 45 overs with 9 wickets. India would have probably won this match based on run-rates.

- In the example new(2), Zimbabwe would have been set a target of 127 off 20 overs with 4 wickets, or 96 runs off 15 overs with 3 wickets in hand.

- In the example new(3), Pakistan would have been set a target of 148 off 35 overs, with 7 wickets available. Pakistan would have won this match since they were already 157 runs at the fall of the sixth wicket.

- In the example new(4), when South Africa came back to bat for the last time they would have needed 5 runs off the last ball to win the match, as opposed to 22 runs. If the match had not resumed England would have won with a run-rate of 5.6, compared to South Africa's 5.52.

It is clear that under our proposed rain-rule these games, used as examples, would have been much more interesting and fairer. It is also apparent that the result in many of these games may have been quite different under the proposed system. In most competitions, changing just one result can completely change the final positions of the teams. This would have been certainly the case in the World Cup competition held in Australia and New Zealand in 1992.

The other important feature in our proposed system is that both teams know precisely what is expected of them to win the game at any stage of the game, irrespective of how many times rain interrupts, or may interrupt play, since all that matters is the overall run-rate, and both teams know that they effectively have one wicket available for every 5 overs. Because of this averaging of wickets, both teams are well placed to pace themselves uniformly throughout the innings. All that they need to focus on is the run-rate. In the Australian and ICC systems it is not possible to foresee what will be required if there are further interruptions to play.

The only question that remains is how to reduce the number of wickets fairly, in particular which batsmen should be chosen as ineligible. Clearly it is unfair to deselect either the top order batsmen or the bottom order batsmen preferentially. In any case if one was to devise some rule, any rule, the teams would be free to manipulate their batting order to exploit that rule. The fairest arrangement is achieved by deselecting a random sample from all of the batsmen. This can be implemented by using computer random number programs, by lot, or by lotto balls, for example. As there may be a tendency not to trust computers (since they are programmed by people), lotto balls may be the most appropriate means to deselect players from a team in a rain reduced match. On very rare occasions it may turn out that a team may be greatly disadvantaged by this deselection process, however on average the random deselection process is the fairest system possible. Another criticism of our proposed system (made by David Richards, Chief Executive, ICC, private communication 1994) is that, in the interests of public entertainment, one or

more of the top players may be deselected from batting in a shortened innings. We believe however that the public would prefer to see a fairer contest, and in any case some of these top order batsman may not get a turn to bat in a shortened match. Another possibility, that would avoid these concerns, is to allow both teams to 'equally' choose which batsman are eligible to bat. If the number of wickets available to a team is reduced to say 5, or in other words to 6 batsmen, then the team batting can choose 3 of them and the team bowling can choose the other 3. If there is an odd number of batsmen then the team batting may choose the extra batsman.

Most of the data used in this paper was obtained from the CricInfo Web Site located at the URL http://www-uk.cricket.org/link_to_database/ARCHIVE/.

ISSUES IN CRICKET AND GOLF

Derek R. Bingham[1], Basil M. de Silva[2] and Tim B. Swartz[1]

Abstract

This paper considers statistical issues related to cricket and golf data. It is a review of the recent work in cricket by de Silva and Swartz [1] and in golf by Bingham and Swartz [2]. With respect to cricket, it is shown that winning the coin toss at the outset of a match provides no competitive advantage in one-day international cricket matches. It is also estimated that playing on one's home field increases the log-odds of the probability of winning a cricket match by approximately .5. In golf, it is shown that the weaker golfer has an advantage in net medal play according to the United States Golf Association handicap system. An alternative procedure which leads to ``fairer'' golf competitions is also presented.

1.   ISSUES IN CRICKET

Over the past 10 years, there have been several papers that have considered various statistical aspects concerning the game of cricket. These include Crowe and Middledorp [3], Kumar [4], Ganesalingam, Kumar and Ganeshanandam [5], Kimber [6], Clarke [7] and Danaher [8].

As presented in de Silva and Swartz [1], we consider one-day international (ODI) matches involving games between the 9 nations belonging to the International Cricket Council (ICC). These games represent the game of cricket played at the highest level with relative stability amongst the teams. We have collected data on the 427 matches played during the 1990's up until the Asia Cup concluding in July 1997. This time period captures the modern game of cricket where the rules have been relatively uniform. It is also the case that recent data is more extensive and reliable. To keep strategies constant, we have limited the data to full 50-over matches and have ignored matches decided by run rates. The data was collected from the comprehensive CricInfo (see www.cricket.org) web page.

At the beginning of a match, a coin is tossed and the team that wins the toss is granted the choice of batting first or second. Some people believe that a team should bat first, establish a number of runs and produce a psychological hurdle for the second team to overcome. Others believe that there is an advantage in batting second as this team knows what score its opponent has produced. This additional information allows the team batting second to adjust their strategy accordingly. Still

[1]   Department of Mathematics and Statistics, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6

[2]   Department of Statistics and Operations Research, RMIT University, GPO Box 2476V, Melbourne, Victoria, Australia 3001

information allows the team batting second to adjust their strategy accordingly. Still others feel that the choice between batting first or second should depend on auxiliary and subjective variables such as the weather, the pitch (ie. field) conditions, the team's health, the team's morale, the opponent, whether the team will bat in daylight or under floodlights, etc. Clearly, this is a topic of considerable interest.

As a preliminary study, Table 1 provides summary data on the 427, ODI matches involving the 9 ICC nations. Here B_1 is the proportion of time that a team chooses to bat first upon winning the coin toss, W_0 is the overall winning proportion and W_h is the winning proportion in games played on a home field. The quantities in parentheses are the number of cases. We see from column B_1 that there is great disparity amongst the various teams with respect to their decision to bat either first or second. For example, upon winning the coin toss, Australia chooses to bat first 87% of the time whereas Sri Lanka chooses to bat first only 36% of the time.

Consider then the data (x_i, y_i), $i = 1,\ldots,n$ where n is the number of games played by the team of interest, $x_i = 1(0)$ if the team wins(loses) the coin toss in the ith game and $y_i = 1(0)$ if the team wins(loses) the game. We have the statistical model $y_i \mid x_i = 1 \sim$ Bernoulli (p_i) and $y_i \mid x_i = 0 \sim$ Bernoulli (q_i) where $P(x_i = 1) = P(x_i = 0) = 1/2$ for $i = 1,\ldots,n$. We are therefore interested in comparing the strategy $\sum_{i=1}^{n} p_i$ versus the strategy $\sum_{i=1}^{n} q_i$. Of the 427 matches in the data set, 8 games resulted in ties. We exclude these matches from the analysis.

Table 1

Summary data for the 9 ICC nations.

Nation	B_1	W_0	W_h
Australia	.87 (68)	.63 (127)	.67 (69)
England	.65 (20)	.36 (45)	.75 (4)
India	.56 (59)	.49 (106)	.72 (39)
New Zealand	.55 (51)	.37 (99)	.50 (44)
Pakistan	.47 (66)	.57 (131)	.60 (20)
South Africa	.69 (42)	.61 (94)	.71 (42)
Sri Lanka	.36 (56)	.46 (100)	.71 (21)
West Indies	.36 (39)	.53 (90)	.57 (21)
Zimbabwe	.58 (26)	.17 (46)	.36 (11)

Our first analysis requires the assumption that $p_i = p$ and $q_i = q$ for all $i = 1,\ldots,n$. This simplistic assumption is clearly unrealistic as it assumes that the opponents are all of equal strength and that the playing conditions are constant over time. However, it is a good starting point and the test of $H_0: p \leq q$ versus $H_1: p > q$ is easily carried out using a two-sample Binomial test. The first binomial variable is the number of wins having won the coin toss and the second binomial variable is the number of wins having lost the coin toss. The p-values are .53, .53, .21, .97, .59, .85, .46, .23 and .50 for each of the ICC nations as listed in alphabetical order. These p-

values are very high; to be significant, p-values must be small (typically $< .05$ if we are testing the hypothesis at 5% level of significance). Therefore, using this method, we observe no evidence of successful strategies for any of the 9 ICC teams. In de Silva and Swartz [1], 3 additional analyses based on weaker underlying assumptions each provide results in the same general direction. We therefore conclude that winning the coin toss has no impact on the outcome of ODI cricket matches.

We now turn to the existence of the home team advantage in ODI cricket matches. From columns , W_h and W_0 of Table 1, we see that every ICC nation has a higher winning percentage during home games. Using the sign test, this is convincing evidence of the existence of a home team advantage (ie. p-value $= 1/2^9 \approx .002$).

To investigate the effect of the home team advantage, we modify our notation and let p_{ijh} be the probability that team i defeats team j at site k where $i, j, k = 1, \ldots, 9$, and, in addition, $k = 0$ denotes a non-ICC site. We introduce the model $\text{logit}(p_{ijk}) = \tau_i - \tau_j + \gamma_{ijk}$ where $\sum_{i=1}^{9} \tau_i = 0$ and

$$\gamma_{ijk} = \begin{cases} \gamma & \text{if team } i \text{ is the home team} \\ 0 & \text{if the game is played on a neutral site} \\ -\gamma & \text{if team } j \text{ is the home team} \end{cases}$$

This is a 9-parameter model where τ_i is a measure of the differential strength of team i. Therefore, the offset $\tau_i - \tau_j$ represents the advantage in log-odds that team i has over team j. The model also assumes that the home team advantage γ is constant over all ICC teams. Note that the logit transformation of p_{ijk} is natural in two respects. Firstly, for teams of equal strength that play on a neutral site, we have $\text{logit}(p_{ijk}) = 0$ which implies $p_{ijk} = .5$. Therefore, there is no need for an intercept term in the model. Secondly, it is sensible to quantify the home team advantage on the log-odds scale since we should expect small relative improvements for strong teams that win most of their games. Conversely, we should expect large relative improvements for weak teams that lose most of their games.

We again exclude the 8 tied games from the 427 matches and fit the model using logistic regression. We obtain $\hat{\gamma} = .53$ with standard error .14. To put this quantity in perspective, a team with a winning percentage of 50% would increase its winning percentage to 63% when playing at home. Therefore, playing on one's home field provides a considerable edge to the home team.

2.   ISSUES IN GOLF

Previous studies on handicapping in golf (eg. Scheid [9] and Pollock [10]) have established that it is the better golfer who has an advantage in net matches between 2 golfers. This sentiment is also echoed by the United States Golf Association (USGA) where they state in section 10-2 of the USGA Handicap Formual manual (see www.usga.org/handicap/manual), "As your Handicap Index improves (gets lower), you have a slightly better chance of placing high or winning a handicap event". As

presented in Bingham and Swartz [2], we consider net medal play between 2 golfers when they are both playing well. From a tournament perspective, this is a most practical question. For a golfer does not expect to win a prize when he plays poorly, but when he plays well, at the very least, he expects a fair chance of winning a prize. Therefore, statistically, our problem reduces to looking at the tails of distributions used to model golf scores.

Data were collected from the computer handicap system at the Pemberton Valley Golf and Country Club, Pemberton, British Columbia, Canada during the 1997 golf season. To keep conditions as constant as possible, only rounds played by male members at Pemberton Valley were considered. We limit our analysis to the 49 male members who had completed 40 or more rounds during the year. We use the first 20 rounds as a tuneup period to allow the golfer to reach ``mid-season'' form. We also restrict our study to the immediate 20 rounds following the tuneup period. We hope that by using a shortened period, golfers will not experience dramatic changes in their skill levels. Each golfer will also have completed the same number of rounds of golf. Therefore our data analysis is based on $49(20) = 980$ scores.

For each golfer, we choose their best m net scores amongst the 20 rounds immediately following the initial tuneup period. Here, the net scores are obtained by subtracting a golfer's handicap as determined by the USGA's slope system from the golfer's gross (i.e. actual) score. With m scores for each golfer, there are $\binom{49}{2} m^2$ possible matches between 2 golfers that can be simulated. The matches are simulated in the sense that the 2 golfers have not directly competed against one another. We consider $m = 2, 3, 4$ as this represents the best 10%, 15% and 20% of net scores (i.e. occasions when the golfers play well). We exclude from the analysis the 5 pairs of golfers that have the same handicap index. In Table 2, we give the results of the

Table 2

Simulated matches between 2 golfers based on their best m out of 20 net scores. The percentages refer to matches won, lost and tied by the lower handicap (i.e. better) golfer.

m	Matches	Wins	Losses	Ties
2	4,684	33.4%	58.2%	8.4%
3	10,539	34.2%	57.4%	8.4%
4	18,736	35.4%	56.0%	8.6%

simulated matches and observe that the weaker golfer enjoys a dramatic advantage when both golfers are playing well. For example, with $m = 2$ the weaker golfer wins 58.2% of the matches. These results may be surprising as they are in the opposite direction of the existing literature.

We now corroborate these empirical findings with theoretical support. Consider that 2 golfers with independent gross scores X_1 and X_2 where H_{X_1} and H_{X_2} are the respective handicap strokes determined by the USGA's slope system. Without loss of generality, we assume that $H_{X_1} < H_{X_2}$ so that X_1 refers to the gross score of the better

golfer. Pollock [10] argues that the normal distribution can be used in modelling gross scores for golfers of varying skill levels. Despite obvious deficiencies in the model such as the underlying assumption of independence between golfers and the approximation of a discrete distribution by a continuous distribution, Pollock's normal model provides insight on a number of handicap issues. In this analysis, we further assume that $X_i \sim \text{Normal}[\mu(H_{X_i}), \sigma^2(H_{X_i})]$ where $\mu()$ and $\sigma()$ are increasing functions. Whereas it is obvious that μ is an increasing function, plots in Bingham and Swartz [2] provide evidence that σ is also an increasing function.

Our interest lies in the investigation of

$$P_k = \text{Prob (the better golfer wins} \mid \text{both golfers play well)}$$

$$= \text{Prob}\left(X_1 - H_{X_1} < X_2 - H_{X2} \middle| X_i - \mu\left(H_{X_i}\right) - k\sigma\left(H_{x_i}\right)\right), i = 1, 2$$

where $k > 0$. Here, $X_i - H_{X_i}$ represents the net score of golfer i and we condition on both golfers playing better than k standard deviations below their average gross score. Using the formula for conditional probability, we have that

$$P_k = \frac{\int_{x_1=-\infty}^{\mu(H_{X_2})-k\sigma(H_{X_2})+H_{X_1}-H_{X_2}} \int_{x_2=x_1-H_{X_1}+H_{X_2}}^{\mu(H_{X_2})-k\sigma(H_{X_2})} f_{X_1}(x_1)f_{X_2}(x_2)\,dx_2\,dx_1}{\int_{x_1=-\infty}^{\mu(H_{X_1})-k\sigma(H_{X_1})} \int_{x_2=-\infty}^{\mu(H_{X_2})-k\sigma(H_{X_2})} f_{X_1}(x_1)f_{X_2}(x_2)\,dx_2\,dx_1}$$

for sufficiently large k where $f_{X_i}(x_i) = \frac{1}{\sigma(H_{X_i})}\phi(\frac{x_i-\mu(H_{X_i})}{\sigma(H_{X_i})})$ and ϕ is the density of the standard normal distribution.

Bingham and Swartz [2] show that $\lim_{k\to\infty} P_k = 0$. Therefore, as both golfers play better (i.e. $k \to \infty$), it becomes impossible for the better golfer to win a match based on net scores. This conclusion is in the same direction as the empirical results.

Mosteller and Youtz [11] considered the scores of professional golfers during the final 2 rounds of PGA tournaments under ideal weather conditions. Under these homogeneous conditions, they found that the scores could be well approximated by a base score plus a Poisson variate. On the other hand, we are faced with heterogeneous conditions (i.e. data involving golfers of varying skill levels playing under various conditions). Furthermore, little is at stake for our golfers and we therefore do not expect their effort to be constant over all rounds. Consequently, we do not expect the Poisson model to provide outstanding fit. Rather, we use it as a rough approximation to reality.

Consider then 2 golfers with independent gross scores X_1 and X_2 where H_{X_1} and H_{X_2} are the respective handicap strokes determined by the slope system. Without loss of generality, let $H_{X_1} < H_{X_2}$. We then assume that the net score $X_i - H_{X_i}$ is such that

$$X_i - H_{X_i} = B_i + W_i$$

where B_i is the constant base score and $W_i \sim \text{Poisson}(\theta_i)$, $i = 1,2$. The base score B_i is meant to represent the ith golfer's idealized or perfect net score.

Under net medal play, let P be the probability that the better golfer wins when both golfers are playing their very best rounds of the year. Assuming that golfer i plays n_i rounds, $i = 1,2$, we have that

$$P = \text{Prob}(X_{1,\min} - H_{X_1} < X_{2,\min} - H_{X_2})$$

where the quantity $X_{i,\min}$ is the lowest of the n_i scores corresponding to the random variable X_i.

Under reasonable conditions, it is shown in Bingham and Swartz [2] that $P \to 0$ as $n_1 \to \infty$ and $n_2 \to \infty$. This establishes again that the better golfer has no chance of winning when the 2 golfers play their best rounds of golf. Bingham and Swartz [2] go on to estimate the probabilities P for golfers of various handicaps. This study also confirms that it is the weaker golfer who has the advantage when both golfers are playing well.

In Bingham and Swartz [2], a new net score

$$T^* = \frac{113(X - R) / S - 2.10 - 1.082I}{2.74 + 0.053I}$$

is proposed based on the normal model where X is the golfers gross score, R is the course rating, S is the slope rating and I is the golfer's handicap index. In Table 3, we repeat the analysis of Table 2 using the new performance measure T^*. We see that the outcomes of the simulated matches are far more balanced than when using traditional net scores. For example, with $m = 4$, the better golfer wins 50.3% of the matches. This is much closer to the idealized value 50% than the value 35.4% which is obtained using traditional net scores.

Table 3

Simulated matches between 2 golfers based on their best m out of 20 scores using the statistic T^*. The percentages refer to matches won, lost and tied by the lower handicap (i.e. better) golfer.

m	Matches	Wins	Losses	Ties
2	4,684	47.9%	52.1%	0.0%
3	10,539	49.0%	50.9%	0.0%
4	18,736	50.3%	49.7%	0.0%

REFERENCES

[1] de Silva B.M. and Swartz T.B., Winning the coin toss and the home team advantage in one-day international cricket matches, *New Zealand Statistician*, 32, 16-22 (1997).

[2] Bingham D.R. and Swartz T.B., Equitable handicapping in golf, Manuscript, (1998).

[3] Crowe S.M. and Middeldorp J., A comparision of leg before wicket rates between Australians and their visiting teams for test cricket series played in Australia, 1977-94, *The Statistician*, 45, 255-262 (1996).

[4] Kumar K., Is cricket really by chance?, in Third Mathematics and Computers in Sport Conference Proceedings (ed. Neville de Mestre), Bond University, Australia, 69-76 (1996).

[5] Ganesalingam S., Kumar K. and Ganeshanandam S., A statistical look at cricket data, in Second Mathematics and Computers in Sport Conference Proceedings (ed. Neville de Mestre) Bond University, Australia, 89-104 (1994).

[6] Kimber A., A graphical display for comparing bowlers in cricket, *Teaching Statistics*, 15, 84-86 (1993).

[7] Clarke S.R., Consistency in sport - with particular reference to cricket, NZOR Proceedings, 27th Annual Conference, Victoria University of Wellington, 30-35 (1991).

[8] Danaher P.J., Estimating a cricketer's batting average using the product limit estimator, *New Zealand Statistician*, 24(1), 2-5 (1989).

[9] Scheid F., An evaluation of the handicap system of the United States Golf Association, in Optimal Strategies in Sports, S.P. Ladany and R.E. Machol, editors, North Holland: Amsterdam, 151-155 (1977).

[10] Pollock, S.M., A model of the USGA handicap system and ``fairness'' of medal and match play, in *Optimal Strategies in Sports*, (ed. S.P. Ladany and R.E. Machol), North Holland: Amsterdam, 141-150 (1977).

[11] Mosteller F. and Youtz C., Professional golf scores are Poisson on the final tournament days, in Proceedings of the Section on Statistics in Sports, American Statistical Association, 39-51 (1992).

SPORT AND PHYSICAL FITNESS : AN APPLICATION OF FACTOR ANALYSIS FOR DEVELOPING A TEST BATTERY

J.P. Verma[1] and Kuldeep Kumar[2]

Abstract

This study was undertaken to develop the test battery for measuring physical fitness of Indian boys among the age group of 9 to 12 years. The sample consisted of 100 boys from different Schools in India. Twentyone test items were selected for this study covering speed, strength, agility, balance, flexibility and endurance. The data obtained from 21 tests were subjected to two types of analysis. Under descriptive analysis, various measures were computed in order to have an idea about the characteristics of all the 21 test items. Secondly, Factor analysis was applied by using the principal component analysis and the final solution so obtained was used to identify the different factors of general fitness. These factors were given an appropriate name depending upon the characteristics of variables it contained. Finally, the test battery for measuring general fitness was developed by picking up one or two variables from each factor, having the highest loading. The battery thus constituted the following test items: 50 Mts. dash for speed, Standing Broad Jump for Power, 1 Min. Situps for Strength, Stork Balance on Bass Stick and 8 Min. Run/Walk test for endurance.

1.   INTRODUCTION

Evaluation of physical fitness depends upon various dimensions relating with the ability of an individual to perform different kinds of day to day functions effectively. Many methods have been suggested to evaluate the physical fitness among the athletes. Physical fitness depends upon many dimensions like work capacity, the total functioning capacity to perform certain specified task, muscular effort of the individual involved, tasks to be performed, quality and intensity of effort. There are conflicting views among the scientists regarding the various parameters which should be used for measuring physical fitness.

Many efforts have been made by the various investigators to develop the test battery for measuring the physical fitness of men and women in different age categories. Studies on construction of fitness batteries were conducted by Start [1], Arnelt [2] and Howell [3], whereas Hall [4], Seashore [5] and Cumbee [6] conducted the studies on construction of test battery related to motor components. Mathews [7] emphasised the capacity of muscles to measure the physical fitness. Further Harris [8], Green [9] and Beckenholdt [10] have suggested various methods of selecting variables for developing test batteries on different aspects, whereas Barry [11] conducted factorial analysis of physique.

---

[1]   Lakshmibai National Institute of Physical Education, (Deemed University), Shakti Nagar, Gwalior-474002 (M.P) India.

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

This study was undertaken by the authors to develop test battery for measuring physical fitness of the Indian boys among the age group of 9 to 12.

2. METHODS AND MATERIAL

A sample of 100 boys from the central schools of India were selected from 9 to 12 years age group randomly. In all 21 fitness test items covering speed, strength, agility, balance, flexibility and endurance were selected for the study. The test items were 40 Mts. Dash (40M), 50 Mts. Dash (50M), 60 Mts. Dash (60M), Squat Thrust (SQ.T), Shuttle Run in standing position (SR1), Shuttle Run by turning (SR2), Sit and Reach test for Hip & Trunk Flxibility (S & R.T), Shoulder Flexibility (SH.F), Bridge up test for Spine Flexibility (SP.F), Pull Ups, Push Ups, 30 Sec. Sit Ups (30 Sec. SU), 1 Min. Sit Ups(1 Min. SU), Maximum Sit Ups (Max SU), Stork Balance on Bass Stick (STORK), Bass Stick Test Lengthwise for Balance (LENGTH), Bass Stick Test Crosswise for Balance (CROSS), Standing Broad Jump (SBJ), Vertical Jump (VJ), 9 min run/walk test (9 Min.) and 8 min run/walk test (8 Min.).

In this study two types of analysis were carried out. Firstly, data obtained on all the 21 fitness parameters were subjected to descriptive analysis. Under descriptive analysis various measures were computed in order to have an idea about the characteristics of variables. Secondly, factor analysis was applied by using the principal component analysis. Final solution so obtained was used to identify the different factors of fitness. These factors were given an appropriate name depending upon the characteristics of variables contained in it. Finally, a test battery for measuring the fitness was prepared by picking up one or two variables having the highest loading from each factor.

3. RESULTS

Various descriptive measures like lowest and highest scores, kurtosis, skewness, mean, standard deviation, standard error and coefficient of variation were computed and these findings are shown in table 1.

Table 1

ITEMS	VARIABLES	LOWEST	HIGHEST	KURTOSIS	SKEWNESS	MEAN	STD.DEV.	STD.ERROR	C.V
40M	X1	5.70	9.50	4.25	1.17	7.03	0.75	0.08	10.62
50M	X2	7.00	11.20	5.09	0.89	8.38	0.69	0.07	8.32
60M	X3	7.90	13.20	4.80	1.09	9.82	0.91	0.09	9.24
Squat Thrust	X4	10.00	24.00	3.89	0.02	18.69	2.58	0.26	13.78
Shu.Run in St.Po.	X5	10.10	15.00	9.10	3.48	11.47	0.80	0.08	6.98
Shu.Run by turning	X6	10.40	15.00	6.94	2.39	11.84	0.79	0.08	6.65
Sit & Reach Test	X7	11.90	16.00	3.10	0.13	13.63	0.79	0.08	5.79
Shoulder Flexi.	X8	26.00	84.00	3.01	0.66	66.29	13.82	1.40	20.85
Spine Flexi.	X9	29.00	72.00	2.83	0.00	48.26	7.81	0.79	16.19
Pull Ups	X10	0.00	13.00	3.06	0.50	3.76	2.93	0.29	77.84
Push Ups	X11	1.00	50.00	5.66	1.53	14.47	08.35	0.85	87.67
30 Sec.Sit	X12	3.00	32.00	3.66	0.59	17.95	5.32	0.54	29.64

Ups									
1 Min.Sit Ups	X13	3.00	58.00	2.47	0.05	27.82	11.14	1.13	40.04
Max. Sit Ups	X14	3.00	76.00	2.79	0.16	30.67	14.87	1.51	48.48
BST Lengthwise	X16	0.50	10.90	5.04	2.06	3.12	2.01	0.20	64.56
BST Crosswise X17		0.40	14.10	9.07	5.06	3.21	2.37	0.24	73.92
Standing Broad Jump	X18	1.40	2.48	2.92	0.00	1.91	0.21	0.02	10.79
Vertical Jump	X19	21.00	54.00	2.45	0.00	37.36	6.86	0.69	18.37
9 Min.Run/ Walk	X20	1100.00	2250.00	2.67	0.01	1656.55	243.09	24.68	14.68
8 Min.Run/ Walk	X21	1180.00	2380.00	3.29	0.21	1664.38	228.14	23.16	13.71

Further data on 21 variables were subjected to correlation analysis. All the correlations among variables are shown in correlation matrix listed in table 2.

Table 2

Correlation Matrix of the Test Items

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21
X1	1.00	.66	.67	-.13	.26	.18	.16	.07	-.17	-.40	-.06	.17	.12	.13	-.11	-.09	-.17	-.33	-.23	-.09	-.39
X2		1.00	.76	-.17	.39	.21	.24	-.08	-.19	-.40	-.21	.10	-.01	-.02	.09	-.01	-.07	-.44	-.38	-.28	-.41
X3			1.00	-.16	.29	.13	.25	-.03	-.16	-.36	-.16	.10	.05	.04	-.06	-.02	-.13	-.42	-.38	-.31	-.48
X4				1.00	-.28	-.20	-.08	-.05	.09	.13	.19	.20	.18	.16	.19	.08	.10	.13	.06	.20	.18
X5					1.00	.55	.14	-.03	-.31	-.29	-.26	-.16	-.20	-.22	.03	.00	.10	-.40	-.31	-.14	-.04
X6						1.00	.04	.02	-.07	-.23	-.08	-.15	-.17	-.19	-.10	-.09	-.06	-.22	-.24	-.08	-.09
X7							1.00	-.03	.13	-.08	.14	-.09	-.10	-.01	.06	.00	.04	-.21	-.15	-.13	.04
X8								1.00	-.14	.12	-.09	.08	.02	.03	-.13	-.20	-.21	.04	.06	-.09	-.14
X9									1.00	-.01	.04	.06	.12	.15	.00	-.12	.05	.25	.09	.25	.26
X10										1.00	.25	.02	-.04	-.07	-.02	.05	.13	.36	.26	.11	.08
X11											1.00	.22	.29	.38	.02	-.19	.13	.22	.29	.21	.12
X12												1.00	.85	.75	-.09	-.09	-.13	.17	.16	.14	-.13
X13													1.00	.93	-.11	-.15	-.15	.17	.06	.19	-.02
X14														1.00	-.11	-.13	-.16	.15	.05	.18	.01
X15															1.00	.26	.38	-.02	-.05	.08	.25
X16																1.00	.22	-.21	-.15	-.17	.09
X17																	1.00	.13	.02	.03	.18
X18																		1.00	.64	.11	.06
X19																			1.00	.10	.19
X20																				1.00	.62
X21																					1.00

X1=40 M.Dash X2=50 M.Dash X3=60 M.Dash X4=Squat Thrust X5=Shuttle Run in Standing Position X6=Shuttle Run by turning X7=Sit & Reach test X8=Shoulder Flexibility X9=Spine Flexibility X10=Pull Ups X11=Push Ups X12=30 Sec.Sit Ups X13=1 Min.Sit ups X14=Max Sit Ups X15=Stork for Balance X16=BST Lengthwise for balance X17=BST Crosswise for Balance X18=Standing Broad Jump X19=Vertical Jump X20=9 Min.Run/Walk test X21=8 Min. Run/Walk test

Correlation matrix so obtained was used in principal component analysis. With the help of principal component analysis, all 21 variables were divided into various factors. With the help of Kaiser's [12] criteria suggested by Guttman only those factors having latent roots greater than one were considered as common factors.

Owing to this criteria 4 factors were retained which are shown in Table 3. Adcock [13] suggested that the variable for which communality is less than .30 should not become the member of the test battery. Such variable whose communality is too low, indicates that the variable is unreliable and hence those variables are not included in the factor analysis.

Table 3

Principal Component Analysis of the Test Items (Unrotated Factor Loadings)

	1 Vect.	2 Vect.	3 Vect.	4Vect.
ROOT	4.41	3.16	1.93	1.55
% Var.Exp	21.01	15.06	9.18	7.38
Cum.Var.Exp.	21.01	36.07	45.25	52.63
40M	0.62	0.47	0.06	0.01
50M	0.77	0.30	0.19	-0.14
60M	0.74	0.38	0.08	-0.17
SQ.T	-0.37	0.11	0.34	-0.22
SR1(Standing)	0.62	-0.15	0.14	0.31
SR2(Turning)	0.44	-0.11	-0.03	0.49
Sit & Reach Test	0.25	0.01	0.27	0.06
Shoulder Flexi.	0.01	0.10	-0.52	0.06
Spine Flexi.	-0.35	0.05	0.21	0.29
Pull Ups	-0.51	-0.18	-0.24	-0.26
Push Ups	-0.43	0.28	0.14	0.01
30 Sec. SU	-0.24	0.83	0.10	-0.11
1 Min. SU	-0.31	0.87	0.16	0.01
Max. SU	-0.31	0.85	0.18	0.03
Stork Balance	-0.06	-0.26	0.58	-0.31
BST Lengthwise	0.08	-0.29	0.36	-0.50
BST Crosswise	-0.14	-0.34	0.46	-0.29
SBJ	-0.67	0.04	-0.30	-0.12
V.J	-0.59	-0.01	-0.27	-0.07
9 Min.R/W	-0.46	0.06	0.38	0.54
8 Min.R/W	-0.48	-0.31	0.49	0.44

Since unrotated factors do not generally represent useful scientific constructs and therefore rotation was necessary if useful and meaningful constructs were to be identified. Due to this fact, this unrotated matrix was subjected to varimax rotation because of its great popularity and usefulness. The rotated factor matrix is given in table 4.

Table 4

Factor Analysis (Varimax Rotation)

	1 Fact.	2 Fact.	3 Fact.	4 Fact.	Communalities
ROOT	3.92	3.61	3.18	1.95	2.31
% Var.Exp.	17.18	15.15	9.29	11.00	
# Var.Exp.	32.65	28.78	17.66	20.90	
40M	0.63	0.25	-0.17	-0.34	0.61
50M	0.74	0.11	0.05	-0.43	0.74
60M	0.67	0.17		-0.49	0.73
SQ.T	-0.20	0.32	0.39	0.14	0.31
SR1(Standing)	0.63	-0.34	-0.09	0.02	0.52
SR2(Turning)	0.47	-0.31	-0.32	0.16	0.45
Sit & Reach Test	0.35	-0.01	0.13	0.04	0.14
Shoulder Flexi.	-0.18	-0.03	-0.45	-0.20	0.28
Spine Flexi.	-0.08	0.16	0.01	0.47	0.26
Pull Ups	-0.64	-0.03	0.08	-0.02	0.42
Push Ups	-0.23	0.42	0.06	0.23	0.28
30 Sec. SU	-0.02	0.87	-0.09	-0.05	0.77
1 Min. SU	0.01	0.92	-0.12	0.09	0.87
Max. SU	0.01	0.90	-0.11	0.12	0.85
Stork Balance	0.05	-0.04	0.70	0.09	0.50
BST Lengthwise	0.02	-0.14	0.64	-0.21	0.48
BST Crosswise	-0.07	-0.13	0.63	0.10	0.43
SBJ	-0.71	0.19	-0.09	0.09	0.56
V.J	-0.63	0.12	-0.09	0.11	0.43
9 Min.R/W	-0.04	0.19	0.00	0.78	0.65
8 Min.R/W	-0.11	-0.09	0.26	0.82	0.77

The latent roots of all the rotated factors have been given in table 4. Only those variables having loading greater than +.40 were retained in the factors. A loading greater than or equal to + .40 usually gives the non overlapping factors.

Each of the four factors obtained in table 4 were interpreted and given names. The four factors obtained in the study accounted for 52.83% of the total common factor variance.

Factor 1

Item No.	Name of the Variable	Factor Loading
1	40 M for Speed	0.63
2	50 M for Speed	0.74
3	60 M for Speed	0.67
6	SR1(Standing) for Agility	0.63
7	SR2(Turning) for Agility	0.47
10	Pull Ups for Strength	-0.64
18	SBJ for Power	-0.71
19	V.J.	-0.63

The above factor is characterised by the speed and power variables. This factor could be named as Speedo - Power. Among the items heavily loaded on the factor were 50 mts. dash, Standing Broad Jump for power, 60 Mts. Dash and Pull Ups. In terms of relative contribution this factor accounted for 32.65 % of the total common factor variance accounted by the 4 factors.

Factor 2

Item No.	Name of the variable	Factor Loading
11	Push Ups for Strength	0.42
12	30 Sec.SU for Strength	0.87
13	1 Min. SU for Strength	0.92
14	Max.SU for Strength	0.90

Factor 2 is characterised by high loadings of those items which commonly measure Strength. Thus the best suited name of this factor might be termed as Strength component. In terms of relative contribution this factor accounted for 28.78 % of the total common factor variance accounted by the 4 factors.

Factor 3

Item No.	Name of the variable	Factor Loading
8	Shoulder Flexibility	-0.45
15	Stork Balance	0.70
16	BST Lengthwise	0.64
17	BST Crosswise for Balance	0.63

In this factor items having highest loadings denotes the balancing variables and hence this could be recognised as Balancing factor. This factor contributes only 17.66 % of the total common factor variance accounted by the 4 factors.

Factor 4

Item No.	Name of the variable	Factor Loading
2	50 M Dash for Speed	-0.43
3	60 M Dash for Speed	-0.50
9	Spine Flexibility	0.47
20	9 Min.Run/Walk Test	0.78
14	8 Min. Run/Walk Test	0.82

Factor 4 consists of five variables only. The loadings were high in those variables which measure endurance thus this component was termed as Endurance factor. This factor accounted for 20.90 % of the total common factor variance accounted by the 4 factors.

4.   DEVELOPMENT OF TEST BATTERY

According to Fleishman (1963), inefficient test batteries are those with too many tests on one factor and none from one or more of the other factors identified. Furthermore,

the addition of more than one test per factor adds relatively little new information about a subject's abilities, relative to the addition of tests from separate factor.

Table 5

Test Battery for Indian Boys for 9 to 12 years of age
Category for Measuring General Fitness

No.	Item No.	Name of the Variable
1	2	50M Dash for Speed
2	18	Standing Broad Jump for Power
3	13	1Min. Situps for Strength
4	15	Stork Balance on Bass Stick
5	21	8 Min. Run/Walk Test for Endurance

Owing to the above mentioned concept, five variables, two from first factor and one each from the remaining three factors having the highest loading were selected to constitute a test battery for measuring the general fitness for the Indian boys in 9 to 12 years of age category.

5.   DISCUSSION

50 metre dash was selected for measuring the speed instead of 40 metre dash and 60 metre dash in the study. Loadings for 40 metre dash and 60 metre dash are lesser than that of 50 metre dash, this indicates that the boys in the age category 9 to 12 years give optimum performance in 50 metre dash. This seems to be quite logical because after attaining the maximum speed by crossing the acceleration phase a boy of nine to twelve years of age may not maintain it for quite a long time. This might be the reason why performance deteriorates in 60 metres. Further it is recommended to test as to whether the boys in the age category less than 9 years give their optimum performance in 30 metres or not.

Standing Broad Jump (SBJ) was selected as the second test item in the test battery. This measures the power component of an individual's fitness. This item was also selected from the first factor. The two items were selected from the first factor because of the fact that they measure the different dimensions of the fitness. If you look at to the correlation between the 50 metre dash and SBJ in the table 2, it is equal to -0.44, which is not a very high association, this indicates that one does not carry the characteristics of others. Further principal component analysis was used in factor analysis thus the first factor so obtained, contributes maximum in measuring the concept of fitness in the target group. Therefore it is justifiable to pick up two test items from the first factor.

Third parameter in the test battery was 1 Min. Situps for measuring strength. This undoubtedly measures the strength of the subject and can serve an index for measuring the strength related fitness.

Stork balance on Bass Stick was selected as fourth item from the third factor, to be included in the test battery. This factor was selected on the basis of its highest loading factor.

Last factor in the test battery was 8 Min. Run/walk test for measuring endurance related fitness. Here again, looking to the values of the loadings it is clear that the performance of the boys was better in 8 Min.Run/Walk test instead of 9Min.Run/Walk test. Here also it may be concluded that probably in the higher age group boys might perform better in 9 Min.Run/Walk test instead of 8 Min.Run/Walk test.

Many research workers around the world advocate the same test items to be used for measuring the fitness level for the 7 to 16 years boys. Thus in the light of the above mentioned facts fitness experts should try to reinvestigate their philosophy of using the same test items for different age groups. On the basis of the above discussion, it is hereby concluded that variables for measuring the speed and endurance might change with age.

Thus through this study, a strategy was developed in order to measure the general fitness of the boys in 9 to 12 years of age on the basis of five representative fitness parameters.

REFERENCES

[1] Start K.B. et.al., A factorial investigation of power, speed, Isometric strength and anthropometric measures in the lower limb. *Research Quarterly*, Vol 37, 553-559 (1966).

[2] Arnelt C., The Purdue Motor Fitness Test Batteries for Senior High School Girls. *Research Quarterly*, Vol 33, 323-328 (1962).

[3] Howell M.L. and Alderman R.B., The Measurement of Physical Fitness by Human Performance. Paper presented at the American College of Sports Medicine, Las Vegas, (March 1967).

[4] Hall D. and Wattenborn J.R., Motor Fitness of Farm Boys. *Research Quarterly*, Vol 13, 432-436 (1942).

[5] Seashore R.H., Buxton C.E. and Mccollom I.N., Multiple Factorial Analysis of Fine Motor Skills. *American Journal of Psychology*, 1940,Vol 53, 251-259 (1940).

[6] Cumbee F.Z., A Factorial Analysis of Motor Co-ordination. *Research Quarterly*, Vol 25, 412-428 (1954).

[7] Mathews D.K., Measurement in Physical Education. Philadelphia, W.B. Saunders Company (1973).

[8] Harris J.E., The Differential Measurement of Force and Velocity for Junior High School Girls. *Research Quarterly*, Vol 8, 114-121 (1937).

[9]   Green K.N., The Development of Battery of Golf Skill Test. *Dissertation Abstracts International*, 32, 5017-A (1972).

[10]   Beckenholdt S.E. and Mayhew J.L., Specificity among anaerobic power tests in male athletes. *Journal of Sports Medicine and Physical Fitness*. Vol 23, No.3, 326-332 (1983).

[11]   Barry A.J. and T.K. Cureton., Factorial Analysis of Physique and Performance of prepubescent boys. *Research Quarterly*, Vol 27, 283-295 (1961).

[12]   Kaiser H.F., The Varimax Criterion for Analytic Rotation in Factor Analysis. *Psychometrika*, XXIII, 187-200 (1958).

[13]   Adcock C.J., Factor Analysis for Non Mathematicians, Melbourne, University Press, Melbourne (1955).

IN VIVO PREDICTIONS OF ACL STRESS AND STRAINS DURING SIDESTEPPING

Scott McLean[1], Robert J. Neal[1], William Daniel[2] and Peter Myers[3]

Abstract

This paper describes a process by which the stresses and strains experienced by the anterior cruciate ligament (ACL) can be predicted for running tasks. The method relies on accurately measuring, using high speed video, the three rotational degrees of freedom at the knee as subjects perform sidestepping (and straight-line running) activities. These data, in conjunction with information on the ligament attachment sites, obtained using magnetic resonance imaging techniques, are then used in a mathematical model of the ACL. This model, which incorporates stress-relaxation and memory effects, as well as linear and non-linear visco-elasticity, is then used to predict the stresses and strains in the ligament by knowing the displacements of the ligament attachment sites. The model data agree well with previously published data on stresses and strains obtained from in-vitro specimens.

1.   INTRODUCTION

Knowledge of the ACL mechanical response to specific joint loads is crucial to the identification of potential injury mechanisms and has previously provided the impetus for improved surgical repair and rehabilitation techniques. The literature pertaining to the evaluation of ACL stress and strain under prescribed movement conditions is predominantly experimental. Previously, ACL strain has been assessed in-vitro using techniques such as simple scaling (Wang [1]), photogrammetry (Blankevoort [2]) and spatial kinematic linkage (Takai [3]). More recently in-vivo ACL strain has been evaluated using a variety of implantable strain gauge devices (Beynnon [4] [5], Pope [6]). While these studies have provided valuable information regarding ligament function and injury, their often invasive approach has questioned their accuracy. As a result, a small, but increasing number of studies have adopted mathematical modeling techniques to evaluate ACL strain and resultant stress. Typically these studies involve measuring joint motions and the location of the ligament attachments on each bone. The joint is then displaced mathematically according to the measured motion, and the distance between the attachment points is calculated for selected joint positions (Crowninshield [7], Edwards [8], Grood and Hefzy [9]). An advantage of analytic knee modeling is that soft tissue structures that are difficult to transduce through experimental studies may be readily investigated, especially when analyses extend to include complex joint movements. The purpose of this study therefore, was to develop an analytic knee model that could accurately predict the in-vivo mechanical response of the ACL to knee movements previously linked to ligament injury.

---

[1]   Department of Human Movement Studies, The University of Queensland.
[2]   Department of Mechanical Engineering, The University of Queensland.
[3]   Brisbane Orthopaedic and Sports Medicine Clinic, Holy Spirit Hospital.

2. Method

Knee joint kinematics

Three-dimensional (3D) knee-joint kinematics were firstly quantified for the stance phase of a "typical" sidestep cutting manoeuvre. The 3D global coordinates of precisely-attached external skin-mounted markers were recorded via a high-speed video (200 Hz) during both stationary and dynamic (moving) trials. These data were then submitted to a custom software package (JTMOTION) which defined the local, anatomically significant orthogonal Cartesian coordinate systems for the femoral $(x^F y^F z^F)$ and tibial $(x^T y^T z^T)$ segments. The origins of each segment LCS ($o_X^F o_Y^F o_Z^F$ and $o_X^T o_Y^T o_Z^T$) were located globally by the position vectors $\bar{r}_o^F$ and $\bar{r}_o^T$. The relative movement between these coordinate systems within the global reference frame subsequently enabled the three rotations at the knee joint to be quantified in clinical terms (flexion-extension, abduction-adduction, external-internal rotation). A more detailed description of this technique can be found elsewhere (McLean [10]).

Identification of attachment locations

A relationship between the mathematical and anatomical locations of the tibial and femoral bundle attachments of the ACL was established using combined high-speed video, magnetic resonance (MR) imaging and structural matrix analysis techniques. Specifically, one of the external reference markers (tibial tuberosity) used in the above kinematic analyses was denoted as the reference marker for the identification of ligament attachment sites. This marker was then left in position in the ensuing MR analyses which enabled a relationship between the external marker coordinates and internal joint geometry to be established. The mathematical theory behind this process is outlined below.

Reference Marker Location

The global (XYZ) 3D position of the reference marker was determined for the subject standing in the anatomical position. Structural matrix analysis techniques were then utilised to define the reference marker in terms of local (femoral and tibial) segment coordinates.

The transformation matrices used to map the global-marker coordinates of the reference marker ($V_X^G V_Y^G V_Z^G$) onto each of the local coordinate systems were given by:

$$^{Femur}T = \begin{bmatrix} i_x^F & j_x^F & k_z^F & o_X^F \\ i_y^F & j_y^F & k_y^F & o_Y^F \\ i_z^F & j_z^F & k_z^F & o_Z^F \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad {}^{Tibia}T = \begin{bmatrix} i_x^T & j_x^T & k_z^T & o_X^T \\ i_y^T & j_y^T & k_y^T & o_Y^T \\ i_z^T & j_z^T & k_z^T & o_Z^T \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (1)$$

Thus, the 3D location of the reference marker with respect to the femoral LCS was given by:

$$
\begin{bmatrix} V_x^F \\ V_y^F \\ V_z^F \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{Femur}T \end{bmatrix}^{-1} \begin{bmatrix} V_X^G \\ V_Y^G \\ V_Z^G \\ 1 \end{bmatrix}. \tag{2}
$$

Similarly, the reference marker is defined with respect to the tibial LCS by:

$$
\begin{bmatrix} V_x^T \\ V_y^T \\ V_z^T \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{Tibia}T \end{bmatrix}^{-1} \begin{bmatrix} V_X^G \\ V_Y^G \\ V_Z^G \\ 1 \end{bmatrix}. \tag{3}
$$

Magnetic Resonance Analyses

Magnetic resonance data were used in conjunction with the above information to identify the geometrical (3D) locations of the femoral and tibial ACL attachments. To accurately determine these locations, the stationary video data and MRI images were required to be recorded for the same knee joint position. This result was ensured by firstly calculating the relative knee rotations displayed by the subject for each of the three degrees of freedom during the stationary video shot using the JTMOTION software (McLean [10]). The static knee flexion angle was then accurately replicated during MR procedures with the use of a 3 degree of freedom goniometer and maintained throughout the protocol.

A water-filled marker was attached directly over the tibial tuberosity (similar to video analyses) for all MR images. This ensured easy identification of the marker for ensuing distance calculations. A series of 2D (sagittal, transverse and coronal) images (TE 22 msec, TR 2500 msec, RARE FACTOR 16) were then taken to determine the 3D location of each bundle attachment with respect to the marker centre (Figure 1).
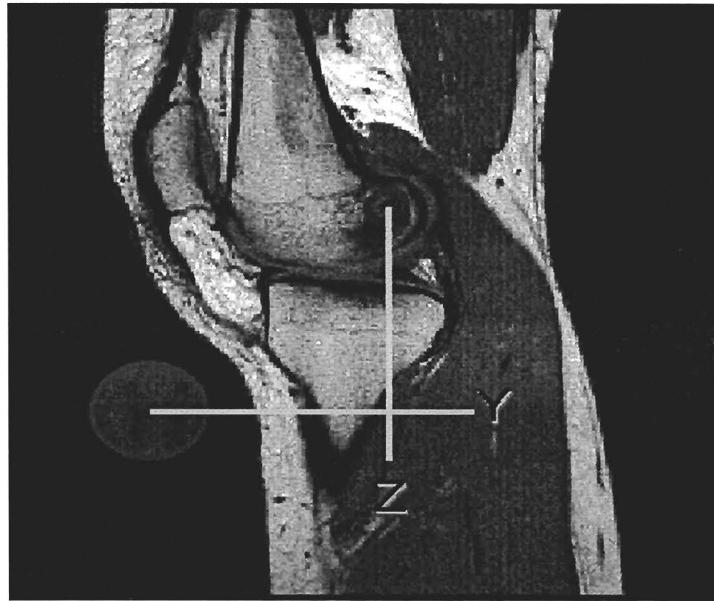
Figure 1: A sagittal image used to determine the 3D locations of the ACL attachment sites. An external water filled marker enabled internal geometry to be determined from external coordinates.

Once the distances from the external reference marker to the femoral (d_x^F, d_y^F, d_y^F) and tibial (d_x^T, d_y^T, d_y^T) ligament attachments were calculated, the anatomical locations of each site were defined in local segment coordinates. Specifically, the local segment coordinates of the femoral $\left(A_{xyz}^F\right)$ and tibial $\left(B_{xyz}^T\right)$ ACL attachment sites were described by:

$$A_{xyz}^F = (V_x^F + d_x^F)\hat{i} + (V_y^F + d_y^F)\hat{j} + (V_z^F + d_z^F)\hat{k} \tag{4}$$

$$B_{xyz}^T = (V_x^T + d_x^T)\hat{i} + (V_y^T + d_y^T)\hat{j} + (V_z^T + d_z^T)\hat{k} \tag{5}$$

The convention of d_x, d_y and d_z in each instance was dependent on the relationship between virtual marker position and the positive orientation of each segments right-hand coordinate system.

3.   ACL LENGTH CALCULATION

In order to determine the length of the ACL at a particular time, each attachment was firstly defined with respect to one local segment coordinate system. For the purpose of the current investigation, the assumption was made that the femur was fixed and displacements were applied to the tibia. Therefore, the instantaneous 3D coordinates of the tibial attachment with respect to the femoral coordinate can be defined using:

$$\left[B_{xyz}^F\right] = \left[\begin{matrix}Femur\\T\end{matrix}\right]^{-1}\left[\begin{matrix}Tibia\\T\end{matrix}\right]\left[B_{xyz}^T\right] \tag{6}$$

A general expression was generated that enabled the location of the tibial attachment to be determined with respect the femoral LCS over a given number of time steps. That is:

$$\left[B^F_{xyz_i} \right] = \left[{}^{Femur}T_i \right]^{-1} \left[{}^{Tibia}T_i \right] \left[B^T_{xyz} \right], \tag{7}$$

where for each iteration ($i = 1$ to n), the unique matrix transformations are calculated.

Thus, the instantaneous length of the ligament (L^F_i), quantified over n time steps, is related to the location of the insertion sites, expressed relative to the femoral coordinate system by the following equation:

$$\left(L^F_i \right)^2 = \left(L^F_{x_i} \right)^2 + \left(L^F_{y_i} \right)^2 + \left(L^F_{z_i} \right)^2, \tag{8}$$

where

$$\begin{aligned} L^F_{x_i} &= A^F_x - B^F_{x_i} \\ L^F_{y_i} &= A^F_y - B^F_{y_i} \\ L^F_{z_i} &= A^F_z - B^F_{z_i} \end{aligned} \tag{9}$$

These equations assume that the ligament lies in a straight line between insertion sites and that the tibial attachment moves in relation to a fixed femoral attachment with respect to time.

Relative elongation of the ACL

Specifically, the relative elongation (λ) of the ligament was determined at each time step using the equation:

$$\lambda_i = \frac{L^F_i}{l_{ref}}, \text{ for } i = 1 \text{ to } n. \tag{10}$$

where L^F_i denoted the instantaneous length of the ligament and l_{ref} corresponded to a pre-determined reference length. For the above expression, the reference length was determined from the previously recorded MR images. When direct measurement is impossible, the reference length of the ACL is typically estimated as the length corresponding to full knee extension (Pioletti [11], Renstrom [12]). It was felt that the length of the ligament in the static MR images was a reasonable estimate of the reference length since knee flexion was less than 5°.

Non-linear visoelastic model of ACL

A model similar to that proposed by Pioletti [11] was implemented to predict the mechanical response of the ACL to the knee joint movements obtained in the kinematic analyses. Specifically, the relative elongation data calculated above were input into the model to determine the resultant instantaneous ligament stress. In

developing this model, incompressibility of the ligament substance was assumed. The motivation behind this assumption was that ligaments and tendons are mainly composed of water, which is known to be nearly incompressible. Pioletti [11] describes the viscoelastic behaviour of the ACL in response to tensile load using the following detailed expression

$$
\sigma = \sigma(\lambda) = \alpha\beta \left\{ \frac{\left(-2\exp\left[\beta\left(\lambda^2 + \frac{2}{\lambda} - 3\right)\right] + \lambda^2 + \frac{1}{\lambda}\right)\frac{1}{\lambda^2}}{+ 2\exp\left[\beta\left(\lambda^2 + \frac{2}{\lambda} - 3\right)\right] - \frac{2}{\lambda}} \right\} \lambda + \eta'\lambda\left(\lambda^2 + \frac{2}{\lambda} - 3\right)\left(2\lambda^2 + \frac{1}{\lambda^4}\right) \quad (11)
$$

For the above equation α and β are constants which were obtained experimentally via traction tests. The short term memory effects of strain rate are modeled implicitly, being viewed as an appropriate means by which to mimic the ligament's viscoelastic behaviours.

Long Term Memory Effects

For the purpose of the current model, a normalised exponential Prony series was used as an appropriate means for the time relaxation identification M(s). It was demonstrated that three exponentials were sufficient to correctly describe the time relaxation behaviour. M(s) thus has the form:

$$
M(s) = \frac{\sum_{k=1}^{3} a_k \exp\left(-\frac{s}{\tau_k}\right)}{\sum_{k=1}^{3} a_k}. \quad (12)
$$

In summary then, the general equation totally describing the viscoelastic behaviour of the ACL was of the form:

$$
S_e = -pC^{-1} + \alpha\beta\left(2\exp\left[\beta(I_1 - 3)\right] - I_1\right)I + \alpha\beta C
$$

$$
S_v = \eta'(I_1 - 3)\dot{C} \quad (13)
$$

$$
\int_\delta^\infty \sum \left(G(t - s), s; C(t)\right)ds = \int_\delta^t S_e\left(C(t - s)\right)\dot{M}(s)ds
$$

where each equation cooresponds to the elastic, short term and long term memory effects respectively.

4.   MODEL LINEARISATION

The model proposed by Pioletti [11] is currently believed to best represent the stress/strain relationship of the ACL, particularly for varying strain rates. Of concern however is the fact that this model does not accommodate for the linear component of the stress-strain relationship that has been shown to exist for the ACL. From the abundance of experimental data available (Butler [13], Noyes [14]) the ACL stress

strain curve can be viewed to consist of two discrete regions. Firstly, a non-linear low stiffness region is evident, where large increases in length are accompanied by relatively small changes in ligament stress. The characteristic crimp of the collagen fibrils that make up the ligament appear to explain this phenomenon. As increased loading causes the fibrils to straighten, a linear high stiffness region exists, where increases in ligament stress are directly proportional to the strain. The exponential function used to model the ACL stress-strain relationship by Pioletti [11] obviously fails to incorporate this second phase.

To remedy the above limitation, we have linearised the model at a pre-determined "reference strain" value, corresponding to the level of strain at which the collagen fibrils become straightened and hence undergo linear and rapid increases in stress as a function of strain. The choice for the reference strain was made based on examination of the abundance of experimental data in the literature. This value was observed typically, to occur at $7.5 \pm 2.1\%$ strain.

The linearisation of the curve was achieved through simple mathematical calculation. Equation 11 can be written in the following form:

$$\sigma = \alpha\beta\left(-\frac{2}{\lambda_1^2}\exp\left[\beta\left(\lambda_1^2 + \frac{2}{\lambda_1} - 3\right)\right] + 1 + \frac{1}{\lambda_1^3} + 2\exp\left[\beta\left(\lambda_1^2 + \frac{2}{\lambda_1} - 3\right)\right] - \frac{2}{\lambda_1}\right) +$$

$$\eta'\dot{\lambda}_1\left(2\lambda_1^4 + \frac{1}{\lambda_1^2} + 4\lambda_1 + \frac{2}{\lambda_1^5} - 6\lambda_1^2 - \frac{3}{\lambda_1^4}\right) \tag{14}$$

where we have used the term λ_1 to define strain rather than λ.

To determine the slope of the tangent at a specific point on this curve (in this case, the point denoting the reference strain), we differentiate with respect to λ_1. That is:

$$sl = \frac{\partial\sigma}{\partial\lambda_1} = \alpha\beta\left(\begin{array}{c}-\frac{4}{\lambda_1^3}\exp\left[\beta\left(\lambda_1^2 + \frac{2}{\lambda_1} - 3\right)\right] - \frac{2\beta}{\lambda_1^2}\left(2\lambda_1 - \frac{2}{\lambda_1^2}\right)\exp\left[\beta\left(\lambda_1^2 + \frac{2}{\lambda_1} - 3\right)\right] - \frac{3}{\lambda_1^4} + \\ 2\exp\left[\beta\left(\lambda_1^2 + \frac{2}{\lambda_1} - 3\right)\right] + 2\lambda_1\beta\left(2\lambda_1 - \frac{2}{\lambda_1^2}\right)\exp\left[\beta\left(\lambda_1^2 + \frac{2}{\lambda_1} - 3\right)\right]\end{array}\right) +$$

$$\eta'\dot{\lambda}_1\left(8\lambda_1^3 - \frac{2}{\lambda_1^3} + 4 - \frac{10}{\lambda_1^6} - 12\lambda_1 + \frac{12}{\lambda_1^5}\right) \tag{15}$$

Now, at $\lambda_1 = 1$ (ie., reference length), $\frac{\partial\sigma}{\partial\lambda_1} = 3\alpha\beta = $ initial slope.

If we then adopt a new origin and use $\lambda_1^* = \frac{\lambda_1}{\lambda_{ref}}$, $\sigma = \sigma_{ref}$ corresponds to λ_{ref} using the

above formula.

That is:

$$\sigma = \sigma_{ref} + \alpha^* \beta \left(\left(-2\left[1 + \beta\left(\lambda_1^{*2} + \frac{2}{\lambda_1^*} - 3\right)\right] + \lambda_1^{*2} + \frac{1}{\lambda_1^*}\right)\frac{1}{\lambda_1^{*3}} + 2\left[1 + \beta\left(\lambda_1^{*2} + \frac{2}{\lambda_1^*} - 3\right)\right] - \frac{2}{\lambda_1^*}\right)\lambda_1^*$$

$$+ \eta'\dot{\lambda}_1\left(2\lambda_1^{*4} + \frac{1}{\lambda_1^{*2}} + 4\lambda_1^* + \frac{2}{\lambda_1^{*5}} - 6\lambda_1^{*2} - \frac{3}{\lambda_1^{*4}}\right)$$

(16)

The non-linear and linear viscoelastic behaviours of the ACL were therefore completely described by equations 14 and 16 respectively. The implementation of each equation was governed by whether the instantaneous strain fell above or below the reference value. Again, to completely define the viscoelastic behaviour of the model, the long term memory effects were necessarily incorporated into each of the above equations as described in equation 13.

5.   RESULTS AND DISCUSSION

The model was subjected to a number of tests to determine the reliability of the predicted viscoelastic response. Cyclic loading and stress relaxation tests revealed that the modelled output was consistent with measured experimental data (Figure 2).
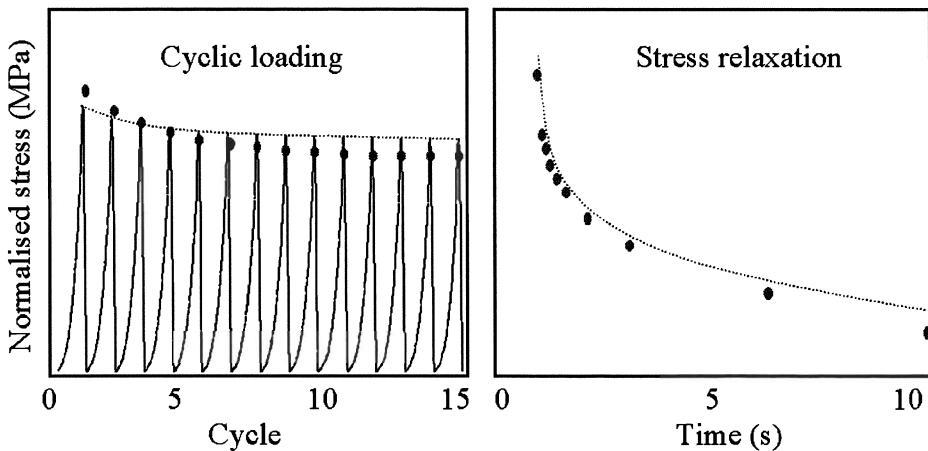
Figure 2: Comparison of model outputs (dashed lines) and experimental data (dots) for standard stress relaxation and cyclic loading of the human ACL.

Maximum strain values for the ACL during the sidestep cutting trials were consistent with those that had been predicted to occur during normal physiological loading and well below hypothesised maximum strain values (Noyes [14]). Furthermore, the stress response of the model was consistent with experimental data and supported the suggestion that the stiffness and ultimate stress of the ligament is dependent on strain rate (Butler [13], Danto and Woo [15]).

From the model outputs, we were able to predict ligament stress as a function of knee angle during the sidestep cut. For a typical sidestep, maximum stress values were within physiologically safe ranges. When joint motions associated with an abnormal sidestep were used to drive the model, the resultant stress-strain response appeared to place the ligament at increased risk of injury and potential rupture (Figure 3).
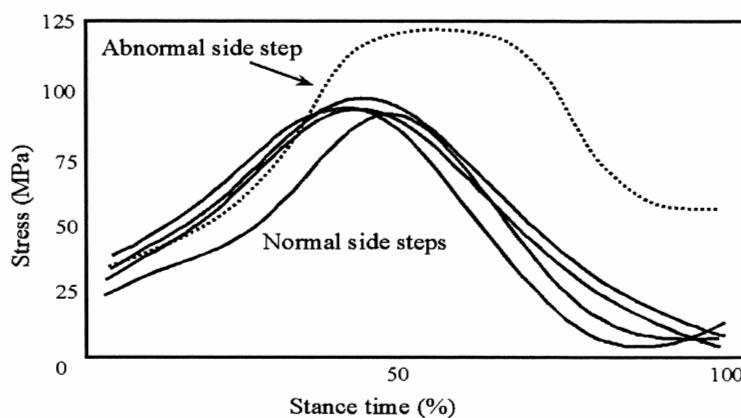
Figure 3: Comparison of predicted stresses in the ACL for normal (solid lines) and abnormal (dashed line) side steps.

6. CONCLUSION

Initial results suggest that the combined model can successfully predict the *in-vivo* stress-strain relationship for the ACL from accurate kinematic and geometric data. Specific model outputs were consistent with previously measured experimental data, with maximum stress/strain values being similar to those proposed to exist under normal physiological loading conditions. An accurate description of the ACL mechanical response to complex joint movements not only assists evaluation of injury potential, but also enables hazardous movement combinations, such as those linked to abnormal cutting techniques to be identified. Despite the success of the current model, further validation and development is proposed, including the effects of ligament wrapping, varying cross-sectional area and increased fibre-bundle numbers. These enhancements will provide valuable information pertaining to injury prevention, rehabilitation and surgical repair.

REFERENCES

[1] Wang C.J. and Walker P.S., et al. The effects of flexion and rotation on the length patterns of the ligaments of the knee. *J Biomech*, 6, 587-596 (1973).

[2] Blankevoort, 1991.

[3] Takai S. and Woo S.L., et al. Determination of the in situ loads on the human anterior cruciate ligament. *J Orthop Res*, 11(5), 686-95 (1993).

[4] Beynnon B.D. and Johnson R.J., et al. The measurement of elongation of anterior cruciate-ligament grafts in vivo. *J Bone Joint Surg Am*, 76(4), 520-31 (1994).

[5] Beynnon B.D. and Fleming B.C., et al. Anterior cruciate ligament strain behavior during rehabilitation exercises in vivo. *Am J Sports Med*, 23(1), 24-34 (1995).

[6] Pope M.H. and Stankewich C.J., et al. Effect of knee musculature on anterior cruciate ligament strain in-vivo. *J. Electromyog Kinesiol*, 1, 191-198 (1991).

[7]   Crowninshield R., Pope M.H. and Johnson, R.J., An analytical model of the knee. *J. Biomech*, 9, 397-405 (1976).

[8]   Edwards R.G., Lafferty J.F. and Lange K.O., Ligament strain in the human knee joint. *J. Basic Eng*, 92, 131-136 (1970).

[9]   Grood E. and Hefzy M.S., An analytical technique for modeling knee joint stiffness - Part 1: Ligamentous forces. *J Biomech Eng*, 104, 330-337 (1982).

[10] McLean S.G. and Myers P.T., et al. A quantitative analysis of knee joint kinematics during the sidestep cutting manoeuvre: Implications for non-contact anterior cruciate ligament injury. *Bulletin for the Hospital of Bone and Joint Disease*, 57(1), 30-38 (1998).

[11] Pioletti D.P. and Heegaard J.H., et al. Experimental and mathematical methods for representing relative surface elongation of the ACL. *J Biomech*, 28(9), 1123-6 (1995).

[12] Renstrom P. and Arms S.W., et al. Strain within the anterior cruciate ligament during hamstring and quadriceps activity. *Am J Sports Med*, 14(1), 83-7 (1986).

[13] Butler D.L. and Kay M.D., et al. Comparison of material properties in fascicle-bone units from human patellar tendon and knee ligaments. *J Biomech*, 19(6), 425-32 (1988).

[14] Noyes F.R. and Grood E.S., The strength of the anterior cruciate ligament in humans and Rhesus monkeys. *J Bone Joint Surg Am*, 58(8), 1074-82 (1976).

[15] Danto M.I. and Woo S.L., The mechanical properties of skeletally mature rabbit anterior cruciate ligament and patellar tendon over a range of strain rates. *J Orthop Res*, 11(1), 58-67 (1993).

"WHO'S ON FIRST!" "WHAT?" "WHAT'S ON SECOND!"
AND HOW 'WHAT' GOT THERE ON AN OPTIMAL BASERUNNING PATH

Chris Harman[1]

Abstract

A model is presented here for finding optimal paths for a baseballer, starting from home and sprinting through first base to second (ie 'stretching a double'). Some progress has previously been made on modelling sprinting on flat circular curves at uniform speed. This has been used to estimate curvature effects for 200 and 400 metre sprint times. However, in the baseball situation, two additional factors are important. The optimal trajectory is evidently not a circular curve and, since the runner starts from rest after hitting the ball, the acceleration phase is crucial. A method is devised here for including acceleration and more general curvature in finding optimal paths for minimising sprinting time when 'stretching a double'.

1.   INTRODUCTION

In baseball, the team that gets more players around the bases to home plate wins the game. However, in the hierarchy of offensive strategies, the most important is to devise ways to get players to second base. Second base is called the `scoring base', since most outfield safe hits will score a runner from second. Consequently there is much effort expended in devising ways to get runners to second. There are two important strategic ways to do this. Firstly a baserunner can `steal' second base from first, in which case the chosen path is trivial, essentially a straight line. Or, on an outfield hit, a runner can decide to `stretch' a double by running from home plate around first base to second. The path chosen by a runner in stretching a double is not at all obvious but is of course crucial to successfully arriving at second base before the ball. Players don't seem to be coached in this particular art of baserunning, the reason being that the optimal path is not known. Good baserunners might naturally choose a path close to optimal, but the situation needs to be analysed in order to provide coaching guidelines.

In recent times, one of the best exponents of stretching doubles was Pete Rose the famous major league baseballer who, even though he was not super fast, made an art form out of turning a single into a double. Pete would now be in the Baseball Hall Of Fame, but he was banned for betting on the game. To model this problem, it is necessary to devise a way to solve the dynamics of a runner accelerating from rest around a general curve. We will call the runner `Pete' - after Pete Rose. `What' is too confusing.

Modelling sprinting in a straight line has been much studied. See for example Fuchs [1], Pritchard [2] and Ward-Smith [3]. But models for running on a curve only seem to have been successfully devised for circular arcs with constant speed (Greene [4]).

[1]   Department of Mathematics & Computing, University of Southern Queensland, Toowoomba, 4350

Greene's model was supported by experimental evidence and compares well with other models (Behncke [5]).

The distance between bases is 30 yards and so by stretching a double, a baserunner runs a little over 60 yards. Velocity profiles against time are well known for the case of accelerating sprinters over short distances (Pritchard [2]) and the approach here is to assume that a baserunner has a sprinter's typical velocity profile in a straight line. This velocity profile will then be moderated by the effect of curvature using a generalisation of the results of Greene. Time of travel can then be calculated on general curves and optimal solutions are found over classes of feasible curves.

2.   SPRINT SPEED PROFILE

It has been shown (Pritchard [2]) that an excellent model for a top sprinter's speed-time profile for 100 metres is given by

$$v(t) = P\tau\left(1 - e^{-t/\tau}\right),$$

where $P\tau$ is the maximum speed of the runner and τ is a constant determined by internal resistances in the runner. Data indicates that τ is usually close to 1.

For simplicity, it will be assumed here that $\tau = 1$ so that the baserunner's natural speed profile v in a straight line is given by

$$v(t) = V\left(1 - e^{-t}\right), \tag{1}$$

where V is the maximum speed in yards per second (since yards are a standard baseball measure). Figure 1 shows this profile for the first 10 seconds where $V = 10$.

Assume now that the path of the sprinter is not a straight line but a curve $y = f(x)$. If the acceleration and speed profiles are assumed to be unaffected by the curvature, the speed profile as a function of x can be solved as follows. Integrating (1) gives the distance s covered in time t

$$s(t) = V\left(t - 1 + e^{-t}\right). \tag{2}$$

Eliminating t from (1) and (2) gives the relation

$$s = -V \log\left(1 - v/V\right) - v.$$

Hence,

$$\int_0^x \sqrt{1 + \left(f'(X)\right)^2}\, dX = -V \log\left(1 - v/V\right) - v.$$
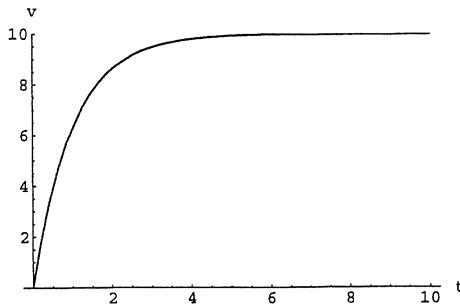
Figure 1: Sprinter's speed profile against time.

Differentiating then gives

$$\sqrt{1+\left(f'(x)\right)^2} = v'(x)\left(\frac{v}{V-v}\right)$$

and so we get the non-linear initial value problem

$$v(x)v'(x) = (V-v(x))\sqrt{1+\left(f'(x)\right)^2}, \quad v(0) = 0.$$

This can then be solved numerically using a Runge-Kutta solver and the solution then interpolated to give the function

$$v = F(x), \tag{3}$$

the speed as a function of position on the curve $y = f(x)$. All computations in this paper were performed using Mathematica. For the purposes of the following analysis, it will be assumed that Pete's maximum speed V is 10 yards per second.

3. RADIUS OF CURVATURE EFFECTS

Greene [4] modelled the mechanical effects of a sprinter running around a flat circular turn of radius R at constant speed. His model took account of the sprinter's straight line top speed V, foot contact-time, ballistic air-time, step length, and stride time. A reciprocal Froude number, or dimensionless radius Rg/V^2 enabled him to compare the theory against experiment for a large number of individuals on the same set of axes. The influence of radius of the turn on subsequent velocity was predicted and tested. The agreement between theory and practice was good and was verified for a range of radii between about 4 and 28 yards. More complex models have since been formulated but Behncke [5] considers that "the simplicity of Green's(sic) result and the apparent ease of its derivation make it - an ideal candidate for the analysis of the track and field situation."

Greene's model requires the solution v of the equation

$$v^6 + (Rg)^2 v^2 - (RgV)^2 = 0.$$

This is essentially a cubic and so Mathematica can be used to find the solution

$$
v = \sqrt{\frac{\left(27Rg^2V^2 + \sqrt{108Rg^6 + 729Rg^4V^4}\right)^{\frac{1}{3}}}{54^{\frac{1}{3}}} - \frac{2^{\frac{1}{3}}Rg^2}{\left(27Rg^2V^2 + \sqrt{108Rg^6 + 729Rg^4V^4}\right)^{\frac{1}{3}}}}, \qquad (4)
$$

where R is the radius of the circular path and V is the top speed of the runner in a straight line.

4.   THE BASERUNNING MODEL

The task of modelling a baserunner's dynamics using basic mechanics would be daunting. The approach to be used here is to assume that (4) applies on the appropriate baserunning curve $y = f(x)$ where the radius of curvature R is given by

$$
R(x) = \frac{\left[1 + (f'(x))^2\right]^{3/2}}{|f''(x)|}.
$$

Furthermore, using (3), it will be assumed that the baserunner's natural top sprinting speed V at any point on his accelerating path will be given by $V(s) = F(s)$, where s is the distance along the curve $y = f(x)$. This takes into account the acceleration from rest of the runner along the curve $y = f(x)$, but in the absence of curvature retarding effects. Hence $V(x) = F\left(\int_0^x \sqrt{1 + (f'(X))^2}\, dX\right)$. Equation (4) can then be used to allow for the effect that curvature has on this natural sprinting speed. In this way the actual speed $v(R(x), V(x))$ is estimated at any point x on the curve $y = f(x)$. Once the speed profile $v(R(x), V(x))$ is known on the curve $y = f(x)$, the time of travel T can then be computed numerically from

$$
T = \int_0^L \frac{ds}{v} = \int_0^{x^*} \frac{\sqrt{1 + (f'(x))^2}}{v(R(x), V(x))}\, dx, \qquad (5)
$$

where L and x^* give the respective extremities of the curve. Note that the problem of sliding into second base is ignored in this paper. It is convenient to just calculate the time taken to reach second base at which time the sprinter will be running at the top speed possible on that curve.

5.   PATH SELECTION

The layout of the bases being considered is shown in Figure 2 with home plate at the bottom right hand corner, first base at the apex of the triangle and second base at the bottom left corner. The distance between bases is 30 yards and the semicircle is indicated as a potential running path.
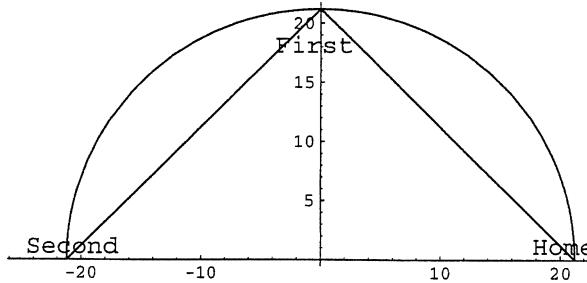
Figure 2: Layout of the bases from home to second showing a runner's semicircular path.

The theoretical optimal path is to run directly at first base, round first with an infinitesimally small curve which slows you down infinitesimally and continue sprinting in a straight line to second. Using (5) it was calculated that Pete would do this in 6.9 seconds over the distance of 60 yards. Of course this optimal strategy fails due to the impossibility of the dynamics at the point on the path with high curvature. However, it is a useful result in that it provides a benchmark lower bound baserunning time.

In order to make things possible, feasible paths will obviously require a constraint on curvature. The path with *minimum* curvature is the semicircle shown in Figure 2. Using (5) and velocity profile (3), Pete's time to run this path was computed to be 9.3 seconds over a distance of 67 metres. This is considerably more than his benchmark time of 6.9 seconds and reinforces the experience that running on a semicircle is a poor strategy.

Now $|f''(x)|$ is an approximation to the curvature $\left|f''(x)\right|\left[1+\left(f'(x)\right)^2\right]^{-3/2}$, and it is well known that the path which minimises $\int_0^x \left[f''(x)\right]^2$ is a cubic spline. Since curvature has a cumulative retarding effect on speed, choosing a cubic spline as a potential running path is evidently a sensible strategy.

In choosing the spline, one thing is clear - the curvature at second base should be zero. Approaching second, the runner is at top speed, ready to go into a slide. The runner is always vertical at this stage indicating zero curvature of the path. What is not certain is the angle at which the runner rounds first base. At first base it will be assumed that the curvature is smooth and that the slope is θ. This information is sufficient to determine the spline as

$$y = f(x) = \begin{cases} 15\sqrt{2} + \theta x + \dfrac{(\theta-1)}{10\sqrt{2}}x^2 + \dfrac{(1-5\theta)}{900}x^3 & , \ 0 \leq x \leq 15\sqrt{2} \\ 15\sqrt{2} + \theta x + \dfrac{(\theta-1)}{10\sqrt{2}}x^2 + \dfrac{(\theta-1)}{900}x^3 & , \ -15\sqrt{2} \leq x < 0 \end{cases} \tag{6}$$

where θ is the slope of the curve at first base.

6.   RESULTS

Using path (6) Pete's running times determined by (5) were computed for a range of θ values. The interpolated results are shown in Figure 3 together with the lengths of the paths.
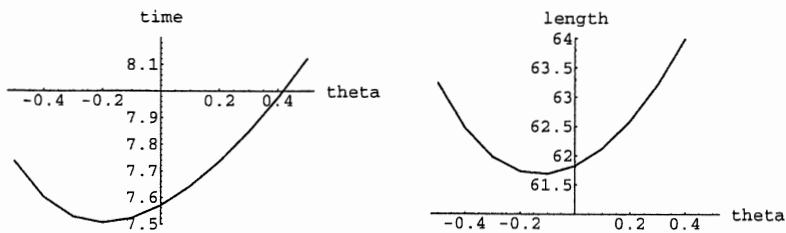
Figure 3:   Running time against θ and running length against θ.

Figure 4 shows the optimal path corresponding to θ = -0.2. The time of travel on this path is 7.506 seconds over a distance of 61.73 yards. For θ = -0.1, the path length is 61.69 yards and is the shortest, but takes 7.523 seconds. The more symmetric path with slope θ = 0 at first base has length 61.83 yards and takes 7.570 seconds.
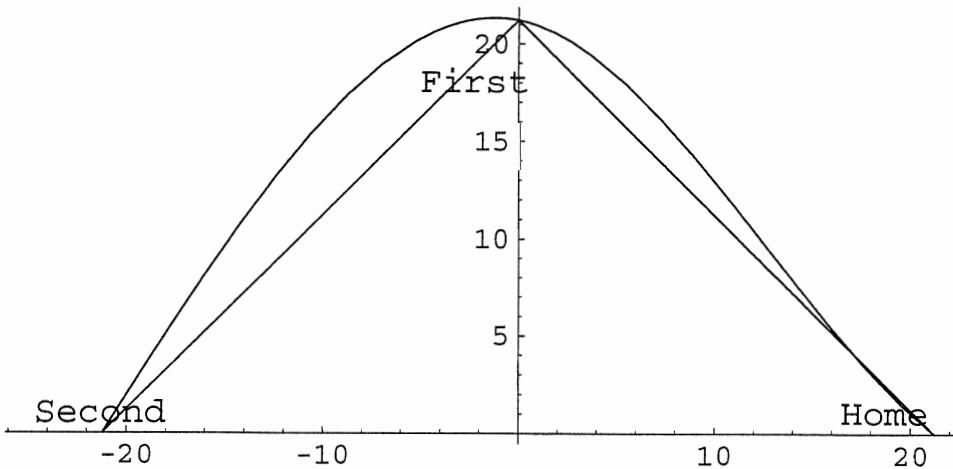
Figure 4:   The optimal running path.

Figure 5 shows Pete's velocity profile against *x* for this spline path with θ = -0.2. It clearly illustrates the effect of curvature as he runs around the bases from right to left.
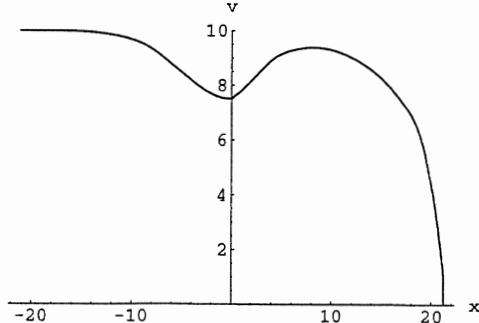
Figure 5:   Speed profile against distance showing influence of curvature.

7. DISCUSSION AND COACHING HINTS

It is interesting to note from above that this optimal time solution path does not correspond to the shortest length spline. The shortest is 61.69 yards, but takes 7.523 seconds. For Pete, the quicker time path would enable him to arrive at second base about 6 inches ahead of the shortest distance path. The more symmetric path, corresponding to a zero slope at first base, takes 7.570 seconds which would leave Pete about 2 feet short of the optimal path arrival at second base. These distances may seem short, but when stretching a double, the plays at second base are usually very close and every inch is important.

In order to exploit the advantage of choosing the optimal path with $\theta = -0.2$, a coach could obviously mark out the trajectory shown in Figure 4 and test it against natural inclinations of the runners. Time trials could be performed with selected runners choosing variations on this path. A `rule of thumb' could be used by just marking the two points where the optimal path is furthest from the baselines. The relevant point between home and first is about 7.4 yards back from first and 1.6 yards out from the line and the point between first and second is about 9.2 yards from second and 3.5 yards out from the line. A runner would then have five marked points (the two marked points plus the three bases) through which to run and this would constrain the path quite nicely as the optimal one.

Pete's maximum speed in a straight line was 10 yards per second. The technique developed here was also tried for faster runners with maximum speeds of 11 and 12 yards per second. These correspond to even-time and champion sprinters respectively. In each case the value of θ for optimal trajectory was again close to -0.2, so that the optimal paths were unchanged from that shown in Figure 4.

It might be possible to improve slightly on this result by using a class of curves more general than cubic splines. But given the mechanical constraints of minimising the effects of curvature, it can be argued that the approach used here would, for all practical purposes, yield the optimal solution. The problem of extending this analysis to running the bases to either third or home is interesting, but not of such importance.

REFERENCES

[1] Fuchs P.M., General mathematical concepts of compensation in sport science with quantitative analysis in the case of sprinting performance. *Mathematical Methods on Applied Sciences*, 18, 473-505, 1995.

[2] Pritchard W.G., Mathematical models of running. *SIAM Review*, 35(3), 359-379, September 1993.

[3] Ward-Smith A.J., A mathematical theory of running based on the first law of thermodynamics, and its application to the performance of world class athletes. *Journal of Biomechanics*, 18, 337-349, 1985.

[4] Greene P.R., Running on flat turns: Experiments, theory, and applications. *Transactions of the ASME Journal of Biomechanical Engineering*, 107, 96-103, May 1985.

[5] Behncke H., Small effects in running. *Journal of Applied Mathematics*, 10, 270-290, 1994.

THE SUBJECT: MATHEMATICS IN SPORT

Graeme L. Cohen[1]

Abstract

Sport is known as an excellent area for mathematical applications and research, and as a source of problems and exercises at all levels of mathematics teaching. In this paper, we describe a full-credit university elective subject, called Mathematics in Sport, delivered for the first time last summer. Topics covered included the assignment problem for team selection, probabilities in tennis arising from the probability of winning a point, the "sweet spot" on snooker balls and cricket bats, the rigging of a rowing eight, and applications of graph theory in tournament scheduling. We also argue here for more attention to be paid to the need for such broad mathematical electives in university curricula.

1. INTRODUCTION

The current interplay between mathematics and sport has the following aspects:

- sport as a resource for new research in mathematics,
- sport as an area of application of existing mathematics,
- sport as a source of applications and exercises in the teaching of mathematics, at primary, secondary and tertiary levels.

There is a natural fourth aspect of this relationship: the development of mathematics in sport as a full-credit subject at university level. The aim of this paper is to describe the subject Mathematics in Sport, which I taught for the first time in the 1997–98 Summer Session at the University of Technology, Sydney. The purpose in developing the subject was to have available in the profile of the School of Mathematical Sciences an elective subject that was equally suitable for, and appealing to, students undertaking a degree majoring in an area of mathematics and students from any other discipline in the university. There were to be no prerequisites other than high school mathematics, including some calculus.

Sixteen students, three of whom were studying computing science and two engineering, took the subject. The remainder were majoring in mathematics. In many cases, the students had one subject to go in their course, and saw this as an attractive option that they could complete over summer. Undoubtedly, this met both the nobler objective of popularising mathematics and the baser objective of gaining further funds for the School. I shall take up these points at the conclusion of the paper.

[1] School of Mathematical Sciences, University of Technology, Sydney, PO Box 123, Broadway NSW 2007

A small amount of funding had been made available to me in order to employ an assistant for literature and web searches, and to make up our own database of resources. This proved to be of immense value. When the decision was finally taken around August 1997 to run the subject, I had most of the material I would need readily at hand.

I had also obtained some material that I was not prepared to use. From the beginning, I was interested only in what might be classified as legitimate uses of mathematics as a means of describing, understanding or predicting sporting achievements. I had no interest in the use of sport as an artificial source of exercises in mathematics, although this is often seen in mathematics texts at all levels. Not knowing the level at which the book was aimed, I sent away for *Yesterday's Sports, Today's Math* [1]. In it, I found the following typical exercise, headed "Prime Time Wrestling" (very slightly reworded to fit this text): "Weight (in pounds) of Mr Perfect is 291, Crush 293, Doink 258, Hulk Hogan 303, Randy Savage 242, Bret Hart 274. Consider these to be actual wrestlers' weights and wrestle this problem to the ground. When two wrestlers meet, the heavier wrestler wins, unless the difference of their weights is a prime number. Which one of the wrestlers can beat all the others?" This is definitely not what I had in mind.

2. THE TOPICS

I shall first describe the major topics that made up the syllabus. The topics will be listed in approximately the order of treatment, with a varying amount of discussion usually dependent on whether I think I have anything half new to say about them. References will be given to allow others to follow up this initiative. Since I am continually finding new sources, it is quite likely that there will be some variation in the topics to be treated next time the subject runs.

Besides these major topics, there were a number of smaller topics, some of which will be given later. These often took just a few minutes to describe.

Assessment principles will be discussed in the next section, except to say here that a major component of the assessment was a large essay. This must be borne in mind since I avoided those topics in the main stream of lectures and seminars that I felt would be more suitable for students to choose for their own research. Statistical descriptions and analyses of various sporting records are the prominent example.

(1) *The Assignment Problem.* This is concerned with the manner in which a team is chosen from a number of candidates. Typically, basketballers are chosen for a team based on a ranking of their performance at different positions. An optimum selection will be one with minimum sum of ranks. This is a standard transportation problem in operations research, and may be treated as a problem in linear programming. There is an alternative and very straightforward algorithmic approach to the solution, called the Hungarian algorithm, described, for example, in Lapin [2].

Some time ago, Machol [3] pointed out the conceptual difficulty of relying on a sum of necessarily approximate rankings, and suggested that choosing a swimming medley team based on actual times in the four events (freestyle, breast stroke, butterfly and backstroke), and minimising the sum of the times, was a more valid

application. Heffley [4] suggested as a second application the choosing of runners in a relay team, since each may perform differently in the separate legs of the race. The class was asked to think of another legitimate application, and one student came up with the choice of a triathlon team for inter-club events, where each event (running, cycling, swimming) has a different competitor.

(2) *A Probabilistic Analysis of Tennis.* If A plays B at tennis, and the probability that A wins a point (based on many matches between them) is p, then what is the probability that A wins a game, a set, a match? What is the difference if tie-breaker sets are played? What is the expected number of points in a game? (Each point is considered to be an independent event.) Such questions have been posed and answered by a number of writers, such as Pollard [5] and Sadovskii and Sadovskii [6].

To determine the probability that A wins a game once the score is deuce seems, for all the writers, to require summing an infinite geometric series. Nowhere did I see the following simple approach. Let that probability be d. From deuce, either A wins the next two points (with probability p^2) or A and B win one point each (with probability $2pq$, where $q = 1 - p$) and then it is back to deuce. Therefore,

$$d = p^2 + 2pqd,$$

from which

$$d = \frac{p^2}{1 - 2pq} = \frac{p^2}{p^2 + q^2}.$$

This is then used in determining the probability that A wins a game by calculating also the probabilities that A wins in four, five or six points, or that A and B reach deuce. Knowing this, the same approach is used for the probability of winning an advantage set from five games all. Also, determining d is the same as determining the probability that A wins a tie-breaker game from six points all.

There is a similar treatment for the expected number of points in a game. We need first the expected number of points from the first deuce until the end of the game. Let this be E. Then

$$E = 2p^2 + 2q^2 + 2pq(E + 2),$$

since either A wins the next two points, or B does, or they win one each, and in the third case the expected number of points to the end of the game is two more than it was from the first deuce. This simplifies to

$$E = 2 + 2pqE,$$

from which

$$E = \frac{2}{p^2 + q^2}.$$

Croucher's derivation of this, in [7], uses series summation that is definitely not available to students who have only a high school background in mathematics.

(3) *Conversion Attempts in Rugby.* The question of how far back from the try line to bring the football so as to maximise the angle for the kick at goal has been treated by a host of writers. Some apparently think they are answering the question for the first time. The sequence of notes on this topic by Hughes [8], Avery [9] and Worsnop [10] can be made into quite a fun lecture. Isaksen [11] extends the notion to American football.

(4) *Oar Arrangements in Rowing Eights.* Following Brearley [12], we showed that the common orthodox rigging of a rowing eight is less efficient than either the German rigging or the Italian rigging (see Figure 1), in that the second and third arrangements produce no turning moment at any stage of the stroke. Actually, Brearley mentions only the orthodox and German arrangements, but does not use these names. They are given by de Mestre [13], who adds also the Italian rigging. It was easy to see that Brearley's analysis applies equally to it.

Orthodox   German   Italian
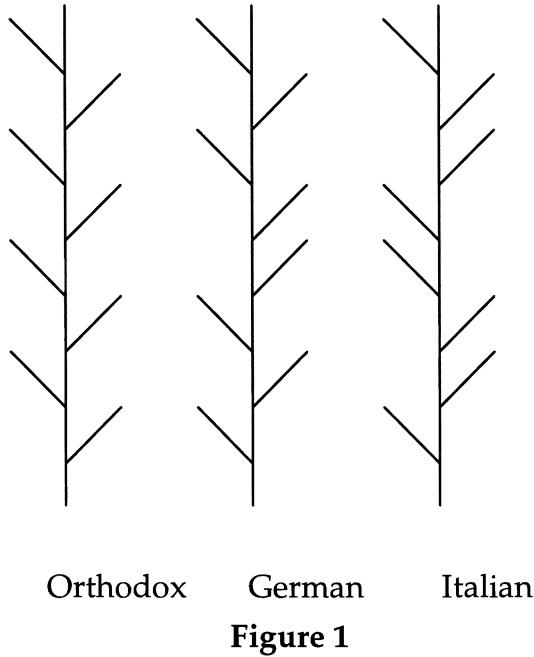
Figure 1
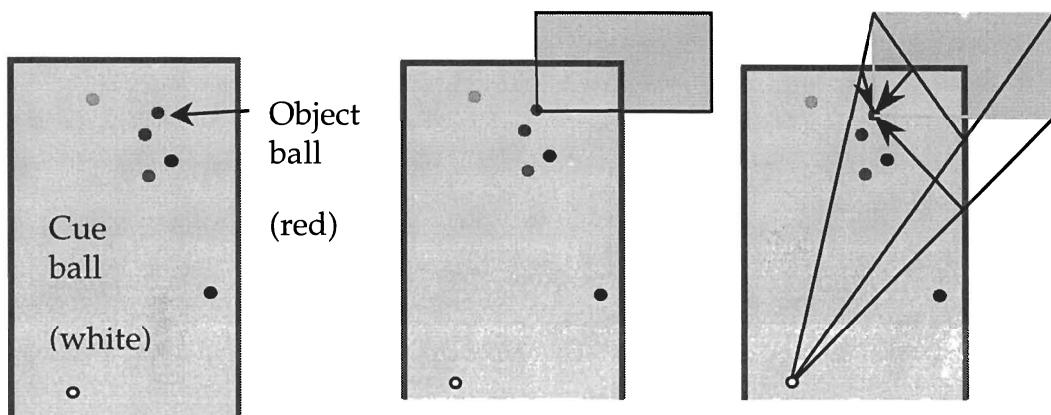
The class enjoyed this topic, and saw it as a superlative use of mathematics to deduce an unexpected but presumably practical result. We found it difficult to track down any actual uses of non-orthodox rigging, besides the reference in Brearley's paper. Then one of the students came up with Jacobsen's history of Australian rowing [14]:

> On arriving at Brisbane for training we learnt that the new boat ordered from George Towns & Sons for the occasion had been smashed and would not be delivered. We therefore borrowed a boat from Queensland University. It had been set up with tandem rigging. Numbers 'four' and 'five' were on bow side, and 'bow' rowed as number 'two' while 'two' rowed as 'bow'. It sounds and felt rather complicated, *and without logical purpose,* and it took some time to become accustomed to the different positions ... [my italics]

(5) *Where To Strike a Snooker Ball.* In order to achieve just the right amount of initial spin, and so as not to cause skidding (when skidding is not wanted), "the cue must strike the snooker ball at a distance above the table equal to seven tenths of the diameter of the ball". This part of the topic was taken from Daish [15]. As did the preceding topic, it required the introduction of a number of physical principles, new to most of the class. It was all much better appreciated when I explained at the end that the centre of percussion that we had found, the place to strike the snooker ball to avoid skidding, was in fact the ball's "sweet spot", a term that everyone knew.

Also appreciated, with some awe, was the following statistic quoted from Lindrum [16]: "... for many years [billiard balls] were made of ivory. Five balls could be made from one elephant tusk and apparently only the female elephant tusk was suitable for this purpose. It has been estimated that something like 12,000 elephants were slaughtered annually to supply billiard balls to Great Britain." The use of ivory finished around the turn of the century.
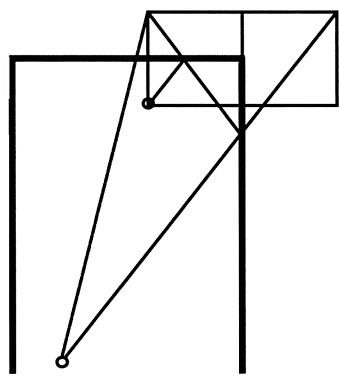
Still on the theme of snooker, in Figure 2 I give copies of the seven transparencies I designed to illustrate the "desnookering rectangle". It has a vertex at the object ball, which is not visible (snookered) from the cue ball, its centre at one of the table's corner pockets and its sides parallel to the table's sides. When straight shots without side-spin will do, shooting for any of the other three imagined vertices may allow an escape from the snooker. In the right situation this really works, although I have not seen any hint of this theme in any book on billiards or snooker. It is of course dependent on the use of congruent triangles and the notion that the angle of incidence equals the angle of reflection.

1. Snookered!
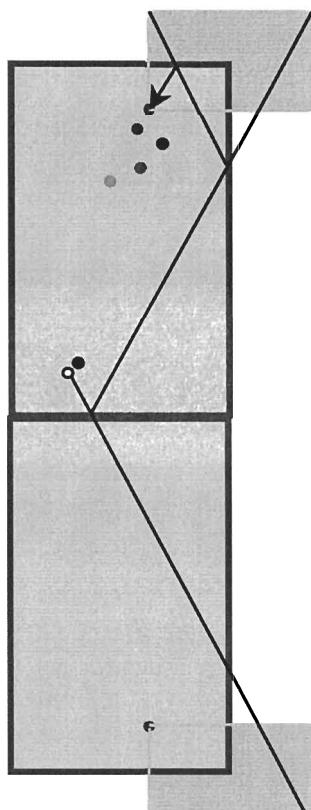
2. The desnookering rectangle!

3. Shoot for any corner of the rectangle.
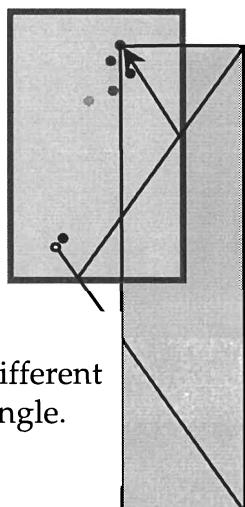
4. Find all the pairs of congruent triangles.

5. Now those shots are blocked. What do you do?

6. Look for reflections of the whole table, complete with desnookering rectangle, and try a corner of the reflected rectangle.

7. Or choose a different desnookering rectangle.

Figure 2

In fact, the angle of incidence is not equal to the angle of reflection. The coefficient of restitution between ball and cushion needs to be taken into account. De Mestre [13] is one source on this. Since a snooker ball is somewhat thicker than a point, assuming the angles to be equal is no doubt perfectly sound as an approximation for getting out of snookers. At least, it is for the kind of snooker I play.

The final item in this topic concerned the game of three-cushion billiards. In this, the "diamond system" appears to allow geometrical reasoning to be used to plan each shot. It is described briefly by Lindrum [16], and in great detail by Byrne [17]. It seems to be very complicated, and in any case, according to Byrne, must have each shot performed with side-spin, or "English" as he calls it.

(6) *The Best Way to Hit a Cricket Ball.* This topic was based on the article by Brearley, Burns and de Mestre [18]. They hoped to show that hitting a cricket ball at the centre of percussion of the bat, besides giving the least jarring, might also be the spot that would allow maximum distance to be attained. More rotational dynamics had to be introduced, and the coefficient of restitution, between bat and ball this time, was again required.

(7) *Introduction to Tournaments.* There are a number of aspects of graph theory relevant to tournaments. Essentially, the vertices of a graph represent teams in a competition, the edges indicate matches between teams, and, if they are directed edges, they show which team won the match.

In graph theory, a complete directed graph is in fact called a tournament. The theorem that every tournament contains a (directed) Hamilton path (Grimaldi [19, p. 581]) allows us to think that there is a team in a round robin event which might be considered to be best (in that there is a team A, which beat team B, which beat team C, …). But then we find that this is rarely plausible. In fact, for tournaments with at least three vertices, if there are two vertices with equal out-degree then there always exists a cycle on three of the vertices. That is, in sporting parlance, if two teams in a round robin event have an equal number of wins then there is always a triangular standoff (where, say, team A beat team B, which beat team C, which beat team A). See Sadovskii and Sadovskii [6, p. 133].

Another possible way to determine a winner in a round robin event might be to seek a king: with three or more teams, there is always a team A (called a king) such that every other team is either beaten by A or is beaten by another team that is beaten by A. See Maurer [20].

On the question of scheduling the different rounds of a round robin, and learning to avoid pitfalls such as taking insufficient care in the early rounds and then finding no feasible arrangement for a subsequent round, we worked through Wallis [21]. Before that though, the class took some delight in de Mestre's approach, in [22], of drawing up the rounds.

(8) *Operations Research in Sports.* Reading through Chapter 14, Operations Research in Sports, of the *Handbooks in OR & MS, Vol. 6* [22] allowed a useful survey of some topics previously treated and still others for which time was not available for a fuller treatment. These included:

- strategies in baseball,
- the value of field position in American football,
- the value of a tie and extra-point strategy in American football,
- when to pull the goalie in ice hockey,
- the validity of winning streaks,
- issues in the draft system in professional sports,
- maximising expected achievements in athletic events of increasing difficulty, such as weightlifting, pole vaulting and high jumping,
- competitive games of boldness, in which the competitor chooses the level of difficulty, such as figure skating, gymnastics and diving, as well as the three events just mentioned,
- handicapping issues.

This chapter of the *Handbooks* also contains a very large list of references to the relevant operations research literature.

(9) *Miscellaneous short topics.* At various times during the course, a few minutes were taken to discuss short topics such as the following: the Fosbury flop in high jumping, the game of nim, the tactics of darts, the possible misleading nature of cricket averages, and the design of a shuffleboard alley.

3.   THE ASSESSMENT

There were two main items of assessment in the subject Mathematics in Sport. The first was an essay. Associated with this were a preliminary paper giving the topic, the aim and the methodology to be adopted, and a seminar presentation. The second item of assessment was a final examination. I intend here to describe the essay briefly.

The Summer Semester involves an accelerated teaching program so the choice of essay topic had to be made quickly. The students were allowed to work in pairs on an essay, but few chose to do so. Many of the essays amounted to the student's own summary and explanation of a paper or two in the literature, often with some assistance from me, and some involved extensive web searches. The students were expected to employ a level of mathematics or statistics in their work commensurate with their attained level of study in these areas, and their essays were graded accordingly.

Some of the topics were my suggestions. These include the following, for which in relevant cases I have given references:

(a) Analysis of the last twenty years track records over different distances.
(b) Comparison of men's and women's track records in selected events.
(c) A rating system for one-day cricket. (See Johnston, Clarke and Noble [24].)
(d) Design and application of a program for the Assignment Problem.
(e) An analysis of handicapping systems in golf. (See Clarke and Rice [25], and Stroud [26].)
(f) Is goal scoring in soccer a Poisson process? (See Avery [9].)
(g) The probability of winning a game in doubles tennis.

From this list, some students chose variants of (a) or (b), and (c), (e) and (f). Other topics chosen by students for themselves were:

(h) Effects of atmospheric changes on trajectories of projectiles in sport. (See de Mestre [27].)

(i) Mathematics in casino gambling.

(j) The probability of winning the Australian Football League finals series. (See Schwertman and Howard [28].)

(k) ATP tennis rankings.

(l) Goal conversions in football (extending the ideas in topic (3) of Section 2, above).

In most cases, the results were very satisfactory, with a few being exceptional in the thoroughness of the work and excellence of presentation. Notable was the essay on the topic (h) (by Peter Krebs). Here is its great introduction:

> In a limited over day and night cricket event recently, New Zealand was chasing 300 runs to win against South Africa. New Zealand needed to get 12 runs off the last over. After four deliveries in that over, New Zealand was still seven runs short to win. Nash managed a sweep shot with a flat trajectory on the next ball. The ball carried towards the boundary and the point of impact was right on top of the boundary rope for four. The distance to the boundary was around 100 metres.

Had the temperature at the Gabba in Brisbane been only one degree (1°C) higher, the ball would have carried a further 50mm to clear the rope for six.

New Zealand lost the match by two runs.

4. BROAD ELECTIVES IN MATHEMATICS

For a few years before the detailed development of the subject Mathematics in Sport, I had been interested in the concept of broad electives in mathematics — subjects which could fit in to the elective program of any course in the university, with appeal to a wide range of students, not requiring extensive prerequisite knowledge, but having respectable mathematical substance. This would give the discipline of mathematics some level of equality with areas such as humanities and language studies, which have always been happy to offer one-off electives to students majoring in mathematical sciences. I had applied, finally with a little success, for grants to help develop this theme.

It seems reasonable to believe that many students, such as those who enjoyed high school mathematics but saw no related career opportunities, would be keen to take incidental studies of this sort. My experience with Mathematics in Sport is that the applicability of mathematics in totally unexpected settings is greatly appreciated (even by the mathematics majors in the class) and should be exploited much more earnestly. No doubt, other university mathematics departments have such electives available for the general body of students, but it has not been easy to find details of their efforts.

In the Australian setting, there is a recent added impetus for broad mathematical electives. Government cuts to recurrent expenditure grants have obliged all publicly

funded university departments to seek their own additional funding. To do so is now itself essentially part of Government policy. The main source of external funds for most universities lies in attracting full-fee paying international students, and full-fee paying postgraduate students. For individual mathematics departments, another source is elective subjects of the type described here. This may simply attract some funds away from other departments in the same university, but apparently this is the game that the Australian Government wants us to play.

This is not the place to canvas this issue any further. I must add though that when I mentioned above that these should be electives "having respectable mathematical substance", I find it hard to think of a better subject than Mathematics in Sport.

REFERENCES

[1]   Fraser D., Yesterday's Sports, Today's Math (Dale Seymour Publications, USA, 1997).

[2]   Lapin L.L., Quantitative Methods for Business Decisions, sixth edition (Dryden Press, USA, 1994).

[3]   Machol R.E., Assigning swimmers to a relay team, in Optimal Strategies in Sports, S.P. Ladany and R.E. Machol, eds (North-Holland, 1977), 168.

[4]   Heffley D.R., Assigning runners to a relay team, in Optimal Strategies in Sports, S.P. Ladany and R.E. Machol, eds (North-Holland, 1977), 169–171.

[5]   Pollard G.H., An analysis of classical and tie-breaker tennis, *Austral. J. Statist.*, Vol. 25, Pt 3, 496–505 (1983).

[6]   Sadovskii L.E., and Sadovskii A.L., Mathematics and Sports (American Mathematical Society, 1993).

[7]   Croucher J.S., The effect of the tennis tie-breaker, *Res. Quart. for Exercise and Sport*, Vol. 53, Pt 4, 124–127 (1982).

[8]   Hughes A., Conversion attempts in rugby football, *Math. Gaz.*, Vol. 62, 292–293 (1978).

[9]   Avery P., Mathematics in sport, *Math. Gaz.*, Vol. 73, 1–6 (1989).

[10] Worsnop G., An aid to conversions in rugby, *Math. Gaz.*, Vol. 73, 225–226 (1989).

[11] Isaksen D.C., How to kick a field goal, *College Math. J.*, Vol. 27, Pt 4, 267–271 (1996).

[12] Brearley M.N., Oar arrangements in rowing eights, in Optimal Strategies in Sports, S.P. Ladany and R.E. Machol, eds (North-Holland Publishing Company, 1977), 184–185.

[13] de Mestre N., Mathematics and sport, *Aust. Math. Teacher*, Vol. 43, Pt 4, 2–5 (1987).

[14] Jacobsen A.N., Australia in World Rowing (Hill of Content, Melbourne).

[15] Daish C.B., The Physics of Ball Games (English Univ. Press, London, 1972).

[16] Lindrum H., Snooker, Billiards and Pool (Summit).

[17] Byrne R., Byrne's Standard Book of Pool and Billiards (Harcourt Brace, Florida, 1987).

[18] Brearley M.N., Burns J.C. and de Mestre N., What is the best way to hit a cricket ball?, *Int. J. Math. Educ. Sci. Technol.*, Vol. 21, Pt 6, 949–961 (1990).

[19] Grimaldi R.P., Discrete and Combinatorial Mathematics, third edition (Addison–Wesley, USA, 1994).

[20] Maurer S.B., The king chicken theorems, *Math. Mag.*, Vol. 53, Pt 2, 67–80 (1980).

[21] Wallis W.D., One-factorizations of graphs: tournament applications, *College Math. J.*, Vol. 18, Pt 2, 116–123 (1987).

[22] de Mestre N., Mathematics applied to sport, in Mathematics and Computing in Sport (Conference proceedings, Bond University, Queensland, Australia, 13th to 15th July, 1992), 137–148.

[23] Handbooks in Operations Research and Management Science, Vol. 6: Operations Research and the Public Sector, S.M. Pollock, M.H. Rothkopf and A. Barnett, eds (Elsevier Science BV, 1994).

[24] Johnston M.I., Clarke S.R. and Noble D.H., Assessing player performance in one day cricket using dynamic programming, *Asia-Pacific J. Op. Res.*, Vol 10, 45–55 (1993).

[25] Clarke S.R. and Rice J.M., How well do golf courses measure golf ability? An application of test reliability procedures to golf tournament scores, *ASOR Bull.*, Vol. 14, Pt 4, 2–11 (1995).

[26] Stroud R.C., Mathematical underpinnings of the slope handicap system, in Science and Golf, A.J. Cochran, ed. (Chapman and Hall, London, 1990), 135–140.

[27] de Mestre N., The Mathematics of Projectiles in Sport (Cambridge University Press, Cambridge, 1990).

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

AN ANALYSIS OF TEAM STATISTICS IN AUSTRALIAN RULES FOOTBALL

Andrew Patterson and Stephen R. Clarke[1]

Abstract

Champion Data has collected Australian Rules Football player statistics since the beginning of the 1996 AFL season. As the data is collected from video replays, the statistics show more detail than those previously published, with the quality and effectiveness of possessions and disposals also being recorded. The statistics include short and long kicks, effective handballs, contested and uncontested marks, hard and loose ball gets, gathers, free kicks and clangers. The data have been analysed to determine the contribution of each factor to the game outcome. The ultimate aim of the investigation is to provide a scientific underpinning for a player-rating system.

1. INTRODUCTION

The game of Australian Rules Football has been around for over a century and is regarded as one of the most popular sports in the country. The 1997 AFL season involved 16 teams who competed against each other with each team playing a total of 22 matches during the home and away season. Although a finals series was conducted at the end of the season this paper only investigates data from the home-and-away matches.

During the 1997 season, each match took place on one of 11 grounds across the country. Each ground has a centre square marked as well as a semicircle of radius 50 metres around the two goal lines at each end of the ground. Each team had a designated "home ground" on which they played between nine and 15 games for the season. A side could use 21 players for a match with 18 players being allowed on the ground at any time. Matches are played over four quarters with each quarter lasting about 30 minutes. At the end of each quarter the two teams exchange scoring ends.

Australian Rules Football is high scoring compared to other team sports. The ball is moved around the ground by methods of kicking or punching, and with no offside rule the game is played at a fast pace. A goal is worth six points and a behind is worth one point. For the 1997 AFL season the average winning score was 16 goals 10 behinds, 106 points with the average losing score being 10 goals 8 behinds, 68 points.

Champion Data has collected the data since the beginning of the 1996 AFL season. The statistics for each match are recorded using two people, one reading out the play and the other entering the relevant information into a laptop computer. In 1997 there were 36 designated statistics that were grouped into three categories - disposals, possessions and attack. For the first time these statistics measured the quality of

possession and disposal. For example, kicks were rated as clangers, ineffective, and effective short and long kicks, according to team rules commonly used in clubs. Some of the definitions of these variables are given in Table 1.

This paper looks at team totals for some of the more important variables, with a view to determining the importance of each variable to the outcome of the game. The competition statistics have been progressively broken down with respect to team, result and opposition. Fisk [1] performed a similar analysis using team data from the American National Basketball Association.

Table 1
Definitions of main player performance statistics.

Variable	Code	Category	Definition
Kick Long	KKL	Disposal	A kick of more than 40 metres to a 50/50 contest or better for the team
Kick Short	KKS	Disposal	A kick of less than 40 metres that results in an uncontested possession for the team
Kick Ineffective	KKI	Disposal	A kick of less than 40 metres to a contest or a kick of more than 40 metres to a worse than 50/50 contest for the team
Kick Clanger	KKC	Disposal	A kick under little or no pressure that goes straight to an opponent
Handball Effective	HBE	Disposal	A handball to a team mate that hits the intended target to the team's advantage
Handball Clanger	HBC	Disposal	A handball under little or no pressure that goes straight to an opponent
Handball Received	HBR	Possession	When a player takes possession of the ball via a handball and a clean disposal follows
Loose Get	LBG	Possession	When a player picks up a disputed ball that has spilled onto the ground and a clean disposal follows
Hard Get	HBG	Possession	When a player picks up a disputed ball that has spilled onto the ground in a pressure situation and a clean disposal follows
Mark Uncontested	MKU	Possession	When a player takes a mark unopposed
Mark Contested	MKC	Possession	When a player marks under pressure or in a pack
Inside-50	I50	Attack	When a player takes or delivers the ball inside the 50 metre area in the team's attacking half
Score	SCR	Attack	The total number of points scored by the team

2. COMPETITION DATA

In 1997 the average number of disposals per match for each team was 266. The competition averages for each of the six methods of disposal is given in Table 2. Kicking is the most common form of disposal with teams averaging twice as many kicks as handballs. Long kicks are performed on more than twice as many occasions as short kicks with clanger/ineffective kicks and handballs (junk disposals) counting for about 15% of all disposals.

Table 2

Competition disposal statistics for the 1997 season.

Variable	1997 Competition Average	% Disposal
Kick Long	99.0	37.2
Kick Short	46.6	17.5
Kick Ineffective	27.7	10.4
Kick Clanger	8.8	3.3
Handball Effective	79.9	30.0
Handball Clanger	4.1	1.6
Total	266.1	100.0

In order for a player to effect a disposal, he must first gain possession of the ball. Since a possession nearly always results in a disposal, the totals of these two categories should always be about the same. The statistics show that football is evenly divided between gaining and retaining possession with about 50% of possessions being won in a contest situation and the other 50% coming from uncontested play.

3. TEAM DATA

In 1997 Carlton averaged the most disposals per match and Melbourne averaged the fewest of the 16 teams in the competition. Table 3 lists the number of disposals each team averaged per match in 1997 as well as the breakdown of each method of disposal. Adelaide and North Melbourne averaged the most effective long kicks while Melbourne and Collingwood averaged the least. The Western Bulldogs had more effective short kicks than any other side and Richmond had the fewest. Carlton, Hawthorn and Essendon averaged a large number of effective handballs while North Melbourne was well below the competition average. The two South Australian sides, Adelaide and Port Adelaide averaged the fewest number of junk disposals for the season.

The percentage breakdown of disposal methods shows the different styles of play used by each team. Adelaide and North Melbourne were long kicking teams that rarely handballed. Collingwood, West Coast and the Western Bulldogs all had high short kick to long kick ratios while Melbourne and Hawthorn were more likely to use handball as a form of disposal.

Table 3

Breakdown of the average number of team disposals per match.

Team	Total	KKL	%	KKS	%	HBE	%	Junk	%
Adelaide	265.9	109.1	41.0	47.3	17.8	73.0	27.5	36.4	13.7
Brisbane	264.8	99.3	37.5	48.5	18.3	75.6	28.6	41.4	15.6
Carlton	283.3	103.7	36.6	49.5	17.5	89.5	31.6	40.6	14.3
Collingwood	261.7	88.2	33.7	50.8	19.4	80.3	30.7	42.4	16.2
Essendon	279.9	101.1	36.1	48.8	17.4	87.3	31.2	42.7	15.2
Fremantle	262.3	97.7	37.3	46.8	17.8	77.0	29.4	40.8	15.5
Geelong	272.1	101.1	37.2	51.6	19.0	81.7	30.0	37.7	13.8
Hawthorn	269.9	100.0	37.1	40.9	15.2	87.9	32.6	41.1	15.2
Melbourne	251.4	85.5	34.0	44.4	17.6	80.3	31.9	41.3	16.4
North Melbourne	252.7	105.7	41.8	40.0	15.8	65.4	25.9	41.6	16.5
Port Adelaide	256.9	103.1	40.1	40.6	15.8	76.7	29.9	36.4	14.2
Richmond	255.3	101.7	39.8	38.0	14.9	75.0	29.4	40.6	15.9
St.Kilda	275.0	101.4	36.9	47.8	17.4	83.7	30.4	42.1	15.3
Sydney	268.5	96.8	36.1	48.2	17.9	81.1	30.2	42.4	15.8
West Coast	270.0	95.5	35.4	49.9	18.5	84.8	31.4	39.9	14.8
Western Bulldogs	270.2	94.5	35.0	53.2	19.7	79.6	29.5	43.0	15.9
Average	**266.3**	**99.0**	**37.2**	**46.6**	**17.5**	**79.9**	**30.0**	**40.6**	**15.3**

4. COMPETITION DATA BY RESULT

In 1997 the team that had the most number of disposals in the game won seventy-five percent of all matches. The difference in the average number of disposals between winning and losing teams as well as the percentage of teams that win given that they outnumber the opposition in a particular disposal method are shown in Table 4.

Table 4

Differences between winning and losing team's disposal averages.

Variable	Winner	Loser	Difference	% Win
Kick Long	104.2	93.9	10.3	83.2
Kick Short	51.5	41.9	9.6	75.6
Kick Ineffective	27.6	27.8	-0.2	52.2
Kick Clanger	8.6	8.9	-0.3	47.2
Handball Effective	83.4	76.5	6.9	60.9
Handball Clanger	4.0	4.3	-0.3	45.5
Total	279.3	253.3	26.0	75.4

On average the winning side had 26 more disposals than the losing side. This included 10 more effective long kicks, 10 more effective short kicks and 7 more effective handballs. The winning side averaged 11% more effective long kicks, 23% more effective short kicks and 9% more effective handballs than the losing side.

Effective long kicks were the best predictor of a match result with 83% of teams who had the majority of effective long kicks in a match winning.

5. TEAM DATA BY RESULT

In 1997 the average side had an extra 10 more effective long kicks when they won compared to when they lost. Table 5 shows the difference and percentage difference comparisons of each team's win/loss averages. The percentage difference is calculated as the increase in percentage between the winning and losing averages. The values that are the most distant from the competition averages are of the most interest because it is these figures that can reflect a team's strength or weakness. Essendon and West Coast had the biggest differences between win/loss averages of any club for effective long kicks with Essendon averaging an additional 24 long kicks and West Coast an extra 22 when they won. Brisbane were up 57% on effective short kicks when they won whereas Western Bulldogs had about the same number of short kicks when they won as when they lost. West Coast averaged 17 more effective handballs when they won whereas teams such as Hawthorn, Melbourne and the Western Bulldogs all averaged more handballs when they lost than when they won.

Table 5

Differences between team's winning and losing averages.

Team	KKL		KKS		HBE	
	Diff	%Diff	Diff	%Diff	Diff	%Diff
Adelaide	3.4	3.1	9.4	22.4	12.9	19.7
Brisbane	5.6	5.8	21.8	57.4	7.8	10.8
Carlton	12.1	12.4	6.4	13.8	12.4	14.8
Collingwood	8.0	9.5	11.1	24.2	10.8	14.3
Essendon	23.9	26.2	11.6	26.3	4.7	5.5
Fremantle	13.5	14.8	4.7	10.6	11.7	16.4
Geelong	7.5	7.8	9.3	20.6	6.2	8.1
Hawthorn	4.6	4.7	11.1	30.2	-2.3	-2.6
Melbourne	6.2	7.3	5.4	12.4	-4.6	-5.7
North Melbourne	4.1	3.9	11.3	33.4	12.7	21.7
Port Adelaide	2.9	2.8	9.9	27.4	13.8	19.8
Richmond	3.0	3.0	16.1	52.6	2.8	3.9
St.Kilda	4.6	4.7	9.4	22.6	5.2	6.5
Sydney	17.3	19.8	5.8	13.0	11.0	14.6
West Coast	22.3	27.1	3.5	7.4	17.2	23.0
Western Bulldogs	12.9	14.9	0.1	0.2	-3.3	-4.1
Average	**10.3**	**11.0**	**9.6**	**22.9**	**6.9**	**9.0**

6. OPPONENT EFFECTS

It is possible to see how the opposition performs against each team in order to get a better understanding of the characteristics of a particular side. Table 6 shows the difference between winning and losing opponents against each team for the 1997 season. Teams that beat Brisbane and Richmond averaged about 19 more effective long kicks than teams that lost to these two clubs. This can be compared with the average difference of 10. Teams that beat Essendon averaged 4 less effective long kicks than the teams that lost. Teams that defeated Brisbane, Melbourne, Sydney and West Coast all averaged a lot more effective short kicks than the teams that were beaten by these sides. Teams that beat North Melbourne averaged slightly less short kicks than the teams that lost. Teams that beat Collingwood, Melbourne and Richmond had a higher handball average than the teams that lost to them. Teams that beat North Melbourne and St.Kilda handballed a lot less than the sides who lost to them.

Table 6
Differences between winning and losing opponent's averages against each team.

Team	KKL		KKS		HBE	
	Diff	%Diff	Diff	%Diff	Diff	%Diff
Adelaide	15.9	16.7	7.5	16.2	7.8	9.4
Brisbane	19.1	20.5	16.0	43.7	12.4	15.3
Carlton	13.7	14.2	4.7	11.3	0.0	0.0
Collingwood	4.6	5.0	8.7	18.8	15.7	22.6
Essendon	-4.1	-4.1	4.6	9.2	4.0	5.3
Fremantle	10.5	10.8	3.0	8.0	5.8	7.1
Geelong	14.4	15.5	5.8	15.9	5.6	7.7
Hawthorn	11.8	12.6	11.2	28.3	12.3	18.5
Melbourne	5.7	6.3	17.1	41.9	21.0	31.6
North Melbourne	16.6	18.0	-0.5	-1.2	-5.3	-6.6
Port Adelaide	13.1	14.1	3.4	6.8	11.4	14.4
Richmond	18.7	19.8	7.1	16.1	15.0	20.6
St.Kilda	14.3	15.5	10.3	28.7	-4.7	-6.7
Sydney	4.2	4.5	17.6	43.6	-2.0	-2.4
West Coast	4.5	4.9	17.0	42.2	1.6	2.0
Western Bulldogs	8.3	8.7	8.6	20.6	8.6	11.5
Average	**10.3**	**11.0**	**9.6**	**22.9**	**6.9**	**9.0**

7. ZONE PROFICIENCY

Because the data records whenever the ball enters a team's attacking or defending 50-metre zone, it is possible to determine the proficiency of each team in certain areas of the ground. The ground has been divided into three zones with the proficiencies of each zone measured in the following way –

Midfield - Ratio of a team's inside-50 and the opposition's inside-50.
Forward Line - Ratio of a team's score and a team's inside-50.
Back Line – Ratio of the opposition's inside-50 and the opposition's score.

Table 7
Ranking of Zone Proficiencies

Team	% Position	Midfield	Forward	Back
Adelaide	1	1	13	7
St.Kilda	2	2	1	10
Geelong	3	3	6	4
Sydney	4	4	9	1
North Melbourne	5	5	11	3
Collingwood	6	9	5	5
West Coast	7	7	7	2
Brisbane	8	8	8	6
Western Bulldogs	9	11	2	12
Carlton	10	12	10	8
Essendon	11	14	3	11
Fremantle	12	6	15	13
Port Adelaide	13	10	14	9
Hawthorn	14	15	4	15
Richmond	15	13	12	14
Melbourne	16	16	16	16

Table 7 shows the zone proficiency rankings for all teams in each of the three zones. Although the AFL Ladder ranks teams on their number of wins, Clarke [2] has shown that percentage is a better indicator of overall team performance. The teams have been listed in their percentage order for the 1997 season. It is interesting to note that the correlation between percentage position and the midfield proficiency ranking is high. The only anomaly appears to be Fremantle who finished 12th on percentage order but had a high midfield ranking of 6. The forward and back line proficiency rankings are not highly correlated with finishing order. Essendon and Hawthorn who finished 11th and 14th on percentage order had the 3rd and 4th most proficient forward lines and Adelaide who had the highest percentage of any team was rated as 13th. The back line rankings were more highly correlated than the forward line rankings but still not significant. St.Kilda who finished 2nd on percentage, was rated as having the 10th best back line in the competition.

8.   REGRESSION ANALYSIS OF TEAM DATA

Table 8
Summary of regression analysis of score difference versus factor difference.

Variable	Coefficient	p-value	R-Sq.
Kick Long	1.89	0.00	45.8%
Kick Short	1.30	0.00	33.6%
Kick Ineffective	-0.22	0.57	0.2%
Kick Clanger	-1.08	0.17	1.1%
Handball Effective	0.66	0.00	12.5%
Handball Clanger	-0.91	0.35	0.5%

Table 8 gives the results of an individual regression analysis performed on each method of disposal with the score difference. Kick Long, Kick Short and Handball Effective were all found to be good predictors of score with their corresponding p-values being close to zero and the R^2 being reasonably high. Ineffective kicks, Kick Clangers and Handball Clangers were all found to be poor predictors of score with all of these variables having high p-values and a low R^2 value. The above coefficients do not change drastically if a multiple linear regression is performed. If all of the disposal methods are regressed together with the score difference an R^2 value of 66.0% is obtained.

The kick long regression coefficient of 1.89 implies that a team would expect to be in front of their opposition by about 2 points for every extra long kick that they had in a match. If a side had 10 more long kicks than their opposition for the game then the estimate of their winning margin would be 10 x 1.89 = 18.9 points. The data suggests that long kicks are worth about one and a half times a short kick and that handballs are worth half a short kick as far as the score difference is concerned.

9.   CONCLUSION

There is a wealth of performance statistics that can be analysed in various ways to determine the bearing that these statistics have on the result of a match. Clearly different teams have different playing profiles and these become more evident when the data is broken down into more detail. Kicks are the most important predictor of match result with effective long kicking seemingly being the key to a team's success. If a handball were to be rated as one point, short kicks would be rated as about two points and long kicks three points. This result could form the basis of a player performance rating system in Australian Rules Football.

REFERENCES

[1]   Fisk C.J., NBA regular season team statistics versus winning. A univariate/multivariate exploratory analysis. *ASA 1996 Proceedings of the Section on Statistics in Sport*, 1996, 60-65.

[2]   Clarke S.R., How fair is the AFL draw?, *Proceedings of the 4th Conference on Mathematics and Computers in Sport*, N. de Mestre and K. Kumar, Editors, Bond University: Gold Coast, 1998, 161-170.

THE MATHEMATICS OF BICYCLING: PART II

W.H. Cogill[1]

Abstract

A previous note Cogill [1] considered the energy of a bicycle and its components. The energy needed to propel a bicycle can be decreased if the mass of the rotating parts is decreased. The present note considers the overturning moment on a bicycle. Apart from imbalance on the part of the rider, the overturning moment is due to the gyroscopic effect caused by turning the front wheel. The effects of the other rotating parts, the rear wheel and the chain wheel, are minor and are neglected at this stage.

1. INTRODUCTION

The object of this note is to identify the restraining forces which act on a bicycle, and which must be corrected by the rider to allow the bicycle to continue moving forwards. The restraining forces are composed of the control by the rider, gradient, wind, frictional and rolling energy losses. The forces vary with the terrain and with the type and condition of the bicycle. Factors are indicated which may influence the stability of a bicycle during steering.

2. NOTATION

L	Lagrangian	P	vector moment of momentum
T	kinetic energy	I	moment of inertia of front wheel, in axial direction
V	potential energy	θ	ratio of lateral to axial moments of inertia of front wheel
N	Torque vector	ω	angular velocity of front wheel as a vector
q	Generalised co-ordinates	$\omega_{z'}, \omega_z, \omega_x, \omega_y$	components of angular velocity
x,y,z	Spatial co-ordinates		
M	vector moment	k'	vectorial direction of original rotation of bicycle wheel
r	radius of bicycle wheel	Ω	vectorial angular velocity of handlebars, front fork and rider

---

[1] 30 Milford Street, Randwick NSW 2031

3.   ENVIRONMENT

A bicycle is constrained vertically by the road surface, and laterally by the restrictions of the available surface. A bicycle therefore moves along the line of intersection of two curved spaces, and is a system having one degree of freedom. An on-road bicycle is restricted by the road pavement, and an off-road bicycle is restricted by the intended course. Random imbalances occur due to unevenness in the surface. The rider corrects the imbalances by turning the handlebar in order to counteract the overturning moment caused by the imbalance. Leaning the bicycle, in the opposite direction to the momentary imbalance, is a valid correction to the imbalance. However, the bicycle and rider respond more quickly to turning the handlebar than to leaning.

4.   DYNAMICS OF THE SYSTEM

The momenta of the bicycle frame and rider depend on their mass and velocity only: the momenta of the rotating parts depend in addition on the angular velocities of rotation of each part. The corrective torque applied to the handlebar depends upon the momentum of the front wheel primarily, but also on the momenta of the rear wheel, the rotating pedals and the chain wheel. Therefore not only the responsiveness of a bicycle depends on the momentum and therefore on the mass of the rotating parts, but also the power supplied by the rider depends on the kinetic energy and therefore on the mass of the rotating parts: kinetic energy lost due to impact with road obstructions and irregularities in the surface is replaced by the rider. This energy is proportional to the mass of the rotating parts. This explains the cyclists' rule, based on experience, that a saving in rotating mass is more to be desired than a saving in the mass of the frame, which undergoes only a forward, non-rotational, movement.

In executing an intended turn, experienced cyclists tend to reverse the turning operation: they first lean the bicycle, then turn the handlebar in order to compensate for the overturning moment caused by the leaning. Once a turn is commenced, minor variations in direction caused by the road-surface irregularities or obstructions can be corrected by turning the handlebar, as in normal straight progression. There is a resistance to turning the handlebar. This resistance is provided almost entirely by the momentum of the rotating front wheel. It is proportional to the moment of inertia of the front wheel, and to the square of its angular velocity. For this reason, cyclists attach more importance to reducing the mass of all rotating parts especially that of the front wheel, than to reducing the mass of the bicycle and rider.

Gallavotti [2] gives expressions by means of which the angular velocity in any one direction of a gyroscopic mass can be expressed in terms of the angular velocties in the remaining two directions.

$$\left.\begin{array}{l} I\omega_z^2 + \theta\, I\omega_x^2 + \theta\, I\omega_y^2 = 2T = \text{constant} \\[2mm] I^2\omega_z^2 + \theta^2 I^2\omega_x^2 + \theta^2 I^2\omega_y^2 = \left[\mathbf{Iw}\right]^2 = \text{constant} \end{array}\right\} \tag{1}$$

We take the moment of inertia I to be that of the bicycle and rider in the transverse direction, the z-direction, in the horizontal direction at right angles to the direction of movement of the bicycle.

$$I_1\dot{\omega}_1 - \omega_2\omega_3(I_2 - I_3) = N_1,$$
$$I_2\dot{\omega}_2 - \omega_3\omega_1(I_3 - I_1) = N_2,$$
$$I_3\dot{\omega}_3 - \omega_1\omega_2(I_1 - I_2) = N_3.$$

In the absence of torque applied to the front wheel, these equations can be written

$$I_1\dot{\omega}_1 = \omega_2\omega_3(I_2 - I_3),$$
$$I_2\dot{\omega}_2 = \omega_3\omega_1(I_3 - I_1),$$
$$I_3\dot{\omega}_3 = \omega_1\omega_2(I_1 - I_2).$$

These equations relate the angular acceleration of overturning the bicycle to the angular accelerations of the front wheel and the handlebar rotation. The equations are integrable, although the result is in terms of elliptic functions and is not illuminating. A graphical construction exists(Norwood [3] and Goldstein [4]) and shows the behaviour of the inertia ellipsoid relative to the invariable plane.

The expression for the relationship between the applied moment and the parameters of a gyroscopic system is as follows

$$\mathbf{M} = \frac{d\mathbf{P}}{dt} = \frac{\mathbf{I}}{\theta}\frac{d\omega}{dt} + \frac{\mathbf{I}}{\theta}(\theta-1)\frac{d\omega_{z'}}{dt}\mathbf{k'} + \frac{\mathbf{I}}{\theta}(\theta-1)\omega_{z'}\frac{d\mathbf{k'}}{dt}$$

A useful form of this expression is given by Joos [5], as follows

$$\mathbf{M} = \frac{I}{\theta}\frac{d\Omega_{n'}}{dt} + I\omega_{z'}\cdot\frac{d\mathbf{k'}}{dt} = \frac{I}{\theta}\frac{d\Omega_{n'}}{dt} + I\omega_{z'}[\Omega_{n'}\mathbf{k'}]$$

The main part of the angular momentum is $I\omega_{z'}$, denoted by $\overline{P}$. The term $\frac{I}{\theta}\frac{d\Omega_{n'}}{dt}$ is the effective moment due to the movement of the bicycle frame and rider. It is negligible under the initial conditions, i.e. before the bicycle has started to overturn. The moment which causes the bicycle to overturn is then given by Joos [5] as follows

$$\mathbf{M} = \overline{P}\frac{d\mathbf{k'}}{dt} = \overline{P}[\Omega\mathbf{k'}] \tag{2}$$

The two components of the overturning moment are directly proportional to (i) the axial angular velocity of the front wheel and (ii) the angular velocity of the rotation of the handlebar. Assume that the front wheel has a mass equal to one kilogramme. The moment of inertia of a standard bicycle wheel having a diameter of 700mm, and of 1 kg mass, is 1.0r^2 = 1.0 (0.35)^2 = 0. 1 kg.-m^2. The overturning moment per unit of 1.0 kg.-m^2 of inertia is shown in Figure 1.

Karnopp [6] suggests that the inverse of Equation (1) can sometimes be obtained, if the system is sufficiently simple. The inverse yields the rates of change of the momentum vectors needed to cause a given overturning moment.
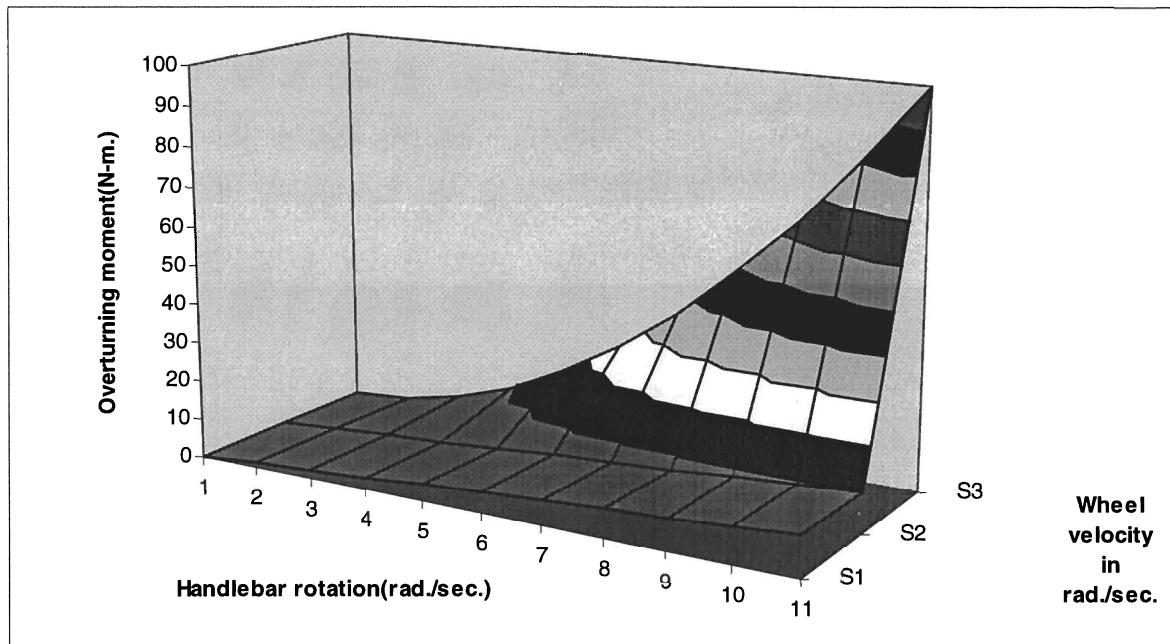
Figure 1: Overturning moment due to handlebar rotation

In Figure 1, the maximum wheel velocity is 10 radians per second. For wheels having a diameter of 700mm., this corresponds with a road speed of 7 m.sec, or 25.2 km/hr.

REFERENCES

[1]   Cogill W.H., The mathematics of bicycling, *Proceedings of 2nd Conference on Mathematics and Computers in Sport*, Bond University, 37-42, (1994).

[2]   Gallavotti G., Elements of mechanics, (Springer, New York) 311, (c1983).

[3]   Norwood J., Classical Mechanics, (Prentice-Hall, Englewood Cliffs, N.J.) 192, Fig. 5.10 (1979).

[4]   Goldstein H., Classical mechanics, Second Edition (Addison-Wesley, Reading, Massachusetts) 206, (1980).

[5]   Joos G., Theoretical physics, (Blackie, London) equation (48) 57, (1950).

[6]   Karnopp D., On inverse equations for vehicle dynamics, *Vehicle System Dynamics*, Vol 20, 371-379 (1991).

OPTIMIZING THE SHOT PUT

Neville J. de Mestre[1], Mont Hubbard[2] and John Scott[3]

Abstract

When a projectile is released from a height above the impact plane, the range depends on this height, the release speed and the release angle. Experiments were carried out on three shot-putters to seek a relationship between these three initial parameters for the trajectory.

Using this relationship as a constraint, the optimal conditions of release can be determined which will produce a maximum value for the range for any particular athlete.

1. INTRODUCTION

Film studies by Cureton [1] and Dessurealt [2] indicate that elite shot putters release the shot at an angle to the horizontal somewhat less than 45°. This is partly because the level of release is approximately 2m higher than the level of impact. There have been a number of articles (Lichtenberg and Wills [3], Trowbridge and Paish [4], Burghes, Huntley and McDonald [5], Townend [6]) which have analysed the situation, but all the models derive a formula for the optimum angle of release based on an assumed independence between the angle of release, the height of release and the speed of release. This assumption does not seem to be physically realistic since it appears to be easier to use the shoulder and arm muscles to project the shot horizontally than vertically.

Experiments were carried out at Davis, California with three college athletes (two of whom were of high standard). From the data, possible relationships between the release conditions were considered. The most appropriate is used as a constraint relation for calculating the angle of release for maximum range. To our knowledge the only investigations carried out previously to find a constraint relation were conducted by Red and Zogaib [7] using a 1.14 kg ball, and Karo [8] using a regular shot. Red and Zogaib found that the release speed decreased linearly with angle, while Karo's experiments were inconclusive. Neither included height of release in their respective relationships.

[1] School of Information Technology, Bond University, Gold Coast Qld 4229
[2] School of Mechanical and Aeronautical Engineering, University of California, Davis, USA
[3] School of Management Systems, Waikato University, New Zealand

2   SHOT-PUT TRAJECTORY EQUATIONS

Consider a shot of mass m and diameter D released from an athlete's hand at a height h above the athletic field with release speed v_0 and release angle α. The point of release is chosen as the origin of a 2-dimensional co-ordinate system with 0x in the horizontal direction of the throw and 0y vertically upwards (Figure 1).
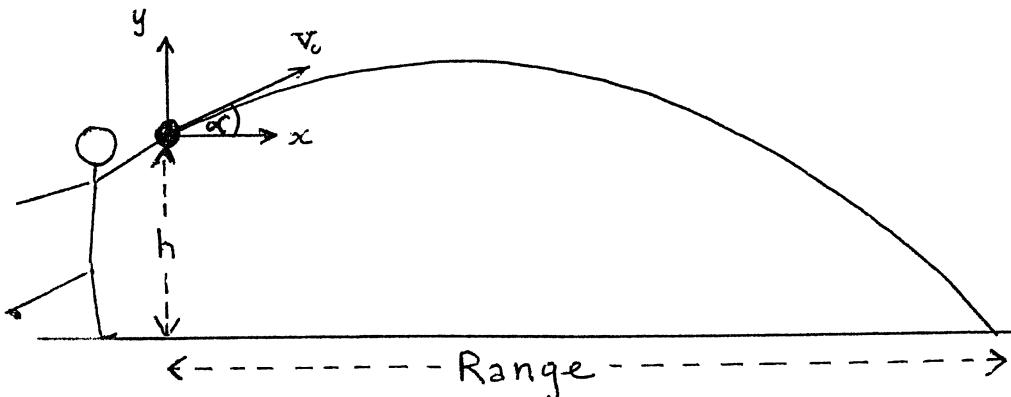
Figure 1: Shot put geometry

The equations of motion for the trajectory of the shot are

$$m\ddot{x} = -\frac{1}{2}\rho A C_D \dot{x}\left(\dot{x}^2 + \dot{y}^2\right)^{\frac{1}{2}} \tag{1}$$

$$m\ddot{y} = -\frac{1}{2}\rho A C_D \dot{y}\left(\dot{x}^2 + \dot{y}^2\right)^{\frac{1}{2}} - mg \tag{2}$$

where ρ is the density of air, A is the cross-sectional area of the shot, C_D is the drag coefficient, g is the acceleration due to gravity and a dot denotes differentiation with respect to time t. For a shot travelling at its usual speeds along its trajectory the drag coefficient can be taken to be a constant. Nevertheless these coupled differential equations are still non-linear and difficult to solve analytically. A solution for the speed $v = \left(\dot{x}^2 + \dot{y}^2\right)^{\frac{1}{2}}$ of the shot can be obtained as a function of its tangent angle $\arctan\left(\dot{y}/\dot{x}\right)$ at any position on the trajectory, but explicit expressions for x and y as functions of t can only be obtained as integrals which have to evaluated numerically over definite limits.

The initial conditions for the problem are that

$$x = 0, \ y = 0, \ \dot{x} = v_0\cos\alpha, \ \dot{y} = v_0\sin\alpha \tag{3}$$

when t = 0. The problem for the athlete is to generate a set of initial values for v_0, α and h so that x is a maximum when y = -h.

Non-dimensionalisation of this differential system using $x = v_0^2 X/g$, $y = v_0^2 Y/g$, $t = v_0 T/g$ yields

$$\ddot{X} = -\varepsilon \dot{X}\left(\dot{X}^2 + \dot{Y}^2\right)^{\frac{1}{2}} \tag{4}$$

$$\ddot{Y} = -\varepsilon \dot{Y}\left(\dot{X}^2 + \dot{Y}^2\right)^{\frac{1}{2}} - 1 \tag{5}$$

with

$$X = Y = 0, \quad \dot{X} = \cos \alpha, \quad \dot{Y} = \sin \alpha \quad \text{when } T = 0 \tag{6}$$

and the drag-to-weight ratio

$$\varepsilon = \rho A C_D v_0^2 / (2mg) \tag{7}$$

where a dot from here on denotes differentiation with respect to T.

3.   EXPERIMENTS

Three shot putters from the University of California (Davis) track-and-field team threw the same shot for 49 different experiments. One athlete (Allan Babayan) was in the top twelve at National level, one (Joey Taylor) was at College level and the third (Jeff Blakefield) competed regularly within the university at local meets. All three used the new rotary action and released the shot without stepping over the foul board. Babayan made 22 measured throws, Taylor 14 and Blakefield 13. Each athlete was asked to produce a set of relatively-high initial-angle throws, normal throws and relatively-low initial-angle throws.

An Expert Vision Motion Analyser was used to determine the initial angle of release, speed of release and height of release for each throw.

The distance of each throw was measured by tape from the estimated position of release. The results are given in Tables 1, 2 and 3, with results presented in increasing order of release height.

Table 1

Throwing data for Allan Babayan

Release height h (m)	Release speed v_0 (ms^{-1})	Release angle α (degrees)	Type of release aimed for	Calculated throwing distance (m)	Measured throwing distance (m)
1.847	11.724	44.369	Normal	15.342	15.34
1.849	11.815	44.872	Low	15.560	15.56
1.862	11.760	42.274	Low	15.550	15.55
1.865	11.805	35.567	Low	15.645	15.65
1.870	11.948	36.664	Low	16.029	16.03
1.893	11.609	37.828	Low	15.331	15.33
1.907	11.986	33.666	Low	16.001	16.00
1.911	11.830	44.517	High	15.774	15.77
1.927	11.667	45.002	Normal	15.297	15.30
1.942	11.615	43.340	Low	15.387	15.39
1.969	11.647	41.132	Normal	15.586	15.59
1.987	11.215	45.460	High	14.408	14.41
2.000	11.221	43.118	High	14.635	14.63
2.005	11.381	46.441	High	14.661	14.66
2.028	11.431	44.988	High	14.991	14.99
2.040	11.837	37.740	Low	16.210	16.21
2.042	11.681	39.598	Normal	15.892	15.89
2.074	11.633	45.491	High	15.483	15.48
2.082	11.102	44.867	High	14.345	14.35
2.087	11.576	38.895	Normal	15.681	15.68
2.103	11.890	38.027	Low	16.480	16.48
2.134	11.591	43.460	Normal	15.607	15.61

Table 2

Throwing data for Jeff Blakefield

Release height h (m)	Release speed v_0 (ms^{-1})	Release angle α (degrees)	Type of release aimed for	Calculated throwing distance (m)	Measured throwing distance (m)
1.811	10.925	31.272	Low	12.770	12.77
1.816	11.245	34.740	Low	13.858	13.86
1.862	10.904	35.689	Low	13.257	13.26
1.955	9.926	48.078	Normal	11.103	11.10
2.002	10.657	41.746	Normal	13.013	13.01
2.009	9.876	39.724	Low	11.395	11.39
2.016	10.913	37.760	Low	13.640	13.64
2.060	10.611	43.356	Normal	12.923	12.92
2.197	10.215	41.163	Normal	12.308	12.31
2.224	10.049	47.447	High	11.580	11.58
2.288	10.099	43.770	High	11.978	11.98
2.371	10.298	48.210	High	12.151	12.15
2.375	9.710	48.410	High	10.911	10.91

Table 3

Throwing data for Joey Taylor

Release height h (m)	Release speed v_0 (ms^{-1})	Release angle α (degrees)	Type of release aimed for	Calculated throwing distance (m)	Measured throwing distance (m)
1.973	11.794	30.270	Low	15.125	15.12
2.025	11.393	40.852	High	14.882	14.88
2.070	11.645	29.868	Low	15.081	15.08
2.083	11.462	30.273	Low	14.675	14.68
2.104	11.139	35.526	Normal	14.387	14.39
2.119	11.335	34.953	Normal	14.863	14.86
2.122	11.137	42.499	High	14.372	14.37
2.125	11.244	42.891	High	14.824	14.82
2.145	11.064	38.529	Normal	14.378	14.38
2.189	10.988	37.454	High	14.224	14.22
2.209	11.035	43.898	High	14.442	14.44
2.210	11.411	34.366	Normal	15.162	15.16
2.213	11.396	30.049	Low	14.783	14.78
2.219	11.607	34.575	Normal	15.743	15.74

The mean throwing distance achieved by Allan Babayan was 15.45m, by Jeff Blakefield was 12.38m and by Joey Taylor was 14.78m. Clearly Jeff Blakefield was not able to match the release speeds of the other two throwers, although he threw over a similar range of angles to Allan Babayan. Joey Taylor was able to throw with release speeds that were just less than Allan Babayan's, but over a smaller range of angles.

The type of release angle aimed for was nominated in advance. The correlation between this and the actual release angle achieved was very poor for Babayan and to a lesser extent Blakefield who seemed not able to differentiate between normal and high throws. On the other hand Taylor was excellent in throwing at the nominated type of release angle, except for one throw.

Taylor was the only athlete whose best throw was a nominated normal angle of release. Blakefield's three longest throws and Babayan's four longest throws were all achieved when they deliberately tried to throw with a lower-than-normal release angle.

4.   OPTIMIZATION WITHOUT AIR DRAG

The shot used in the experiments had a mass (m) of 7.328kg and a circumference of 0.397m. Air density at Davis on the day of the experiment was 1.226 kg m^{-3} with g = 9.7935 ms^{-2}. For a spherical shot the drag coefficient is C_D = 0.45 for speeds in the range of the shot put. With a maximum speed of 12 ms^{-1} (almost achieved by Babayan) the value of ε = 0.006942 $\approx$ 0.007, hence ε is very small for all experimental throws.

Therefore X and Y can be expanded in a perturbation series as follows

$$\left. \begin{array}{l} X = X_0 + \varepsilon X_1 + 0\!\left(\varepsilon^2\right) \\ Y = Y_0 + \varepsilon Y_1 + 0\!\left(\varepsilon^2\right) \end{array} \right\} \tag{8}$$

The zeroth-order approximations to equations (4)-(6) are

$$\ddot{X}_0 = 0$$

$$\ddot{Y}_0 = -1$$

with $X_0 = Y_0 = 0$, $\dot{X}_0 = \cos\alpha$, $\dot{Y}_0 = \sin\alpha$ when T = 0. These show that air drag is not included in the zeroth-order approximations, and equations (8) indicate that it will have less than 1% effect on the range.

The solutions are

$$X_0 = T\cos\alpha \tag{9}$$

$$Y_0 = T\sin\alpha - \tfrac{1}{2}T^2 \tag{10}$$

Now $y = -h$ transforms to $Y = -H$ with $h = v_0^2 H/g$ and so the zeroth-order time of flight for the shot is obtained from equation (10) as

$$T_0^{(f)} = \sin\alpha + \left(2H + \sin^2\alpha\right)^{\frac{1}{2}}$$

The corresponding range is obtained from equation (9) as

$$X_0^{(f)} = \cos\alpha\left[\sin\alpha + \left(2H + \sin^2\alpha\right)^{\frac{1}{2}}\right] \tag{11}$$

To throw the shot, each athlete used a rotary motion starting in a slightly bent position with the shot resting on his hand which itself was placed at the junction of his neck and shoulder. The athlete then moved across the throwing circle using $1\frac{1}{2}$ rotations about a vertical axis superimposed on vertical and horizontal velocity components. These raised him upwards and forwards to the edge of the throwing circle. During this time his arm extended during the last $\frac{1}{2}$ revolution to its fullest extent and then released the shot. Before release, the shot itself travelled in an oblique upwardly-spiralling curve.

If h_0 is the height of the throwing shoulder at release, L is the length of the throwing arm, and θ is the angle of the extended throwing arm to the horizontal, it is clear (Figure 2) that
$$h = h_0 + L\sin\theta$$

Note that h_0 and L can be considered as constant for each throw by a specific athlete, and that the release height h only changes because θ changes.
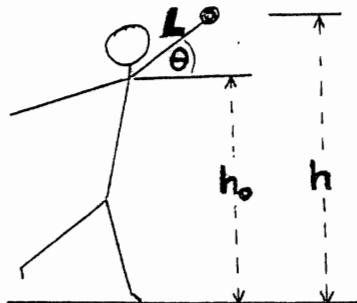

Figure 2: Release height geometry

Now the angle of release α would be the same as the angle of the arm θ if the athlete threw with his shoulder stationary. Based on this observation a simple approximation to the relationship between α and H $(= gh/v_0^2)$ would be

$$H = K + B\sin\alpha \tag{12}$$

A least-squares fit of the data for each athlete produces the following values:

Babayan	,	K = -0.1844	,	B = 0.4934
Blakefield	,	K = -0.1506	,	B = 0.5141
Taylor	,	K = 0.0400	,	B = 0.2700

With equation (12) substituted into equation (11) the non-dimensional range is given by

$$X_0^{(f)} = \cos\alpha\left[\sin\alpha + \left(2K + 2B\sin\alpha + \sin^2\alpha\right)^{\frac{1}{2}}\right]$$

Differentiation with respect to α yields

$$\frac{dX_0^{(f)}}{d\alpha} = 1 - 2\sin^2\alpha + \frac{\left\{B + (1 - 2K)\sin\alpha - 3B\sin^2\alpha - 2\sin^3\alpha\right\}}{\left(2K + 2B\sin\alpha + \sin^2\alpha\right)^{\frac{1}{2}}}$$

and so the critical point for optimum range is given by the solution by

$$(2\sin^2\alpha - 1)(2K + 2B\sin\alpha + \sin^2\alpha)^{\frac{1}{2}} = B + (1 - 2K)\sin\alpha - 3B\sin^2\alpha - 2\sin^3\alpha$$

This equation was solved by Mathematica for a numerical solution near $\sin\alpha = 0.6$ using the values of K and B for each athlete. The results are given in Table 4.

Table 4 Optimum Release Characteristics

Athlete	Optimum H	Optimum α (degrees)	Distance for max v_0(m)
Babayan	0.1656	45.18	16.78
Blakefield	0.0294	42.99	15.32
Taylor	0.1881	44.46	16.54

It has not been established that the maximum achieved throw is possible for the optimum conditions, but the results seem to indicate that the athletes are capable of extending their personal best throws for the experiments.

Further work is needed to establish a more complicated constraint relation similar to equation (12), and this should be based on the dynamics of the shot put throw before release.

ACKNOWLEDGMENTS

REFERENCES

[1]   Cureton T.K., Elementary Principles and Techniques of Cinematographic Analysis, *Res. Quarterly* 10, 3 (1939).

[2]   Dessurealt Jacques, *Selected Kinetic and Kinematic Factors Involved in Shot Putting*, Ph.D. thesis, Indiana University, (1976).

[3]   Lichtenberg D.B. and Wills J.G., Maximising the range of the shot put, *Am. J. Phys.* 46, 546-549 (1978).

[4]   Trowbridge E.A. and Paish W., Mechanics of Athletics, *The Inst. of Maths and its Applic.*, 17, 144-146 (1981).

[5]   Burghes D.N., Huntley I. and McDonald J., Applying Mathematics, Ellis Horwood, Chichester, (1982).

[6]   Townend M.S., Mathematics in Sport, Ellis Horwood, New York, (1984).

[7]   Red W.E. and Zogaib A.J., Javelin dynamics including body interaction, *J. Appl. Mech.* 44, 496-498 (1977).

[8]   Karo C., *Putting a shot, what angle to throw at*, Report No 17, Maths Department, Australian Defence Force Academy, (1989).