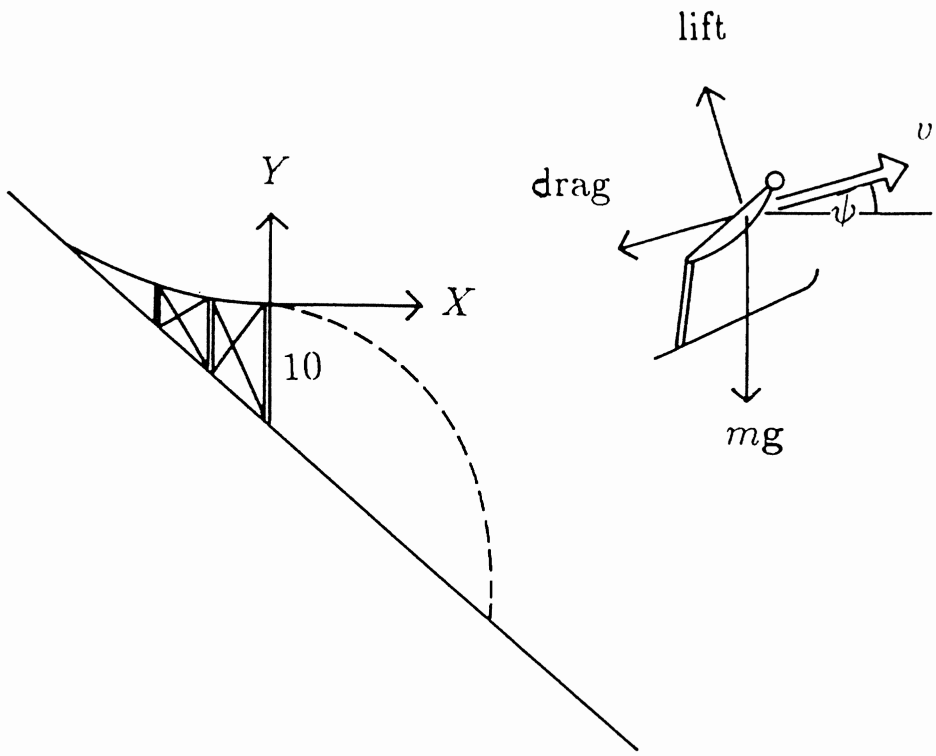
Second Conference on:

MATHEMATICS AND COMPUTERS IN SPORT



Held at Bond University, Queensland, Australia
11th to 13th July, 1994

Sponsored jointly by

The Australian Mathematical Society

and

The Australian Sports Commission

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

CONTENTS (Vol 2)

GENERAL

Welcome to the second conference held in Australia on the use of the mathematics and computers in sport. As with the first conference its aim is to bring together researchers in these aspects of sports science.

There have been many enquiries about this conference and also many individuals who could not make the first conference have written to obtain copies of the initial proceedings. As I am expecting similar enquiries about the second conference proceedings I should state at the outset that I will have extra copies of the second conference proceedings printed for sale to those who could not attend.

This conference has received more papers than the previous ones and there are slightly more participants. However only six of the attendees to the first conference of two years ago have returned — an interesting point.

I would like to thank the four invited speakers — Mont Hubbard (University of California, Davis, USA), Stephen Clarke (Swinburne University of Technology), John Scott (University of Waikato, New Zealand) and Maurie Brearley (Emeritus Professor, University of Melbourne) for agreeing to participate in the conference and lead discussions. To the other speakers and attendees I am also grateful for their contributions towards making the conference successful.

Finally, I must place on record my appreciation for the help given by my secretary, Jeanette Niehus, for her invaluable and cheerful assistance in preparing this booklet for publication and in assisting me with other organisational matters.

Neville de Mestre
July 1994

PROGRAM

Monday 11th July

8.45 am	Opening.
9.00 am	Mont Hubbard (University of California) – "Simulating sensitive dynamic control of a bobsled".
10.00 am	Hugh Morton (Massey, N.Z.) – "The relationship between power demand and endurance time".
10.30 am	*Morning Tea.*
11.00 am	Geoffrey Paull (Curtin) – "Interactive video simulation for research".
11.30 am	John Cogill (UNSW) – "The mathematics of bicycling".
12.00 noon	Stephen Clarke (Swinburne) – "An adjustive rating system for tennis and squash players".
12.30 pm	*Lunch*
2.00 pm	Rod Barrett (Griffith–Gold Coast) – "A method for determining sand-running efficiency".
2.30 pm	David Lamble (Swinburne) – "The dynamics of the lawn bowl revisited".

Tuesday 12th July

9.00 am	Stephen Clarke (Swinburne) – "Variability of scores and consistency in sport".
10.00 am	Pam Norton (Monash) – "Softball statistics".
10.30 am	*Morning Tea.*
11.00 am	Selvanayagam Ganesalingam (Massey, N.Z.) – "A statistical look at cricket data".
11.30 am	Rod Weber (ADFA) – "Finals draws".
12.00 noon	Richard Monypenny (James Cook) – "Useful systems references".
12.30 pm	*Lunch.*
2.00 pm	Deborah Hoare (Aust Sports Commission) – "Sports search".
2.30 pm	Rick Baker (Hi-Tech Video) – Video presentation.
3.00 pm	John Baker – Tour of Golf Laboratory.

Wednesday 13th July

9.00 am	Maurie Brearley (Clifton Springs) – "Driving a ball with a golf club".
9.30 am	John Scott (Waikato, N.Z.) – "Direct golf putting dynamics and strategies".
10.00 am	John Baker (Bond) – "Towards a preferred golf swing and its application to teaching golf".
10.30 am	*Morning Tea.*
11.00 am	Heather Whitford (Flinders) – "Golf scores and exploratory data analysis".
11.30 am	Tony Whitford (South Australia) – "The Australian golf handicapping system".
12.00 noon	Ian Collings (Deakin) – "Low-angle golf trajectories".
12.30 pm	*Close and lunch.*

PARTICIPANTS

Mr John BAKER	(Bond)
Mr Ric BAKER	(Hi-Tech Video)
Mr Rod BARRETT	(Griffith–Gold Coast)
Prof Maurie BREARLEY	(Clifton Springs)
Assoc Prof Stephen CLARKE	(Swinburne)
Dr John COGILL	(UNSW)
Dr Ian COLLINGS	(Deakin)
Mr John CORMACK	(Charles Sturt–Mitchell)
Mr Noel COVILL	(QUT)
Assoc Prof Neville DE MESTRE	(Bond)
Dr Selvanayagam GANESALINGAM	(Massey)
Mr Trevor GILHAM	(Guilford Grammar)
Prof Brian GRAY	(Sydney)
Assoc Prof Chris HARMAN	(Southern Queensland)
Dr Deborah HOARE	(Aust Sports Commission)
Prof Mont HUBBARD	(California, Davis)
Dr Kuldeep KUMAR	(Bond)
Mr David LAMBLE	(Swinburne)
Mr Michael LYNCH	(QUT)
Dr Richard MONYPENNY	(James Cook)
Assoc Prof Hugh MORTON	(Massey)
Dr Pam NORTON	(Monash)
Mr Geoffrey PAULL	(Curtin)
Assoc Prof John SCOTT	(Waikato)
Dr Rod WEBER	(ADFA)
Mr Bernie WILKES	(Charles Sturt–Mitchell)
Dr Heather WHITFORD	(Flinders)
Dr Tony WHITFORD	(South Australia)

SIMULATING SENSITIVE DYNAMIC CONTROL OF A BOBSLED

M. Hubbard[1]

Abstract

A simulator can be a cost-efficient method of teaching a sensitive dynamic skill, that of driving a bobsled. Mathematical descriptions of the three-dimensional track surface, differential equations which describe the motion of the sled on the surface, and dynamic optimization techniques for derivation of optimal trajectories as solutions of two-point boundary-value problems, are the main areas of mathematics which form the theoretical basis of the simulator. Considerable physiology and engineering underlie the design and mechanization of hardware to provide the visual, tactile, vestibular and auditory feedback to maximize the realism of the system. Quantitative feedback of variables for performance improvement is possible since all variables are computed in real time.

1. INTRODUCTION

Bobsledding is a notoriously expensive sport, both in the design and construction of world class tracks (which can cost in the range of $15-30 million), in the design, fabrication and testing of competitive vehicles (typically >$25,000), and in the requirements for costly international travel to the relatively few tracks in existence. All but one (Calgary) of the tracks presently certified by the FIBT for World Cup competitions are in Europe, although there will soon be a track constructed for the 1998 Winter Olympic Games in Nagano, Japan. Thus travel expense for a team becomes a burden piled on top of the cost of competitive sleds. These are all financial incentives for the use of a simulator, similar in design to aircraft flight simulators, to teach the intricacies of bobsled driving.

2. MATHEMATICAL MODELS

Although there have been several useful one-dimensional studies of bobsled dynamics and sled performance [1,2], only recently has a model been developed [3,4,5] which includes the fully three-dimensional shape of the track surface and the possibility of free sled motion on this surface. This general track surface model is based on bi-cubic splines [6] and, using actual track shape data, can generate a surface approximation whose deviation from the actual surface is of the order of 0.004 m, certainly considerably more accurately than the position of the actual ice surface is known.

Since a body moving on a two-dimensional surface in three space has one rotational and two translational degrees of freedom, three differential equation are required to

---

[1] Department of Mechanical and Aeronautical Engineering, University of California, Davis, California, 95616 USA

describe the motion [7]. These equations contain the general track shape through the surface spline coefficients, various vehicle parameters (mass, runner-ice coefficients of friction, moments of inertia, aerodynamic drag and lift coefficients, etc.), as well as the time-varying steering profiles executed by the driver in his or her attempt to control the vehicle.

Controlling a bobsled is an extremely sensitive task since only the front runners are steerable, the drag coefficient of friction m of the steel runners on ice is only approximately 0.015, and the lateral steering coefficient of friction [2] is about five times as large. The runner steering angle about the track normal direction (typically considerably less that 5 deg), which creates a lateral force at the front runners to accelerate the vehicle from side to side in the tangent plane of the track, also creates a moment about the track normal which causes an angular acceleration of the vehicle in heading about this axis. Thus the competing tasks of lateral position and heading control must be accomplished with the single control input, much akin to the simultaneous tracking and balancing tasks executed by a bicycle rider. Furthermore, the difficulty occasioned by this inherent coupling of control tasks is compounded by the fact that the control forces generated by turning the runners are so small.

The measure of performance in the bobsled is elapsed time from start to finish. After the pushers generate as large an initial velocity as possible, the driver's task is to control the sled so as to minimize the distance travelled while simultaneously minimizing the additional steering-induced ice friction and thus preserving as large a speed as possible for the remainder of the run.

This is a classical optimal control problem [8]. The state equations become the motion differential equations, control constraints are given by limitations on the forces generated as a function of steering angle, and the performance index is the total elapsed time. Classical numerical optimal control techniques [8] can therefore be applied to calculate the optimal driver actions required to minimize the terminal time.

Finally, since the driver begins controlling the vehicle only after the push phase of the event is completed, a valid and accurate model of the push process is also required so that the optimal control problem can be provided with realistic initial conditions (speed at the 50 m mark as a function of velocity). At this juncture it appears that a functional model for the bobsled acceleration produced by the pushers which is similar to that described for sprinters by Vaughn [9] is adequate.

At every stage of model development, it is important to incorporate actual data as often and as completely as possible. For example, the vehicle parameters used in the simulation come from wind tunnel tests, ice friction studies, vehicle mass and inertia measurements, as well as the measurements of track shapes. In addition, after the simulation is producing meaningful results (e.g. split times and velocities on a certain run), comparing these with actual measurements can help to validate the model and its computations and to identify areas where the model might be improved. This inclusion of experimental data and testing of the model against experiment cannot be overemphasized, since without these steps the model can rapidly lose touch with reality, and generate numerical results which have little basis in actuality.

3. REAL TIME SIMULATION

A bobsled simulator is an electro-mechanical system designed and constructed in an attempt to mimic the actual experience as closely as possible. The human interacts with the bobsled environment via controlling actions taken (steering) and via sensations received by his or her sensory systems. A realistic simulator computes where it is commanded to go and provides the driver with sensations as similar to those perceived in the actual task as possible. It is important to involve as many of the human's sensory subsystems as possible in the experience in order to heighten the feeling of realism. The four senses which are addressed by the bobsled simulator hardware and software are the visual, tactile, vestibular and auditory. Two of these (visual and vestibular) are the most important in generating the perception of motion.

The heart of the simulator is a high speed graphics workstation and the differential equations which model sled dynamics on the track. The driver's steering input is measured using an accurate optical encoder and fed directly into the computer to affect the solution of the equations of motion. These are solved in real time at approximately 100 Hz. At a somewhat lower rate, a continuously changing visual image is generated which would be seen by the driver from her correspondingly continuously changing track position. Visual feedback accounts for perhaps 60-70% of the perception of motion, but the angular motion of the driver's otolith organs is also important. Even though actual sleds experience angular accelerations in all three directions, the roll component is by far the largest and most violent, with angular accelerations frequently exceeding 60 rad/sec^2. Thus, in order to mimic this most important motion cue a high bandwidth roll controller for the simulator cockpit is essential. This is provided by a DC motor and associated gearing.

In addition to those sensations provided by the visual and vestibular systems, the "feel" of the steering system is important as well. Calculations are made continuously of the forces on the runners from the ice, and the forces which these cause through the steering linkage on the drivers' hands. These hand forces are then produced through an active electro-mechanical control system in the steering linkage so that the steering tactile sensations of the driver are realistic.

Track surfaces are rarely smooth. The interaction of surface irregularities with the runners produces an extremely severe vibratory environment, including substantial auditory noise. A final touch of realism in the simulator is provided by playing recorded acoustic noise from actual track runs.

Although every effort is made to have the sensory feedback as accurate and realistic as possible, it is important to realize the limits of simulation. For example, the actual bobsled event involves extremely high specific forces. Indeed, rules for track design [10] prohibit the imposition of more than 5 g's, and for a period not to exceed 2 seconds. Nevertheless, it is clearly impossible to simulate these high g's without traversing the same trajectory in inertial space as the actual track itself. Thus the specific forces in the simulator are by necessity limited to 1 g.

4. QUANTITATIVE FEEDBACK

Because everything must be computed in the real-time numerical solution of the differential equations, virtually every variable of interest can be displayed to the driver after the run for the purposes of learning what went right and what went wrong. This is clearly a great advantage of the simulator over even the actual experience, since similar information about the relevant variables during an actual run could be achieved only with exorbitant cost and effort. This immediate quantitative feedback to the driver is extremely important in the evolution and improvement of driving technique.

A great deal of effort has been made in the software development to make this information transfer as useful and simple as possible. For example, graphic plotting of post-run variables (e.g. speed vs distance) is crucial in making the information more quickly assimilable. In addition, certain information (e.g. the previous path) can be provided graphically during the next simulation for purposes of real-time comparison with the present path. In essence, the simulator becomes a relatively inexpensive measurement device with extremely high repeatability and accuracy.

Repetition of a particular run is also possible. On a real track, because of competition for starting slots and logistical concerns, bobsled drivers are sometimes lucky to achieve four runs, each of one minute duration, in a single day. In the simulator, however, it is easy to get roughly a factor of ten or twenty more. Another obvious advantage of the simulator is that of safety. It is possible to experiment with driving strategies which might be too risky and dangerous to try in practice.

5. USE OF THE SIMULATOR IN TRAINING

A fixed-base version of the bobsled simulator (in which only visual feedback was provided) was completed in time for use by the US Bobsled Team in preparation for the 1992 Winter Olympic Games in Albertville, France. Since then, enhancement of the hardware and software has resulted in the addition of a motion base and other improvements. This second generation simulator was utilized in January 1994 in preparation for the Lillehammer Olympic Games and will soon be installed permanently at the headquarters of the US Bobsled Federation in Lake Placid, New York. Although the simulated experience can never completely replace actual sledding, elite and neophyte drivers alike are uniform in their praise for its high degree of realism, its substantial efficacy in capturing the essential features of the driving task, and its effectiveness in increasing their familiarity with particular track layouts.

In Section 2 above, we have already emphasized that the use of experimental data in the model building and validation stages is crucial. Ideally it would be desirable to be able to prove that using the simulator improves a bobsledder's driving skills by measurement of this improvement when transferred to actual training or competitions in real sleds. Unfortunately, it is nearly impossible to test experimentally the effectiveness of the simulator in a controlled manner on real tracks. The presence of many uncontrollable factors make this prohibitively difficult.

Driving effectiveness is only one, and not even clearly the most important, of the three major determinants of finish time in actual competitions. The two other important factors are push effectiveness (i.e. velocity and time at the end of the push phase), and sled aerodynamic and frictional efficiency, both of which are completely unrelated to driving skill.

Additionally, the driver's steering angle is a function of time, and a characterization of its "goodness," other than as measured by the ultimate finish time, is not immediately obvious. But there may be cases where the finish time does not accurately measure the "goodness" of an entire driving run. For example, a single driving error early in the course, where the sensitivity to errors on the final time is highest, can completely overwhelm a sterling effort throughout the remainder of the course.

We have seen that uncontrollable factors may limit our ability to measure the improvements in driver performance in actual practice. On the other hand, when using the simulator it is possible to control completely for the effects of these factors mentioned above which are uncontrollable in actual races. Differences in finish time in the simulator are due only to differences in driving technique since all other factors are specified, and therefore held absolutely constant, in the simulation. It is clear that driver performance, as measured by the times achieved on the simulator over a long period, does gradually improve.

The simulator will be installed permanently at the training complex of the US Bobsled Federation in Lake Placid, New York and it will be used year round in the preparations for the World Cup circuit and the 1998 Olympic Games. Although the simulator has already proved effective as a tool in the tuning of the skills of world class drivers, we believe that an equally large potential impact of the simulator lies at the other end of the experience spectrum where it may aid in the rapid development of novice drivers.

6. Acknowledgment

Much of the work reported in this paper was supported by a grant from the US Olympic Committee through the Sport Science and Technology Grants program. Additional in-kind and financial support by Silicon Graphics, Inc. and the Flight Simulation Division of NASA Ames Research Center is also gratefully acknowledged.

7. References

[1] Baumann W., The influence of mechanical factors on speed in tobogganing, *Biomechanics III*, Vol. 8, pp. 453-459, (1973).

[2] Ulman D.G. and Cross C., Engineering a new-generation bobsled, *ASME paper 79-DE-E-5*, (1979).

[3] Hubbard M., Kallay M. and Rowhani P., Three-dimensional bobsled turning dynamics, *Int. J. Sport Biomechanics*, Vol. 5, pp. 222-237, (1989).

[4] Hubbard M., Kallay M., Joy K., Reus J. and Rowhani P., Simulation of vehicle and track performance in the bobsled, *Proceedings 3rd ASME/ASCE Mechanics Symposium*, San Diego, July, (1989),

[5] Huffman R.K., Hubbard M. and Reus D., Use of an interactive bobsled simulator in driver training, *Presented at ASME Winter Annual Meeting*, New Orleans, Dec., (1993).

[6] Faux I.D. and Pratt M.J., Computational Geometry for Design and Manufacture, (John Wiley, New York, 1979).

[7] Kane T.R., and Levinson D.A., Dynamics: Theory and Applications, (McGraw-Hill, New York 1985).

[8] Sage A.P. and White C.C., III, Optimum Systems Control, (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1977).

[9] Vaughn C.L., Simulation of a sprinter. Part I. Development of a model, *Int. J. Bio-Medical Computing*, 14, 65-74, (1983).

[10] International Rules, Federation Internationale de Bobsleigh et de Tobogganing, via Piranesi 44/B, Milano, 20127 Italy, (1992).

THE RELATIONSHIP BETWEEN POWER DEMAND AND ENDURANCE TIME

R Hugh Morton[1]

Abstract

It is axiomatic that for work requiring high power output, endurance time is short, and vice versa. Parameters describing this relationship are important in characterising work performance and the capacity of humans as a source of mechanical power. Early graphical and curve fitting analyses have largely been superseded by attempts to model the human bioenergetic processes, to varying degrees of sophistication and with varying levels of success. This paper critically examines published work and offers some previously unpublished variations on the theme in an attempt to strike a balance between realism and parsimony.

1. INTRODUCTION

Consider the question "Starting from a fully rested state, how long does it take for a human to become exhausted at different levels of power output?" It is of course obvious that for work requiring sustained high power, endurance time is short, and that lower power outputs can be maintained for extended periods of time. More specifically one should ask whether, given our knowledge of the chemical and mechanical processes involved, the relationship between power output (or total work performed as the integral of power) can be represented mathematically by a systems model with meaningful parameters. Such a relationship is important, not only to ergonomists in characterising work performance, but also in the ability of humans as a source of mechanical power (Wilkie [1]) in activities such as human powered flight (Nadel [2]).

The purpose of this paper is to provide a brief review of the available literature relevant to the investigation of this relationship and its parameters. Initially we shall consider graphical analyses and empirical curve fitting. Systems modelling includes the development of the critical power concept and a variety of extensions, several to account for its short comings and others to broaden its scope to other forms of exercise.

2. THE EMPIRICAL APPROACH

Graphics have often been used simply for illustration when no particular analysis is intended (Bellamare and Grassino [3], Ettema [4], Knuttgen et al [5, 6], Maughan et al [7], Petrofsky and Phillips [8], Stegmann [9]). More usually since the intention is to investigate the relationship between endurance time to exhaustion, t, for various

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

constant power settings, P (which may also be reflected through the total work performed W = P.t), some form of mathematical curve is fitted to the data.

The power curve (in a variety of forms)

$$\log t = a \log P + b$$
$$\log W = c \log P + d \quad (c = a + 1 \text{ and } d = b)$$

seems to have been the earliest approach (Kennelly [10]) and has been used sporadically since (Grosse-Lordemann and Müller [11], Aunola et al [12], Pavlat et al [13]). Harman et al [14] have attempted to improve the inter-individual fitting of the power model by introducing an individual scaling factor as a standardising tactic.

The exponential or (natural) logarithmic curves of form

$$\log t = a - bP$$
$$P = a + b\, e^{-ct}$$

have also been used (Aunola et al [12], Bigland-Ritchie and Wood [15], Hopkins et al [16], Mayhew et al [17]) as an empirical model for the relationship. They too have been subject to scaling procedures (Gleser and Vogel [18], McLellan and Skinner [19]). Even double exponential equations have been used (Clarke [20]).

Some quite elaborate empirical models have been proposed, for example

$$t = kp(F/F_{max} - CF)^{-n} \qquad \text{(Monod and Scherrer [21])}$$

for intermittent static endurance, where p is the proportion of time spent in contraction, F is the force maintained, F_{max} is the force of maximal voluntary contraction (MVC) and CF is a critical force below which endurance is (in theory) infinite; or

$$t = -1.5 + \frac{2.1}{F/F_{max}} - \frac{0.6}{(F/F_{max})^2} + \frac{0.1}{(F/F_{max})^3}$$

due to Rohmert [22] and also for static endurance. Apart from F_{max} and CF in the above, only in the case of Gleser and Vogel [18] has any real attempt been made to interpret the parameters of the curves physiologically.

3. Systems Modelling

A system model is one established *a priori*, to theoretically describe the process that produces the data. If this is a sufficiently accurate description then a good fit by data can logically be expected. The parameters of such a model represent descriptions of aspects of the metabolic processes in action during exercise. The simplest and most widely known such model is the critical power (CP) concept reviewed by Hill [23].

3.1 The Critical Power Concept

This concept, and the resulting critical power test is based on the following two simple assumptions of human bioenergetics

i) that there is an upper limit to the power output that can be maintained "indefinitely" without exhaustion, depending only on renewable aerobic energy sources, known as the critical power

ii) that there is a second fixed capacity energy source available anaerobically, known as the anaerobic work capacity (AWC), which is available on demand at any required rate. When this source is totally consumed, exhaustion immediately occurs.

Thus for any fixed power setting P > CP, the total amount of work performed until exhaustion is given by

$$W = P.t = AWC + CP.t$$

This is the original linear formulation used by Monod and Scherrer [21]) and used by many others (De Vries et al [24], Moritani et al [25], Nagata et al [26], Poole et al [27], Whipp et al [28]). Dividing by t yields an alternate linear formulation

$$P = AWC.t^{-1} + CP$$

used by several authors (Housh et al [29], Pepper et al [30]). More correctly perhaps one should use the hyperbolic formulation

$$t = AWC/(P-CP).$$

In all cases good fits by data have been reported though it isn't clear which variant is regarded as preferable [Gaesser et al [31], Smith and Hill [32]).

The CP test is the exercise testing procedure utilised by many of the above researches to provide data for estimating CP and AWC. Several tests each at a different constant power are conducted to exhaustion, usually on the cycle ergometer and in random order with sufficient rest intervals. One or other of the above equations are fitted to obtain the parameter estimates.

3.2 Parameter Interpretation

The parameters CP and AWC as described above can both be easily understood as descriptors of a simple two-component bioenergetic system, and the CP test is regarded as a valid and reliable means of assessing aerobic and anaerobic capacity (Green et al [33], Nebelsick-Gullett et al [34]).

CP is regarded by Wilkie [35] as the maximal aerobic power, $\dot{V}O_{2\,max}$, and Housh et al [36] has found no significant difference between the two. However it is well established that exercise demanding an oxygen consumption of $\dot{V}O_{2\,max}$ cannot be maintained very long, say up to 8 minutes in normal individuals. Even the interpretation of CP as a maximal rate of fatigueless work has not been validated since several studies demonstrate that exercise at CP cannot be sustained much beyond about half an hour (Housh et al [29], Pepper et al [30], Jenkins and Quigley [37], McLellan and Cheung [38], Overend et al [39]). Although the interpretation of CP as a "threshold" marker has been questioned (Housh et al [36]), it correlates well with most of the so-called threshold measures; ventilatory anaerobic threshold (De Vries et al [24], Moritani et al [25], Nagata et al [26], Talbert et al [40]); lactate anaerobic threshold including the onset of blood lactate accumulation, the individual anaerobic threshold, and maximal lactate steady state (Housh et al [36], McLellan and Cheung [38], Ginn and Mackinnon [41], Wakayoshi et al [42-45]); and the fatigue threshold determined from integrated EMG data (De Vries et al [24, 46]). Since CP has been found to increase in response to 6-8 weeks of regular continuous (Gaesser and Wilson [47], Poole et al [49, 50]), but not in response to high intensity strength training (Jenkins and Quigley [51], Stokes et al [52]); and since CP increases with inspiration of an oxygen-enriched mixture (Whipp et al [28]) and decreases with inspiration of an oxygen-deficient mixture (Moritani et al [25], Whipp et al [28]); and since prior exhaustive exercise has no effect on CP (Swanson et al [53]); it is quite clear that CP in general terms has an interpretation closely linked to the aerobic fuel supply mechanism.

AWC in the model represents the aggregate work performed by non-renewable fuel supplies regardless of their rate of utilisation (Whipp et al [28]). As such it is clearly a measure of anaerobic capacity though not mentioned specifically in a review of standard anaerobic exercise tests (Vandewalle et al [54]). It correlates well with other measures of anaerobic capacity; Wingate test performance (Nebelsick-Gullett et al [34], Vandewalle et al [55]); the ability to perform repeated bouts of high intensity exercise (Jenkins and Quigley [56]); and a running ability test (Bulbulian et al [57]). Although the maximal accumulated oxygen debt (MAOD) (Medbø et al [58]) correlates with Wingate test performance (Vandewalle et al [54]), correlations between AWC and MAOD are unclear (Jenkins and Quigley [56], Hill and Smith [59]). AWC is unaffected by continuous or interval training (Gaesser and Wilson [47], Poole et al [49, 50]), but increases with high intensity strength training (Jenkins and Quigley [51], Stokes et al [52]). AWC is unaffected by inspiration of hyperoxic or hypoxic mixtures (Moritani et al [25], Whipp et al [28]), but decreases in response to prior exhaustive exercise (Swanson et al [53]). Nevertheless, AWC is still questioned as a measure of anaerobic capacity (Housh et al [60], Johnson et al [61]). Some other properties of the power-endurance relationship and its parameters are of interest. Males have higher maximum strength and absolute endurance than females, but females have longer relative endurance (Clarke [20]). Stronger subjects of either sex have longer absolute but shorter relative endurance also (Carlson [62], Carlson and McGraw [63]). Both CP and AWC are lower in older individuals but higher in relative terms; absolute endurance is less but relative endurance longer in older individuals also (Overend et al [39]).

3.3 The CP Concept in Other Exercise Forms

The CP concept and test has mainly been applied to constant power dynamic work on the cycle ergometer where work and power can be directly evaluated. Incremental work on the cycle ergometer can easily be dealt with (Morton [64]). Here the subject pedals against power increasing linearly from zero at a rate s watts per second until exhaustion. The work done is evaluated as an integral which can be equated to AWC plus a simple function of CP and t. As a result we can obtain

$$t = CP/s + \sqrt{(CP/s)^2 + 2\ AWC/s}$$

There are no published reports using a rowing ergometer, though Ginn and McKinnon [41] have applied the concept to kayaking. For swimming [Wakayoshi et al [42-45], Biggerstaff et al [65]) and running (Hughson et al [66], McDermott et al [67], Sid-Ali et al [68]), the concept modifies to a critical velocity and an anaerobic distance capacity. Hopkins et al [16] considers a running test variant using constant velocity by increasing inclines on the treadmill. Another application of the idea (rather than the CP concept itself) is to weightlifting (Mayhew et al [17]) which leads to the concept of a critical weight and an anaerobic lift capacity. An exponential equation is used rather than the simple hyperbola, and the number of repetitions at constant cadence until exhaustion takes the place of endurance time.

For static work (Monod and Scherrer [21], Rohmert [22]) the concept would appear to become a critical force and an anaerobic tension capacity. However this simple transposition is not fully appropriate because at higher tensions, intramuscular pressure causes bloodflow (and thus aerobic supply) to become progressively and then totally occluded. This necessitates a modification of the concept. If static work is intermittent, then clearly endurance is prolonged, the more so the contraction time relative to the relaxation time is decreased. Only Monod and Scherrer [21] appears to have investigated this setting experimentally. Eccentric exercise is only considered very briefly (Knuttgen et al [5, 6]).

4. Variations On A Theme

Most authors recognise that the CP concept, though appealing, is somewhat too simplistic. Wilkie [35] has attempted to extend the concept by allowing for an exponential kinetic delay in the availability of the aerobic power at its maximal level, of around 2 to 3 minutes. Without this, which is well established in the study of oxygen-uptake kinetics, the estimates of CP and AWC are biased, as evaluated by Vandewalle et al [55]. Nevertheless the whole concept is only regarded as acceptable for endurance times which are neither too short (under 1 minute) nor too long (in excess of 20 minutes). Outside these limits the concept breaks down.

Morton [69] has attempted to extend the short endurance/high power situation by proposing the existence of a third parameter, P_{max}, such that just beyond that power level the subject cannot turn the pedals and accomplishes no work. As $P \rightarrow P_{max}$ from below, less and less of AWC is utilised, whereas $P \rightarrow CP$ from above more and more is utilised. This leads to an equation taking the form of a general rectangular hyperbola

$$(t + k) \cdot (P - CP) = AWC$$

where t can be expressed as $AWC/(P_{max} - CP)$.

A more comprehensive extension at both ends of the time spectrum is due to Peronnet and Thibault [70,71]. They allow both the kinetic delay in aerobic power availability, and the progressive reduction in aerobic power able to be sustained as exercise duration exceeds beyond 7 minutes. Their extension is somewhat complex and leads to an equation of form

$$P = [S/T(1 - e^{-t/k_2})] + \frac{1}{T}\int_0^T [BMR + B(1 - e^{-t/k_1})]dt.$$

When applied to running races rather than cycle ergometry, the model was extremely accurate in predicting world running records from 100m to the marathon.

Another complex extension producing a hyperbolic type of relationship is due to Morton [72, 73]. This derives from a three component bioenergetic system comprising (anaerobic) phosphagen utilisation, anaerobic glycolysis, and aerobic power, originally intended as a representation of the bioenergetic process occurring during exercise and recovery. Like the models of Wilkie and Peronnet it also includes a kinetic delay, and like Peronnet's model it recognises that sustainable aerobic power declines with longer endurance. However the decline is ultimately to a non-zero asymptote, which may be interpreted as analogous to CP, rather than declining ultimately to zero as in the Peronnet model. Morton's model also includes the existence of a P_{max} and recognises that AWC isn't necessarily totally depleted at exhaustion. More controversially it includes a parameter representing the existence of the much-debated anaerobic threshold.

Another complex model (Ward-Smith [74]) uses biomechanical modelling of the force and power requirements for running, coupled with a model of the human energy system based on the first law of thermodynamics, in order to derive a relationship between race distance and the time taken to run that distance. The model fits data from 100 to 10,000 metres very well.

As far as static work and bloodflow occlusion is concerned, Morton [75] has proposed a model incorporating progressive bloodflow occlusion. As with the bioenergetic model there is also a kinetic delay in the aerobic supply via the circulation and allowance for exhaustion without total depletion of the anaerobic tension capacity. Two critical forces were defined. CF1 is the upper limit of force which can be sustained continuously, and CF2 is the force at which blood flow occlusion becomes total. Between CF1 and CF2, steady state blood flow declines (linearly) to zero. Despite the complexity, a fit by published data from the literature is extremely good.

Another series of detailed models deserve mention (Behncke [76], Cooper [77], Keller [78], Maronski [79, 80], van Ingen Schenau et al [81], Woodside [82]). They are not specifically aimed at the power-time relationship per se, but rather at optimising performance in running, swimming, skiing, cycling, speed skating and wheelchair propulsion. The objective is to exercise control over velocity such that a fixed distance is covered in a minimum time by using optimal control methods applied to a force/power/energy model.

5. METHODOLOGICAL QUESTIONS

The asymptotics of several of the power-time models are of interest. Most have the reasonable property that for ultra-endurance as t becomes very large, the power asymptote is non-zero. At the other time extreme, only the exponential models (Hopkins et al [16], Mayhew et al [17]) and Morton's models [69, 73, 75] predict finite P for $t \to 0$. In the case of the 3-component model (Morton [73]) $t \approx 6$ seconds for the maximum attainable anaerobic power.

In the case of the CP concept one faces the choice of which equation to use; one of several linear variants or the non-linear hyperbolic equation. The former have been preferred in the past, but more recently the latter seems to be the equation of choice. Arguably it is the one that is intuitively more appropriate, and probably satisfies the usual regression assumptions to a better degree. Whichever, the evidence in the literature is unclear (Hill [23], Gaesser et al [31], Smith and Hill [32]).

One regression assumption which is most obviously questionable is that of homoscedasticity, for the longer endurance times at low power have been found to exhibit a greater degree of variability. Whether this supports using say a P versus t^{-1} linear relationship or whether the error structure could be discerned enabling a more refined fitting procedure to be used, is yet to be determined. For the moment, using a weighted regression procedure is probably advisable.

Irrespective of the model used, questions of design arise in the choice of power settings used. Clearly as wide a spread as is feasible ought to be the preferred choice in order to improve accuracy and precision of estimates. Most studies use four or five settings such that exhaustion occurs between one and ten minutes. Only one study (Aunola et al [12]) has collected data at substantially longer endurance times. Wakayoshi [44, 45] has argued that only two settings are sufficient for swimming settings, as has Clingeleffer [83] for kayaking. While this will enable parameter estimation, it will not permit standard error estimates. Housh et al [84] has shown that two settings do provide very similar estimates as do four (provided the two are spread as widely as possible), but the absence of any measure of precision must be viewed with suspicion.

6. SUMMARY

Earlier empirical modelling yielded a variety of sometimes complex formulations of the relationship between power demand and endurance time in both dynamic and static work. In recent years the most widely used model is the two component (aerobic + anaerobic) hyperbolic systems model, characterised by a critical power (or maximal rate of fatigueless work) and an anaerobic work capacity. This model has a simple appeal, its parameters are well understood and has always been well fitted by data over a somewhat restricted range. Extensions to the hyperbolic model incorporate a more realistic representation of the human bioenergetic system, and encompass a wider range of power and duration.

REFERENCES

[1] Wilkie D.R. Man as a source of mechanical power, *Ergonomics* 3, 1-8 (1960).

[2] Nadel E.R. The Daedalus project: Overcoming the physiological limits of human-powered flight, *Med. Sci. Sports Exerc.* 20 (Suppl.), 584 (1988).

[3] Bellamare F. and Grassino A. Effect of pressure and timing of contraction on human diaphragm fatigue, *J. Appl.Physiol.* 53, 1190-1195 (1982).

[4] Ettema J.H. Limits of human performance and energy production, *Int. Z. angew Physiol.* 22, 45-54 (1966).

[5] Knuttgen H.G. Human performance in high intensity exercise with concentric and eccentric muscle contractions, *Int. J. Sports. Med.* 7 (Suppl.), 6-9 (1986).

[6] Knuttgen H.G., Patton J.F. and Vogel J.A. An ergometer for concentric and eccentric muscular exercise, *J. Appl. Physiol.* 53, 784-788 (1982).

[7] Maughan R. J., Harmon M., Leiper J.B., Sale D. and Delman A. Endurance capacity of untrained males and females in isometric and dynamic muscular contractions, *Eur. J. Appl. Physiol.* 55, 395-400 (1986).

[8] Petrofsky J.S. and Phillips C.A. The strength-endurance relationship in skeletal muscle : Its application to helmet design, *Aviat. Space Environ. Med.* 53, 365-369 (1982).

[9] Stegmann J. *Exercise Physiology : Physiologic Bases of Work and Sport.* (George Thieme: New York 1981).

[10] Kennelly A.E. An approximate law of fatigue in the speeds of racing animals, *Proc. Amer. Acad. Arts Sci.* 42, 275-331 (1906).

[11] Grosse-Lordemann H. and Müller E.A. Der einfluss der leistung und der arbeitsgeschwindikeit auf das arbeits maximum und den wirkangsgrad beim radfahren, *Arbeits Physiol.* 9, 454-475 (1973).

[12] Aunola S., Alanen E., Marniemi J., and Rusko H. The relation between cycling time to exhaustion and anaerobic threshold, *Ergonomics* 33, 1027-1042 (1990).

[13] Pavlat D.J., Housh T.J., Johnson G.O., Schmidt R.J. and Eckerson J.M. An examination of the electromyographic fatigue threshold test, *Eur. J. Appl. Physiol.* 67, 305-308 (1993).

[14] Harman E.A., Knuttgen H.G., Frykman P.N. and Patton J.F. Exercise endurance time as a function of percent maximal power production, *Med. Sci. Sports Exerc.* 19, 480-485 (1987).

[15] Bigland-Ritchie B. and Woods J.J. Changes in muscle contractile properties and neural control during human muscle fatigue, *Muscle and Nerve* 7, 691-699 (1984).

[16] Hopkins W.G., Edmond I.M., Hamilton B.H., MacFarlane D.J. and Ross B.H. Relation between power and endurance for treadmill running of short duration, *Ergonomics*, 32, 1565-1571 (1989).

[17] Mayhew J.L., Ball T.E. and Bowen J.C. Predictions of bench press lifting ability from submaximal repetitions before and after training, *Sports Med. Training Rehab.* 3, 195-201 (1992).

[18] Gleser M.A. and Vogel J.A. Endurance capacity for prolonged exercise on the bicycle ergometer, *J. Appl. Physiol.* 34, 438-422 (1973).

[19] McLellan T.M. and Skinner J.S. Submaximal endurance performance related to the ventilation thresholds, *Can. J. Appl. Sports Sci.* 10, 81-87 (1985).

[20] Clarke D.H. Sex differences in strength and fatigability, *Res. Q. Exerc. Sport* 57, 144-149 (1986).

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

[22] Rohmert W. Ermittlung von erholungspausen für statische arbeit des menschen, *Int. Z. angew. Physiol.* 18, 123-164 (1960).

[23] Hill D.W. The Critical power concept: A Review, *Sports Med.* 16, 237-254 (1983)

[24] De Vries H.A., Moritani T. Nagata A. and Magnussen K. The relation between critical power and neuromuscular fatigue as estimated from electromyographic data, *Ergonomics* 25, 783-791 (1982).

[25] Moritani T., Nagata A. , De Vries H.A. and Muro M. Critical power as a measure of physical work capacity and anaerobic threshold, *Ergonomics* 24, 339-350 (1981).

[26] Nagata A., Moritani T. and Muro M. Critical power as a measure of muscular fatigue threshold and anaerobic threshold. In: *Biomechanics* VIII, H. Matsui and K. Kafayassi (Eds) Human Kinetics: Champaign IL, 312-320 (1983).

[27] Poole D.C., Ward S.A., Gardner G.W. and Whipp B.J. Metabolic and respiratory profile of the upper limit for prolonged exercise in man, *Ergonomics* 31, 1265-1279 (1988).

[28] Whipp B.J., Huntsman D.J., Storer T.W., Lamarra N. and Wasserman K. A constant which determines the duration of tolerance to high-intensity work, *Fred. Proc.* 41, 1591 (1982).

[29] Housh D.J., Housh T.J. and Bauge S.M. The accuracy of the critical power test for predicting time to exhaustion during cycle ergometry, *Ergonomics* 32, 997-1004 (1989).

[30] Pepper M.L., Housh T.J. and Johnson G.O. The accuracy of the critical velocity test for predicting time to exhaustion during treadmill running, *Int. J. Sports Med.* 13, 121-124 (1992).

[31] Gaesser,G.A., Carnevale T.J., Garfinkel A. and Walter D.O. Modeling of the power-endurance relationship for high-intensity exercise, *Med. Sci. Sports Exerc.* 22, S16 (1990).

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

[33] Green S., Goodman C., Carey M.F., Cable N.T. and Dawson, B. The maximal work duration relationship and aerobic-anaerobic metabolism in cyclists, *Med. Sci. Sports Exerc.* 25, S105 (1993).

[34] Nebelsick-Gullett L.J., Housh T.J., Johnson G.O. and Bauge S.M. A comparison between methods of measuring anaerobic work capacity, *Ergonomics* 31, 1413-1419 (1988).

[35] Wilkie D.R. Equations describing power input by humans as a function of duration of exercise. In : *Exercise Bioenergetics and Gas Exhance,* P. Cerretelli and B.J. Whipp (Eds). (Elsevier/North Holland Biomedical: Amsterdam, 75-80 1980).

[36] Housh T.J., Johnson G.O., McDowell S.L., Housh D.J. and Pepper M. Physiological respones at the fatigue threshold, *Int. J. Sports Med.* 12, 305-308 (1991).

[37] Jenkins D.J. and Quigley B.M. Blood lactate in trained cyclists during cycle ergometry at critical power, *Eur. J. Appl. Physiol.* 61, 278-283 (1990).

[38] McLellan T.M. and Cheung K.S.Y. A comparative evaluation of the individual anaerobic threshold and the critical power, *Med. Sci. Sports Exerc.* 24, 543-550 (1992).

[39] Overend T.J., Cunningham D.A., Paterson D.H. and Smith W.D.F. Physiological responses of young and elderly men to prolonged exercise at critical power, *Eur. J. Appl. Physiol.* 64, 187-193 (1992).

[40] Talbert S.M., Smith J.C., Scarborough P.A. and Hill D.W. Relationships between the power asymptote and indices of anaerobic and aerobic power, *Med. Sci. Sports Exerc.* 23, S27 (1991).

[41] Ginn E.M. and MacKinnon L.T. The equivalence of onset of blood lactate accumulation, critical power and maximal lactate steady state during kayak ergometry. *Proc. First IOC Cong. Sports Sci.*, Colorado Springs, 34 (1989).

[42] Wakayoshi K., Ikuta K., Yoshida T., Udo M., Moritani T., Mutoh Y. and Miyashita M. Determination and validity of critical velocity as an index of swimming performance in the competitive swimmer, *Eur. J. Appl. Physiol.* 64, 153-157 (1992(a)).

[43] Wakayoshi K., Yoshida T., Udo M., Kasai T., Moritani T., Mutoh Y. and Miyashita M. A simple method for determing critical speed as a swimming fatigue threshold in competitive swimming, *Int. J. Sports Med.* 13, 367-371 (1992(b)).

[44] Wakayoshi K., Yoshida T. Udo M., Harada T., Moritani T., Mutoh Y. and Miyashita M. Does critical swimming, velocity represent exercise intensity at maximal lactate steady state? *Eur. J. Appl. Physiol.* 66, 90-95 (1993a).

[45] Wakayoshi K., Yoshida T. and Moritani T. Does critical swimming velocity represent exercise intensity at maximal lactate steady state?, *Med. Sci. Sports Exerc.* 25, S66 (1993b).

[46] De Vries H.A., Tichy M.W., Housh T.J., Smyth K.D., Tichy A.M. and Housh D.J. A method for estimating physical working capacity at the fatigue threshold, *Ergonomics* 30, 1195-1204 (1987).

[47] Gaesser G.A. and Wilson L.A. Effects of continuous interval training on the parameters of the power-endurance time relationship for high-intensity exercise, *Int. J. Sports Med.* 9, 417-421 (1988).

[48] Jenkins D.J. and Quigley B.M. Endurance training enhances critical power, *Med. Sci. Sports Exerc.* 24, 1283-1289 (1992).

[49] Poole D.C., Ward S.A. and Whipp B.J. Effects of training on the metabolic and respiratory profile of high-intensity exercise, *The Physiologist* 29, 161 (1986).

[50] Poole D.C., Ward S.A., and Whipp B.J. The effects of training on the metabolic and respiratory profile of high-intensity cycle ergometer exercise, *Eur. J. Appl. Physiol.* 59, 421-429 (1990).

[51] Jenkins D.J. and Quigley B.M. The influence of high-intensity exercise training on the $W_{lim} - T_{lim}$ relationship, *Med. Sci. Sports Exerc.* 25, 275-282 (1993).

[52] Stokes D.G., Shwanawed T.W., Womack C.J., Weltman A. and Gaesser G.A. Effects of strength training on the lactate treshold and parameters of the power-endurance hyperbola, *Med. Sci. Sports Exerc.* 25, S163 (1993).

[53] Swanson G.D., Holland R.A. and Van Walleghem E.B. Effect of prior leg exercise on high-intensity arm ergometry, *The Physiologist* 35, 199 (1992).

[54] Vandewalle H., Peres G. and Monod H. Standard anaerobic tests, *Sports Med.* 4, 268-289 (1987).

[55] Vandewalle H., Kapitaniak B., Grün S., Raveneau S., and Monod H. Comparison between a 30-s all-out test and a time-work test on a cycle ergometer, *Eur. J. Appl. Physiol.* 58, 375-381 (1989).

[56] Jenkins D.J. and Quigley B.M. The y-intercept of the critical power function as a measure of anaerobic work capacity, *Ergonomics* 34, 13-22 (1991).

[57] Bulbulian R., Wilcox A.R., and Darabos B.L. Anaerobic contribution to distance running performance of trained cross-country athletes, *Med. Sci. Sports Exerc.* 18, 107-114 (1989).

[58] Medbø J.I., Mohn A.C., Tabata I., Bahr R., Vaage O. and Sejersted O.M. Anaerobic capacity determined by maximal accumulated O_2 deficit, *J. Appl. Physiol.* 64, 50-60 (1988).

[59] Hill D.W. and Smith J.C. A comparison of methods of estimating anaerobic work capacity, *Ergonomics* 36, 1495-1500 (1993).

[60] Housh T.J., Johnson G.O., McDowell S.L., Housh D.J. and Pepper M. The relationship between anaerobic running capacity and peak plasma lactate, *J. Sports Med. Phys. Fitness* 32, 117-122 (1992).

[61] Johnson G., Housh T., McDowell S., Housh D. and Pepper M. Validity of a treadmill test of anaerobic capacity, *Med. Sci. Sports Exerc.*, 22, S16 (1990).

[62] Carlson B.R. Level of maximum isometric strength and relative load isometric endurance, *Ergonomics* 12, 429-435 (1969).

[63] Carlson B.R. and McGraw L.W. Isometric strength and relative isometric endurance, *Res. Q.* 42, 244-250 (1971).

[64] Morton R.H. *The Critical Power Test for Ramp Exercise*, School of Mathematical and Information Sciences Preprint Series 94/4, Massey University (1994a).

[65] Biggerstaff K.D., Hill D.W., Jackson S.L. and Sams B.R. Use of the critical power concept to evaluate anaerobic capacity in swimmers, *Med. Sci. Sports Exerc.* 24, S74 (1992).

[66] Hughson R.L., Orok C.J. and Staudt L.E. A high velocity treadmill running test to assess endurance running potential, *Int. J. Sports Med.* 5, 23-25 (1984).

[67] McDermott K.S., Forbes M. R. and Hill D.W. Application of the critical power concept to outdoor running, *Med. Sci. Sports Exerc.* 25, S109 (1993).

[68] Sid-Ali B., Vandewalle H., Chair K., Moreaux A. and Monod H. Lactate steady state velocity and distance-exhaustion time relationship in running, *Archiv. Int. Physiol. Biochim. Biophys.* 99, 297-301 (1991).

[69] Morton R.H. *The 3-Parameter Critical Power Model*, School of Mathematical and Information Sciences Preprint Series 94/5, Massey University (1994b).

[70] Peronnet F. and Thibault G. Analyse physiologique de la performance en course a pied: Revision du model hyperbolique, *J. Physiol . (Paris)* 82, 52-60 (1987).

[71] Peronnet F. and Thibault G. Mathematical analysis of running performance and world running records, *J. Appl. Physiol.* 67, 453-465 (1989).

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

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

[74] Ward-Smith A.J. A mathemtical theory of running based on the first law of thermodynamics, and its application to the performance of world-class athletes, *J. Biomech.* 18, 337-349.

[75] Morton R.H. A simple model to link hemodynamics, fatigue and endurance in static work. *J. Biomech.* 20, 641-644 (1987).

[76] Behncke H. Optimisation models for the force and energy in competitive sports, *Math. Meth. Appl. Sci.* 9, 298-311 (1987).

[77] Cooper R.A. A force/energy optimisation model for wheelchair athletics, *IEEE Trans. Systems, Man, Cybernet.* 20, 444-449 (1990).

[78] Keller J.B. Optimal velocity in a race, *Amer. Math. Monthly* 81, 474-480 (1974).

[79] Maronski R. On optimal running downhill on skis, *J. Biomech.* 23, 435-439 (1990).

[80] Maronski R. On optimal velocity during cycling, *J. Biomech.* 27, 205-213 (1994).

[81] van Ingen Schenau G.J., de Koning J.J. and de Groot G. Optimisation of sprinting performance in running, cycling and speed skating, *Sports Med.* 17, 259-275 (1994).

[81] Woodside W. The optimal strategy for running a race (A mathematical model for world records from 50m to 275km), *Math. Comput. Modelling* 15, 1-12 (1991).

[82] Clingeleffer A., McNaughton L., and Davoren B. Critical power may be determined from two tests in elite kayakers, *Eur. J. Appl. Physiol.* 68, 36-40 (1994).

[83] Housh D.J., Housh T.J. and Bauge S.M. A methodological consideration for the determination of critical power and anaerobic work capacity, *Res. Q. Exerc. Sport* 61, 406-409 (1990).

ECOLOGICAL VALIDITY IN SPORTS RESEARCH THROUGH VIDEO SIMULATION.

Geoffrey Paull and Ian Case[1]

Abstract

A major contributor to expertise at any skill is the knowledge formed for the competitive environment and the optimum response actions that apply at any point in the state of play. If applied sports researchers are to examine this knowledge store, they must establish the contextual reality of the experimental paradigm to ensure that data reflects the true factors of expertise. The development of an inter-active video simulation for sports skills analysis may provide such ecological validity for research, and decrease the experimenter intrusion that detracts from the reality of the laboratory environment for athletes. This paper describes the simulator, and outlines two studies that demonstrate the face validity of the equipment to the usual competitive situation. The first study compared expert baseball batters to a group of novice batters and demonstrated the differences in cognitive operations between the groups. The next study examined the use of visual cues by base runners attempting to steal a base. Differences were illustrated in the spatial arrangement of cues from the body of the pitcher as to the forthcoming alternative actions. The simulator is shown as a cost-effective paradigm for laboratory testing of sport skills. Further development is possible with greater funding, and the potential future stages of technological implementation are discussed.

1. INTRODUCTION

The mastery of skills shown by athletes can be understood in part from the knowledge they acquire for the environment they must perform in. The structure of this memory store in elite level athletes will include knowledge for specific cues from the sensory array, and associated specific output motor actions for skilled performance (Anderson [1]). The challenge for applied sports researchers is to provide an experimental paradigm that provides an environment requiring such task specific knowledge to be accessed in providing measures for analysis. This concern for ecological validity in the methodology of experiments is addressed here in describing a configuration for computer controlled inter-active video simulation for applied sports research.

In the sports domain, it is typically found that experts at a skill will show enhanced cognitive functioning from that of novices (eg, reduced response latencies and finer attentional operations) to relevant stimuli for the skill of interest (eg, Abernethy [2]; Glencross & Paull [3]; Houlston & Lowes [4]). This assumes that the *relevant stimuli* can be provided in the experimental procedure, therefore eliciting the optimum

[1] School of Psychology, Curtin University of Technology.
Kent St, Bentley, Western Australia, 6102.

response learned from practice at the actual task in training. If this contextual relevance is not provided, there must be attenuation of ecological validity in the paradigm, leading to invalid data and perhaps misleading information for athletes (Paull & Fitzgerald [5]).

The requirements, then, for an experimental paradigm to examine visual perception in a sports skill must include:

a) An environment that reproduces the visual information cues usually available in the competitive event.

b) Utilisation of realistic response actions that have, through practice, become coupled in memory to the perceptual system.

c) The ability to manipulate visual cues in the display with the expectancy that this will create differences in the responses provided by subjects.

d) The recording of control, timing and response data without interference to subjects in executing their usual sports actions.

Only after fulfilling these requirements can researchers can be confident in inferring that the recorded cognitive functioning truly represents that utilised in the competitive situation. Development of inter-active video for such research can be a cost effective method of meeting these requirements. The technological configuration for the system under development at Curtin University is outlined here, and results from two studies are used to illustrate the paradigm.

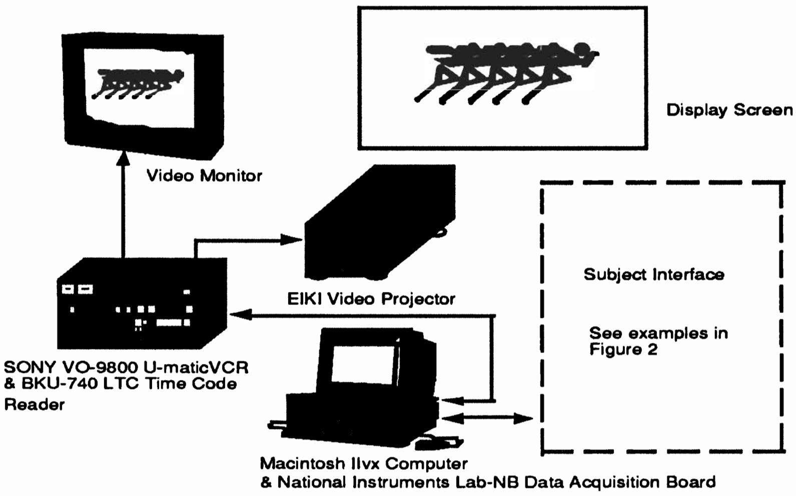


Figure 1: Pictorial layout of the technological configuration

2. BUILDING BLOCKS OF AN ECOLOGICALLY VALID VIDEO BASED SPORT ANALYSER

The basic tenet of a sport analyser providing a high degree of ecological validity is that the experimental subject can perform the usual actions of a sporting skill with a minimal amount of interference from the system. In this respect, for the sport analyser shown in Figure 1, the following features can be observed:

a) Control of all equipment is centralised in one computer, thereby reducing experimenter intrusion into the testing environment

b) The life-size projection of sport action scenarios is provided using high resolution video equipment.

c) The data collection points for subjects' responses are embedded into a natural part of the game apparatus or sporting action.

d) The central computer is used to coordinate the display of the video images and the acquisition of response signals, and to maintain time phasing between the two.

e) Data files generated during the real-time performance by experimental subjects are stored to disk for later analysis by the researchers.

f) A communications bus, with minimum cabling, links all necessary components and sub-systems for control and data signals.

The sport analyser developed at the School of Psychology at Curtin University uses Hi-band/SP U-matic format for playback and display. The U-matic (PAL @ 25 frames/sec or 50 fields/sec) format provides a display resolution of 300 lines and permits a longitudinal time code to be recorded onto the video tape for frame-by-frame and within-frame analysis, and editing of sequences. The pre-recorded video material is professionally edited to produce relevant displays of sports actions in blocks of time-coded sequences to reduce tape search transitions during playback.

The SONY U-matic VO-9800P VCR used for playback incorporates a remote connection RS-422 serial port operating at 19.2K bits/sec. The serial port provides access for an external computer to monitor and control all of the major VCR functions including cue forward, play, pause, cue backward, rewind, and stop(see Sony [6] for the full specification of this function). The ability to also read time code information via the serial port allows the controlling computer to synchronise video events (in this case information 'cues' previously identified by the researchers and designated by time code) with a master time clock and any of the subject's response data acquired via the sport apparatus interface.

The EIKI video projector connects to the VIDEO OUT terminal of the VCR and projects life size video images onto a suitable screen. This semi-portable projector allows flexibility in the experimental layout. It can be placed in any suitable location and provides an effective zoom range to adjust the projected display size. In this style of equipment, the initial image for projection is formed on a three layer colour LCD screen in front of a single high intensity lamp. The three layers of LCD represent each of the primary colours of blue, yellow and red, and a combination of these held 'ON' provides the spectrum of colours required for the images. For even less intrusion by equipment into the experimental environment, the projector has a 'reverse image' mode that allows for back-lit projection from behind the screen.

The controlling computer for the system is a Macintosh IIvx fitted with a National Instruments Lab-NB data acquisition board. The Lab-NB board has 24 TTL-compatible digital ports, a 12-bit analog-to-digital converter with 8 analog ports, and six 16-bit counter/timer channels. This card provides the hardware interface for the transducers in the modified sporting apparatus. The software program, which provides the experimenter interface to the system, oversees the operation of the VCR and synchronises the collection of data. The program uses proprietary library routines for the Lab-NB operations and specially written software libraries in either C or LabView* (refer National Instruments [7]) for the VCR operations.

The transducers used in the modified sport apparatus for the two studies presented in this paper are shown in Figures 2a & b.

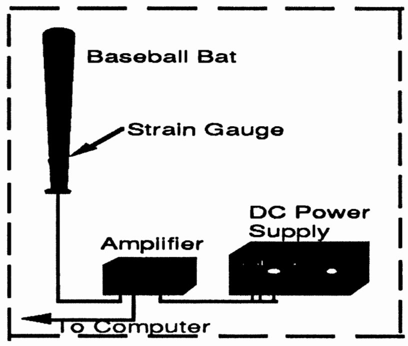


Figure 2a Baseball batting interface

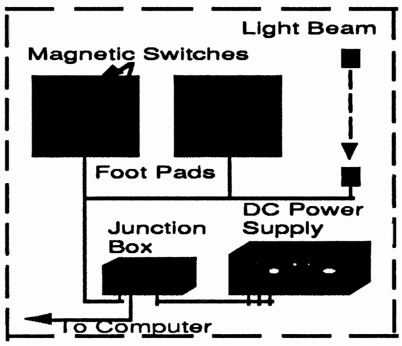


Figure 2b Base Stealing interface

3. STUDY 1: BASEBALL BATTING

The first study utilising the inter-active video simulator investigated the operations of baseball batters. The object of the study was to identify the visual information in the pitch sequence that batters use to make their decisions about the time and location of the ball as it passes home plate (see Glencross and Paull [3] for a full explanation of this study). There were two experiments in this study which compared the performance of a group of expert batters with that of a novice group.

Experiment one examined the decision times and accuracy of batters viewing pitches on a video monitor (note that full size video projection was introduced after this research - see study 2 for details). Further, their ability to set probabilities about the forthcoming pitch was tested by providing strategic game information for half of the trials. The hypothesis for experiment one expected that: (a) expert batters will have faster decision times and be more accurate than novices, and (b) when game strategic information is presented, experts will use this information to set subjective probabilities about forthcoming action and hence have improved decision time and accuracy scores, while novices will not have knowledge to use this information and their decision time and accuracy will not be significantly improved.

The second experiment used video that was manipulated to determine the value of information throughout the sequence of a pitch. The display consisted of video which was edited to provide an occlusion, whereby, at five designated points in the pitch delivery, the pitcher and ball disappeared and blank stadium background remained. Editing provided five displays of each pitch with each display occluded at a different point in time as follows:

(i) T_1 - two frames (-80msecs) before moment of release of the ball from the hand (MOR)

(ii) T_2 - at MOR

(iii) T_3 - two frames(+80msecs) after MOR

(iv) T_4 - four frames (+160msecs) after MOR

(v) T_5 - six frames (+240msec) after MOR

It was expected that expert batters would utilize visual cues provided in advance of MOR and from early in the pitch delivery. As novices are not expected to use this early information, it was hypothesised that experts will have greater accuracy than novices when occlusion occurs earlier in the pitch delivery.

4. System Operation and Results for Study 1

The video tape recording of baseball pitching used pitchers from the Australian Baseball League. The pitcher was video taped at a shutter speed of $^1/_{500}$th of a second (to provide crisp, distortion free images) from the aspect of a right-handed batter. This video sequence was displayed to the batter on a high resolution 64cm monitor. Subjects were positioned at a distance from the monitor such that the pitcher's image subtended an angle to the eye of approximately 6º 17', which is the approximate size for an 180cm pitcher at a distance to the pitcher's plate of 18.44 metres. The video monitor location was elevated equivalent to the view of a pitcher positioned on a standard 35cm mound. Subjects could step and swing the bat if this helped to provide a more natural approach to the trials.

The earliest natural physical response that best represented a decision by a batter to swing at a pitch is a tightening of the grip on the bat. In the experimental situation, this squeezing on the strain gauge in the handle of the bat caused the video signal to the monitor to be broken, thus precluding any further information of ball flight. After the monitor blanked, subjects provided a number from a grid of the strike zone which best represented their prediction of the ball destination over the plate. In the case of strategic scenario pitches, the subject was shown a game card depicting the state of a hypothetical game before each trial.

For the decision-time response outlined above, the baseball bat in Figure 2a is fitted with a strain gauge. A change in grip pressure by the batter is sensed by the strain gauge connected as a quarter bridge and amplified by a stable balanced strain gauge amplifier with gain of A_V=1000. This signal is then feed to the analog input of a 12 bit analog-to-digital converter in the data acquisition card. A sampling rate of 1.0 msecs is generally used but is adjustable as a software parameter. This signal is in the range of 0-100 mV depending on grip pressure. The batter response is detected in software when a preset threshold is reached. Given the different resting grip pressures of individual batters, the threshold can be calibrated before trials are run.

In the occlusion study (experiment 2) no decision time was recorded as the stimulus (information from the pitcher's actions and the ball flight) was removed at a preset time in the pitch. Figure 3 details the major experimental sequence of events for the two experiments.

Each trial sequence was started with the video tape cued to the designated start-point. All trial start-point and frame information of the video tape is read from a data file containing the time codes designating these points. Figure 4 details a time line of the events in each trial. When starting a VCR 'PLAY' sequence from a cued position the servo-mechanism controlling the playback video heads of the VCR is already energised and the heads are in position. This means that there is a minimum, and repeatable, time required for the tape-drive mechanism to get the tape to play speed. A very stable transition from a paused or still video signal to normal video signal is also achieved.

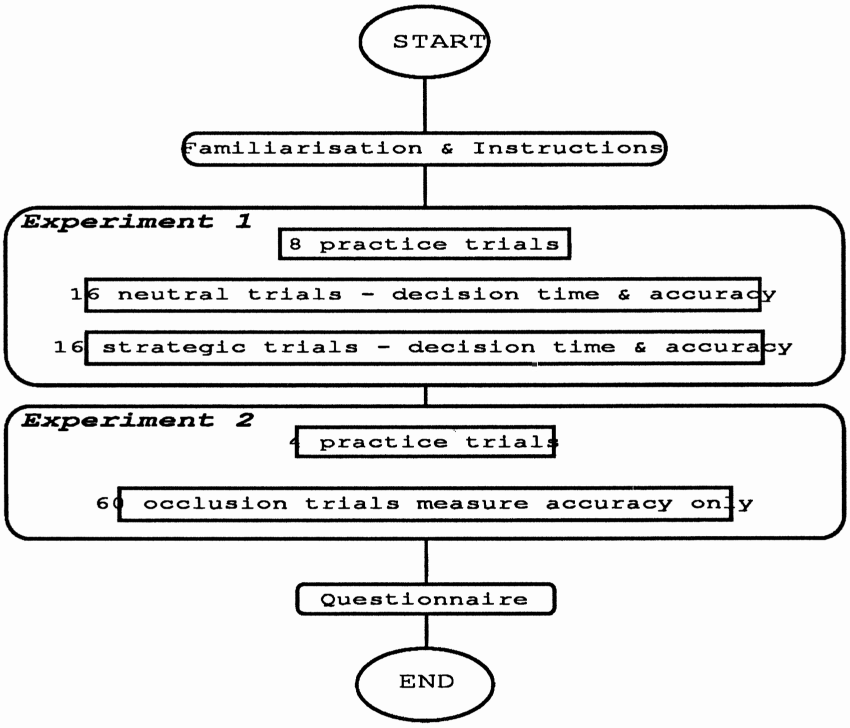


Figure 3: Sequence chart of equipment events

30 right handed subjects participated; 15 in the expert group (players with a batting average >.300 for the previous season) and 15 novices (players who have never played higher than B grade (the fourth level in their competition), but with a minimum of three years experience). Scoring for each subject comprised 'decision time' (DT) in milliseconds, and predictions of ball location (calculated to an 'error' score). Decision times were automatically recorded during each trial while pitch prediction was recorded from the batter's subjective assessment of the expected position of the ball as it would have passed home plate.

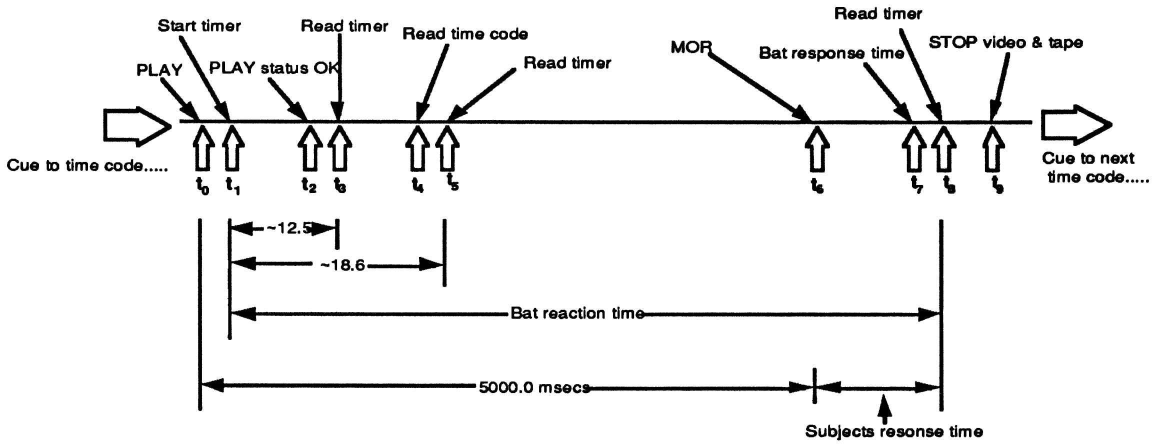


Figure 4: Time line of experimental procedure

With 30 subjects, a total 3000 trials were carried out (2860 experiment, 240 practice). With reference to figure 4, the measurement of the tape play command to a tape play status command duration (t_1 to t_3); the tape play command to read time code command duration (t_1 to t_5); and the difference in actual time code read at t_5 to the cued time code at t_0 were used to measure the repeatability of the trial process. Normal video accuracy is quoted by the manufacturer as +/– 1 frame (40 msec) (Sony [8]), however, testing of the VCR used and monitoring over the 2860 experiment trials showed that the time from cue to tape play at normal speed (t_1 to t_3) is a highly repeatable duration with mean = 12.5 msecs and standard deviation = 1.37 msecs. The tape cue to time code read time (t_1 to t_5) was nearly as repeatable with mean of 18.6 msecs and standard deviation of 2 msecs. Both time durations were very repeatable when compared with the specified absolute accuracy, which implies a high degree of reliability of time measurement. The average time code read command duration had a mean of 6.15 msecs and standard deviation of 1.38 msecs. Thus in general, at least 6 time code reads can be accomplished using software polling of the video frame LTC time code. This resolution of time code reading allows synchronisation of video events with external events to be in the order of milliseconds. It was also found, in all but a few cases (11 out of 2860 or 0.003%), that the first time code read frame value did equal the cue frame value. It can be safely assumed that normal tape play speed is reached within the first two frames of playback, and that a highly reliable time duration measurement can be made with a resolution in the order of milliseconds, and that any absolute error in time measurement accuracy when compared with an accurate timing standard is repeatable.

Figures 5 and 6 summarise the results of the first experiment described above for decision times (DT) and errors scores. It was found that experts were significantly faster in response (expert mean = 463msecs, novice mean = 571msecs) with $F_{(1,28)}=10.45$, $p<0.01$. This was not at the expense of accuracy, with experts being significantly more accurate than the novices with $F= (1,28)=9.96$, $p<0.01$. The effect of strategic information to facilitate faster and more accurate decision making was also found. On both DT and accuracy, subjects improved their scores with F_{dt}

(1,28)=55.37, p<0.001 and F_{error} (1,28)=7.90, p<0.01. The expected interaction for these factors (ie, experts improving but not novices) was not found (F_{dt} (1,28)=<1, and F_{error} (1,28)=<1).

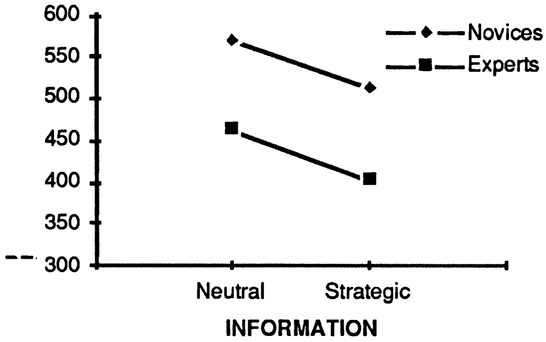


Figure 5: Decision times for Expert and novice batters

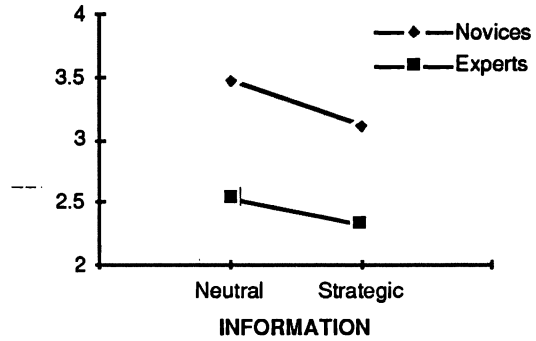


Figure 6: Error for expert and novice batters

Part of the results of experiment two are given in Figure 7 (for full results refer to Glencross & Paull [3]). It was hypothesised that experts would be able to use information from earlier in the pitch than novices . As shown, experts showed generally more accurate scores than novices after the moment of delivery of the pitch - this difference between groups, however, was not significant (F (1,28) = 1.93, NS). Improvement in accuracy scores by both groups was not shown until after T_3 - that is, after approximately 3 metres of ball flight. From T_3, the improvement in scores indicates some greater ability of experts to use early ball flight cues in determining the trajectory of the ball. The difference between groups approaches significance at the last two stages of occlusion (F_{T4} (1,28) = 3.90, p = .058; F_{T5} (1,28) = 4.24, p = .049), however, considering experimentwise error, these marginal results may be difficult to support.

The divergence in scores over the flight of the ball does not provide a significant interaction. In any event, this trend for differences in scores later in the ball flight does not support the hypothesis that examined the use of advanced cues by expert batters. If this effect had been found (ie, the expert group provided significantly better scores at T_1 and T_2 than novices), it would have been that novices' scores would more likely become closer to experts' scores later in the ball flight (ie, T_3 - T_5).

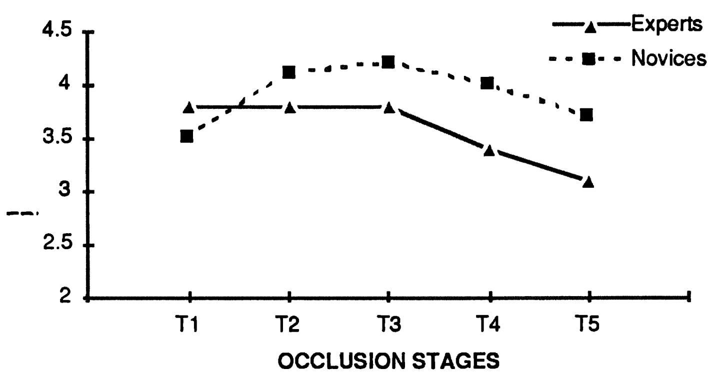


Figure 7: Error Scores Across Stages of Occlusion.

5. STUDY 2: BASEBALL BASE STEALING

The second in the series of studies examines the perceptual features that a base runner uses in deciding to steal a base or return to the base. These decisions are based on the actions of the pitcher in either pitching to the batter, or 'picking off' to the base the runner has led off from. In general coaching practice, there is discussion about the first pertinent cues provided on these two alternative actions by the pitcher (Grove [9]), and this experiment addresses this question. In the situation of a runner at first base attempting to steal a base, he or she tries to run to second base in the time it takes for the pitcher to pitch a ball to the catcher, and the catcher throw to a fielder at second base (in elite competition, something less than 4 seconds for the 25 metre distance after a lead-off from first base). Alternatively, if the pitcher attempts to catch the runner off base by throwing to the first base fielder, the runner needs to make contact with the base in less than around half a second. In these time frames, it becomes obvious that the initiation of these responses becomes critical to a successful move.

The task for base runners here is not to determine where in a sequence of action the information stipulates a response of a certain type, but more a matter of determining which of two information possibilities (originating in different parts of the body) are presented. As visual attention forms a 'spotlight' of about 1° (Hoffman & Nelson [10]), the relevant strategy for the runner is where to direct attention on the body of the pitcher - who presents an image spanning about 6° on the eye (Shank & Haywood [11]).

This experiment is concerned, then, with the response times of runners to the cues from regions of the body of a pitcher who could either pitch or pick off. The video display is edited into five arrays of pitches to direct attention to segments (S_1 - S_5) of the display as follows:

- S_1. The first block of pitches shows a random order of the full view of pitchers in their usual actions of pitching or picking off.
- S_2. The next block incorporates a mask edited into the images that covers the pitcher's body from the waist down forcing attention only to information from the upper part of the body.
- S_3. The second block incorporates a mask that covers the pitcher's body from the waist up - now forcing attention only to information from the lower part of the body.
- S_4. The third block provides a mask that covers the rear half of the pitcher's body down the seam of the pants. Runners now can only see cues from the front half of the body.
- S_5. The fourth block provides a mask that covers the front half of the pitcher's body down the seam of the pants. Only cues from the rear half of the body are now available.

The blocks are randomly displayed to subjects in a counterbalancing order, except that the full display is always shown first to obtain a free choice of attention baseline measure for each subject.

It is expected that, in this experiment, there will be differences in the response times of baseballers across the regions of information provided in the display. As the motivation for the study is the lack of concurrence as to the optimum attentional focus for a basestealer (Grove [9]), no indication is hypothesised here as to which region will provide the earliest cues from the pitchers' actions.

6. System Operation and Results

For the base stealing study, pitchers from the Australian Baseball League were video taped at a shutter speed of $^1/_{500}$th of a second viewed from a base runner's position when leading off towards second base - that is, approximately 2.5 metres from first base . In the laboratory, the video material was projected onto a screen 7 metres from the subject in a direction equivalent to the position of the pitcher on an infield diamond. The projection of the display as a life size image at a distance to the subject was provided to overcome some of the difficulties in presenting a three dimensional event as a two dimensional image. Evans [12] reports that viewing the display at the actual distance of the filming (video taping) preserves some of the distance cues (binocular convergence of the eyes which agrees with perspective constancies in the image). Attention to these factors of perception further increase the ecological validity of the paradigm. As in the previous study, the size of the pitcher's image was determined to create the same image size (approximately 6°) as in the real life situation.

The foot pads shown in Figure 2b are each fitted with two magnetic reed proximity switches, one near the ball of the foot and one at the heel. The reed switches are closed when magnets attached to the subject's feet are in close proximity, and release when the subject launches from the pads (decision time) in response to the video action. The four switch values feed through the junction box to digital ports in the data acquisition card. The through-scan optical switch is positioned on the path of the runner when returning to first base (ie, on a pick-off trial). Accuracy of the choice of direction on any trial is determined from (a) the registration in software of whether each trial is a pitch or a pick-off, and (b):

i) On a pitch: a signal from the optical switch = ERROR (subject should have run towards second base); no signal in the time-out period (2 seconds) = CORRECT

ii) On a pick-off: signal from optical switch = CORRECT; no signal in 2 seconds = ERROR.

The output of this switch is latched in the junction box before being directed to a digital port in the data acquisition card.

Data collection is continuing in this study, and final analysis is yet to be conducted. Some preliminary descriptive results are provided below to illustrate the operation of the configuration for this experiment.

The data from six baseballers is presented here. The group has a mean playing experience of 10.1 years, and all have represented Western Australia in junior and/or senior baseball. Four have playing experience in the Australian Baseball League and two of these have played professionally in the USA.

Subjects were required to watch the video display of right handed pitchers (left handed pitchers have been video taped and will be tested later) and respond to the pitcher's action by 'stealing' toward second base or returning to first base. In each block, five pitchers are shown throwing one pitch and one pick-off each. The ten trials are randomly ordered within the block. Before each display block (ie, different occlusion segments) two practice trials are given at the forthcoming type of display. Decision times (DT) and accuracy scores of each subject were automatically recorded by the computer as described above.

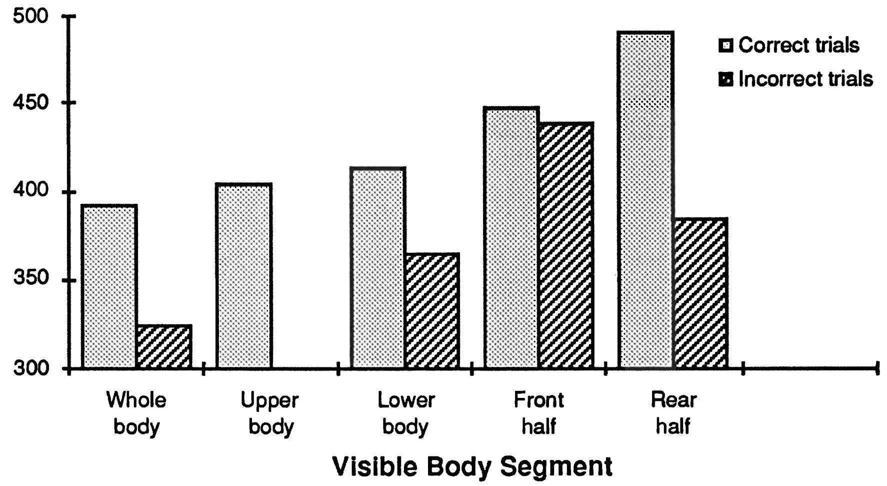


Figure 8: Decision times for the five displays of pitchers (note - there are no incorrect trials in upper body displays)

Figure 8 shows the response times for the base runners. While the low numbers of subjects, and particularly the few incorrect trials, preclude a detailed statistical analysis here, overall it is seen that the fastest response times occur when the whole body is displayed in the video. A small increase in times for the top half and bottom half views was found, and greater increases in response time above the full view are provided from front half/back half views.

This would suggest some holistic strategy by runners to view the body of the pitcher as a whole. This is dubious in terms of the 1∞ 'spotlight' of visual attention (Hoffman & Nelson [10]) that suggests that peripheral vision in humans is not suitable for such fine perceptual detection of cues. A more plausible explanation is that runners have a preferred point for attention on either the top or bottom half of the pitcher, and this generally varies equally across subjects (this would support the unequivocal views of coaches about the best point to fixate during the pitch). This provides a lower mean response time for the full body view when each runner focuses on their favoured point on the pitcher's body. The scores for top/bottom view will then only marginally increase as only half of the subjects will be disadvantaged on each display. It may well be that no runners usually attend exclusively to a front/rear point on the pitcher. Therefore, when this view is presented all runners have difficulty in determining the first cues provided in the display.

Of interest are the response times for incorrect trials. Generally, these times are faster than correct trials, indicating a speed/accuracy trade-off by runners. This means that any runner can be 100% accurate if they delay their decision long enough. In the quest for speed in the actions to return or steal, the risk of an incorrect response increases as the time to decide on the actions of the pitcher shorten.

7. FUTURE DEVELOPMENTS FOR TECHNOLOGY-BASED APPLIED SPORTS RESEARCH.

The video-based simulation for baseball skills described above has overcome some of the difficulties of contextual relevance previously encountered in experimentation in applied sport performance . This configuration, in particular, demonstrates a cost effective use of technology where more elaborate systems are not available. It is expected that higher level systems will become within the reach of smaller research centres in time, however, as funding is applied in this direction, some considerations are worthy of attention.

Firstly, *video disk media* offers greater potential for the simulation of sports situations. There are at least three advantages of this video format over video tape:

(i) the need to edit raw video of sports action is reduced. With fast transit across the disk surface, start and end points of sequences are easily determined and accessed, leaving unwanted frames isolated on disk.

(ii) reading off the disk at any segment eliminates the tape transit time that occurs when cueing to start-points in a sequence.

(iii) disk access is virtually silent, eliminating tape pick-up, motor winding and reversal noise that intrudes on the contextual reality of the experimental environment.

Another enhancement, and reduction of experimenter intrusion can be provided by *including instruction and non-visual information in the video display*. For the strategic information utilised in study 1, this could have been presented as a broad view of the field with the base runners included and showing the scoreboard in the scene. Other contextual information may include crowd noises, umpire's messages, etc.

The images provided can be improved by the use of *high speed cameras* to present a more continuous display of visual information. Standard video (25 frames per second) refresh the display each 40 milliseconds. While this might be adequate for general viewing, it means that in high speed action (eg, the rotation of the pitcher's arm) at $^{1}/_{500}$th of a second shutter speed, there are gaps of 38 msecs of 'information' occluded from the display. If the differences in responses to visual information of the most elite athletes and the next rankings are in the order of a few milliseconds, this may not be detected with standard video refresh times. The hundreds of frames per second provided by high speed cameras may be necessary to finally examine the subtle differences that define the performance variations among elite athletes.

The use of *full-size images* is necessary to the paradigm. While it has been explained above that perceptual cues are maintained with this style of display, real-life relationships are also facilitated in full size displays. For example, in facing baseball pitches, the trajectory of a curve ball in reality may be relatively high, and then 'break' to pass around calf height. On a small screen, this motion is depicted, but the ball after breaking is still somewhere around waist height and does not provide the 'feel' of the actual batting situation.

Moving from the perceptual input for subjects, there can also be enhancements to the *analysis of subjects' output responses*. This would address such factors as:

a) verbalised statements from subjects about operation, accuracy, etc in the experiments. There is no real support that elite athletes can translate what is usually a skeletal-muscular response into a verbal description and maintain reliability of the data provided (Berry & Broadbent [13]; Nisbett & DeCamp Wilson [14]).

b) allow greater flexibility for subjects to act in their usual manner for the skill.

As examples, from the studies above, the use of a passive motion analysis system could have; (i) tracked the bat head in study 1 (experiment 1) as the batter 'hit' the pitch. The position of the bat head over the plate may have been a more precise measure than the derived grid error score used; (ii) in study 2, a similar motion analysis system could have monitored any body part of the subject and recorded the first movement of sufficient magnitude to indicate the decision. This would allow the subject to make minor movements in the 'ready' position without tripping the feet movement switches. More importantly, it would overcome 'false starts' where the subject makes an early decision - but in the wrong direction. If he corrects within the 2 second time-out period of the optic switch monitor, this can be designated as a correct and very fast response (therefore classed wrongly as strong positive data).

Finally, the inter-active video *simulation of sport has a place in training sessions*. A major determinant of skilled performance is the amount of consistent practice undertaken (Schneider & Fisk [15]). In many bat and ball sports, the provision of such consistent visual information for a whole team on one training night is a major handicap to the development of elite skills. It is conceivable that an inter-active video simulation as described in this paper may be one way to overcome this problem. In one baseball scenario, a player could set up in a batting situation and ask the computer to present a random array of pitch types from a library of displays on a video disc. Feedback on performance can be provided at pre-determined intervals, and after twenty or so pitches the batter may ask for a selection of pitches at which he has performed poorly in the first bracket (eg, left handed pitchers, curveballs, etc). With the ability to train in any weather and any time without the help of an array of expert pitchers, it may be expected that a developing batter could perform enough repetitions to accelerate the acquisition of the skill beyond the rate found from the usual training provided by live batting practice only.

8. CONCLUSION

This paper has outlined the development of a cost effective inter-active video simulator for applied sport research. At this stage, the centralised control and interfacing technology is able to provide contextually relevant displays of sports 'stimuli' in a stable control, timing and communication configuration. With this technological platform, various subject interface mechanisms can be designed to transmit response data to the computer from the natural actions of the athlete while operating to the video display.

The studies described in the paper illustrate the success to date in achieving ecological validity for examining baseball skills. In both batting and base stealing, the operational specifications for researchers and subjects suggest that the environment was a fair representation of the actual sport environment. Further, the illustration of the expected differences in cognitive operation in study 1 (experiment 1), indicate the valuable potential for examining the differences that exist between experts at a skill and novices at the skill who are yet to learn the subtle cues that facilitate expert level performance. Such face validity for the paradigm is the start of more rigorous assessment of the technology that will include utilisation in more sports.

The future development of the simulator is assured as the demand for continuous access to training resources increases. The suggestions for refinement that are outlined above are already feasible additions to the configuration of equipment. The degree of funding is probably a greater constraint of progress than any engineering shortcomings that may be present in the suggestions. Certainly the simulator shows the potential to be a viable inclusion in sport development programs to accelerate the learning by athletes of the essential visual information that facilitates elite performance in sports.

REFERENCES

[1] Anderson J. R., Cognitive Psychology and Its Implications (3rd ed). (New York: W. H. Freeman 1990).

[2] Abernethy B., Anticipation in squash: differences in advance cue utilisation between expert and novice players. *Journal of Sports Sciences, 8*, 17-34 (1990).

[3] Glencross D. J. & Paull G. C., Expert Perception and Decision Making in Baseball. In S. Serpa (Ed.), *Proceedings of the 8th World Congress of Sports Psychology*, pp 356-359, (Lisbon, Portugal, 1993).

[4] Houlston D. R. & Lowes R., Anticipatory cue utilisation processes amongst expert and non-expert wicketkeepers in cricket. *International Journal of Sport Psychology, 24*, 59-73 (1993).

[5] Paull G. C. & Fitzgerald D., Addressing Ecological Validity Through Inter-Active Video Based Sport Simulation. In S. Serpa (Ed.), *Proceedings of the 8th World Congress of Sports Psychology*, pp 175-179, (Lisbon, Portugal, 1993).

[6] Sony, *Video Recorder VVH-2XXX Series: Protocol of Remote - 2 (9 pin) Connector*, 1st edition - revised. Japan; Sony Corporation (1990).

[7] National Instruments, *LabVIEW 2 User Manual*. Austen (Texas); National Instruments Corporation (1991).

[8] Sony, *Technical Manual for VO-9800P VCR*. Japan; Sony Corporation (1989).

[9] Grove J. R., Visual processes in baseball: A glance at the literature and a comment on research possibilities. *Journal of Applied Research in Coaching and Athletics, 4,* 176-194 (1989).

[10] Hoffman J. E. & Nelson B., Spatial selectivity in visual search. *Perception and Psychophysics, 30,* 283-290 (1981).

[11] Shank M. D. & Haywood K. M., Eye movements while viewing a baseball pitch. *Perceptual and Motor Skills, 64,* 1191-1197 (1987).

[12] Evans L., Automobile-speed estimation using movie-film simulation. *Ergonomics, 13,* 231-237 (1970).

[13] Berry D. C. & Broadbent D. E., On the relationship between task performance and associated verbalisable knowledge. *The Quarterly Journal of Experimental Psychology, 36,* 209-231 (1984).

[14] Nisbett R. E. & DeCamp Wilson T., Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259 (1977).

[15] Schneider W. & Fisk A. D., Degree of consistent training: Improvements in search performance and automatic process development. *Perception & Psychophysics, 31,* 160-168 (1982).

THE MATHEMATICS OF BICYCLING

W.H. Cogill[1]

Abstract

The total energy of a bicycle can be written as the sum of its potential and kinetic energies. It is shown that the potential energy is always the larger of the two. The kinetic energy is negligible under ordinary circumstances: a plot of the total energy during a road traverse is not distinguishable from the plot of the altitude against the distance along the road. The forms of the major resistance are different from uphill to downhill. The resistance during a climb is almost entirely due to the gradient. During a downhill surface rolling and wind resistance predominate. Finally the gyroscopic effect of the wheels is considered. The reaction from the road surface is altered when turning. This is an effect which is not significant under ordinary conditions of road riding, but which may have an application under close competitive conditions.

1. INTRODUCTION

The object of this note is to identify the ways in which a cyclist's energy is expended. The significance of various energy sinks - the gradient, wind, frictional and rolling energy losses - 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. This note is written from the point of view of a mountain cyclist who is mainly in off-road terrain.

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
Q	Lagrangian generalised force	w	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

[1] School of Civil Engineering, UNSW

3. TERRAIN

A typical road profile in the Blue Mountains, New South Wales, Australia, is shown in Figure 1. This figure is replotted from reference [1]. The road, a fire trail, is unsealed except where it approaches a built-up area. The extremes in altitude are approximately 300 m apart. If the terrain chosen were flat and open, the potential energy due to altitude can be replaced by potential due to wind - energy expended when riding against the wind can be recovered when riding with a following wind. At a speed of 10 km/h, the energy due to motion corresponds with an additional 0.4 m in potential, which is negligible on any scale which it is practicable to plot.

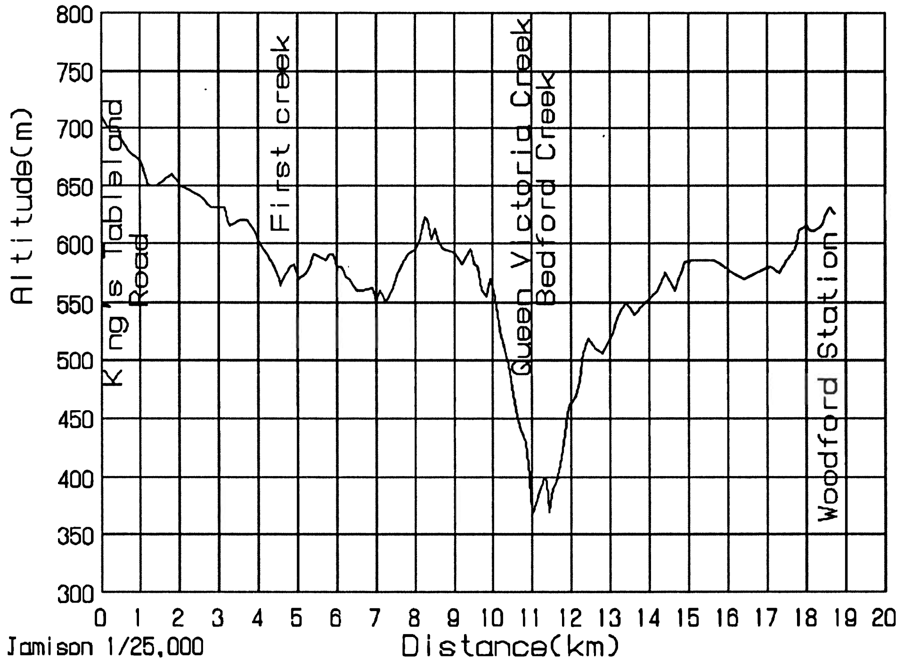


Figure 1: Profile of typical terrain, Anderson's Fire Trail, Blue Mountains, New South Wales, Australia[1].

The power needed at a point can be estimated from data such as that given in Table 1. The rolling friction can be calculated by measuring the distance required to stop without braking from a known velocity while travelling on a flat surface. The gradient can be measured directly, and the wind resistance can be estimated using Stokes' or other formulae.

Table 1

Friction: smooth surface	0.05g
Gradient, 10 %	0.1g
Wind, 20 km/h head at 5 km/h	0.1g

The power output for a load of 100kg moving at 5km/h is $100 \times \frac{5 \times 1000}{3600} \times 0.25 \times 9.8 = 340\text{W}$.

The total energy expended during a day's ride, and the approximate output power for some grades of one-day cycle rides is shown in Table 2 and in Figure 2. The classification is based on that used by the Bicycle Institute of New South Wales in classifying rides[2].

TABLE 2

RIDE CLASSIFICATION (using notation employed by the Bicycle Institute of New South Wales[2])	MEDIAN TOTAL ENERGY OF RIDE (JOULES)	OUTPUT POWER (WATTS)
E - EASY: Any healthy person. Little cycling experience	1.0E+05	40 - 100
M - MEDIUM: Teenage and healthy adult with geared bikes	2.5E+05	80 - 200
MH - MEDIUM-HARD: Fit teenagers, athletic adult	6.25E+05	150 - 300
H - HARD: Experienced cyclists of above average ability	1.25E+06	200- 400
XH - VERY DIFFICULT: Fit cyclists with good stamina	6.25E+06	300 - 800

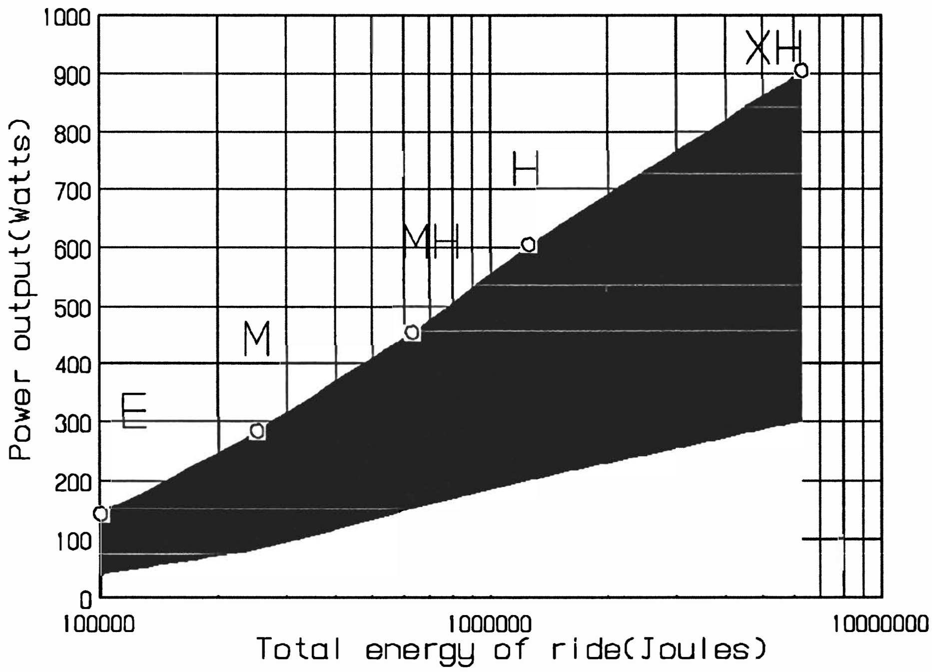


Figure 2: Expected range of power output for a load of 100kg (980 N)

The output power shown in Table 2 and in Figure 2 is required for up to an hour in the MH to XH, (Medium Hard to Very Difficult) ranges. In the E and M (Easy and Medium) rides the power is maintained for up to half an hour. The values shown were estimated by comparison with a calibrated indoor bicycle.

4. ANALYSIS

Let the kinetic energy of a bicycle be T and the potential energy be V. Write the Lagrangian as $L = T - V$, and hence the Hamiltonian equation is

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = Q,$$

where Q is the generalised force on the bicycle. It is more informative to write the Lagrangian as

$$\frac{d}{dt}\left(\frac{dT}{d\dot{x}}\right) + \frac{dV}{dx} = F.$$

Here F is the propulsive force on the bicycle, and T and V are, as before, the kinetic and potential energies of the bicycle, and x is the distance measured along the road. The potential energy term is far larger than the term which involves the kinetic energy T. A cyclist exerts effort not to create kinetic energy (to go fast), but to overcome the resistances of gradient, wind and friction. The relative importance of these factors varies of course under different terrain, road and weather conditions.

A bicycle consists of a single mass (the rider and bicycle) supported by two rotating masses (the two wheels). The wheels have a moment of inertia in the direction of the axles, and two more moments of inertia at right angles to the axles. The latter two moments of inertia are equal. The wheels possess axial symmetry, characteristic of a gyroscope. The angular momentum and rotational kinetic energy of the wheels are conserved. In straight travel, the momentum vector of the wheels is at right angles to the direction of travel. The momentum vector of the rider and the bicycle frame is in the direction of travel. The rotation of the wheels exerts a stabilising effect on account of their gyroscopic effect. Steering causes a change in the direction of the front wheel. The gyroscopic effect resists this. It is interesting to consider the overturning moment on the frame and rider caused by steering the front wheel.

The moment applied to the front wheel can be written as a vector quantity as follows[3]:

$$M = \frac{dP}{dt} = \frac{I}{\theta}\frac{dw}{dt} + \frac{I}{\theta}(\theta - 1)\frac{d\omega_{z'}}{dt}k' + \frac{I}{\theta}(\theta - 1)\,\omega_{z'}\,\frac{dk'}{dt} \qquad (1)$$

This moment is applied to the frame as a result of steering the front wheel. This alters $\omega_{z'}$, and the moment is balanced by heeling the bicycle, which alters the orientation vector $\boldsymbol{k'}$. If the bicycle is not heeled, the overturning moment applied to the frame and rider causes the bicycle to overturn. This occurs merely by turning the handlebar, even before the bicycle frame has started to turn. The mechanics of the steering movement is expressed by Goldstein[4] as follows:

Thus the w vector will always move such that the corresponding normal to the inertia ellipsoid is in the direction of the angular momentum. In the particular case under discussion the direction [..of w] is fixed in space and it is the inertia ellipsoid(fixed with respect to the body) that must move in space in order to preserve this connection between w [..and k'].

When the bicycle and rider enter the turn, a radial centrifugal force is developed. It operates in the same direction and causes the same overturning effect on the bicycle as does the gyroscopic moment.

Gallivoti[5] 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 velocities in the remaining two directions.

$$\begin{aligned} I\omega_z^2 + \theta I\omega_x^2 + \theta I\omega_y^2 &= 2T = \text{constant} \\ I^2\omega_z^2 + \theta^2 I^2\omega_x^2 + \theta^2 I^2\omega_y^2 &= [Iw]^2 = \text{constant} \end{aligned} \qquad (2)$$

Equations (2) represent the fact that the kinetic energy and the angular momentum of each of the two wheels, the front wheel in particular, are constant. As any one of the components of angular velocity is changed by steering the front wheel, the change must be accompanied by a change in each of the other two components. In particular, a change must occur in the component of momentum at right angles to the plane of symmetry of the bicycle. This causes the bicycle to overturn and is normally counterbalanced by heeling in the direction opposite to the overturning moment.

Karnopp[6] treats the problem of inverting equations such as equation (1) in order to obtain the components of the vector *w* in terms of the applied moment caused by either superelevation of the roadway or by heeling of the bicycle. This problem is more general than the problem treated using Gallivoti's expressions. Its use may be justified as data become available.

REFERENCES

[1] Central Mapping Authority of New South Wales, Bathurst, New South Wales. Topographic Map 1:25,000 Second Edition. Jamison 8930-II-N.

[2] Bicycle Institute of New South Wales, Sydney, Australia. *Push On,* Vol 18, No 1, p. 7, 1994.

[3] Joos, G., Theoretical physics, p. 150 ff, (Blackie, London, 1950).

[4] Goldstein H., Classical mechanics., Second Edition, p.206, (Addison-Wesley, Reading, Massachusetts, 1980).

[5] Gallivoti G., Elements of mechanics, p.311, (Springer, New York, 1983).

[6] Karnopp D., On inverse equations for vehicle dynamics, *Vehicle System Dynamics*, Vol 20, 371-379 (1991).

AN ADJUSTIVE RATING SYSTEM FOR TENNIS AND SQUASH PLAYERS

Stephen R. Clarke[1]

Abstract

The advantages of a rating system for tennis and squash players, incorporating elite down to club players, are discussed. Some problems with rating systems currently in use are highlighted. The use of exponential smoothing as a rating method is proposed, and its possible application to tennis and squash demonstrated. Possible solutions to some minor difficulties are suggested.

1. INTRODUCTION

Imagine my difficulties in organising a tennis match while I am at the conference.

Steve: Bill, would you like a game of tennis?
Bill: I might be interested, but how good are you?
Steve: Well I play section A1 mixed on Thursday night Blackburn district competition, but section A4 on Saturday Eastern Suburban competition. Years ago I played No 3 in C special pennant.
Bill: I play veterans pennant section 2 in Victoria, but I have no idea how that compares - we will just have to hope we are the same standard.

As a contrast consider golf:

Bill: My handicap used to be 1, but over the last few years I have dropped to 5. What's yours?
Steve: I don't have an official handicap, but I play to a handicap of about 28. I heard that American say he is off 3, why don't you try him for a game?

Golf has a recognised system of rating all players from different courses, states and countries. The system uses the same scale for the best golfers in the world down to the weekend hacker. A golfer can always strive to improve or hold his handicap. He can compare his current handicap with his best-ever handicap. He can compare his best with another player's best, even if they played in different eras.

On the other hand, tennis players have trouble determining relative standards even within a club. It would be nice to have a system that rated tennis or squash players, from the ordinary club player to the best in the world, in a similar manner to the handicap system for golf.

This article broaches the possibility of such a rating and suggests some possibilities.

---

[1]Mathematics Department, Swinburne University, Hawthorn, Vic.

2. A Current Squash Rating System

Various rating systems already exist in racquet sports. There is of course a computer tennis ranking system for elite players. Players accumulate points depending on how far they progress in each tournament, but not on whom they defeat and the margin of victory. A win over a strong opponent carries the same reward as a win over a weak opponent. A close loss to a strong opponent earns no more points than a thrashing by a weak opponent. Stefani [1] calls this a cumulative scheme, as distinct from an adjustive scheme whereby players' ratings are adjusted up or down depending on their performance relative to that predicted by their rating. Hoffman [2] uses the terms 'performance' to describe what is measured by a cumulative scheme, and 'ability' to describe what is measured by an adjustive scheme. We follow that terminology here.

Another example of a cumulative scheme is that used by the Victorian Squash federation to rank its players in pennant competition. Players are awarded rating points according to the position at which the match was played and the result of the match. The rating points allocated to the players at the four different positions within each team are outlined below:

Position Played	Maximum Points	Match Outcome	Maximum Allocation
1	29 Points	3/0	100%
2	20 Points	3/1	90%
3	14 Points	3/2	70%
4	10 Points	2/3	30%
		1/3	10%
		0/3	0%

For example, a player who wins 3 games to 1 at number 2 will earn 90% x 20 = 18 points while his opponent earns 10% x 20 = 2 points.

While such a system might give a reasonable measure of a player's performance over the year, there are several 'problems'.

- The current rating points are being applied to each individual grade in the competition, and therefore only players within the same grade can be compared.

- Because the system is cumulative, players who play more will receive more points. Of course this may be what the organisers wish to do. However a player who plays half the matches, and wins them all will be ranked no higher than a player who plays all the matches and wins half. Similarly, a good player in a lower section who is called on for half the matches to fill in at a higher grade will be ranked low in both grades.

- In common with most adjustive schemes the scheme also suffers in that a good effort can go unrewarded. Unlike the ATP tennis ranking, a player can earn points for a loss, but an extremely close 3 set loss still earns zero points.

- This rating table has a very strong tendency to cluster players together according to the position at which they are playing. For example, players performing at position one would tend to be rated all together at the top of the table, whilst players performing at position four would tend to be clustered together at the bottom of the rating table. In practice we would expect player's abilities to be spread more uniformly.

- The points are awarded on the basis of position, not quality of opponent. Similar wins (or losses) against the same opponent are given more points when that match is played at a higher position.

- The association has a lot of difficulty incorporating tournament results into the rating. Good players, absent from the competition to play tournaments, fall down the rating because they are not accumulating points. They do not receive any recognition for good performances in tournaments.

Some of these points could be rectified by making adjustments to the scheme. However such adjustments usually tend to be arbitrary and not based on any underlying theory. A Table Tennis Association has a ranking scheme which takes 4 pages to describe, that incorporates standard points (ranging from 1 to 12 depending on the difference between the opponents ratings), match weighting factors (a multiplication factor ranging from 1 to 5 in steps of .5 depending on the importance of the match), tournament bonus points (ranging from 0 to 80 depending on weighting and how far a player progresses), pennant bonus points (depending on grade), round robin bonus points (depending on the number in the tournament and the finishing position), negative points (from 5 to 15, and only deducted from the loser if they are ranked well above the winner) and dummy players (players not ranked, but given a dummy ranking by the organisers).

However all cumulative schemes essentially measure a players overall performance in the competition. They are not useful for prediction, and do not measure ability. There is a better (and simpler) way.

3. EXPONENTIAL SMOOTHING SYSTEM

The exponential smoothing method overcomes most of these problems. Players ratings are adjusted depending on their expected performance based on their current ratings. The only data that needs to be stored is the player's single rating.

The system works on the margin between the two players. In football the margin might be the difference in points, in squash the difference in games (we will discuss this later). For the moment suppose we have some way of determining a margin that reflects the closeness of the match. Suppose player A is rated 200, player B rated 170. Then we expect A to beat B by 30. If in fact player A beats B by 50, that implies either A is rated too low or B too high. The difference between the actual margin and the expected margin is 20, so we adjust each player's rating by some percentage (say 20%) of this. So Player A goes up by 4 to 204 and player B goes down by 4 to 166. The basic formula is

New Rating of Player
= Old rating +0.2 *[actual margin - predicted margin]
= Old rating +0.2 *[(player score- opponent score)
- (player rating - opponent rating)]

Again in the above if A loses by 10 then we have:
New rating of A = 200 + 0.2*[-10 - 30] = 192.
New rating of B = 170 + 0.2*[10 - (-30)] = 178.

Such a system is used by Clarke [3], [4] to rate and predict football. He shows it is just as accurate as expert tipsters. Strauss & Arnold [5] suggest its use for racquetball. In chess it is called the ELO system and has been used for many years to rate all chess players, and compare players from different eras. Elo [6] gives a lot of the theory underlying the system but the book is rather difficult to obtain. The mean and the range of the ratings can be chosen arbitrarily at the expense of slight complications in the formula. There are some problems in operating such a system, particularly to preserve its integrity in rating players over a long period of time. For example the treatment of additions through new players and deletion through retirement or disinterest has to be handled carefully. However such a system has many advantages.

- It is completely objective. There is no subjective input required from organisers at all. For example, you do not have to decide how position 1 compares with position 2 - in general players with higher ratings will play in position 1. Administrators do not need to decide the importance of tournaments. Tournaments are treated the same way as any other match - the more important the tournament then the higher will be the ratings of the participants.

- The system has an underlying theoretical base. A given difference in the rating of any two players translates to an expected margin, and a probability of winning. Administrators have an idea of the difference in match performance between any 2 players. Players know how much a given rating improvement will affect their court performance.

- The system needs to store only the rating of players and so is suitable for handling a large number of players. Compare this with the ATP tennis ranking which requires the storage of the last 12 months' results.

- The system makes every match important - and every game important. A player faced with a certain loss will still strive for every possible game. Superior players will not just take it easy if winning comfortably, as the margin of victory will affect their rating. This should improve the overall standard of the game.

4. APPLICATION TO TENNIS OR SQUASH

For squash and tennis we need to investigate how 'performance' is measured - by sets, games, points etc. For squash, games seems too crude, so we need to incorporate points. This technique must ensure that the player actually winning the match is always allocated more performance points, even in the situation where the losing player actually wins more points than the winning player (ie. Player A beating Player B 0/9 0/9 10/9 10/9 10/9). I will use SPARKS (For Set - Point mARKS), but something like 'performance points' could be used. 20 sparks are allocated for winning a game and 1 spark is allocated for winning a point.

Case 1: A wipeout in 3 Sets (Player A beats Player B 9/0 9/0 9/0)
Player A Score: (3 Games x 20) + (27 Points x 1) = 87 sparks
Player B Score: (0 Games x 20) + (0 Points x 1) = 0 sparks
so margin for rating purposes is 87

Case 2: A Very Close 3 Setter (Player A beats Player B 10/9 10/9 10/9)
Player A Score: (3 Games x 20) + (30 Points x 1) = 90 sparks
Player B Score: (0 Games x 20) + (27 Points x 1) = 27 sparks
so margin for rating purposes is 63

Case 3: A one sided 4 Setter (Player A beats Player B 9/10 9/0 9/0 9/0)
Player A Score = (3 Games x 20) + (36 Points x 1) = 96 sparks
Player B Score = (1 Games x 20) + (10 Points x 1) = 30 sparks
so margin for rating purposes is 66.

Case 4: A Close 4 Setter (Player A beats Player B 9/5 6/9 10/9 9/7)
Player A Score = (3 Games x 20) + (34 Points x 1) = 94 sparks
Player B Score = (1 Game x 20) + (30 Points x 1) = 50 sparks
so margin for rating purposes is 44

Case 5: Closest possible 5 Set Win (Player A beats Player B 0/9 0/9 10/9 10/9 10/9)
Player A Score = (3 Games x 20) + (30 Points x 1) = 90 sparks
Player B Score = (2 Games x 20) + (45 Points x 1) = 85 sparks
so margin for rating purposes is 5

Since the maximum margin is 87 sparks the rating formula would need to be adjusted slightly to allow for this. If in going from a beginner to a world champion we have (say) 10 levels where a higher level player defeats a lower level player about 9-0, 9-0, 9-0, this implies a range of ratings about 900.

For tennis, most matches are of only 3 sets, and usually games are recorded but not points. While a more accurate measure may be obtained by including points, there is no use having a system that uses data not normally available. A possible performance measure for tennis could be 6 for each set and one for each game. Thus

Case 1: 6-0, 6-0 victory gives 2 sets 12 games to 0 sets 0 games for 24 sparks to 0

Case 2: 6-4, 6-4 victory gives 2 sets 12 games to 0 sets 8 games for 24 sparks to 8

Case 3. 6-5, 5-6, 6-5 gives 2 sets, 17 games to 1 set 16 games for 29 sparks to 22.

Case 4: 6-5, 0-6, 6-5 to give 2 sets, 12 games to 1 set, 16 games for 24 sparks to 22.

Note – the system doesn't need the actual scores, only the total number of sets, games etc. ie 2 sets to 1, 15 games to 12. Many squash and tennis score sheets require the totals to be entered in this manner anyway.

5. Some Minor Problems

English and American Scoring in squash

In English scoring, games are played to 9 points and players only win a point when they win a rally in which they serve. Winning their opponent's serve only gives them the right to serve the next rally. Thus a player who loses 9-0 may have won several rallies. In American scoring a game is up to 15 and every rally won scores a point. Thus a 9-3 win in English may be equivalent to 15-9 in American. If this complication needed to be accounted for there are several papers that address the relationship between English and American scores. eg Clarke & Norman [7], Strauss & Arnold [5], Goldstein [8].

Doubles

A lot of club tennis is played as doubles consisting of only one set. A player would need a separate doubles rating, and the rating of the pair determined as the average rating. Each individual player would then be adjusted in the same manner as above. Hoffman [2] gives details of using an adjustive scheme to rate individuals in a team competition. For one set matches, some adjustment to the predicted margin would be necessary.

Testing

Naturally such a system would need to be tuned and tested. There are various parameters that can be altered (the smoothing constant, the relative importance of sets, games and points etc). In a project for The Victorian Squash Association some testing of rating systems was performed by Clarke, Bucci et al [9]. There appears to be two ways of testing.

- Use simulation to see how well various rating schemes optimise the correlation between the ratings produced and the known ratings incorporated in the simulation.

- From actual match results, use the rating to predict results of future matches. Clarke [3] describes how several years data was used to optimise the value of the parameters in applying the method to football.

6. Implementation

There are various ways such a system could be implemented. At one level it could be used by a single competition to rate their players. Thus in pennant squash (or tennis) it could be run if necessary in tandem with a cumulative scheme. The cumulative scheme could still be used to decide 'most successful player' etc, but the smoothing scheme would be invaluable in deciding tournament seedings, promotion/relegation of teams, grading of new teams or other matters.

Note that the correct data must be recorded. The system requires the date of the match (probably optional, but strictly speaking ratings should be updated in order of match date; in most competitions, round number would do just as well), the names (or unique identifiers) of the two players and the match result. While this seems a minimum requirement for any match record system, different forms of recording results may mean results are not in this form. For example the computer system for the Victorian squash results kept records on a club, grade and position basis, so it was not easy to match opponents Clarke, Bucci et al [9].

Once the rating system ranges over a number of competitions, there is a necessity for unique lifelong registration numbers. There may be a possibility of nesting, with each level (state, competition, club) handling matches between their own players. While a player's rating actually moves up and down slightly with each match, it may only be necessary to run through the matches and publish updated ratings at set intervals of (say) every 3 or 6 months.

How such a system would be funded I have not investigated. However with increasing use of computers to store and process competition results, along with the increasing use of modems and electronic means of transferring data, the means of setting up a nationwide rating scheme for tennis or squash players is available. I would certainly be willing to help any competition administrators to institute such a system. I would be particularly interested in hearing from anyone with competition data in a form suitable for testing the above system.

In attempting to interest administrators in such a system, I am aware of the words of David Strauss in a private communication on his efforts to introduce a similar scheme in the U.S.A. "We didn't get too far. I didn't persuade the national people to go for it. We also developed a tennis system but didn't get it adopted. The whole thing needs marketing skills. Good Luck". While I am sure my marketing skills are no better than David's, it seems that at a conference on Mathematics and Computers in Sport, collectively we have the technical ability and the interest in the area to begin its implementation. I would be interested in discussing the possibilities with anyone at the conference.

REFERENCES

[1] Stefani R. Applying least squares to team sports and Olympic winning performances, *Mathematics and Computers in Sport*, N. de Mestre, editor, Bond University, Qld. 43-59, (1992).

[2] Hoffman D. A team rating system, ASOR 11th National Conference, Gold Coast, Qld. p 363, (1991).

[3] Clarke S.R., Computer Forecasting of Australian Rules football for a daily newspaper, *Journal Operational Research Society*, **44** (8) 753-759 (1993).

[4] Clarke S. Computer and human tipping of AFL football - a comparison of 1991 results, *Mathematics and Computers in Sport*, N. de Mestre, editor, Bond University, Qld. 81-93, (1992).

[5] Strauss D. and Arnold B.C., The Rating of Players in Racquetball Tournaments, *Journal of Applied Statistics*, 36 (2) 163-173 (1987).

[6] Elo, A.E., *The Rating of chess players, past and present*, (Batsford, London, 1978).

[7] Clarke S.R. and Norman J.M., Comparison of North American and International Squash Scoring Systems - Analytical Results, *Research Quarterly*, 50 (4) 723-728 (1979).

[8] Goldstein L., Squash. What difference does a Scoring System Make? *Australian Senior Mathematics Journal*, 6 (1), 50-52 (1992).

[9] Clarke S., Buchi D., Bellizia M., Peart A., Patterson B., Plant K., Rating Victorian squash players, (Department of Mathematics, Swinburne University of Technology, 1993).

A METHOD FOR DETERMINING SAND-RUNNING EFFICIENCY

Rodney Barrett[1] and Robert Neal[2]

Abstract

The aim of the present study is to measure the mechanical efficiency of sand running by developing a technique to quantify the total work done at a range of running speeds for which the metabolic cost has been determined. Three methods for calculating work done were identified in the literature. The dynamics method was chosen so as to properly represent the sources of the mechanical energy expenditure. Mechanical work done will be calculated from the time integral of the muscle power curves where muscle power is defined as the product of the net muscle moment and joint angular velocity. The time history of the ground reaction force during running will be determined using a transfer function derived from acceleration-time records.

1. ENERGETICS OF RUNNING

Full investigations of the energetics of running require the simultaneous measurement of biomechanical and physiological variables. Physiological measures are made in order to estimate the metabolic cost of running. Measures of submaximal metabolic energy expenditure during running, termed running economy or VO2submax, appear to have gained widespread acceptance for this purpose (Williams [1]). The difference between running economy and resting metabolic cost is believed to reflect the increase in muscular activity and some increase in demand by supporting metabolic structures (Williams [2]). Due to ethical and methodological difficulties, biomechanical measures of muscular effort are generally restricted to the use of mechanical models governed by equations of motion that estimate joint forces and torques.

Biomechanical measures of muscular effort have be classified in the literature (Andrews [3]) as (i) instantaneous measures such as maximum or minimum of force, torque or power values, and (ii) interval measures such as average force, torque or power values, work done, change in mechanical energy and linear and angular impulses. Due to the cyclic nature of running, instantaneous measures of muscular effort do not correlate well with running economy (Martin et al. [4]). The most common means of quantifying total muscular effort cited in the literature has been through the estimation of mechanical work done, which has enabled researchers to calculate the mechanical efficiency of a runner as the ratio of mechanical work done to the metabolic energy expended (Winter [5]). An efficient runner will therefore be able to perform more mechanical work for the same energy expenditure, or will be able to perform the same mechanical work at a lower energy cost.

1 School of Exercise Science, Griffith University, Queensland.
2 Department of Human Movement Studies, University of Queensland.

2. Measuring Mechanical Work Done

There is considerable confusion as to how mechanical work done should be measured (Putnam & Kozey [6]). Some authors (Aleshinsky [7]) claim that the concept of mechanical work as it is defined in mechanics, should not be used to estimate mechanical energy expenditure (MEE) of human movement, since the completion of any number of cyclic motions, such as in running at constant velocity, results in no work done. This gives rise to an anomalous situation, since there is clearly an energy cost associated with the motion (zero work paradox). The folly of this approach is further demonstrated in downhill gait, where negative work and hence negative efficiencies are reported (Cavagna et al. [8]). Attempts have consequently been made in the literature to appropriately define global measures of muscular effort. The term mechanical energy expenditure (MEE) has been used to differentiate the proportion of total mechanical work that is due to muscular effort and hence incurs a metabolic cost (Aleshinsky [7]).

Estimates of MEE during locomotion are generally classified according to three different analysis methods.

1. Sum of internal and external work method
2. Instantaneous segment energy method
3. Dynamics method

Sum of internal and external work method

Early attempts at measuring mechanical work during running assumed that the energy changes of the whole body could be represented by the energy changes of the runners total body centre of gravity (Cavagna and Kaneko [9]; Gersten [10]; Datta [11]). This method was criticised for neglecting the energy contributions of individual body segments since reciprocal body segment movements (such as the arms and legs in running) result in changes in energy but have no net effect on the motion of the body's centre of gravity (Winter [12]). Errors associated with these techniques were estimated to be 16 % (Winter [13]). In order to account for the movements of individual segments mechanical work has also been estimated as the sum of work on individual body segments (internal work) and the total body centre of gravity (external work). It has since been demonstrated that the assumed independence of body segment motions from each other and from the total body centre of gravity is inappropriate for calculating work done in running (Aleshinsky [14]).

Instantaneous segment energy method

The work-energy relationship states that the amount of work done by a force is equivalent to the change in energy experienced by the system on which the force acts. The total mechanical energy of an individual body segment at an instant in time is defined as the sum of its kinetic (translational and rotational) and potential energies. To reveal the work done by the whole body, it is necessary to add the absolute values of the changes in energy of its composite segments. The result of these calculations will depend on what energy changes (passive transformations within segments or exchanges between segments) are allowed.

Energy transformations within a segment are evident during simple pendulum-type movements. For example during running the trunk acts as a conservative system whereby about half of its changes in potential and kinetic energy are interchanged (Winter [13]). Energy exchanges between segments occur as a result of the power flows generated by muscle moments and joint reaction forces. The deceleration of the knee during the non-support phase of running causes a simultaneous increase in the angular velocity of the shank. The transformation of mechanical energy from translational to rotational during this whip-type action can occur independent of a knee extensor control moment. Estimations that ignore either of these mechanisms will significantly overestimate the internal work done (Pierrynowski et al. [15]) as they are both known to reduce the requirement for muscular effort (Ralston [16]).

Robertson and Winter [17] identified the possible work functions that can occur between two segments connected by an active muscle. Muscles were shown to be capable of generating mechanical energy during concentric contractions or absorption for eccentric contractions. Transfer of mechanical energy between adjacent segments was shown to occur only when both segments rotated in the same direction.

Application of the instantaneous segment energies method, where consideration has been made for energy transfers within and between segments, have also been subject to criticism on the basis that they only reflect muscular MEE under certain movement conditions (Aleshinsky [18]).

Dynamics method

The probable reason for the limited use of the dynamics method in estimating mechanical work is the relative difficulty of determining joint kinetics compared with segment energies. The advantage of this method is that MEE is determined on the basis of joint power so as to properly reflect the origin of the MEE, the force moments action (Aleshinsky [7]). The power developed by the source is calculated as the product of the joints force moment and angular velocity. MEE for a running cycle can subsequently be expressed as the time integral of the sum of respective joint powers. Calculated values of MEE depend on how the influence of: (i) concentric and eccentric muscle contractions, (ii) elastic energy storage, and (iii) the activity of multijoint muscles are incorporated into the computations.

During a stretch-shorten cycle like that associated with the support phase of running, eccentric muscular contractions result in the storage of strain energy in the elastic components of the musculotendinous tissues which is returned to the system upon movement reversal. For a given movement, the effect of elastic energy storage during negative work is to decrease the muscular effort required in the concentric work phase (Asmussen and Bond-Petersen [19]). As the metabolic cost of eccentric contractions are considerably lower than those for concentric work (Abbott, Bigland and Ritchie [20]), any mechanism that leads to a decrease in concentric muscle activity will result in performance efficiencies. Two joint muscles have been proposed as a means for reducing the energy cost of a system compared to the same system with one joint muscle only (Wells [21]).

3. THE ENERGETICS OF SAND RUNNING

Beach running is an important feature of surf competition, and competitors testify to its unique demands when compared to running on less compliant surfaces. This presumption was tested by Zamparo et al. [22] who were able to demonstrate that the energy cost of sand running was increased over that for firm terrain by 15-40% at speeds ranging from 6-16 km/hr. Reduced capacity to utilise stored elastic energy and larger fluctuations in potential energy at each stride were cited as possible mechanisms underlying these differences. However, the finding that the relationship between stride frequency and running speed was the same for both sand and firm surfaces appears to contradict current thinking in practice (albeit anecdotal) which suggests that stride frequency is increased and stride length reduced when running on sand, especially for high running speeds. It is consequently of interest to the sports scientist to examine how technique is modified when running on soft sand and how this may alter mechanical energy expenditure (MEE).

Procedures for measuring the MEE of sand running using the dynamics method are outlined in the following sections.

4. MEASURING JOINT FORCES AND TORQUES

With current technology it is not possible to measure muscle forces directly, so they are quantified in terms of the net moment of all muscles crossing individual joints (Putnam and Kozey [6]). The contribution to this moment by passive joint structures such as ligaments and the joint capsule during gait are generally considered negligible. Resultant joint moments (RJMs) have been reported in the literature at a full range of running speeds for conventional (rigid) running surfaces (Winter [23]; Simonsen et al. [24]). RJMs for these studies have been determined using the inverse dynamics approach outlined in Winter [5] in which ground reaction forces (GRFs) are measured using a force platform and subsequently entered into dynamic equations of motion in conjunction with anthropometric and kinematic input (Fig 1). Segment lengths, masses and moments of inertia are based on representative data available in the literature (Zatsiorsky and Seluyanov [25], Dempster [26]). Segment displacements are recorded using a high speed video camera and the data smoothed and differentiated to provide segmental velocities and accelerations. Equations of dynamic equilibrium defined by d'alembert's principle ($\Sigma F_x=0$, $\Sigma F_y=0$, $\Sigma M_{COG}=0$) are used to solve for joint reaction forces and moments on a joint-by-joint basis from the ground up.

Measuring GRF's experienced by the runner during sand running is confounded by the physical properties of the running surface. The following section outlines a method for measuring the GRF experienced by the runner using data from a force plate buried beneath the sand surface.

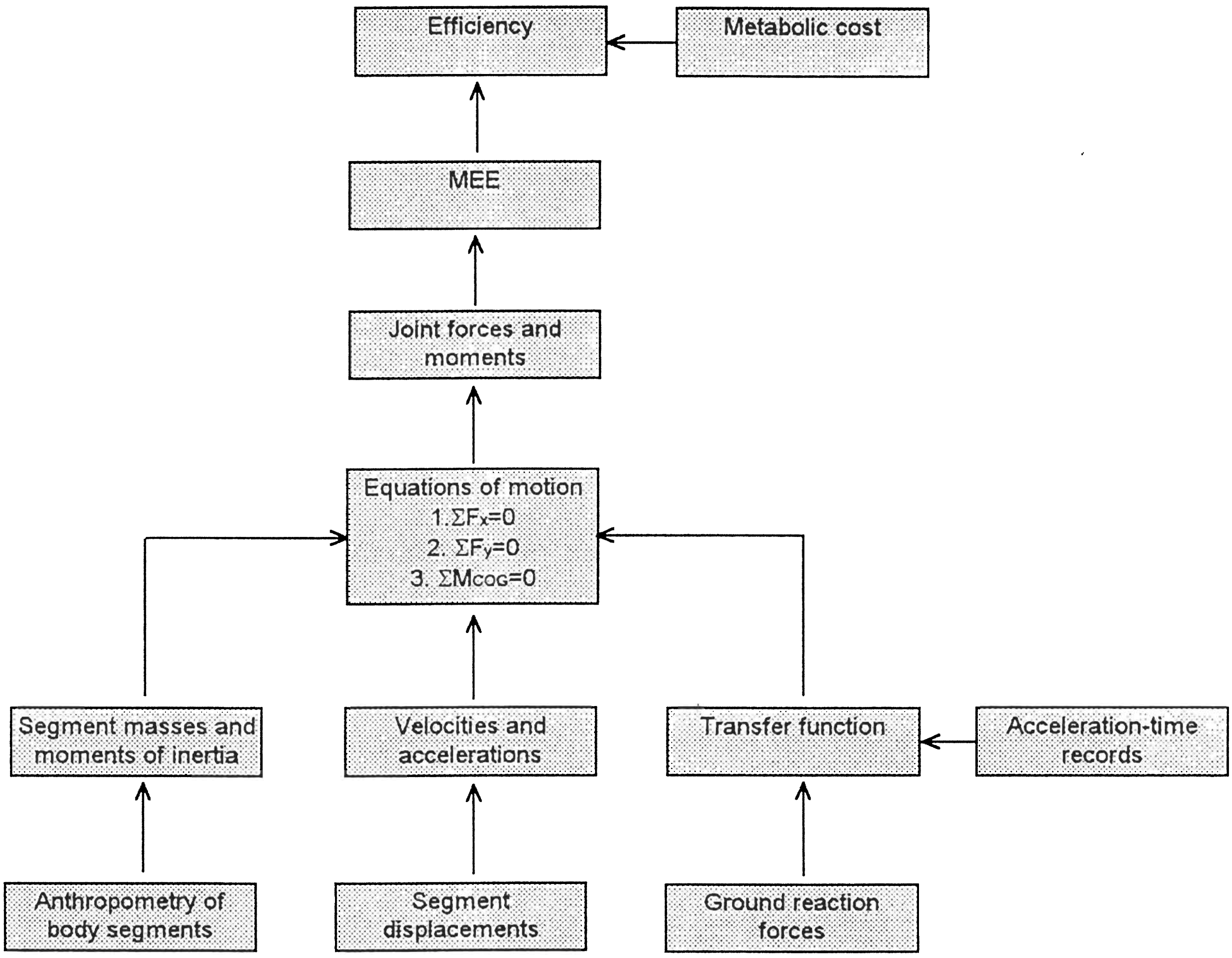


Figure 1: A method for determining the efficiency of sand running

5. MEASURING GRF'S IN SAND

McMahon and Greene [27] showed that the GRF exerted on a runner is affected by the physical properties of the surface across which the runner moves and postulated that for optimum performance track surfaces should be designed so that the track stiffness is compatible with the spring stiffness of the leg muscles. GRFs in sand running have not previously been assessed, but are likely to differ from those measured for less compliant surfaces, and will therefore have implications for running performance.

The GRF exerted against the sole of the foot in beach running will contain a wealth of information about running mechanics. For example, the time history of the horizontal GRF will provide information describing the nature of the transition from braking force to propulsive force application during the support phase of running. Vertical GRFs provide information with respect to fluctuations in kinetic and potential energy of the total body centre of gravity and the passive and active components of vertical force attenuation. Of particular interest in this study is the time history of the GRFs and the location of the centre of pressure of the vertical GRF during the support phase of sand running so that joint kinetics (forces and moments) can be determined.

GRFs are readily measured during locomotion using a force plate. However, during locomotion over soft sand the values recorded on a force plate do not reflect those experienced by the runner. Accelerometers have been used in running research to measure the 'shock' transmitted along the leg during footstrike. Nigg [28] used an accelerometer mounted on the tibia of subjects that ran over a force platform to show that the shape of the GRF and acceleration curves were similar for the initial impact peak, which indicates that accelerometers can be used to approximate the GRFs exerted beneath the subjects feet during locomotion. The current study aims to develop a transfer function that will allow GRFs experienced by the subject to be estimated from the partially attenuated forces (due to the relatively low force transmissibility of sand) exerted on a force platform buried beneath the sand.

The Standards Association of Australia [29], [30] have published methods for testing synthetic sports surfaces to assess a variety of mechanical properties. Of relevance to the present study are the tests for determining surface stiffness and slip resistance. These will be modified for the purpose of the current study as follows:

Determination of stiffness

An accelerometer mounted on a test specimen of known mass and geometry will be impacted at known velocities onto a variety of sand and reference surfaces. The axial acceleration-time history will be recorded, stored and the following dependent variables calculated: peak acceleration during impact, time to peak acceleration, severity index. Nigg (1986) has measured peak axial tibial accelerations of $200ms^2$ for one foot contact during heel-toe running. GRFs and centre of pressure of the vertical ground reaction will also be measured on an underlying force plate so that the relationship between acceleration-time and force-time records can be determined.

A displacement-transducer will be used to measure the indentation of the impact with respect to time. The displacement-time history will provide data describing peak and time-to-peak indentation. Dynamic hardness will be assessed as the product of peak acceleration, surface thickness and load weight divided by the product of contact area and maximum penetration.

Gatto, Swannell and Neal [31] employed the above methods to investigate the mechanical properties of a series of gymnastic mats and were able to simulate the GRFs during underfoot impact with these surfaces by collecting acceleration and indentation records simultaneously. In the present study force-indentation records will be supplemented by the time history of the GRF in order to further elucidate the dynamic loading response associated with the support phase of running.

Determination of slip resistance

A Road Research Laboratories skid tester will be fitted with an artificial foot and the apparatus adjusted so that the pendulum slide length meets the prescribed guidelines. Efforts will be made to ensure the consistency of the sand surface and friction readings will be made for controlled dry and wet surface conditions. These values will be used to assess the effect of various surfaces on the braking and propulsive GRFs measured on an underlying force platform.

6. POWER, WORK AND EFFICIENCY

Muscle powers will be calculated as the scalar product of the RJM and its corresponding joint angular velocity vector. The internal work done by an RJM over a stride will be quantified by integrating the power-time function of the RJM. Ignoring air resistance, the total work done by the RJMs will be calculated as the sum of the integrals of the absolute values of each RJM power-function during a complete stride, in order to account for the equal amounts of positive and negative work associated with constant velocity running on a level surface.

The energetics of running can be studied by measuring both mechanical work and energy cost to calculate mechanical efficiency (mechanical work done per unit of energy expended). Methods for evaluating the metabolic energy cost of running are also known to contribute to variations in values for efficiency (Kaneko [32]), and generally relate to the way in which the 'base-line' cost is defined and how the relative costs of positive and negative work are partitioned. While these issues are beyond the scope of the present paper, it should be stated that further research in this area is also required to ascertain its influence on running efficiency.

7. CONCLUSIONS

By assessing the effect of surface and running speed on the energetics of running it is envisaged that specific kinematic and kinetic differences will be discerned, which will have implications for performance enhancement amongst beach-running competitors and lead to a greater understanding of the causes and effects of motor control and human movement.

REFERENCES

[1] Williams K.R., The relationship between mechanical and physiological energy estimates, *Med. Sci. Sp. Ex.*, 17(3), 317-325 (1985).

[2] Williams K.R., Relationships between distance running biomechanics and running economy, in Biomechanics of distance running, (Human Kinetics, Illinois, 1990).

[3] Andrews J.G., Biomechanical measures of muscular effort, *Med. Sci. Sp. Ex.*, 15(3), 199-207 (1983).

[4] Martin P.E. Heise G.D. & Morgan D.W., Interrelationships between mechanical power, energy transfers, and walking and running economy, *Med. Sci. Sp. Ex.*, 25(4), 317-325 (1993).

[5] Winter D.A., Biomechanics and motor control of human movement (2nd ed.). (John Wiley and Sons: New York, 1990).

[6] Putnam C.A. & Kozey J.W., Substantive issues in running, In C.L. Vaughan (ed.) Biomechanics of sport, (CRC Press, Boca Raton, 1989).

[7] Aleshinsky S.Y., An energy 'sources' and 'fractions' approach to the mechanical energy expenditure problem-I. Basic concepts, description of the model, analysis of a one link system movement, *J. Biomech.*, 19(4), 287-293 (1986).

[8] Cavagna G.A., Saibene F.P. & Margaria R., External work in walking, *J. Appl. Phys.*, 19, 249-256 (1963).

[9] Cavagn G.A. & Kaneko M., Mechanical work and efficiency in level walking and running, *J. Appl. Phys.*, 268, 467-481 (1977).

[10] Gersten J. W., Orr W., Sexton A.W. & Okin D., External work in level walking, *J. Appl. Phys.*, 26(3), 286-289 (1969).

[11] Datta S.R., Chatterjee B.B., & Roy B.N., Maximum permissible weight to be carried on the head by a male worker from Eastern India, *J. Appl. Phys.*, 38, 132-135 (1975).

[12] Winter D.A., Calculation and interpretation of mechanical energy of movement, *Ex. Sp. Sci. Rev.*, 6, 183-201(1978).

[13] Winter D.A., A new definition of mechanical work done in human movement, *J. Appl. Phys.*, 46(1), 79-83 (1979).

[14] Aleshinsky S.Y., An energy 'sources' and 'fractions' approach to the mechanical energy expenditure problem-II. Movement of the multi-link chain model, *J. Biomech.*, 19(4), 295-300 (1986).

[15] Pierrynowski M.R., Winter D.A. & Norman R.W., Transfers of mechanical energy within the total body and mechanical efficiency during treadmill walking, *Erg.*, 23(2), 147-156 (1980).

[16] Ralston H.J. & Lukin L., Energy levels of human body segments during level walking, *Erg.*, 12(1), 39-46 (1969).

[17] Robertson D.G.E. & Winter D.A., Mechanical energy generation, absorption and transfer amongst segments during walking, *J. Biomech.*, 13, 845-854 (1980).

[18] Aleshinsky S.Y., An energy 'sources' and 'fractions' approach to the mechanical energy expenditure problem-IV. Criticism of the concept of 'energy transfers within and between links', *J. Biomech.*, 19(4), 307-309 (1986).

[19] Asmussen E. & Bonde-Petersen F., Apparent efficiency and storage of elastic energy in human muscles during exercise, *Acta Physiol. Scand.*, 92, 537-545 (1974).

[20] Abbott B.C., Bigland B., & Ritchie J.M., The physiological cost of negative work, *J. Physiol*, 117, 380-390 (1952).

[21] Wells R.P., Mechanical energy costs of human movement: an approach to evaluating the transfer possibilities of two-joint muscles, *J. Biomech.*, 21(11), 955-964 (1988).

[22] Zamparo P., Perini R., Orizio C., Sacher M. & Ferretti G., The energy cost of walking or running on sand, *Eur. J. Appl. Physiol. Occ. Physiol.*, 65, 183-187 (1992).

[23] Winter D.A., Moments of force and mechanical power in jogging, *J. Biomech*, 16(1), 91-97 (1983).

[24] Simonsen E.B., Thomsen L. & Klausen K. ,Activity of mono- and biarticular leg muscles during sprint running, *Eur. J. Appl. Physiol.*, 54, 524-532 (1985).

[25] Zatsiorsky V. & Seluyanov V., The mass and inertial characteristics of the main segments of the human body, in H. Matsui & K. Kobayashi (Eds.) Biomechanics VIII-B, pp. 1152-1159. (Human Kinetics, Illinois, 1983).

[26] Dempster W.T., Space requirements of the seated operator, WADC Technical Report, (Wright Patterson Air Force Base, 1955).

[27] McMahon T.A. & Greene P.R., The influence of track compliance on running, *J. Biomech.*, 12, 893-904 (1979).

[28] Nigg B.M., Biomechanics of running shoes, (Human Kinetics, Illinois, 1986).

[29] Standards Association of Australia. Australian Standard AS 2983.4-1988. Methods of test for synthetic sports surfaces. Method 4: Test for slip resistance (SAA, Sydney, 1988).

[30] Standards Association of Australia. Australian Standard AS 2983.5-1988. Methods of test for synthetic sports surfaces. Method 5: Determination of stiffness (SAA, Sydney, 1988).

[31] Gatto F., Swannell P. & Neal R., A force-indentation relationship for gymnastic mats, *J. Biomech. Eng.*, 114, 338-345 (1992).

[32] Kaneko M., Mechanics and energetics in running with special reference to efficiency, *J. Biomech.*, Suppl 1, 57-63 (1990).

THE DYNAMICS OF THE LAWN BOWL REVISITED

D.H. Lamble[1] and A.K. Easton[2]

Abstract

The mathematical model describing the dynamical behaviour of a bowl rolling on a deformable plane surface is discussed by Brearley and Bolt (1958). Since that time, there have been a number of advances in computer technology which can be exploited and developments in the sport of lawn bowls which can be investigated. Video imaging techniques and computer analysis have been used to determine the mechanical parameters of bowls with a view to simulating the bowl trajectories. A new method for the timing of the green has been developed as an outcome of investigation into the characteristics of the green.

1. INTRODUCTION

Lawn bowls are manufactured from uniform material as solids of revolution with no added weight or insert. They possess a bias determined by the distribution of mass with respect to the running surface. This distribution can be described by the profile of the bowl with respect to the axis of revolution expressed through the moments of inertia about the axes.

The path or trajectory of the bowl on a green primarily depends on

(i) the distribution of mass of the bowl and

(ii) the speed of the green.

The effect of rolling surface in retarding the forward motion can be represented by a couple K whose axis is parallel to the ground and normal to the direction of motion. The speed of the green determines the value of this couple which is constant for a particular bowl on a particular surface.

Brearley and Bolt (1958) discuss the dynamics of the Classic or standard Henselite bowl which was the most widely used bowl at that time. Developments over the past five years or so have led to the introduction of many bowls with different characteristics. These fall broadly into two categories, namely

(i) the "arcing trajectory" bowls as manufactured by Taylor's in Scotland, Drake's Pride and GreenMaster in England and the Maestro bowl of Henselite and

(ii) the "hooking trajectory" bowls such as the lesser bias Classic2 of Henselite.

---

[1] Physics Department, Swinburne University of Technology, Hawthorn, Vic, 3122.
[2] Mathematics Department, Swinburne University of Technology.

These newer bowls have different profiles and mass distribution and, in particular, they have a narrower trajectory compared with the standard Henselite bowl.

This paper describes procedures used to measure the profiles of bowls to determine their mechanical parameters and a new method for the timing of the green. These investigations allow the value of *K*, the retarding couple, to be determined. This allows extension of the earlier work to simulate the trajectories of different bowls.

2. DIFFERENT BOWLS

The calculation of the mechanical parameters of different bowls have been determined by using video-imaging techniques combined with separate accurate measurement followed by analysis using "MATHEMATICA".

The video imaging has an available resolution of 512 by 512 pixels for a "hemispherical section" of the bowl of approximately 125 mm diameter which implies a measurement accuracy of the order of 0.25 mm per pixel. Removal of 0.5 gram of material from a bowl of total mass 1525 grams has been clearly demonstrated on the testing table at Henselite to have a significant effect on the trajectory of the bowl. Hence this resolution is unsatisfactory where accurate trajectory determination is desired, but is adequate for initial validation of the technique.

The second procedure measured the different profiles directly providing high accuracy (within 0.01 mm on any axis) using the Ferranti 3-axis co-ordinate measuring system available within the Engineering Faculty of Swinburne. This enables the video profiles to be accurately scaled, thus allowing for camera error resulting from the imaging not taking the full profile. This is illustrated in Figure 1.

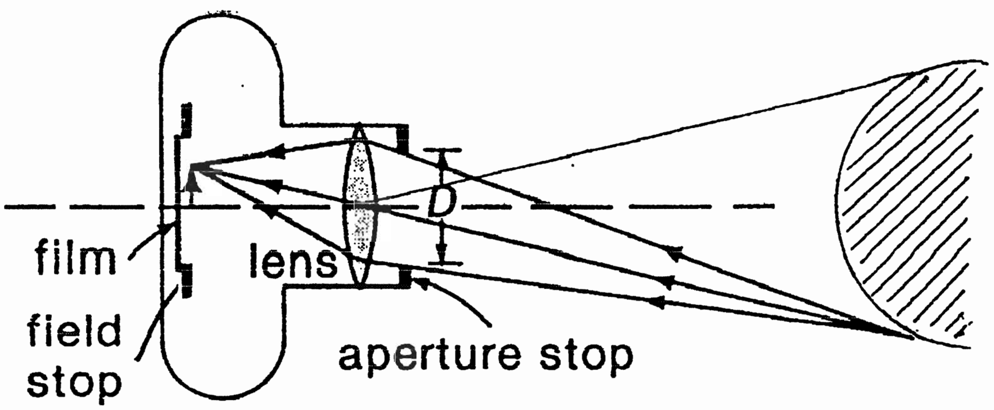


Figure 1: The camera image of the bowl

The MATHEMATICA software package was used to obtain a parametric fit of elliptical segments to the profile. A separate measurement of the mass of the bowl on an accurate balance is followed by calculation to determine the volume, density, the moments of inertia and the co-ordinates of the centre of mass for the different bowls. The mathematical basis of the calculation is detailed in the appendix.

3. TRAJECTORIES

Bowls from different manufacturers were purchased as close as possible to the same size and weight. Using the R.V.B.A. chute in order to achieve the same initial velocity on the green, the trajectories of the different bowls have been measured on a green timed at 14.25 seconds using the accurate laser-based theodolite survey equipment available at Swinburne. It is a characteristic of the lawn green running at less than 15 seconds that the bowls have a tendency to "track" i.e. an indentation in the surface is left and this influences the path of the next bowl. To avoid this, the chute was moved slightly after each set of bowls had been measured. The individual trajectories have been translated to a common origin, and then rotated to finish on the axis in order to provide the comparison of trajectories in Figure 2.

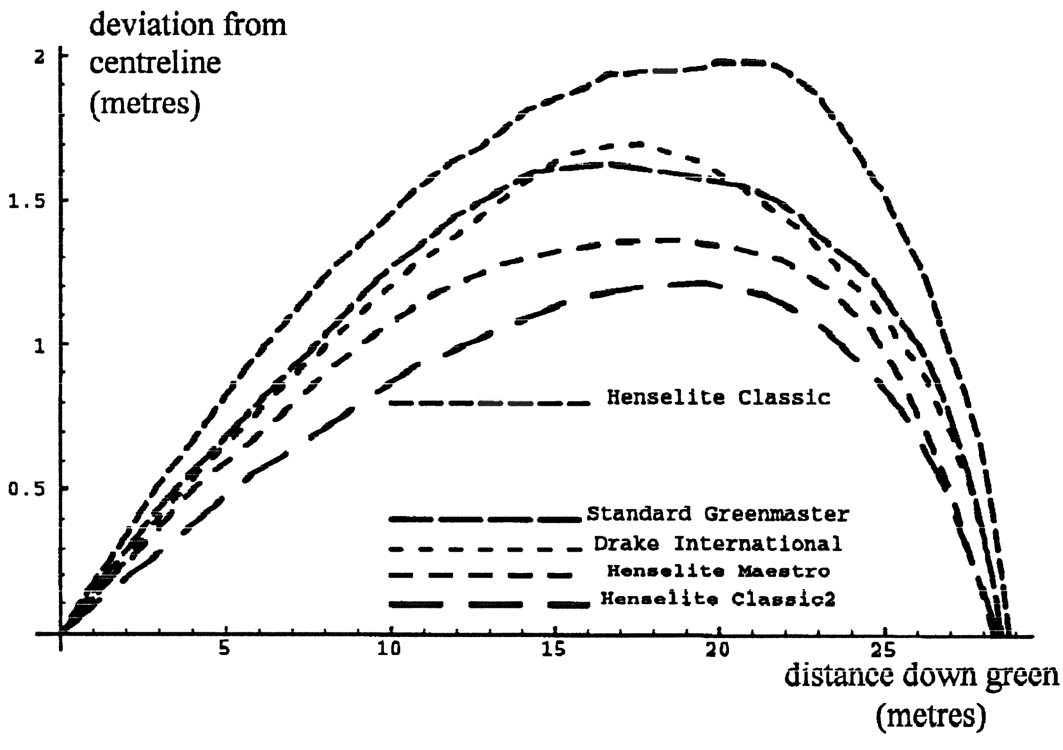


Figure 2: Trajectories of different bowls

For the "hooking trajectory" bowls, including the standard bowl, there is good agreement between the measured and simulated trajectories using the overturning moment $M * g * c * Sin(\Theta - \theta)$, as in the analysis presented in Brearley and Bolt (1958). Here, M is the mass of the bowl, g the acceleration due to gravity, c the perpendicular distance of the centre of mass from the running surface axis, Θ the final value of θ the inclination of the axis to the vertical. Brearley (1961) replaces the overturning moment by a form appropriate to the bowl shape and justifies the use of $M * g * c * Sin(\Theta - \theta)$ for the standard bowl. However, this does not provide a good fit for the trajectories of all bowls. A correlation exists between the profile of the bowl and the path it takes and the form of the overturning moment for the "arcing trajectory" family of bowls is being investigated.

4. THE TIMING OF THE GREEN

The current method of timing the speed of the green involves the use of a stopwatch to measure the time interval between the instant of delivery and the time the bowl comes to rest 27 metres away on the centreline of the rink. The accuracy of this method is limited because of the different trajectories of different bowls, which cover a greater or lesser distance before coming to rest, and the variation in the use of the stopwatch by the different accredited umpires or team managers.

In order to obtain consistency in measurement of this important parameter (the speed of the green), a new technique has been developed and implemented in conjunction with Mr. D. Merrie, Chairman of the Greens Committee, Royal Victorian Bowls Association. This procedure can be explained in the following way. A sphere of known mass and diameter is rolled down a chute of pre-determined height and the distance it travels along the green is measured. In this way, the measurement of time has been changed to a measurement of distance, which has significantly less variation and consequently greater accuracy. This removes the differences in human response time from the measurement of the speed of the green ensuring that any comparison between greens at different clubs is more accurate.

The analysis of a rolling sphere down an incline can be found in any standard mechanics text which includes a discussion of combined rolling and translational motion. The sphere comes to rest as a result of the frictional couple K which is related to the speed of the green. The loss of energy $= K * \Theta$, where Θ is the angle through which the sphere rotates before coming to rest. The graphs in Figure 3 show the variation of distance travelled for

(i) a jack, a sphere of diameter 63.5 mm and mass about 250 gram, and

(ii) a sphere of diameter 83.5 mm and mass 450 gram.

for different speeds of the green. These graphs were determined from measurements of the speed of many greens by both methods which enables the variation in the K value to be determined using $K * \Theta = m * g * h$ assuming no frictional loss down the chute, with m being the mass of the sphere rolling down the chute from an initial height h.

Brearley and Bolt (1958) state "K may be taken constant for the same bowl and bowling surface". This is true. However, this investigation of the timing of greens clearly demonstrates that the value of K on a given green depends on the size and mass of the object under consideration, as well as varying with the speed of the green.

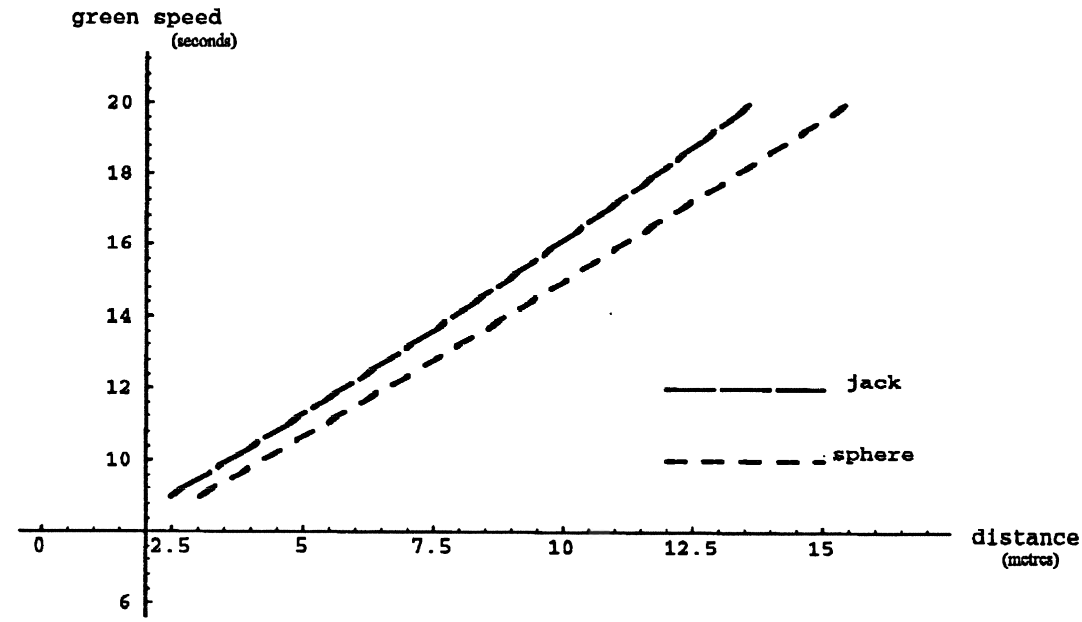


Figure 3: Distance travelled by the spherical objects on different greens

5. CONCLUSIONS

The new method for measuring the speed of the green has been adopted by the R.V.B.A.. The computer simulation for the "hooking" family of bowls has been confirmed for both the standard bowl and the narrower Classic2 bowl at different green speeds.

Further investigation will be undertaken with respect to the variation of K on a given green for the bowls of different size and weight and into the form of the overturning moment applicable to the "arcing" family of bowls before the trajectory simulation is complete.

REFERENCES

[1] Brearley M.N. and Bolt B.A., The dynamics of a bowl, *Quart.Journ.Mech. and Applied Math.*,Vol XI, Part 3, pp 351 -363, (1958).

[2] Brearley M.N., The motion of a biased bowl with perturbing projection conditions, *Proc. Camb. Phil. Soc.*, Vol 57, Pt.1 , pp131 -151, (1961).

ACKNOWLEDGMENTS

The assistance of the various bowls' manufacturers for supplying sets of bowls to my specification is acknowledged. In addition, Henselite provided access to the testing table at their premises.

The Royal Victorian Bowls Association, through the offices of a councillor, Mr. D. Merrie, provided the bowls' chute, and purchased a chute to my specification prior to the experimental evaluation of the technique for timing of the green.

The Box Hill Bowling Club allowed the use of its green, which was specially prepared for the measurements of the trajectories.

The authors gratefully acknowledge Swinburne, for providing the facilities and time for the investigation to be undertaken.

APPENDIX

This appendix summarises the formulas used to calculate the volume of a bowl, its moments of inertia, and the position of the centre of mass.

Consider a solid of rotation whose rotation axis is coincident with the x axis and profile in the $z = 0$ plane given by

$$x = f(t)\ ,\quad y = g(t)\ ,\quad t \in [0,1]\ . \tag{1}$$

where $f(t)$ and $g(t)$ are continuous, smooth functions of the parameter t.

The volume of the solid is given by

$$V = \iiint_V dV = \int_{x(0)}^{x(1)} \pi(y^2 + z^2)dx = \pi\int_0^1 g^2(t)\frac{df}{dt}dt \tag{2}$$

The moment of inertia about the symmetry axis is calculated by splitting the solid into disks centred on the axis and summing the moments of these discs. Hence:

$$I_{xx} = \iiint_V r^2 dm = \int_{x(0)}^{x(1)} dI_x(x) = \frac{\pi M}{2V}\int_0^1 g^4(t)\frac{df}{dt}dt \tag{3}$$

The moment about the axis perpendicular to the symmetry axis and passing through the centre of mass of the solid can be found by splitting the solid into discs perpendicular to the symmetry axis and writing the total moment as the sum of the moments of these laminae

$$I_{yy} = \int_{x(0)}^{x(1)} dI_y(x) \tag{4}$$

The lamina moment can be obtained from the perpendicular axis theorem plus the parallel axis theorem

$$dI_y(x) = \tfrac{1}{2} dI_x(x) + (x - x_c)^2 dm = \frac{\pi M}{4V} g^4(t) dx + (f(t) - x_c)^2 \frac{\pi M}{V} g^2(t) dx \tag{5}$$

so that

$$I_{yy} = I_{zz} = \frac{\pi M}{V} \int_0^1 g^2(t) \left(\tfrac{1}{4} g^2(t) + (f(t) - x_c)^2 \right) \frac{df(t)}{dt} dt \tag{6}$$

The centre of mass lies on the symmetry axis with x position given by

$$x_c = \frac{1}{M} \iiint_V x dm = \frac{\pi}{V} \int_0^1 f(t) g^2(t) \frac{df(t)}{dt} dt \tag{7}$$

VARIABILITY OF SCORES AND CONSISTENCY IN SPORT

Stephen R Clarke[1]

Abstract

This paper discusses the concept of consistency in sport, and its relationship with the variability of the outcomes. Using examples from golf, cricket and field events, it is demonstrated that highly variable results will produce more wins, and in some cases better average scores than 'consistent' results. It is also shown that consistent behaviour can produce highly variable outcomes. This implies that if coaches or sports followers use subjective feelings about the actual scores of players or teams to judge consistency, they are likely to be severely misled. Given the importance of variability to the outcome of sporting events, the evaluation and publication of suitable measures such as variance is suggested.

1. INTRODUCTION

Mathematicians interested in sport can probably be divided into two main groups – those interested in deterministic aspects such as dynamics and mechanics, and those into stochastic aspects. I am definitely in the latter category, being particularly interested in using probabilistic models to assist with strategy or tactics. In this paper I wish to use some simple models to explore the notion of consistency in sport, and its relation to the variability of the outcomes or scores.

It is a basic tenet of sport that an essential characteristic of excellence is consistency. Sportsmen are taught to strive to improve consistency as this will improve performance, and this is often equated with low variation in outcomes. For example Pollock [1] in discussing the mean and standard deviation of scores in relation to consistency of golfers says 'it seems reasonable that a better player would have low values for both'. This paper demonstrates that this attitude is often not the case. In sports such as golf, where one player competes simultaneously against many others, highly variable scores (ie ones with a large standard deviation) will result in more wins than less variable scores with the same mean level. Furthermore, it is shown that for competitions where the final score is the best of several attempts, the variability in the individual attempts contributes positively to the final score. In judging the consistency of athletes, coaches and followers usually look at their results. This can be misleading, as it is shown that highly variable results will often arise from consistent behaviour. Indeed, in some cases, the more consistent the behaviour the more variable are the results. Given its importance it would be beneficial if statistics that measure variability were published.

---

[1]Mathematics Department, Swinburne University, Hawthorn, Vic.

2. GOLF

A golfer's score on each hole X_i can be considered as so many shots over par (negative if under). Since par takes into account the difficulty of the hole, this should produce X_i's reasonably independent and identically distributed. Thus the total S for 18 holes is given by

$$S = X_1 + X_2 + \ldots\ldots + X_{18} \tag{1}$$

and so by the central limit theorem should be normally distributed. The above normality and independence assumptions are supported by Scheid [2] who reports the net scores of all but the worst scores of 3000 Massachusetts players to be normal, with a very small correlation between scores on one hole and the next. Similarly, Wilson [3] examined the scores of many players of differing ability and found the better 40% of scores for each player approximated the normal distribution. Thus the two parameters of a normal distribution are appropriate, and we can use μ as a measure of a golfer's ability and σ as a measure of the variability of the scores (however as I will show later, perhaps not of consistency).

2.1 Consistent scores mean fewer wins.

This section aims to show that of two players with the same average, the more variable will win more tournaments. In fact, a player of lesser ability can win more tournaments than a better player if they are also more variable. A typical top class tournament has over 100 players, all of varying abilities, and is played over 72 holes, the lowest total winning. We need to find the probability that a given player will have a lower score than the minimum score of all the other players, but here we make some simplifying assumptions. Also in practice, only those under the cut-off score after the first 36 holes complete the tournament. This is typically about 60 players. However, since those players cut would have little chance of winning the tournament if they were allowed to continue, we can ignore this complication. Since the score necessary to win is the best performance of many players, it will always be far better than the average performance of any individual player. Below we assume that 10 under (–10) will win the tournament, but the argument is the same for any chosen figure. Note that in some tournaments players scores are generally lower than for other tournaments. However by replacing par with the average that the course actually played, the argument remains valid. For example, in the 1992 US masters, the winner Fred Couples needed to score a total of 9.8 less than the average total of the 56 qualifiers. Given μ and σ for a golfer, it is a simple task to find their chance of achieving a certain score for the tournament.

A player with a mean score μ and standard deviation σ over 18 holes will have a mean 4μ and standard deviation 2σ over 72 holes. Suppose a score of w is required to win.

Then $\Pr(\text{Score} \le w) = \Pr[\, Z \le (w - 4\mu + 0.5) / 2\sigma \,]$
$= \Phi[\, (w - 4\mu + 0.5) / 2\sigma]$

where Z is the standard normal and $\Phi(z)$ is the area under the normal curve to the left of z. The 0.5 is needed as a correction for continuity.

Wilson [3] states that standard deviations for golfers' scores over 18 holes ranged from 2 to 6, with lower-handicap golfers having the lower values.

For example, if w = –10, μ = –1 and σ = 2,

$$\Pr(\text{Score} \leq -10) = \Phi[(-10 - 4.-1 + 0.5) / 4] = \Phi(-1.375) = 0.085$$

Thus a golfer who each round averages 1 under par with a standard deviation of 2 will win 8.5% of tournaments, assuming 10 under is needed to win.

The results of this calculation for various values of μ and σ are shown in Table 1. This shows the percentage chance of winning a tournament, for values of μ from +1 to –3, and σ from 1 to 4.

Table 1

Percentage chance of score better than 10 under

		σ per round						
		1.0	1.5	2.0	2.5	3.0	3.5	4.0
	1.0	0.0	0.0	0.0	0.3	1.2	2.7	4.6
	0.5	0.0	0.0	0.2	1.1	2.8	5.0	7.5
μ	0.0	0.0	0.1	0.9	2.9	5.7	8.7	11.7
per	-0.5	0.0	0.6	0.3	6.7	10.6	14.7	17.4
round	-1.0	0.3	3.3	8.5	13.6	17.9	21.6	24.6
	-1.5	4.0	12.2	19.1	24.2	28.0	30.8	33.1
	-2.0	22.7	30.8	35.4	38.2	40.1	41.5	42.1
	-2.5	59.9	56.6	55.0	54.0	53.3	52.8	52.5
	-3.0	89.4	79.7	73.4	69.2	66.1	63.9	62.4

Note firstly that the increase in σ is only helpful in winning more tournaments if μ is less than the value required to win. The last 2 rows show a decreasing chance of winning as σ increases. Thus for players so much better than everyone else that their average is better than the winning score (in golf there aren't any), it is in their interests for their scores to be as consistent as possible. For everyone else, the reverse is true.

A player averaging par, with σ = 2 will barely win 1% of tournaments. By reducing their average by 1/2 a stroke, or increasing σ by 1/2 a stroke, they increase their tournament wins to just on 3%. If they do both they increase their winning percentage to 6.7%.

Clearly a large variation in scores is advantageous. What are the ramifications for golfers? Clearly, given a choice, a player should choose strategies that maximise variation. For example, if faced with a choice of playing over a bunker, with a 50-50 chance of the ball getting on the green with a consequent birdie, and a 50-50 chance of finding the bunker and a certain bogie, or playing around the bunker for a certain par, the player should choose the first alternative. Similarly, players should avoid any tendency to play the first round of a tournament 'steadily' if that means producing an average round with no disasters and no great rounds.

Note that the argument applies not just to winning tournaments. The prize-money for most tournaments is not a linear function of position, with the incremental prize-money increasing as you move up the finishing order. Thus a 10th and 12th will return more in prize-money than coming 11th twice.

If golfers put such a store in 'consistency' (even though they have it back to front), and the variance of a player's rounds is important to their chance of winning, why is it not published as a performance statistic? It would be interesting to check if top golfers do vary significantly in their round-by-round standard deviations. Rotella and Boutcher [4] use regression analysis on the playing statistics of professional golfers to predict money earned. Over 13 published statistics were used, but standard deviation of scores was not one of them. However birdies divided by greens in regulation was the second most important variable behind scoring average. This statistic may be a surrogate for variance, as it goes up when the number of bogies goes up as well as when the number of birdies increases. Hale & Hale [5] find that for the leading money winners the performance statistics are not a good predictor of success - perhaps this is further evidence that some others are needed.

This argument can apply to many sports. In an earlier article on cricket Clarke [6] showed that success in Sheffield Shield is dependent on having highly variable results - viz lots of outrights even if many are losses. A team can actually lose more than half their matches and still finish on top of the shield table. In discussing the triathlon, de Mestre [7] suggests that the influence of an event in determining the winner is proportional to the variance.

2.2 Consistent behaviour can mean variable results

While the above seems to go against the traditional view that consistency of performance is a desirable trait, this apparent contradiction can be resolved, as it is shown that the most variable scores are often produced by consistent behaviour or play. The consistency an athlete attempts to achieve is consistent behaviour. Thus a golfer's swing should be the same all the time, and not alter under pressure. A cricketer is required to treat every ball on its merits, to concentrate equally throughout an innings, and not allow lapses in effort. Similar golden rules exist in most sports that fundamentally require consistency of behaviour.

However, in judging the consistency or otherwise of an athlete, the performance (or set of scores) of the athlete is looked at, and consistency judged by the range of performance - the smaller the range the more consistent. Players with a large range of scores may be judged to be inconsistent, and in need of some help to improve the consistency of their behaviour. In fact the reverse is often true - a small range in scores can be indicative of inconsistent behaviour, while a large range may be generated from consistent application.

While the following is developed for golf, it applies to sports like golf, rifle shooting and archery, where a player has a series of attempts which can be classed as a success (par, birdie, bulls-eye etc) or a failure.

Let 1 represent a success, and 0 a failure, and suppose there are (a constant) n attempts. Thus in golf, for a top golfer 1 might represent a birdie, 0 a par and we have 18 attempts in one round. Suppose attempt i has a probability of success p_i.

$$\text{Let} \quad X_i = \begin{cases} 1 \text{ with probability } p_i \\ 0 \text{ with probability } (1 - p_i) \end{cases}$$

Then X_i (either from first principles or as a binomial variable with $n=1$, $p=p_i$) has mean p_i and variance $p_i(1 - p_i)$

As before, let the score $$S = X_1+X_2+X_3 +\ldots+ X_n = \sum_{i=1}^{n} X_i \qquad (2)$$

Then the mean or expected value of the score E(S) is given by

$$E(S) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p_i$$

Assuming independence, the variance of the score Var(S) is given by

$$Var(S) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p_i(1-p_i)$$

$$= \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} p_i^2$$

$$= E(S) - \sum_{i=1}^{n} p_i^2$$

Thus for a given mean score E(S), the variance of the score is a maximum when $\sum_{i=1}^{n} p_i^2$ is a minimum. Using calculus and the Lagrange multiplier technique, Bellman and Dreyfus [8], this is easily shown to be when all p_i are equal.

Thus the maximum variation in total scores occurs when all attempts have the same chance of success.

For example, in golf, a player who always begins slowly but improves their chance of a birdie throughout the round will show less variation in scores than one who has a constant chance of success.

While this model would need to be greatly improved to form a realistic model of golf, the results of this simple model are very clear. A player whose concentration wavers, or who tries harder some holes than others, will produce scores that are less variable, and so appear more consistent than one who plays each hole the same. A coach who berates a rifle shooter for inconsistency for scores ranging between 85 and 95 may be better off concentrating on the shooter who always scores 90.

2.3 Hole-by-hole variances

Consider 2 golfers: one short but straight, a cautious putter who invariably gets par on a hole; the second a long but sometimes wayward hitter and bold putter who has a good chance of a birdie but also a good chance of a bogie. In a round both players could both get par, but one may get it by having 18 pars, while the other may get 6 pars, 6 birdies and 6 bogies. This difference could be measured by the variance or standard deviation of the X's, the hole-by-hole variance, where the first would show up with a low value and the second with a high value. Now in general Var S = ΣVar X_i + $\Sigma\Sigma$cov(X_iX_j). If holes are independent, then Var S = 18 Var X. A player who plays better when up and worse when down (cov(X_iX_j) > 0) would have a Var S > 18 Var X, while a player who loses concentration when up and tries harder when down (cov(X_iX_j) > 0) would have Var S < 18 Var X. Thus again an insight into the psychology or mental attitude of the player might be gained by the extra statistics, with consistency being measured by the closeness of VarS/(18 Var X) to 1.

For example, consider the following four golfers' scores on 2 rounds of 9 holes. Each has the same total score of zero over par.

		Hole									
Player	Rnd	1	2	3	4	5	6	7	8	9	S
a	1	1	0	0	0	1	0	-1	0	0	1
	2	0	1	0	0	-1	0	-1	0	0	-1
b	1	1	1	1	1	1	-1	-1	-1	-1	1
	2	-1	-1	-1	-1	-1	1	1	1	1	-1
c	1	1	1	1	1	1	1	1	1	0	8
	2	-1	-1	-1	-1	-1	-1	-1	-1	0	-8
d	1	1	1	1	0	1	-1	-1	1	-1	2
	2	-1	1	-1	0	1	1	-1	-1	-1	-2

The statistics for each player are shown below

Player	Mean score	VarX	VarS	$\frac{\text{VarS}}{\text{18VarX}}$
a	0	0.36	2	0.615
b	0	1.11	2	0.200
c	0	0.11	128	28.000
d	0	0.94	8	0.941

Note the difference in the extra statistics. Golfers b and d are shown as 'attacking' players by the high variance of X, golfer d is shown as consistent by a ratio near 1, c is shown as highly variable, and b & c as inconsistent.

2.4 Rating the Golf course

In this section the point of view is switched to the golf course. A lot of work has been done on rating the difficulty of a golf course, but little on the discrimination powers of individual holes or the course as a whole.

Clarke and Rice [9] have looked at rating courses. Hole-by-hole data of players' scores in the 1992 U.S. Masters' tournament obtained from the organisers included the average score on each hole, but the only indication of the variation of scores was the number of eagles, birdies and bogies. These are very difficult to interpret. Table 2 shows the mean and standard deviation of scores on each hole. It is clear that holes 2, 8, 13 and 15 (all par 5's) were easier to play than the others. However it is not only the ease or difficulty of a hole that is important. If everybody takes 1 over par on a hole, that hole is not separating players. The standard deviation gives a measure of the ability of a hole to discriminate between players. Note that hole 10, clearly the hardest hole over the 4 rounds, was on only one day in the top half of the holes in order of discrimination. In this regard holes 12 and 13 clearly stand out above the rest, between them sharing the highest and second highest standard deviation in every round. Hole 12, a par 3, is not unduly difficult, but clearly produces highly variable scores. In Parsons [10] the section of the Augusta course from holes 11 to 13 is described thus "The Masters championship has been won or lost so often between the 11th and 13th that this three hole stretch has become known as Amen corner". In Parsons 10], Jack Nicklaus describes the 12th as " the most demanding tournament hole in the world". In this case, the standard deviations reflect golfers' views of the holes. Although the standard deviation of the scores on a hole contains much more information about the importance of the hole to the result of the tournament, this statistic is never quoted. It should be:

Table 2

Descriptive Statistics for the 63 Qualifiers in the US Masters

		ROUND							
		ONE		TWO		THREE		FOUR	
Hole	Par	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1	4	-0.05	0.55	-0.02	0.68	0.05	0.55	0.02	0.63
2	5	-0.22	0.73	-0.56	0.56	-0.21	0.72	-0.29	0.79
3	4	0.00	0.48	-0.08	0.55	0.21	0.72	0.10	0.59
4	3	0.11	0.63	0.16	0.60	0.22	0.66	0.14	0.50
5	4	0.17	0.58	0.03	0.44	0.19	0.56	0.13	0.66
6	3	0.08	0.52	-0.03	0.65	0.06	0.54	0.10	0.56
7	4	0.05	0.58	0.05	0.61	-0.24	0.76	0.19	0.72
8	5	-0.25	0.65	-0.29	0.52	-0.17	0.58	-0.40	0.58
9	4	-0.06	0.47	0.10	0.64	0.03	0.59	0.06	0.54
10	4	0.17	0.61	0.11	0.57	0.22	0.58	0.40	0.55
11	4	0.11	0.65	0.11	0.63	0.03	0.44	0.22	0.55
12	3	-0.03	0.78	0.21	0.70	0.17	0.93	0.41	1.10
13	5	-0.41	0.73	-0.35	0.77	-0.46	0.80	-0.44	0.96
14	4	-0.21	0.45	0.11	0.57	0.14	0.64	-0.08	0.52
15	5	-0.24	0.69	-0.54	0.64	-0.57	0.76	-0.65	0.63
16	3	0.00	0.54	-0.02	0.58	0.02	0.55	-0.05	0.68
17	4	-0.02	0.58	-0.19	0.59	0.02	0.55	0.11	0.57
18	4	0.02	0.49	-0.11	0.54	0.19	0.59	0.00	0.54
TOTAL	72	-0.78	2.50	-1.30	2.15	-0.10	2.76	-0.03	3.17

However a hole should also discriminate between golfers appropriately. The better players should do well and the poor do badly. Clarke and Rice [9] have investigated this by applying some standard psychological procedures to golf scores.

3. FIELD EVENTS: BETTER SCORES FROM INCONSISTENT PERFORMANCE

In many sports the penalties for poor attempts are not great. In the long jump the athlete's score is the longest out of 3 attempts. Similar rules apply to the triple jump, javelin, discus, shot put and hammer throw. Thus unlike golf, poor attempts do not bring down the score. In such cases, variation in jumps (inconsistent jumping?) will not only increase the chance of winning the event, but will actually produce a greater score.

The statistics of extremes is quite complicated, particularly the distribution of maximum values. However Gumbel (1958) gives some graphs and formulas for the average maximum of n standard normals from which the table below has been derived.

Table 3

Expected maximum of n normal variates with mean μ and standard deviation σ

Number of variates	Expected maximum
1	μ
3	$\mu + 0.85\,\sigma$
6	$\mu + 1.25\,\sigma$
100	$\mu + 2.50\,\sigma$

Ladany [11] collected data for an athlete, showing the actual length of jump from take-off to landing was normally distributed with a mean 701.23 cm with a standard deviation of 20.44 cm. We will ignore problems associated with the aiming line and disallowed jumps, and take this to be the distribution of the recorded jump. Thus the expected score for this jumper is 701.23 + 0.85*20.44 = 701.23 + 17.30 = 718.53 cm. Note that over 17 cm of his score arises through the variation in the jumps.

Note also that every increase of 10 cm in the variation results in an increase of 8.5 cm in the expected score. Thus for example, a 10 cm deficiency in the average jump will be made up for by a 12 cm increase in the variation.

In some competitions, the top competitors are allowed a further 3 jumps giving 6 jumps in all and even more advantage to the inconsistent jumper. A jumper now gains 25% more from an increase in variability than the same increase in average. Note also, that the actual maximums of the inconsistent jumper would be more variable (about their greater mean) than for the consistent jumper. Thus we also have the previous effect operating as well, giving the inconsistent jumper even more chance of winning an event.

When records are considered, we are looking at a jumper's personal best. An increase of 10 cm in the variation of jumps would add 25 cm to the best of 100 jumps. This has ramifications for selection, as some selections are effectively done on a personal best (eg Olympic qualifying times). This may disadvantage the consistent performer, who may be expected to do relatively better in the actual competition when only 3 or 6 jumps are allowed. On the other hand, the person with the large variance may produce the statistical outlier that produces a medal. In discussing Beamon's famous long jump, Brearley [12] concludes it was a 'statistical miracle'.

Do field athletes have different standard deviations? Unfortunately, once again, there is little data allowing an investigation of these effects. In most cases, only the athlete's best effort is recorded.

4. CRICKET

As another example of consistent behaviour producing seemingly inconsistent scores let's look at a batsman's scores at cricket. Consider J.D Siddons, who batted about number 6 for Victoria in the Australian Sheffield Shield competition in 1985/6. His scores for the year were

33, 17, 76, 5, 74, 7, 7, 107, 1, 45, 17, 2, 36.

Many cricket followers would say that is an inconsistent set of results, since they expect a consistent batsman to have scores with a small standard deviation, like 51, 55, 52, 53, 54. However scores like this mean that a batsman has no chance of going out until he/she reaches 50, and is almost certain to go out soon after. So in terms of probability of dismissal they are very inconsistent. A consistent batsman who has a 30% chance of making 50, should turn 30% of those 50's into centuries, and 30% of his/her centuries into 150's etc.

This assumption of a constant probability of dismissal leads to a geometric distribution (or its continuous counterpart, the negative exponential) for scores. The negative exponential distribution is common as the distribution of waiting times for random events - in this case it is the waiting time (measured by score) until a dismissal. A histogram of Siddons' scores and the geometric distribution with p = 1/33 are shown in Figure 1. The two are virtually identical.

Clearly Siddons' scores follow closely what theory suggests a consistent player with an average of 33 should produce. The standard deviation of Siddons' scores is 34, again agreeing to that predicted by an exponential distribution whose standard deviation is equal to the mean. Followers who judge Siddons to be inconsistent on the basis of his scores would be doing him a great injustice.

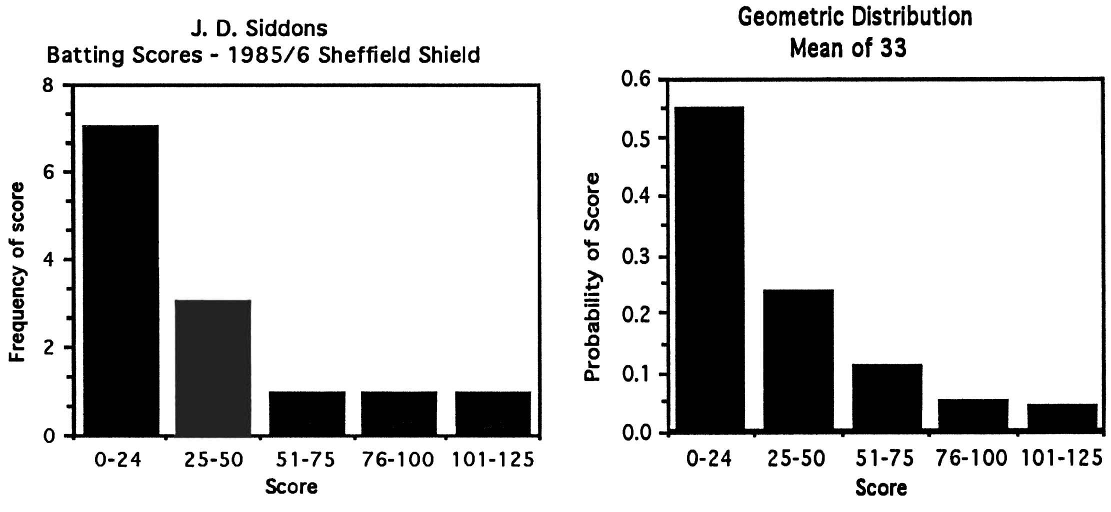


Figure 1: Comparison of batting scores with Geometric distribution

Elderton [13] and Wood [14] discuss further the meaning of consistency when applied to cricketer's scores, and the fitting of the geometric distribution. They show that not only are there theoretical reasons why the standard deviation of players scores should be equal to or greater than the mean, but that an analysis of actual scores of many players confirms this to be so. It is players with a small variation in their scores who should be the prime targets for the criticism of supporters or remedial coaching for lack of consistency in application. Wood suggested using the

Coefficient of Variation (CV) as a measure of consistency, with the closer to 100, the more consistent a batsman. Pollard [15] claims a high CV indicates a batsman has problems early, but scores runs more easily later in the innings. However, Clarke [16] showed that under a model similar to equation (2) perfectly consistent batsmen will have CV's greater than 100 and that perfectly consistent batsman, but with a different scoring profile, will have different coefficients of variation between 100 to 105. Thus it is not possible to have a single measure (closeness to 100) which indicates perfect consistency for all batsmen as claimed by Wood. The CV can be approximated by

$$CV(S) = 100 + 50(R/m)(CV(X)^2 - 1) \tag{3}$$

where R is the Run rate, m is the average score and CV(X) is the coefficient of variation of X, where X is the score each ball. This gives the coefficient of variation of scores in terms of the 3 parameters that describe a batsman's scoring profile: m is their average score and describes how many runs they get, R is the rate and describes how fast they get them, and CV(X) describes how the runs are distributed between singles, fours etc. Clarke [16] also suggests several other measures, including the mean and standard deviation of the number of balls faced, and the standard deviation of X to further describe batsmen, and gives these statistics for a 1-day series.

At the very least the publication of batsmen's standard deviation of scores would give some idea of their consistency - a standard deviation about the same level (or slightly higher) indicates consistency. Anything less does not.

5. Variability of the Team's Score

Just as individuals are often unjustly criticised for inconsistency, so are whole teams. It seems there is a lack of understanding in the general population as to the inherent variability in a lot of games. The variability in a cricket innings is easily investigated. The innings score is the sum of 10 partnership scores. Assuming each partnership is exponential with mean 30 (and variance 30^2) then the innings score is mean 300 and variance 9000, or standard deviation of about 95. Thus the innings could be anything from 100 to 500 - highly variable. A more realistic model assuming partnerships of different expected lengths only increases the variation. Johnston [17], Johnston et al [18] simulated one-day cricket innings using optimal batting rate policies and obtained a mean score of 215 with a standard deviation of 45. The lowest of 1000 simulations was 75, the highest 322. Clearly much of the variation we see in cricket can be explained by the highly variable nature of the game and is not necessarily due to good or bad play.

6. Conclusion

Sports followers need to be careful in assessing the attribute of consistency, and in extolling its virtues. While consistency in behaviour is to be encouraged this will not necessarily reflect itself in low variability of scores or results. However, low variability in results is often not even a desirable outcome. In some events it will decrease the actual score, while in others it will reduce the number of events won.

If coaches use subjective feelings about the actual scores of players to judge consistency, they are likely to be severely misled. Players who have small variation in results may be the prime targets for remedial coaching or the attention of psychologists, not the players with large variation in outcomes.

With the increasing use of computers, the evaluation of other statistics which measure the variability of players and teams performance is made simpler. The calculation and publication of these statistics could assist players and coaches to a better understanding of their performances, and at the least would provide followers with some interesting data. With a suitable choice of names, these could be interpreted by the general public.

REFERENCES

[1] Pollock S.M., A Model of the USGA Handicap System and Fairness of Medal and Match Play, *Optimal Strategies in Sport*, S.P. Ladany and R.E. Machol., Editor, 141-150, (North Holland, 1977).

[2] Scheid F., On the Normality and Independence of Golf Scores with Various Applications, *Science and Golf*, A.J. Cochran, Editor, 147-152, (Chapman and Hall, London, 1990)

[3] Wilson P., The Competition Scratch Score, *Science and Golf*, A.J. Cochran, Editor, 129-134, (Chapman and Hall, London, 1990)

[4] Rotella, R.J. and Boutcher, S.H., A Closer Look at the Role of the Mind in Golf, *Science and Golf*, A.J. Cochran, Editor, 93-97, (Chapman and Hall, London, 1990).

[5] Hale, T. and Hale, G.T., Lies, Damned Lies and Statistics in Golf, *Science and Golf*, A.J. Cochran, Editor, 159-167, (Chapman and Hall, London, 1990).

[6] Clarke, S.R., Another Look at the 1985/86 Sheffield Shield Competition Cricket Results. *Sports Coach*, **10**(3) 16-19, (1986)

[7] de Mestre, N., Mathematics Applied to Sport, *Mathematics and Computers in Sport*, N. de Mestre, Editor, 137-148, (Bond University, 1992).

[8] Bellman, R.E. and Dreyfus, S.E., *Applied Dynamic Programming*. (Princeton University Press, Princeton, 1962).

[9] Clarke, S.R. and Rice, J.M., An application of Test Reliability Procedures to golf Tournament Scores. (submitted)

[10] Parsons, I., Augusta: Dream course in a dream setting, *The World Atlas of Golf*, I. Parsons, Editor, 116-119, (Mitchel Beazley, London, 1976).

[11] Ladany, S.P., Humes, J.W. and Sphicas, G. P., The Optimal Aiming Line. *Operational Research Quarterly*, **26**(3,i) 495-506 (1975).

[12] Brearley, M.N., The Long Jump Miracle of Mexico City, *Optimal Strategies in Sport.*, S.P. Ladany and R.E. Machol, Editor, 162-167, (North Holland, New York, 1977).

[13] Elderton, W.E., Cricket Scores and some Skew Correlation Distributions. *Journal of the Royal Statistical Society A*, **108** 1-11 (1945).

[14] Wood, G.H., Cricket Scores and Geometrical Progression. *Journal of the Royal Statistical Society A*, **108** 12-22 (1945).

[15] Pollard, R., Cricket and Statistics, *Optimal Strategies in Sport*. S.P. Ladany and R. E. Machol, Editor, 129-130, (North Holland, New York, 1977).

[16] Clarke, S.R. Consistency in Sport - with particular reference to Cricket. 27th NZOR conference, 30-35, (1991).

[17] Johnston, M.I., Assessing team and player performance in one day cricket using Dynamic Programming, (MAppSc Thesis, Swinburne University, 1992)

[18] Johnston, M.I., Clarke, S.R. and Noble, D.H., An Analysis of Scoring Policies in One Day Cricket, *Mathematics and Computers in Sport*, N. de Mestre, Editor, 71-80 (Bond University, 1992).

SOFTBALL STATISTICS

Pam Norton[1]

Abstract

An investigation is made to find a way of assessing individual batting and pitching performances in junior league softball. It was found that the statistics of top league softball (or baseball) were inappropriate. This paper gives two ways of assessing a batter's performance, by means of a run rate and the percentage of times the batter made first base. Assessing a pitcher's performance is more difficult, although the number of runs scored per innings by the opposition may give some measure of a pitcher's performance.

1. INTRODUCTION

Softball is a sport growing in popularity among young girls. It is very much a team sport, compared to sports such as tennis and swimming, for which the individual performances are easy to quantify. At the end of each season, one can look at the results and assess how the team performed compared to the other teams in the grade. Last summer my daughter played in a school team in the Waverley Women's Softball Association Pixie A1 division (for girls in grade 6, aged 11-12 years). Our problem was to find some way of assessing individual performance over the season, and use this information to organise the playing order for finals, and possibly for team selection next season..

Why is this desirable? For semi-finals, it is clearly in the team's best interests if they get off to a good start, and our aim was to maximise the chances of this.

It is also desirable to be able to assess individual performances, for even at quite junior levels, special teams are selected, such as "District Teams" and special school teams. At our local district level, the selectors walk around and watch various games, and invite certain of the players to come to a trial to select the district team. At school, a few teachers are involved in the selection. They may spend one (or two?) afternoons watching the girls practice, perhaps throwing a few balls at them. At no stage in this process do the selectors consider individual performances in previous games. One must question the fairness of this type of selection.

J. Neville Turner [3] says: "I am now prepared to predict an assault on another bastion of legal immunity - the inviolability of selectors. When Merv Hughes was omitted from the World Series matches in 1992-93, after performing in a sterling way in the test matches, on the fatuous basis that he was not a good 'one-day cricketer', I consider that he might have sued the selectors. He must have lost a good deal of the

[1]Department of Mathematics, Monash University, Clayton, Vic. 3168

monetary rewards that he deserved, and these would have been the claimable damages."

The stakes are not so high in the junior league softball, but there is a lot of prestige for those selected in top teams, and a feeling of rejection perhaps otherwise.

2. STATISTICS OF TOP LEAGUE SOFTBALL

It is very difficult to find any information on either Australian softball or baseball statistics. However there is a fair amount of information on American baseball, namely the American League and the National League. So it is to these sources, for example, Ritter [2], that I turned to find out the type of information collected.

(a) Batting

A player's batting average is the number of hits he gets divided by the number of times at bat, to three decimal places. Every single, double, triple and home run counts as just one hit.

When calculating batting averages, bases on balls do not count as times at bat, so they don't affect a player's batting average. This is important at this level, as recognised "top batters" may be deliberately pitched four balls to prevent them getting a good hit. However, if a player gets on base because of an opposing player's fielding error, it does count as a time at bat - and since it isn't a hit, his batting average drops.

Only the very best baseball players achieve a batting average greater than 0.3 over a season. Table 1 gives the batting average of the top batter in each of the National League and the American League respectively from 1980 - 1990.

Table 1

Batting Average of the Top Batter in each of the National League and the American League, 1980-1990. (From Meserole [1])

	National League		*American League*	
	Top Batter	Batting Average	Top Batter	Batting Average
1980	Bill Buckner, Chi	.324	George Brett, KC	.390
1981	Bill Madlock, Pit	.341	Carney Lansford, Bos	.336
1982	Al Oliver, Mon	.331	Willie Wilson, KC	.332
1983	Bill Madlock, Pit	.323	Wade Boggs, Bos	.361
1984	Tony Gwynn, SD	.351	Don Mattingly, NY	.343
1985	Willee McGee, St L	.353	Wade Boggs, Bos	.368
1986	Tim Raynes, Mon	.334	Wade Boggs, Bos	.357
1987	Tony Gwynn, SD	.370	Wade Boggs, Bos	.363
1988	Tony Gwynn, SD	.313	Wade Boggs, Bos	.366
1989	Tony Gwynn, SD	.336	Kirby Puckett, Min	.339
1990	Willee McGee, St L	.335	George Brett, KC	.329

Other information tallied for the season are runs batted in, home runs and stolen bases.

(b) Pitching

The best measure of a pitcher's ability is his ERA (Earned Run Average), which is the average number of earned runs he/she allows the opposition over nine innings. For purposes of figuring a pitcher's ERA, "unearned" runs (runs due to a fielding error, so not the pitcher's fault) are separated from "earned" runs. The final figure is given to two decimal places.

An ERA under 3.00 is considered excellent.

Table 2 gives the earned run average of the best pitcher in each of the National League and the American League respectively from 1980 to 1990.

Table 2

Earned Run Average (ERA) of the Best Pitcher in each of the National League and the American League, 1980 - 1990. (From Meserole [1].)

	National League		*American League*	
	Best Pitcher	ERA	Best Pitcher	ERA
1980	Don Sutton, LA	2.21	Reedy May, NY	2.47
1981	Nolan Ryan, Hou	1.69	Steve McCatty, Oak	2.32
1982	Steve Rogers, Mon	2.40	Rick Sutcliffe, Cle	2.96
1983	Atlee Hammaker, SF	2.25	Rick Honeycutt, Tex	2.42
1984	Alejandro Pena, LA	2.48	Mike Boddicker, Bal	2.79
1985	Dwight Gooden, NY	1.53	Dave Stieb, Tor	2.48
1986	Mike Scott, Hou	2.22	Roger Clemens, Bos	2.48
1987	Nolan Ryan, Hou	2.76	Jimmy Key, Tor	2.76
1988	Joe Magrane, St L	2.18	Allen Anderson, Min	2.45
1989	Scott Garrelts, SF	2.28	Brett Saberhagen, KC	2.16
1990	Danny Darwin, Hou	2.21	Roger Clemens, Bos	1.93

Other statistics quoted for pitchers are the win-loss game record (number of games won/number of games played) and the number of strikeouts.

3. STATISTICS OF JUNIOR LEAGUE SOFTBALL

The game is substantially different at the Junior League level. Firstly, a game does not involve a fixed number of innings, but is played to a time limit. This has been either an hour or an hour and ten minutes. There are usually only three innings completed, although on occasion there will sometimes be only two and on others there may be four. Secondly, an innings is completed when either three players are out or all nine players have batted. If the ninth player has batted and made it safely to first base without three players being out, then all players remaining on bases are awarded a run. The pitchers and catchers have to be changed after each two innings.

There are also intrinsic differences in the tactics of the games. For example, in junior league no batter is deliberately pitched four balls, and it is a relatively easy task to steal from first to second base.

The only information we had was that recorded in the scorebook.

Very limited information about a game was actually collected, although the scoresheet does have the capacity to record a lot more information. Unfortunately the scorers at this level are very "amateur", and the only information recorded for a batter is whether they were out or what base they had got to.

(a) Batting

The only statistics we were able to collect for a batter were

(i) the number of times she batted;

(ii) the number of times at bat that she made first base safely;

(iii) the number of times that she made a run;

(iv) the number of times she was out.

We were unable to determine how the batter got to first base - by a hit, a misfield, four balls or dropped strike three. Thus it was impossible to calculate a traditional batting average. We were also not able to tell how a batter was out - struck out, fielded out, failed base stealing or through no fault of her own.

In the absence of further information, we decided to take two measures of each batter's performance. These were the run rate (which is the total number of runs made divided by the number of times at bat) and the percentage of times the batter made first base. The information for the team members is given in table 3.

Table 3

Batting Performance of Junior League Softballers

Player	No. of times at bat	Run rate	% times made 1st base
1	32	.63	.78
2	34	.50	.65
3	36	.61	.78
4	23	.57	.78
5	35	.63	.80
6	35	.69	.89
7	31	.58	.87
8	35	.69	.71
9	36	.64	.78
10	31	.55	.71

The run rates are quite high - varying between 0.50 and 0.69, with an average of 0.61. Even the lowest run rate above would make a professional softballer (or baseballer) green with envy. The team had been together for two seasons, although there had been a couple of changes. In the previous season, the average run rate was 0.48, so the girls had clearly improved. Indeed the improvement was quite surprising when the general level of improvement in the fielding was considered from one year to the next. Generally, as the teams go up in grade, so the number of runs scored drops.

The percentage of times they made first base varied from 0.65 to 0.89, with an average of 0.77. (This was not worked out for the previous season). These figures also seem to be quite high, but cannot be compared to the batting average of the top league players, because it is impossible to determine how many times they hit their way to first base. But once again, top league players would no doubt be thrilled to see first base on such a large percentage of occasions.

No other information on the players batting performance is likely to be obtainable until the girls rise to the level with "professional" scorers.

(b) Pitching

At this level pitchers do not have the same effect on the game. Even a good pitcher's efforts are often dulled by poor fielding! Balls hit by the batter are frequently poorly fielded. Third strikes may be dropped, and very few catchers can accurately throw to second base. So it is virtually impossible to come up with any information as a basis for comparison of pitchers.

The only information we were able to determine was the number of runs scored by the opposing team in an innings. Table 4 gives the results.

Table 4

Distribution of the Number of Runs Scored per Innings

No runs per innings	Our team 1st pitcher	2nd pitcher	Opposition teams 1st pitcher	2nd pitcher
0	6	2	5	1
1	2	2	3	
2	4	2	5	3
3	6	1	2	1
4	1	1	2	
5	2	1		
6				1
7	3	4	4	5
8	2	2	5	2
9	2		2	2
Total No. of Innings	28	15	28	15
Average	3.46	4.13	4.03	5.60
Std Dev	2.97	3.02	3.31	2.95

There is clearly a wide variability in the number of runs scored per innings. An interesting feature is the "hole" in the middle of each table. If these figures were multiplied by 9, the best result would be about 29 runs conceded per nine innings game. In a major league game, the pitcher would be sacked for such an effort. However, there is no way that these figures can be compared to those for major league games, for many of the runs are due to poor fielding, no fault of the pitcher. Clearly there is some reflection of the pitcher's ability in these scores, but there is also a sizeable reflection of the standard of fielding of the rest of the team, so this figure should be used cautiously.

4. CONCLUSION

Given the available information, there are two useful measures to compare players' batting abilities. The first is the run rate, which is the number of runs scored divided by the number of times batted. This measure may reflect in part the batter's ability to steal bases. The second is the percentage of times at bat that the batter made first base. Assuming that things such as the number of walks given and the number of poorly fielded hits are approximately proportional to the number of times at bat, then this second measure probably gives a better picture of the batting ability of the player.

It is very difficult to come up with any appropriate measure for comparison of pitchers, for at this level, the runs scored off them are more often due to poor fielding than good hitting. However the average number of runs scored per innings by the opposition may be considered to reflect a little of the pitcher's ability.

REFERENCES

[1] Meserole M., editor: The 1991 Sports Almanac., (Houghton Mifflin Co., 1990).

[2] Ritter L.S., The Story of Baseball, (William Morrow and Co., New York, 1983).

[3] Turner J.N., The Culture of Australian Sport, in Montage (Monash University), p8, (March 1994).

A STATISTICAL LOOK AT CRICKET DATA

S. Ganesalingam[1], K. Kumar[2] and S. Ganeshanandam[1]

Abstract

Historically the game of cricket holds a unique position in the development of the scientific study of sports and games. The question that naturally arises is "what in the world do quantitative techniques have to do with sports"?

This paper examines the performance of a few distinguished multivariate techniques such as discriminant analysis, cluster analysis and principal component analysis in identifying sportsmen according to their skills and special interests. In cricket, for instance, we may classify a cricketer either as a 'one day' player or a 'test' player. This approach could be used for selecting players for future games. Also one can identify himself with a particular discipline for his future training. It is shown here that with the use of multivariate techniques we may form an index which can be used to assess the skills of a cricketer.

1. INTRODUCTION

Multivariate analysis is the analysis of observations on several correlated random variables on a number of individuals. Such analysis becomes necessary when one deals with several variables simultaneously. For example, suppose a cricket team manager wished to select a team for the current season. There will be a number of distinguished players such as bowlers, batsmen and all rounders with various other skills. It is the task of the manager to assess the players for their selection into the team.

There are a number of variables one can consider for this process. They are namely, runs made by the players, the time a player spent on the crease, number of balls played, number of fours and sixes scored, number of overs bowled by a player, number of catches taken by a player, etc.

The objective of this paper is to examine how some multivariate statistical techniques such as principal component analysis and discriminant analysis can be used to form an index for rating a person's sports skills. We may also check the existence of natural groupings, if any, among players using cluster analysis. One of the first books on statistical methods (Elderton [1]), made extensive use of cricket to illustrate elementary theory. Some years later G.H. Wood published a series of articles in "The Cricketer", which culminated in the reading of a paper to the Royal Statistical Society (Wood [2]). At the same time Sir William Elderton also read a paper on cricket

---

[1] Department of Statistics, Massey University, Private Bag 11222, Palmerston North, New Zealand

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

(Elderton [3]) so that he and Wood must share the distinction of producing the first full quantitative paper on sport.

No satisfactory distribution has been found to describe the frequency of scores of individual batsmen in cricket. The use of the Type X Pearson curve, which is exponential, the Geometric distribution, or the negative exponential have been recommended. Reep et al [4] showed that the negative exponential distribution produces a close fit to some series of scores. However, it was observed that the variance of the observed scores was higher than that of the geometric and negative exponential distributions. This confirms a common belief in cricket: there is a high likelihood of a batsman being dismissed early in his innings before he is settled; but once settled; adding to an already existing score becomes easier. It is to be noted that there exists some correlation between the two batsmen's scores since the individual innings of these two batsmen are placed under identical conditions. Elderton [1] noted that since the distribution of a single batman's scores was markedly skew (roughly geometric) and certainly not normal, any correlation would be hard to discern.

One of the main objects of this paper is therefore to apply a distribution free technique such as principal component analysis to develop an index which in some sense indicates the performance of a player. Further our analysis extends in using a "Mahalanobis distance" based discriminant technique for classifying a cricketer either as a test match player or one-day limited-over match player.

2. DESCRIPTION OF TECHNIQUES

2.1 Principle Components Analysis

When a larger number of measurements are available, it is natural to enquire whether they could be replaced by a fewer number of the measurements of their functions, without loss of much information, for convenience in the analysis and in the interpretation of data. Principal components, which are linear functions of the measurements are suggested for this purpose. It is therefore relevant to examine in what sense principal components provide a reduction of the data without much loss of the information we are seeking from the data. At first, principal components are most likely to be regarded by non-mathematicians as a highly arbitrary set of manipulations. Such a reaction should be dismissed, as soon as the geometrical meaning is considered: principal component analysis merely leads to new angles of viewing data analysis best suited to disclose the nature of size and shape variation. Often, the principal component analysis is first undertaken without any clear objective and then an attempt is made to interpret the derived results. The basic idea can in fact be well illustrated if we consider the case where $p = 2$ and a concentration ellipse as shown in Figure 2(a, b) below.

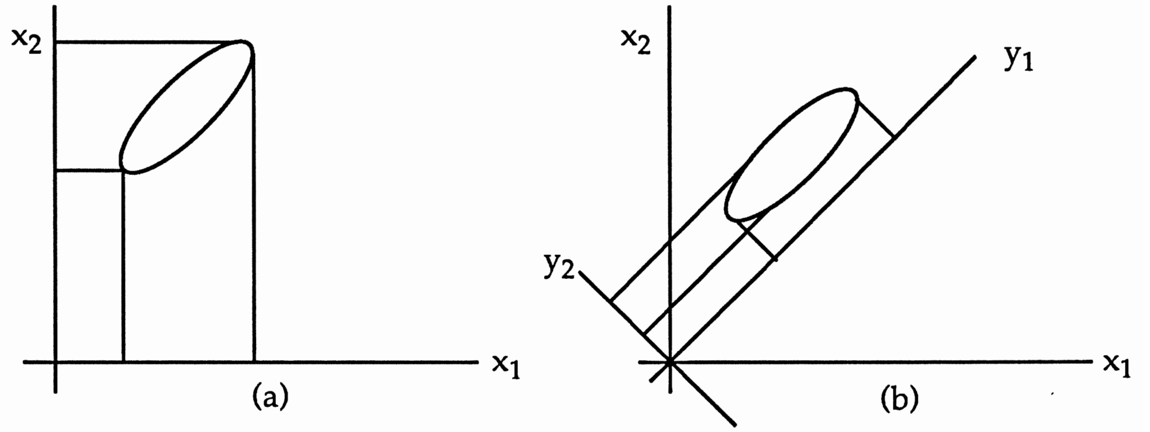


Figure 2: (a) Concentration ellipse for bivariate populations; (b) principal components analysis with two variables

Notice that there is a fair amount of dispersion in either x_1 or x_2 considered alone but that, due to the correlation between the two variables, much of the dispersion in their joint distribution occurs along the major axis of the ellipse. In fact, if we rotate the x_1, x_2 axes so that they are parallel to the axes of the ellipse and call the new axes y_1 and y_2, we can see from Figure 2(b) above that most of the variation is accounted for by y_1 and we might perhaps think of dropping y_2. In other words, since there is very little variation in y_2 we might eliminate it, thereby reducing the number of dimensions from two to one without a substantial loss of information.

In the general p-variable case we may find that only a relatively few dimensions are necessary, after an appropriate rotation, and that we can extract the necessary information from the first few components. Such components are used as indices to rate a player as a batsman, a bowler or an all rounder.

2.2 Discriminant analysis

Discriminant analysis as a whole is concerned with the relationship between a categorical variable and a set of interrelated variables. More precisely, suppose there is a finite number, say, g, of distinct populations, categories, classes or groups, which we shall denote here by $G_1, ..., G_g$. We will henceforth refer to the G_i as groups. Note that in discriminant analysis, the existence of the groups is known as *a priori*. An entity of interest is assumed to belong to one (and only one) of the groups. We let the categorical variable z denote the group membership of the entity, where z = i implies that the entity belongs to group $G_i(i = 1, ..., g)$. Also we let the p-dimensional vector **x** $= (x_1, ..., x_p)'$ contain the measurements on p available features of the entity. There is an enormous bulk of literature available in this area. The reader is referred to McLachlan [5] for an extensive bibliography.

In this framework, the topic discriminant analysis is concerned with the relationship between the group membership label z and the feature vector **x**. Within this broad topic there is a spectrum of problems, which corresponds to the inference-decision spectrum in statistical methodology. At the decision end of the scales, the group membership of the entity is unknown and the intent is to make an outright assignment of the entity to one of the g possible groups on the basis of its associated measurements. That is, in terms of our present notation, the problem is to estimate z solely on the basis of x. In this situation the general framework of decision theory can be involved. An example in which an outright allocation is required concerns the selection of players for future matches where the final decision to admit players as a one-day player or a test player is based on their current performance in the played matches. For this decision problem, there are two groups $G_i(i = 1, 2)$. The feature vector x for a player contains his past performance. A rule based on x for allocating a player to one of the two groups can be formed from an analysis of the feature vectors of known players from each of the two groups. A suitable allocation rule is given in the subsequent seasons.

A closely related approach, but with a different philosophy, is called canonical discriminant analysis. The simplest approach here is to construct a linear combination of the variables which clearly separates the groups, if the mean value of this new variable changes considerably from group to group with the value within a group being fairly consistent. One way to choose the coefficients in this linear combination is therefore to maximise the between-group variation, while simultaneously minimising the within-group variation. When this approach is used, it may be possible to determine several linear combinations for separating the groups.

These new variates (or dimensions) are usually referred to as canonical variates; where the first variate reflects as much inter-group difference as possible, the second captures the maximum of the balance and so on. The first few (one or two) canonical variates are generally sufficient to account for almost all of the group differences. One of the major attractions of this procedure is that if only one or two canonical variates are needed, then a simple graphical representation of the relationship between the various groups can be produced by plotting the values of these variates for sample observations. For further information readers are referred to Manly [6], Digby and Kempton [7], Dillon [8], Krzanowski [9] and Majer et al [10].

2.3 Cluster Analysis

The multivariate techniques we discussed so far have looked for a reduction in the dimensionality of the data to describe the individuals or units in a lower dimensional space. An alternative approach is to look for groups of individuals such that individuals within a group are relatively similar and at the same time dissimilar from individuals in other groups. In effect we want to group individuals rather than variables. Cluster analysis is an individual oriented technique and the basic aim is to find the "natural groupings" if any, of a set of individuals – such that members in a homogeneous cluster are similar to each other, but members of that cluster differ considerably from those of another. Thus the concept of cluster encompasses the duality of homogeneity within a cluster and heterogeneity between clusters.

The problem that cluster analysis is designed to solve is the following one: Given a sample of n objects, each of which has a score on p variables, devise a scheme for grouping the objects into classes so that "similar" ones are in the same class. Here the number of classes are unknown and the method sought must be completely numerical.

Grouping or clustering is distinct from the classification methods discussed. Classification pertains to a known number of groups and the operational objective is to assign new observations to one of these groups. Cluster analysis is a more primitive technique in that no assumptions are made concerning the number of groups or the group structure. Groupings are done on the basis of similarities or dissimilarities which are generally known as the measure of resemblance of individuals in the whole group.

In some investigations cluster analysis methods may be used to produce groups which form the basis of a classification scheme useful in later studies for predictive purposes of some kind. For example, a cluster analysis applied to a data consisting of a sample of cricketers may produce groups of players who react differently when treated with different type of training; thus enabling the manger to decide whether a particular type of training is suitable for a particular type of player. Here in our analysis of the cricket data we have identified such groupings.

3. AIM, CHOICE OF VARIABLES AND DATA

The data in this study refers to cricket test matches and limited-over one-day matches held between Australia and New Zealand, New Zealand and South Africa and Australia and South Africa. These matches were played in Australia between November, December 1993 and some one-day matches in January 1994. The data on various players of Australia and New Zealand have been collected on the following variables.

- x_1: Game played
- x_2: Type of game: test first inning, test second inning, one-day inning
- x_3: Runs made by the player in a particular inning
- x_4: Time player spent on the crease
- x_5: Number of balls faced by the player
- x_6: Number of fours scored by each player
- x_7: Number of sixes scored by each player
- x_8: Number of overs bowled by a player
- x_9: Number of maiden overs bowled by a player
- x_{10}: Runs scored against by a player (bowler)
- x_{11}: Number of wickets taken by the player
- x_{12}: Number of no-balls bowled by the player
- x_{13}: Number of wide-balls bowled by the player
- x_{14}: Number of catches taken by each player
- x_{15}: Classification of the player: Batsmen (B), Bowler (C) and All Rounder (A).

The data consists of 33 players, 14 from Australia and 19 from New Zealand. Note here that some of these variables are highly correlated, for example x_3, x_4 and x_5. Profiles of x_3, x_4, x_5, x_6, x_7 refer to batsmen, profiles of x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13} refer to bowlers; x_{14}, the number of catches taken by a player refers to a good fielder. A good all rounder will have qualities of both a good batsman as well as a good bowler.

The aim of this study is to characterise a player as a one-day match player or a test player on the basis of his past performance and also to identify which variables characterise them as a batsman, bowler or an all rounder. For this we have used the data of Australian and New Zealand players and their known classification as batsman, bowler and all rounder. This study could be helpful in classifying a new player based on his profiles in to one of the categories. The players can concentrate more on that area knowing in which category he falls. Also the team selectors can use the rule or index to select the best players.

4. RESULTS

The various statistical analyses were employed using the statistical computing software package SAS. The results and conclusions are reported separately in the following sections. In this study, we have ignored the difference between the first and the second innings of test matches, hence the "type of match" played by a player constitute only of two categories: One-day innings (0), Test innings (T).

4.1 Principal Component Analysis (PCA)

Three sets of PCAs were performed, one for each category: Batsmen, Bowlers, All rounders, using the characteristics that best describe them. Noting the fact that these characteristics are observed on quite different scales or units (runs scored, overs bowled etc.), the PCAs were carried out on the standardised data. First consider the results of the PCA on "All rounders". Table 4.1(a) shows the variables/characteristics used, the coefficients of the first four principal components (PC1, PC2, etc.) and the percentage of variation explained by each PC.

It can be observed in the analysis that the first four principal components (PCS) absorbed 83.2% of the total variance. Furthermore, the first principal component (PC1) has relatively larger coefficients attached to variables such as runs scored (x_3), time on crease (x_4), number of balls faced (x_5) and number of fours scored (x_6) than with the other variables. Though PC1 explains 38% of the total behaviour of an all rounder it measures more of his batting ability than his bowling ability. Further the coefficients of PC1 associated with x_7 (no. of sixes) is small (0.185) indicating that this PC1 is a measure of "steady scoring ability". Note, on the other hand the coefficients associated with the "bowing variables", except for x_{13} (# wides), though not as large as the "batting" variables indicate to a reasonable extent the quality of a bowler. Hence, PC1 gives an ideal index of a good all rounder.

On the other hand, if we look at PC2, it is clear that this component has relatively large coefficients associated with variables such as x_8 (no. of overs bowled by a player), x_{10} (runs given) and x_{11} (no. of wickets taken). Therefore this component is a measure of "bowling ability". Also note that PC3 has large coefficient associated with x_{13} (no. of wide balls bowled) and hence a smaller PC3 is a good measure of a bowler.

Further, PC4 has a large coefficient attached to x_7 (no. of sixes scored) and hence measure the "big hitting" ability of a player. Thus a good all rounder must have a high value for PC1, PC2 and PC4 but a low value for PC3. Hence these PC scores can be used as indices to identify the skills of players and also for the selection process. The mean scores of PC2 were plotted against those of PC1 and is given in Figure 4.1(a). The symbol in the plot is the value of x_2, the type of the game played, which is either a test match (T) or a one-day match (O). Note that there is a clear distinction along PC1 in the performance of players between the one-day and test matches. Similar results are observed with the PCA for the batsmen, and bowlers of all matches. The result for the batsmen are given in Table 4.1(b) and that for the bowlers are given in Table 4.1(c) along with the mean PC score plots respectively in Figure 4.1(b) and Figure 4.1(c).

It can be seen from Table 4.1(b) that the first two principal components explain about 90% of the total variance, hence only these two components are used in the discussion. PC1 seems to be a weighted average of all the variables (x_3 to x_7). Observe that the PC1 for the batsmen has relatively large coefficients with all these variables except for x_7 (no. of sixes). Therefore PC1 is a measure of a "steady scoring" batsman. PC2 on the other hand, is a contrast between x_3, x_4, x_5 and x_6, x_7, and has a larger coefficient with x_7. Hence PC2 may be a measure of a "big hitting" batsman. Therefore, larger values of PC1 and PC2 may identify a good batsman. These indices could be used for selection procedures. It is noticeable from Figure 4.1(b) that there is a clear distinction in performance between one-day matches and test matches. A similar result was observed for bowlers of all matches. We can see from Table 4.1(c), that PC1 has relatively large coefficient with variables x_8 to x_{12} except with x_{13}, where a negative coefficient is sighted. Thus PC1 measures a good bowling ability, and hence a higher value of PC1 will indicate a good bowler. PC2 on the other hand has a negative coefficient attached to x_{10} (runs given), x_{12} (no balls) and therefore it is a contrast between x_{11}, x_{13} and x_{10}, x_{12}. Again from Figure 4.1(c) we could see a clear distinction in performance among bowlers between one-day and test matches.

Table 4.1(a)

PCA of All Rounders of all matches (both one-day and test innings)

	Eigen Vectors			
Variables	PC1	PC2	PC3	PC4
x_3: Runs scored	0.421	-0.251	0.064	0.094
x_4: Time on crease	0.411	-0.271	0.099	-0.024
x_5: Balls faced	0.403	-0.277	0.125	-0.008
x_6: Fours	0.411	-0.215	0.085	-0.056
x_7: Sixes	0.185	0.141	-0.308	0.817
x_8: Overs bowled	0.250	0.459	0.062	-0.201
x_9: Maidens	0.272	0.220	-0.223	-0.489
x_{10}: Runs against	0.168	0.474	0.146	0.063
x_{11}: Wickets	0.279	0.365	0.082	0.094
x_{12}: No balls	0.193	0.392	-0.112	0.030
x_{13}: Wides	-0.088	0.129	0.879	0.163
% variation explained	38.1	27.3	9.5	8.3
Eigen values	4.19	3.00	1.05	0.92

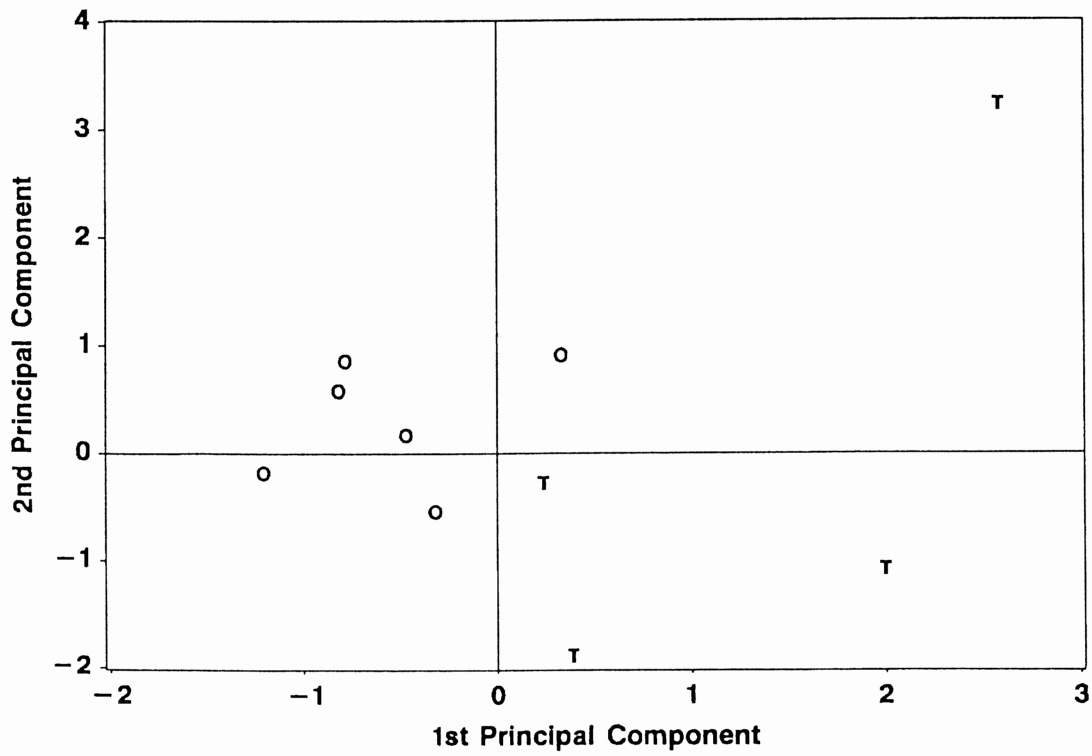


Figure 4.1(a) Plot of PC1 and PC2 (mean of PC scores) of all rounders of all matches

Table 4.1(b)

PCA of batsmen of all matches

Variables	Eigen Vectors			
	PC1	PC2	PC3	PC4
x_3: Runs scored	0.517	-0.034	0.163	-0.839
x_4: Time on crease	0.518	-0.154	0.273	0.400
x_5: Balls faced	0.523	-0.138	0.237	0.352
x_6: Fours	0.432	0.278	-0.852	0.090
x_7: Sixes	0.053	0.937	0.339	0.060
% variation explained	68.6	21.5	7.5	1.9
Eigen values	3.43	1.08	0.38	0.10

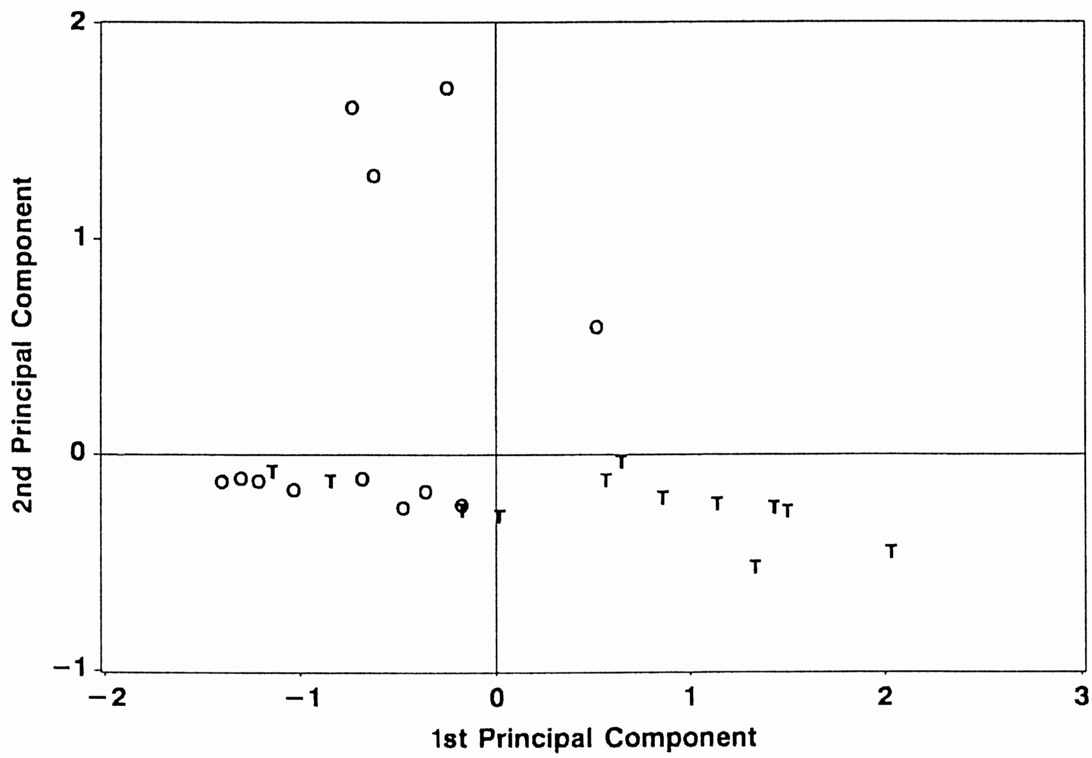


Figure 4.1(b) Plot of PC1 vs PC2 (mean PC scores) of Batsmen of all matches

Table 4.1(c)

PCA of bowlers of all matches

Variables	PC1	PC2	PC3	PC4
	Eigen Vectors			
x_8: Overs bowled	0.586	0.008	-0.111	0.209
x_9: Maidens	0.486	0.128	-0.362	-0.136
x_{10}: Runs against	0.501	-0.082	0.097	0.502
x_{11}: Wickets	0.254	0.642	0.056	-0.589
x_{12}: No balls	0.280	-0.278	0.843	-0.272
x_{13}: Wides	-0.159	0.698	0.364	0.515
% variation explained	45.8	20.2	14.3	11.5
Eigen values	2.75	1.21	0.86	0.69

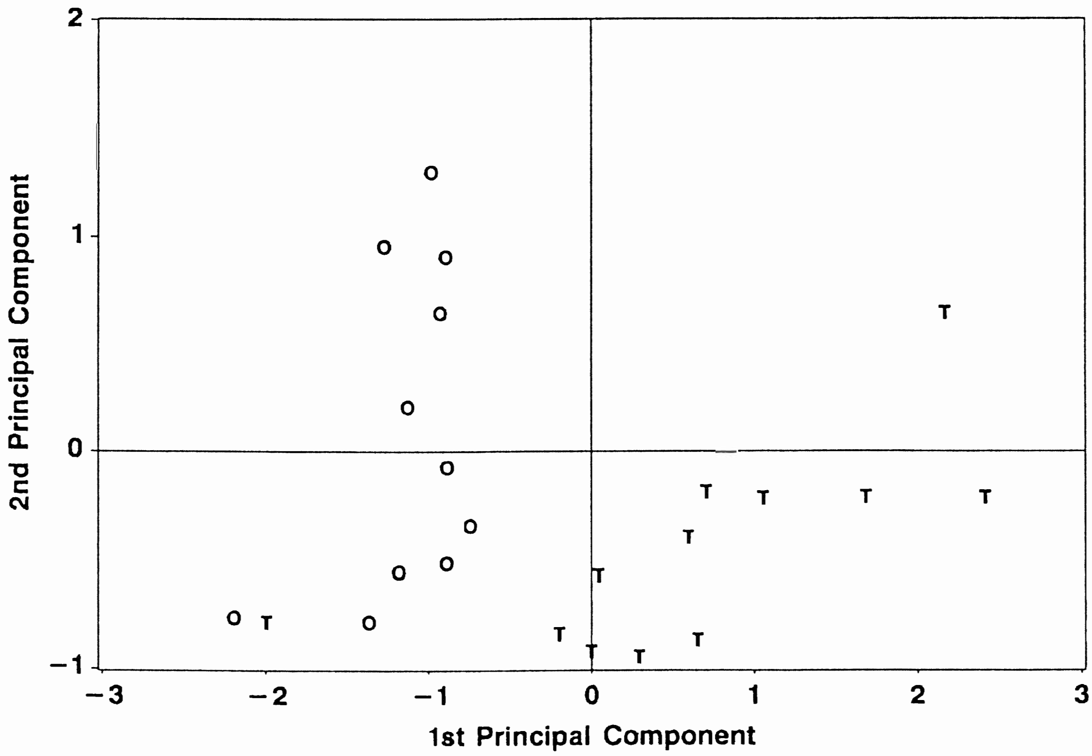


Figure 4.1 (c) Plot of PC1 vs PC2 (mean PC scores) of bowlers of all matches

4.2 Discriminant Analysis

Here our interest was focused on classifying a cricketer as a one-day limited-over match player or a test match player. As introduced in Section 2.2 we let G_1 denote the group of one day matches and G_2 denote the group of test matches. In our data those who were identified as all rounders, played 48 one-day matches and 29 test matches. The assumption of equal covariance matrices was found to be reasonable by the comparison of sample covariance matrices. Further, under the assumption of multivariate normality for the variables considered and equal priors, we undertook a linear discriminant analysis approach and the results are summarised in Table 4.2(a). The allocation of a player to G_1 or G_2 depends on which group he is closer to. Using the coefficients under each main category and the sub-category, [that is, for example, under batsmen (main category), the sub-categories are test match and one-day match]. One can find the distance of an individual from the centroid of each group G_1 and G_2. The individual is then allocated to G_1 or G_2 according as to which centroid he is closer to. The cross-validated error rates associated with each rule are found reasonably small and hence the rules are reliable.

With canonical discriminant analyses, since we have only two groups the first canonical variate, CAN1, only is meaningful. The coefficients for the all rounders, batsmen and bowlers are given in Table 4.2(b) along with the plot of CAN1 against CAN2 (where CAN2 represents a random perturbation given to each observation). The corresponding plots of CAN1 against CAN2 are given in Figures 4.2(a), (b) and (c), respectively for all rounders, batsmen and bowlers. We can see a clear separation between one-day match players and the test match players. It appears that players with larger CAN1 score are test match players where as those with smaller CAN1 score are one-day players.

Table 4.2(a)

Linear discriminant function coefficients, the Mahalanobis distance between G_1 and G_2 and the number of matches played

	All rounders		Batsmen		Bowlers	
Coefficients	**one day match**	**test match**	**one day match**	**test match**	**one day match**	**test match**
constant	-1.584	-2.179	-0.341	-1.367	-1.239	-2.617
x_3	0.094	-0.037	0.017	-0.056		
x_4	-0.011	0.010	-0.016	-0.001		
x_5	-0.006	0.036	0.027	0.050		
x_6	-0.414	-0.036	-0.061	0.162		
x_7	-0.438	0.111	1.909	-0.773		
x_8	0.341	0.026			0.064	0.189
x_9	-0.782	1.035			-0.220	0.046
x_{10}	0.007	0.038			0.022	0.005
x_{11}	0.186	-0.790			0.564	0.031
x_{12}	-1.273	-1.096			-0.191	0.494
x_{13}	0.987	-0.106			0.974	0.046
# matches played	48	29	74	73	59	71
Squared Mahalanobis distance	5.107		1.635		3.826	
Total Cross validated error rate	0.093		0.232		0.136	

Table 4.2(b)

Canonical discriminant coefficients of the first canonical variate CAN1

Coefficients	All rounder	Batsmen	Bowler
x_3	-2.005	-1.917	
x_4	0.774	1.017	
x_5	1.333	1.164	
x_6	0.669	0.625	
x_7	0.104	-0.382	
x_8	-0.809		0.732
x_9	1.336		0.565
x_{10}	0.297		-0.271
x_{11}	-0.379		-0.400
x_{12}	0.043		0.547
x_{13}	-0.245		0.394

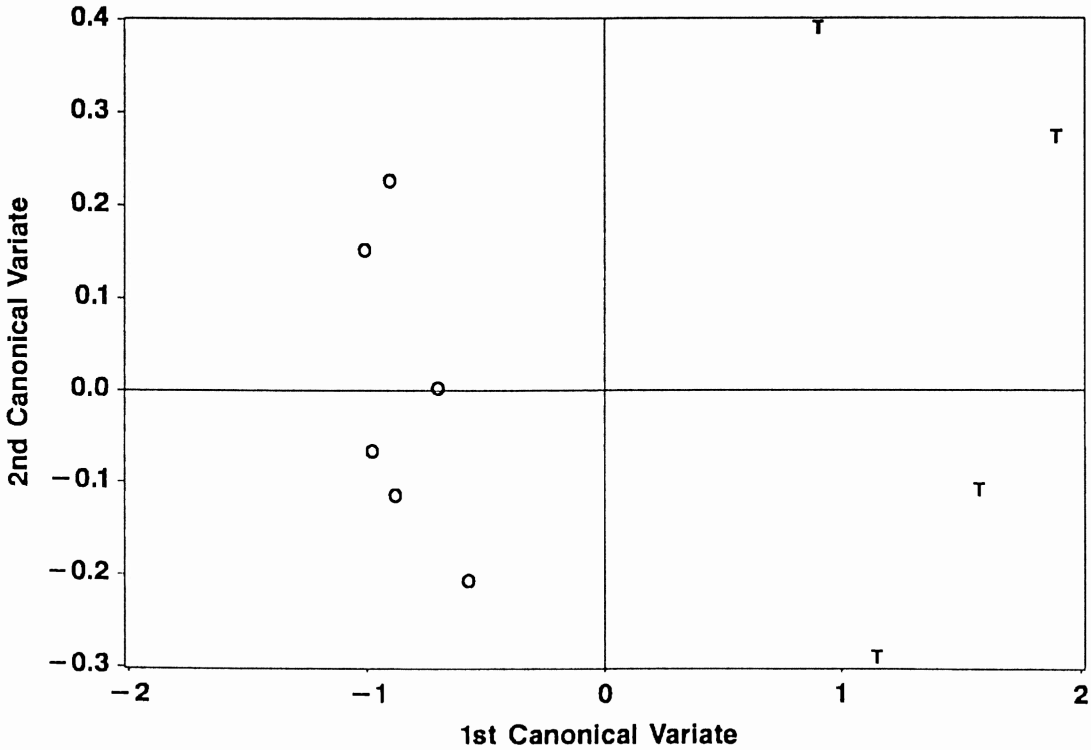


Figure 4.2 (a) Plot of mean canonical variates scores; CAN1 against CAN2 for the all rounders

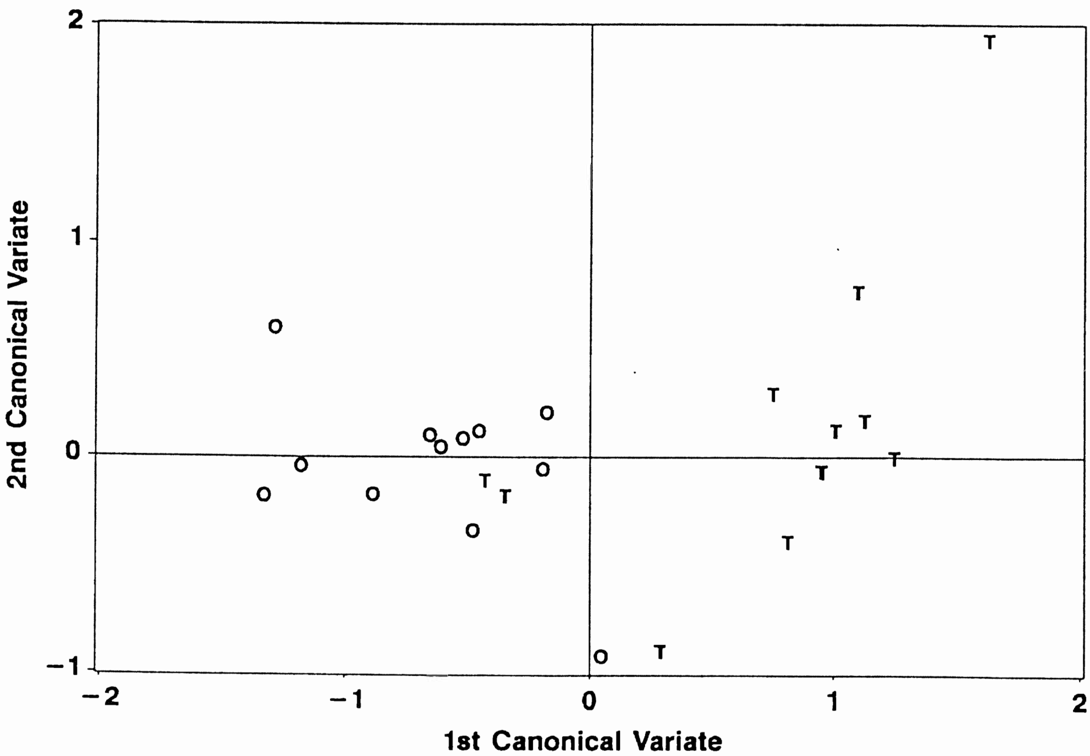


Figure 4.2(b) Plot of mean canonical variates scores CAN1 against CAN2 for the batsmen

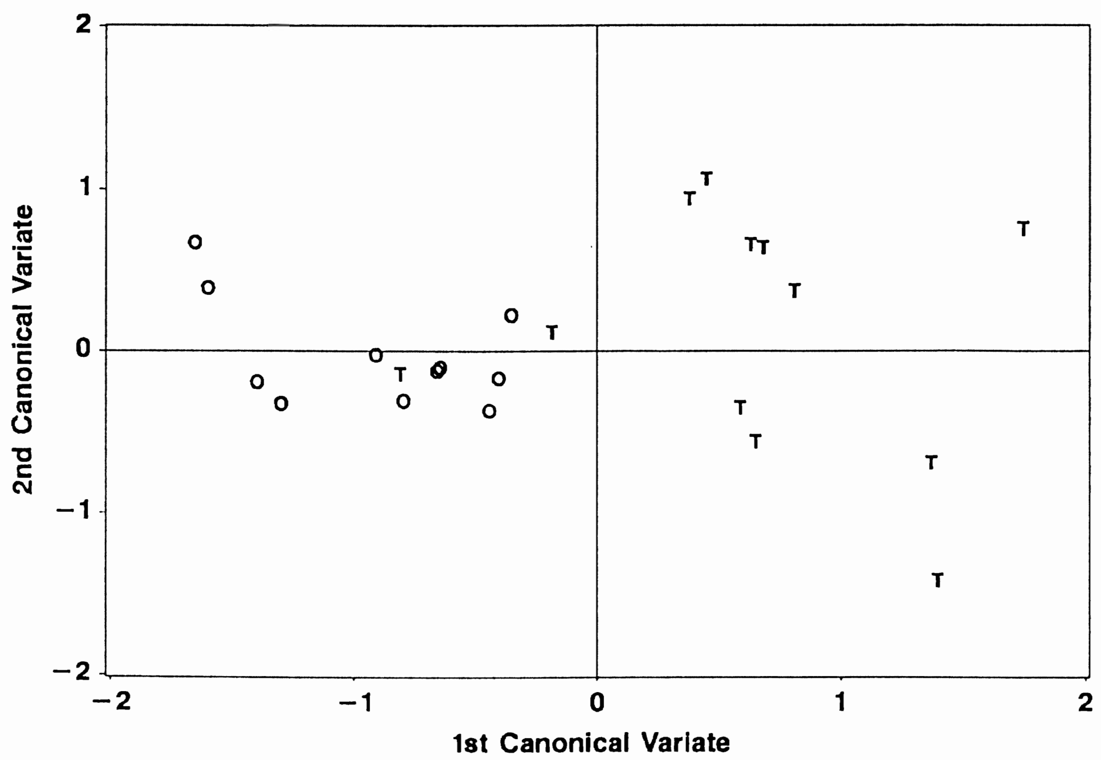


Figure 4.2(c)

Plot of mean canonical variates scores CAN1 against CAN2 for the bowlers

4.3 Cluster analysis

Here our interest was to see whether we could recover the (known) groupings using cluster analysis. This is generally used at an exploratory stage. In our data we know that there are three categories of players namely, All rounders, Batsmen and Bowlers, identified by earlier records. For the purpose of the analysis we pretend that the whole data is a mixture of all type of players. Then single-linkage cluster analysis was undertaken, and from the dendrograms given in Figures 4.3(a) and 4.3(b), we can see the existence of 3 distinct groups. It is difficult to identify as to what type of group each player falls if we don't have any information on the players. If we can identify one in a group as a batsman then the homogeneity of the group will imply that that particular group consists of all batsmen.

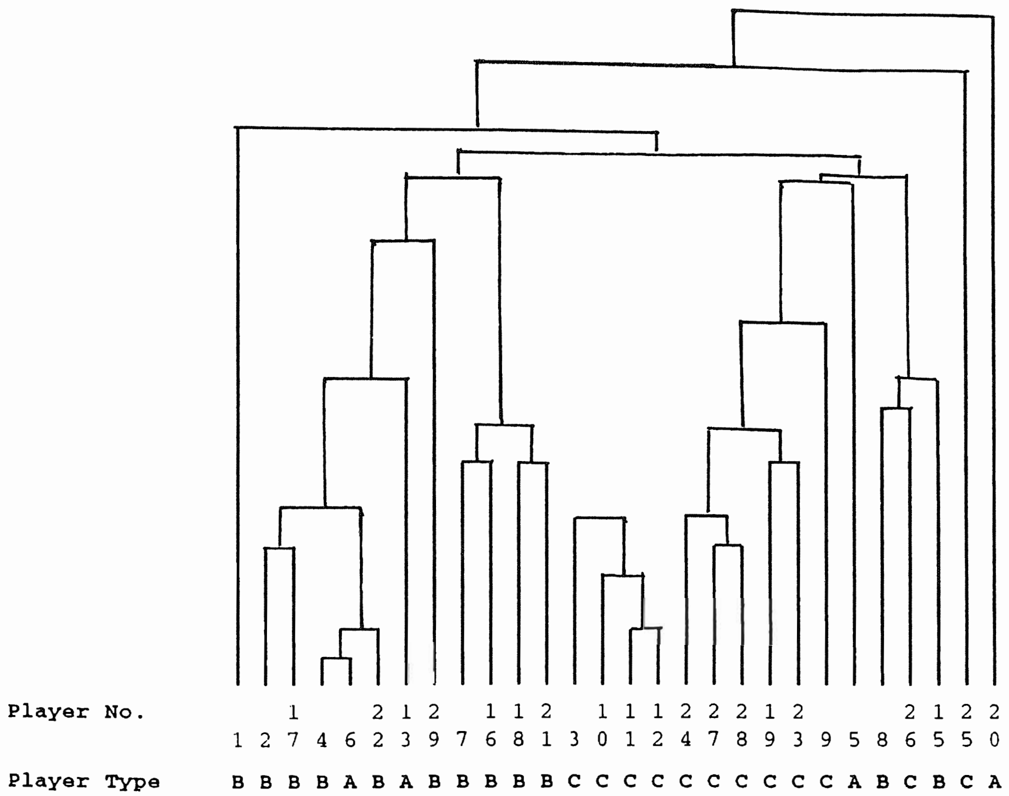


Figure 4.3(a) Dendrogram produced for the test match players

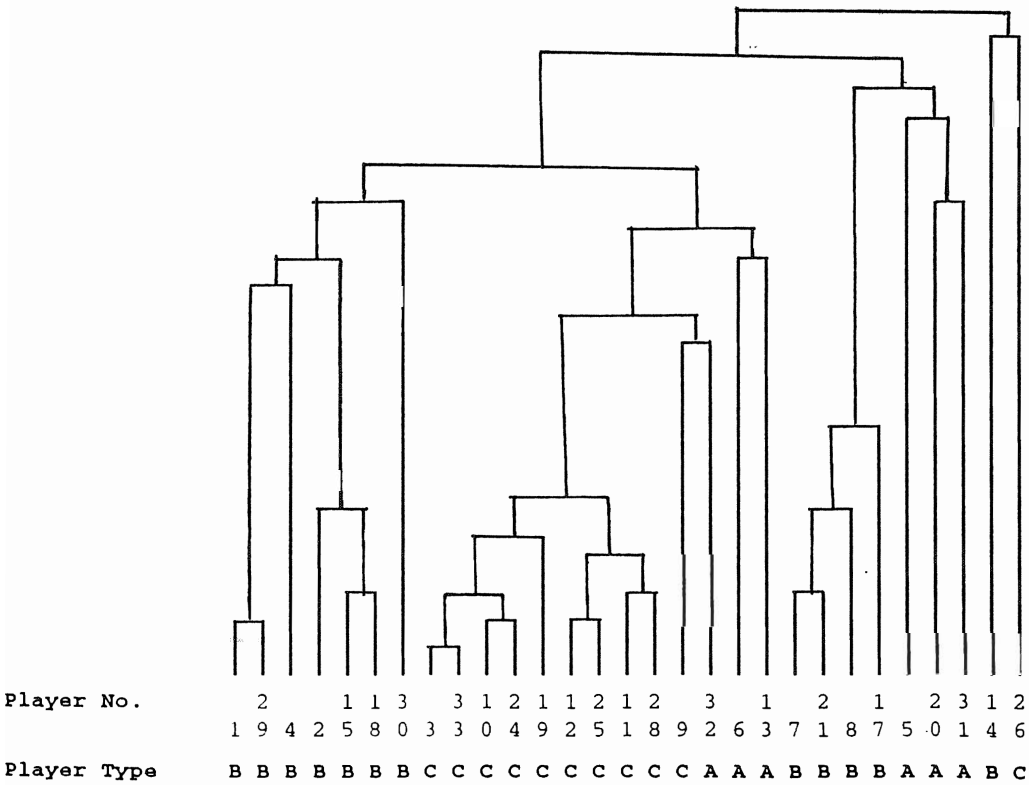


Figure 4.3(b) Dendrogram produced for the one-day match players

5. CONCLUSION

Methods of handling multivariate statistical data are constantly being improved as computers increase in sophistication. Large multivariate data sets can prove difficult to comprehend and methods of summarising and extracting relevant information are necessary. In this paper we have shown how some multivariate statistical techniques such as principal component analysis, discriminant analysis and cluster analysis can be used in the context of sports, in particular for the cricket data for rating cricketers as one-day match players or test players and what special characteristics identifies them as all rounders, batsmen or bowlers. The indices developed are useful for the selecting process as well as for an individual player to know in which category he faces for his future training. Thus quantitative techniques have some appeal in the analysis of sports data.

REFERENCES

[1] Elderton W.P., Primer of Statistics (Black: London, 1909).

[2] Wood G.H., Cricket Scores and Geometrical Progression, *J.R. Statist. Soc. A.* 108, 12-22, (1945).

[3] Elderton W.P., Cricket Scores and Some Skew Correlation Distribution, *J.R. Statist. Soc. A.*, 108, 1-11, (1945).

[4] Reep C., Pollard R., and Benjamin B., Skill and Chance in Gall Games, *J.R. Statist. Soc. A.*, 134, 623-629 (1971).

[5] McLachlan G.J., Discriminant Analysis and Statistical Pattern Recognition, (John Wiley, 1992).

[6] Manly B.F.J., Multivariate Statistical Methods a Primer, (Chapman and Hall, London, 1986).

[7] Digby P.G.N. and Kempton R.A., Multivariate Analysis of Ecological Communities, (Chapman and Hall, London, 1987).

[8] Dillon W.R. and Goldstein M., Multivariate Analysis Methods and Applications, (John Wiley, 1984).

[9] Krzanowski W.J., Principles of Multivariate Analysis, (Oxford University Press, Oxford, 1988).

[10] Majer J.D., Recher H.F. and Ganeshanandam S., Variation in foliar nutrients in Eucalyptus trees in eastern and Western Australia, *Australian Journal of Ecology 17*, 383-393 (1992).

FINALS DRAWS

R.O. Weber[1] and N.J. de Mestre[2]

Abstract

Since 1931 the McIntyre system in one form or another has been used to prepare the finals draws for the Victorian and Australian Football Leagues. In 1994 the McIntyre system is to be modified so that eight teams are arranged in a draw to produce a winner over four week-ends, with the top teams having extra chances. The new system is analysed, and probabilities associated with it are considered.

1. INTRODUCTION

In 1931 the VFL (now the AFL) adopted the McIntyre final four for its finals series, played at the end of the home-and-away series. In 1972 this was extended to the McIntyre final five, played over four weeks. For the finals series in 1991, 1992 and 1993, a modified McIntyre final six operated, still over four weeks, but subject to considerable scepticism and uncertainty from the public. For the 1994 finals campaign there will be eight teams (from a league of only fifteen) in a four-week series based on a double McIntyre final four. It remains to be seen if the new system gains wide acceptance.

Mathematically, it is of interest to try to enumerate all the possible final draws for n teams held over m weeks. The merits of each draw can then be assessed, with a view to finding the 'fairest' finals draw. Of course it is vital to ensure that teams have an incentive to finish 'high on the ladder', i.e. highly ranked after the home-and-away series. Generally, enumerating all possible final draws is a huge task. However, when there is a large number of teams in a small number of weeks, the possibilities are fewer.

As an extreme case, in four weeks the maximum number of teams which can be accommodated is sixteen. As each game at each stage will eliminate one team, this would not be a suitable option for the AFL or similar leagues.

In the following sections the original McIntyre system and the modified McIntyre systems will be examined. Clarifying the operation of these final draws (which are more complex than a knockout competition) is of interest in itself. Furthermore, it is vital to anyone wishing to wager on the eventual winner.

[1] Department of Mathematics, University College, UNSW, ADFA, Canberra ACT 2600
[2] School of Information Technology, Bond University, Gold Coast Qld 4229

2. THE MCINTYRE SYSTEM

As introduced for a final four in 1931, the McIntyre system [1] has one elimination game each week. The diagram in Figure 1 illustrates a match and its result geometrically. The left-hand horizontal lines represent the two teams that are to play in a match, while the right-hand horizontal lines represent, from the top, the winner and loser respectively. When there is no bottom right-hand horizontal line that team has been eliminated. A single horizontal line indicates that the team is resting during that week.

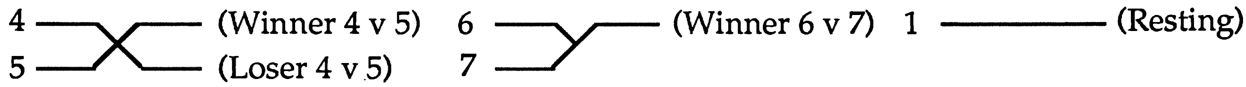


Figure 1: Basic geometric match representation

The general operation of the original McIntyre finals systems [1] is illustrated geometrically in Figure 2 with the number of teams (n) at the beginning of the finals series being whittled down as the series progresses from left to right. The diagram illustrates the pattern for n = 7.

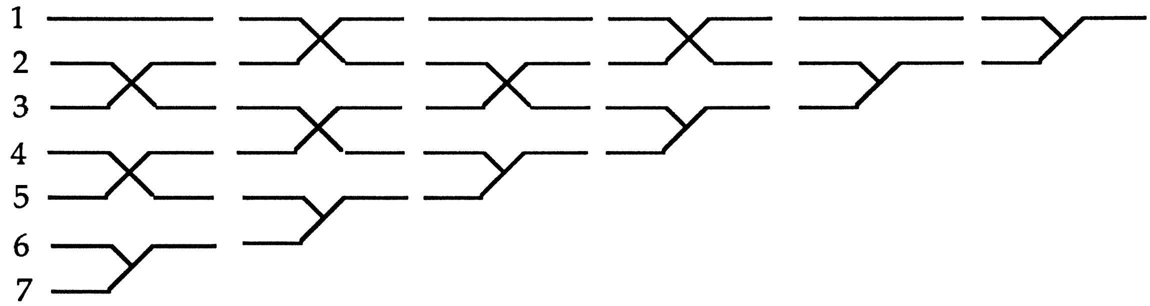


Figure 2: Original McIntyre finals draw for n = 7

Noteworthy features are that the system is suitable for any number of teams (odd or even) and that it always confers an advantage to teams with a higher ranking at the beginning of the finals campaign. In this original form it has been abandoned by the AFL for its finals series with six or eight teams because it is deemed to require too many weeks to complete.

The essential features shown in Figure 2 are:

(i) Each column represents the draw for a week-end.

(ii) The number of weeks needed is always one less than the number of teams, since only one team is eliminated each week.

(iii) Only the last two teams do not have a double chance.

(iv) It can be truncated from the left to give the draw for any number of teams less than 7.

(v) It can be extended to the left to give a reasonable draw for any number of teams greater than 7.

(vi) Some matches may be repeated. For example, considering the case n = 7 it is possible for team 2 to play team 3 in week one with the former winning. Then in week two team 2 loses to team 1 and team 3 defeats the winner of 4 v 5. Hence in week three there is a re-match of 2 v 3.

Therefore if the original McIntyre finals system was to be used for the final eight there would have to be a seven-week series and the possibility of considerable repetition of games.

3. THE MODIFIED MCINTYRE SYSTEMS

The modified McIntyre final six, and now final eight, have been quite controversial with the football-interested public. This is principally because they are significantly more complex than the original McIntyre system; necessarily so to play a finals series in only four weeks. Additionally an incorrect use of the system in 1991 (when third and fourth had little advantage over fifth and sixth) did little to endear the 'new' system to the public.

Of immediate interest, however, is the modified system to be applied in 1994. This is known as the McIntyre final eight system [2], [3] and is based on the Double McIntyre Final Four because of the restriction on the number of weeks available for the finals.

Essentially nine matches are required over four weekends to reduce the final eight to a single winner. A Double McIntyre Final Four could be a solution as indicated in Figure 3.

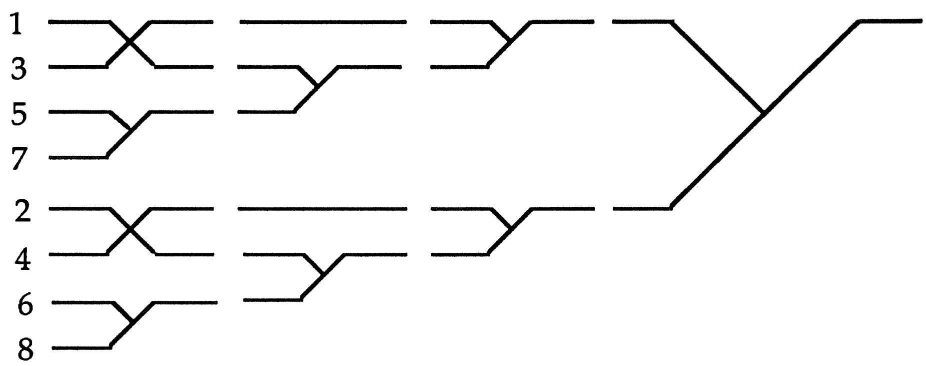


Figure 3: Double McIntyre Final Four

As constituted above this system has one main drawback. It does not distinguish between the finishing positions of the first four teams.

To try to rectify this situation, McIntyre suggested that the first week of the two halves of the Double McIntyre Final Four draw be rearranged as

1 v 7, 3 v 5, 2 v 8, 4 v 6

where 1 through 8 denotes the ranked order of the teams at the end of the home-and-away series.

Consider in detail the first half containing the teams 1, 3, 5, 7, as the system for the second half with teams 2, 4, 6, 8 will be the same.

For the modified McIntyre system there will be two winners and two losers after the first week of matches. The higher ranked winner has a rest the following week, while the lower ranked loser drops out. The remaining two teams play-off in the second week for the right to progress to the third week.

For the matches 1 v 7, 3 v 5 the system depends heavily on the outcome of 1 v 7. If team 1 defeats team 7 the diagram (Figure 4) becomes

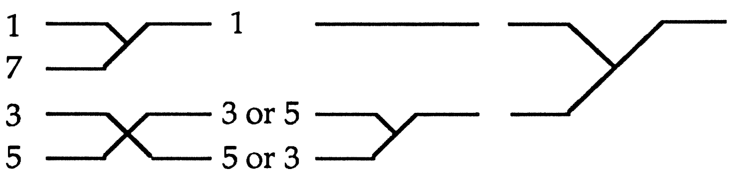


Figure 4: First-half modified McIntyre (team 1 defeats team 7)

If team 7 defeats team 1 the diagram (Figure 5) becomes

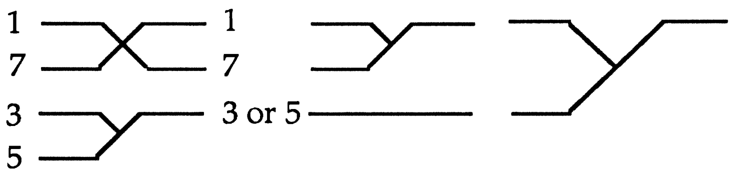


Figure 5: First-half modified McIntyre (team 7 defeats team 1)

Note that it is possible for team 3 to be eliminated in the first week. This would happen if team 7 defeated team 1, and team 5 defeated team 3. Team 3 would then become the lower-ranked loser, and would therefore not have a double chance. Therefore only teams 1 and 2 would be guaranteed a second chance if defeated in the final series.

The worst feature of the draw associated with Figures 4 and 5 is that all situations produce a repetition of at least one previous week's match. In other words, two of week one's matches will be repeated in week two – a highly unsatisfactory situation.

To eliminate this problem McIntyre [2] suggests cross-divisional (or cross-halves) adjustment in the second week. As an example, consider team 1 defeating team 7, team 2 defeating team 8, and the remaining matches (3 v 5, 4 v 6) going any way. Then 3 v 5 and 4 v 6 will be repeated in week two if no crossing over is allowed. As expected from the rankings – and produced by the results – teams 1 and 2 will obtain a rest in week two, while teams 7 and 8 drop out. McIntyre suggests that the opponents in the matches for week two be crossed over to yield 3 v 6 and 4 v 5, with both losers dropping out.

The cross-divisional process is more complicated if team 1 defeats team 7 but team 8 defeats team 2. Then team 7 and the lower-ranked loser of 3 v 5 and 4 v 6 drop out, while team 1 and the higher-ranked winner of 3 v 5 and 4 v 6 gain the rest week.

McIntyre also points out that there are 105 possible options for the first week alone. Besides the two mentioned so far:

1 v 3	5 v 7	2 v 4	6 v 8	(Figure 3)
1 v 7	3 v 5	2 v 8	4 v 6	(Figure 4 or 5)

he suggests

1 v 8   2 v 7   3 v 6   4 v 5,

and this has been adopted by the AFL for the 1994 season.

As an illustration of what will happen consider the following scenario where teams 2, 7, 4, 5 are in the first half of the draw and the lower-ranked teams both win, while teams 1, 8, 3, 6 are in the second half of the draw and the higher-ranked teams both win (see Figure 6). After the first week teams 4 and 8 drop out, being the lower ranked losers in their halves of the draw. Also teams 1 and 5 progress directly to week three, as they are the higher-ranked winners in their respective halves of the draw.

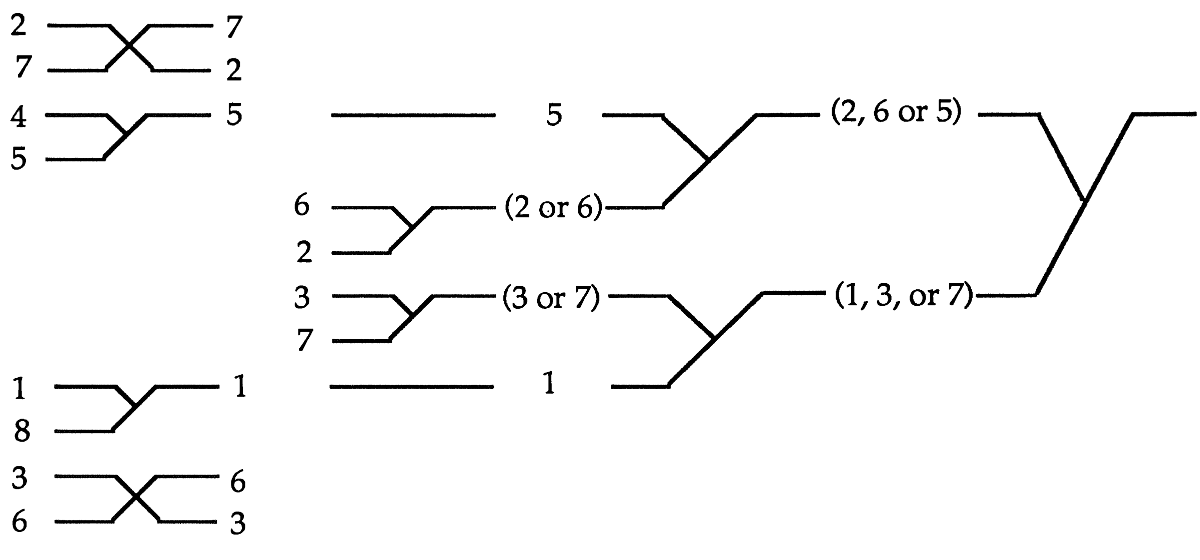


Figure 6: A possible outcome

For week two the cross-divisional matches are played to eliminate repetition of 2 v 7 and 3 v 6. The new draw is 6 v 2 and 3 v 7, and the losers drop out. This reduces the finals series by week three to four teams. These play two semi-finals to produce two winners who eventually meet in the Grand Final. One weakness of the above system is obvious from the example chosen in Figure 6. Team 1 can lose the second semi-final and not be able to invoke its double chance. Thus it will be out of the competition after losing only once in the finals. There are other anomalies, but this is the most blatant. After week one all double chances are removed.

4. BETTING PROBABILITIES

Probability models for the Australian Football League (AFL) finals series have been produced on the original McIntyre final five by Schwertman and Howard [4], [5] and on the modified McIntyre final six with preferential ranking by Clarke [6].

Schwertman and Howard initially assume independence and that in each game in the finals series each team has an equal chance of winning. They then proceed to consider the more realistic situation of incorporating the relative strengths of the teams involved into the probabilities. Two models were considered. One was based on the teams' percentages at the end of the home-and -away series, while the other was based on the seedings 1 to 5. Comparing with historical data for the final five series from 1972 to 1989, they found that both models gave a reasonable fit, but that the seedings' model produced a smaller chi-square value indicating a better fit.

Clarke [6] considered the more complicated situation of the modified McIntyre final six with preferential rankings used by the AFL for the 1992 and 1993 seasons. The method involved using a spreadsheet to designate all possible ultimate finishing orders. There are $2^7 = 128$ possible sequences and, because of the preferential ranking, the pattern for the final order is not easy to determine. However, once all the 128 possible sequences are written down, the probability of a team ultimately finishing at the top can be obtained simply by counting. Clarke assumed that the probability of a team winning any match in the finals series was $\frac{1}{2}$, and therefore did not include calculations for the Schwertman and Howard seedings' model.

For the modified McIntyre final eight system to be used in 1994, with a probability of $\frac{1}{2}$ being assumed for any team to win a finals' match, the probabilities of teams winning the Grand Final are shown in Table 1.

Table 1

Team	Final Eight Probabilities	Final Six Probabilities
1	$\frac{3}{16}$	$\frac{4}{16}$
2	$\frac{3}{16}$	$\frac{4}{16}$
3	$\frac{2}{16}$	$\frac{3}{16}$
4	$\frac{2}{16}$	$\frac{2}{16}$
5	$\frac{2}{16}$	$\frac{2}{16}$
6	$\frac{2}{16}$	$\frac{1}{16}$
7	$\frac{1}{16}$	
8	$\frac{1}{16}$	

The middle four teams have an identical chance which doesn't give sufficient incentive, in our view, to these teams to finish strongly. Furthermore, there is really only a small spread in chances between the top and bottom teams in the final eight compared with the final six of previous seasons (also included in Table 1). This means that we are more likely to see teams from the bottom half of the seedings win the Grand Final under the expanded final eight series than in the final six series. ($\frac{6}{16}$ compared with $\frac{5}{16}$). A corollary to this is that teams 1 and 2 have lower probabilities of winning under the 1994 system than they did in previous years.

An alternative to Table 1 can be developed if the 1994 McIntyre Final Eight system is used in conjunction with the Schwertman and Howard seedings' model. For eight teams the seedings' model would produce 56 different probabilities for the probability P(i,j) that team i defeats team j. This is based on

$$P(i,j) = \frac{j}{i+j} \quad (i \neq j;\ i = 1, \ldots, 8;\ j = 1, \ldots, 8)$$

Using these probabilities, the total probability of each team subsequently winning the Grand Final can be computed if all the possible paths to winning are known for that team. These paths were reasonably simple in the original McIntyre finals draw but are much more complex in the modified McIntyre finals draw with preferential ranking. The possible outcomes could be evaluated using a spreadsheet as suggested by Clarke [6]. This has not yet been completed by us. We are saving it as an exercise for our students.

5. Acknowledgements

ROW wishes to thank K.G. McIntyre (a delightful octogenarian) for discussions on his many ideas, ranging from finals draws to the Portuguese discovery of Australia 460 years ago. Col Hutchinson, the AFL statistician is to be thanked for providing valuable background material.

References

[1] McIntyre K., Report to Victorian Football League, 1931.

[2] McIntyre K., Report No 1 to the Australian Football League, 12 May, 1992.

[3] McIntyre K., Report No 2 to the Australian Football League, 20 May, 1992.

[4] Schwertman N.C. and Howard L., A probability model for the Victorian Football League finals series, *Aust. Math. Teacher*, 45 (2), 2–3 (1989).

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

[6] Clarke S.R., Analysing football finals with a spreadsheet, *Aust. Math. Teacher*, 7 (10), 35–44 (1993).

SURVEY OF USEFUL SYSTEMS REFERENCES

Richard Monypenny[1]

Abstract

A survey of four groups interested in systems (Circles readers, readers of the Bulletin of ISSS, ASSCT members and people interested in Mathematics and Computers in Sport) has identified about 300 references which respondents have found to be "useful systems references", but failed to identify why most of these references are seen to be useful.

In order to achieve our aim of providing useful references to the next generation of non-specialist users of systems concepts:

First, we need to understand the reasons for the low response rate to the survey.

Second, we need to determine why most of these references are seen to be useful.

Third, we need to distinguish which are references of system content or knowledge; that is, references that describe what is known about a given system.

Fourth, we need to distinguish which are references of systems process; that is, references about the process used to develop solutions in a given system. And

Fifth, we need to distinguish which are references of generic systems processes; that is, references about processes that can be used to develop solutions in many, most or in all systems.

We plan to make these results and the useful references widely available. At this stage various printed and electronic means are being considered. In the meantime please contact the author on Facsimile (077) 81 4149.

1. INTRODUCTION

Systems theory has developed over the years into many specialist areas, each with its own specialist literature. For example: agricultural systems, soft systems methodology, systems analysis, systems engineering, etc. Many of us have colleagues or acquaintances with such training.

However, there is another aspect to systems; that is, non-specialists who turn to systems, searching for help in attempting to solve their problem. Often they have turned to systems precisely because their problem is seen to be difficult, or even intractable because of the perceived "systems" aspects of their problem.

[1] Department of Economics, James Cook University, Townsville Qld 4811.

Many non-specialists have found this contact with systems time-consuming, frustrating and unproductive; and have moved on to something else without taking advantage of what systems could have contributed to the solution of their problem.

There are however many others who have succeeded. It would be very useful to tap this experience and make it readily available to the next generation of non-specialist users of systems concepts. The results of a survey of four groups interested in systems (Circles readers, readers of the Bulletin of ISSS, ASSCT members and people interested in Mathematics and Computers in Sport) reported in this paper is part of such an attempt.

2. THE SURVEY

The survey will ultimately cover five different groups of systems users. A survey of readers of the QDPI (Queensland Department of Primary Industries) newsletter Circles was mailed in early August 1993. A survey of readers of the Bulletin of ISSS (International Society for the Systems Sciences) was mailed in mid September 1993. A survey of members of the ASSCT (Australian Society of Sugarcane Technologists) was mailed in mid November 1993. A survey of people interested in Mathematics and Computers in Sport was mailed in mid February 1994. A survey of Australian residents who have published in the journal Agricultural Systems is planned for mid 1994. Each of these groups represent individuals with different backgrounds and different interests in systems. This diversity will help identify common features of the useful references and of respondents and will also help highlight the differences. The readers of the Circles newsletter were chosen because Circles focuses on the use of systems methodology in QDPI. The readers of the Bulletin of ISSS were chosen because it focuses on the application of systems theory to issues of science in the service of humanity. Members of the ASSCT were chosen because of the wide coverage of the members and because of the frequency of interaction between biological and physical systems. The people interested in Mathematics and Computers in Sport were chosen because of their need to integrate different types of systems in order to produce recommendations for sports people. The people who have published in the journal Agricultural Systems were chosen because the journal focuses on the behaviour of integrated agricultural systems.

All surveys were basically the same. The survey asked about systems references that the individual had found useful. The following questions were asked:

1. The full reference.
2. The main points made in the reference.
3. What made the reference useful?
4. Where is there a copy?
5. Name and address to send a copy of the results.

3. THE CIRCLES RESULTS

The references that Circles respondents found useful were extremely varied in terms of their systems content, in the breadth of the domain and in the actual knowledge domain covered.

In terms of systems content, they ranged from references with little systems content, but that people "found useful", as for example a book with an excellent compendium of data; to references that provided a systems coverage of a given knowledge domain. However most of these references would not have been retrieved in an on-line or CD literature search with a systems focus. They would only have been retrieved using terms related to the specific knowledge domain.

There were references in action research, community land management, hard systems, organizational development, rapid rural appraisal, and in soft systems methodology. There were 320 Circles readers surveyed, there were 7 responses. This is a response rate of 2.2%. Given the readership and the aim of Circles this low response rate is very disappointing. The references provided by Circles respondents are available on request.

4. THE BULLETIN RESULTS

The responses from readers of the ISSS Bulletin were all long lists of references [7]–[56], rather than just a few provided for in the survey that was sent out. These references related to subjects taught by the respondent or to their area of research. However, but only very little information about why the references were seen to be useful. There were 1000 Bulletin readers surveyed, there were 4 responses. This is a response rate of 0.4%. Given the high public profile of many ISSS members and the high international standing of the ISSS this low response rate is very disappointing.

5. THE ASSCT RESULTS

The references that ASSCT respondents found useful were similar to those from Circles respondents in that they were also extremely varied in terms of the systems content, in the breadth of the domain and in the knowledge domain covered. The main difference with the references from the Circles respondents was that the knowledge domains were different.

In the breadth of the domain covered, they ranged from very narrow to broad. However, most were relatively narrow. In the actual knowledge domain covered, most were relatively specific. There were references in agricultural chemicals, biochemistry, computer modelling in agriculture, environmental assessment and management, plant biology, soil fertility and fertilizer management, sugarcane breeding and genetics, sustainable pasture systems, and in systems analysis and agricultural management. There were 193 ASSCT members surveyed, and there were 17 responses. This is a response rate of 8.8%. The references provided by ASSCT respondents are available on request.

6. THE SPORT RESULTS

The references that people interested in Mathematics and Computers in Sport found useful were similar to those from Circles and ASSCT respondents in that they were extremely varied in terms of the systems content, and in the actual knowledge domain covered, but they were different in that the breadth of the domain covered was very narrow. There were references in teaching and learning, applied operations research, modelling of golf, and in soft systems methodology. There were 50 people interested in Mathematics and Computers in Sport surveyed, but there were 8 responses. This is a response rate of 16%. References provided by Sport respondents are attached as Appendix 1.

The two surprising differences with Sport responses was the number that did not understand what was being asked (4%) and those that had not used systems texts (2%). One positive difference with Sport responses was the considerable detail provided about the main points about the reference and why the respondent considered the reference to be useful. Some of this detail is very useful in that it can be seen as part of the experience that could be made available to the next generation of non-specialist users of systems concepts. Some of this detail is:

- Provided fundamental equations.
- Provided experimental data.
- Has many references to particular problems.
- Provocative text that makes one think about systems issues.
- Techniques for working with unstructured problems and for problem structuring.
- A good summary and a case study for each methodology.
- A refreshingly different approach to problems and their management.
- Useful reference for many systems issues.
- For a process problem I do a literature search; for a content problem I only deal with real problems so I talk to the people who have the problem.
- Provides data necessary for mathematical modelling.

7. AVAILABILITY OF RESULTS

We plan to make the results of this survey and the useful references widely available. At this stage various printed and electronic means are being considered. In the meantime please contact the author on Facsimile (077) 81 4149. To date about 300 references, excluding duplicates, have been received. These references are being validated by using ABN (The Australian Bibliographic Network). There are some references that we have not been able to validate. A sample of those references, from Bulletin respondents, that ABN indicates as having at least one holding in Australia are attached as Appendix 2.

8. DISCUSSION

The analysis of the respondents actual references has been relatively straightforward and we can conclude that there are some similarities and some differences. This is what we would have expected, anyway. However there are two points about the responses, rather than about the actual references, that merit discussion. These are: the low response rate and the difference between system content or knowledge and systems process.

Low response rate. The response rates from all four groups (Circles readers, readers of the Bulletin of ISSS, ASSCT members and people interested in Mathematics and Computers in Sport) were so low that they merit further investigation. Some follow up work with Circles readers has identified some problems.

Circles readers underestimated or misunderstood their importance in the survey. First, some readers did not realize that the aim of the survey was to collect useful systems references. Second, many readers assumed that the survey was directed only at "experts" and as they did not see themselves as experts, they took their experience as irrelevant and thus did not respond. In an attempt to remedy this misunderstanding, a note of clarification was sent to all Circles readers in late November 1993. It indicated that the reality is precisely the opposite; that is, Circles readers were surveyed for four reasons. First, because many readers were not experts but rather that they were non-specialists looking for help from systems. Second, even though many Circles readers have left systems in frustration, many have made significant progress towards using systems to help solve their problem. Third, many of them have done so after using significant effort exploring the literature for "useful" references. Fourth, and very importantly in terms of making useful systems references more accessible, because many Circles readers are only recent users of systems concepts. The response to this note of clarification was lower than the response to the original survey. A telephone follow up of Circles readers who would have been expected to respond, and did not, has been prepared. The selection of readers was done by the editor of the newsletter and was based on their systems experience, and perceived respect within the readers.

The reason for the low response rate of ASSCT members is unlikely to be the same as that of Circles readers. It is more realistic to assume that it is due to a perfectly reasonable lack of interest in the topic. A telephone follow up of ASSCT members, similar to that of the Circles readers, to obtain some validation of this assumption, has been prepared. The selection of members was based on the likely systems content of their occupation.

The reason for the low response rate from readers of the ISSS Bulletin is very difficult to explain. One hypothesis is that the survey had very low priority given the turmoil within the ISSS related to its direction. An individual follow up of ISSS members, similar to that of the Circles readers, to obtain some validation of this hypothesis, has been prepared for the members attending the recent ISSS Conference in Asilomar, California.

The reason for the low response rate of people interested in Mathematics and Computers in Sport is unlikely to be the same as that of Bulletin readers. It is more realistic to assume that it is similar to that of ASSCT members; that is, due to a perfectly reasonable lack of interest in the topic. An individual follow up of Sport people, similar to that of the Circles readers, to obtain some validation of this assumption, has been prepared for the people attending the conference at Bond University.

The difference between system content or knowledge and system process. Whether or not non-specialists are successful in using systems to solve their problem, can depend on many factors. For the purpose of this paper they can be grouped into two: content or knowledge and process. Content or knowledge references are one or more books that describe what is known about a given system. For example, for a sugarcane plant breeder one such book could be: Sugarcane improvement through breeding, edited by Don J. Heinz, Elsevier Amsterdam, 1987.

Process, on the other hand, is more dynamic and experiential. Process is the way in which we go about solving a problem. Process includes the realization that many real world problems do not have clear names or descriptions. Rather that the real world is nearly always a "mess" and that often when we think that we have sorted it out, later on we find out that we have actually only just begun. Process includes accepting that for many problems you cannot find the solution. Rather you have to actually search out, develop, and put together the solution. Process is the way you combine bits of content, of conversations, from all sorts of sources to arrive at your solution.

When we designed this survey we assumed that the references that people had found useful would cover both the system content or knowledge; that is, what is known about a given system and system process; that is, the process by which the person goes about solving a problem in a given system. For example a book with data on agricultural chemicals and the process by which you integrate the data, the person's experience and the required outcome in order to select the appropriate chemical for a given application. When we tested the questionnaire, we had no problems with this assumption. However, most of the respondents' references concentrated on the system content or knowledge. There were very few references on system process. There were, however, several comments penciled in by the respondents, about the importance of process.

Some respondents' comments related to content and process were:

- Stick with the standard references from the specialist literature.
- Process is what you do to solve the problem, you talk with people, seek out or ferret out the solution.
- A librarian provides the process, I just ask the questions.
- The questionnaire is too vague, the term systems needs to be defined.
- Before I can answer your questions I need to know more about what you consider to be a system as opposed to a process.
- There are no references, you just have to solve the problem the best way that you can.
- There are too many references, it depends on what you want (which is the specific knowledge domain of your problem).

- I have not worked with systems.
- I do not know what a system is.
- You just have to find the answer yourself.
- My reference is experience and life (age 74).
- I am baffled by computers, but I am interested.

The significance of the distinction between content and process is related to us being able to achieve our aim of providing useful references to the next generation of non-specialist users of systems concepts in that a new user may be seeking references on system content or knowledge and/or references on system process.

There is a third type of reference, a reference on generic systems processes; that is, references about processes that can be used to develop solutions in many, most or in all systems.

The ultimate aim of this survey is to provide references on generic systems processes, however a less optimistic and more realistic aim is to draw attention to the usefulness of references on generic systems processes and to provide examples of references on system content or knowledge and system process for some systems.

9. Acknowledgements

The author wishes to acknowledge the financial support of the Department of Economics, James Cook University in implementing this research. The co-operation of people who responded to the survey, is gratefully acknowledged.

APPENDIX 1: References from Sport Respondents

These are references [1]–[6] provided by the Sport respondents. Journal references have not been validated. Monograph references have been validated by using ABN (The Australian Bibliographic Network) and are listed as having at least one holding in Australia. There were five more references, that were incomplete and/or could not be validated, that are not included.

[1] Brearley M.N. and Bolt B.A., Dynamics of the lawn bowl, *Quart. J. Mech. and Applied Math*, Vol XI, Pt 3, 351-363 (1958).

[2] Cochran A.J. (Editor), Science and golf, Proc. First World Sci. Congr. Golf, Scotland July 1990, (Chapman and Hall, London, 1990).

[3] de Mestre N.J., The mathematics of projectiles in sport, (Cambridge University Press, Cambridge UK, 1990).

[4] Mehta R.D., Aerodynamics of sports balls, *Ann. R. Fluid Mech.*, 17, 151-189 (1985).

[5] Rosenhead J., Rational analysis in a problematic world: problem structuring methods for complexity, uncertainty and conflict, (Wiley, New York, 1989).

[6] Senge P., The fifth 5th discipline: the art and practice of the learning organization, (Doubleday, New York, 1990).

APPENDIX 2: A SAMPLE OF USEFUL SYSTEMS REFERENCES

This is a sample of the references [7]–[65] received from Bulletin respondents. All these references have been validated by using ABN (The Australian Bibliographic Network) and are listed as having at least one holding in Australia.

[7] Ackoff R.L., The art of problem solving, (Wiley, New York, 1978).

[8] Ackoff R.L. and Emery F.E., On purposeful systems, (Aldine-Atherton, Chicago, 1972).

[9] Allen T.F.H. and Starr T.B., Hierarchy: perspectives for ecological complexity, (University of Chicago Press, Chicago, 1988).

[10] Allen T.F.H. and Hoekstra T.B., Toward a unified ecology, (Columbia University Press, New York, 1992).

[11] Allport G.W., Becoming: basic considerations for a psychology of personality, (Yale University Press, New Haven, 1955).

[12] Ashby W.R., An introduction to cybernetics, (Chapman and Hall, London, 1956).

[13] Beer S., Brain of the firm: a development in management cybernetics, (Allen Lane the Penguin Press, London, 1972).

[14] Beer S., Cybernetics and management, (English University Press, London, 1959).

[15] Beer S., The heart of enterprise, (Wiley, New York, 1979).

[16] Berrien F.K., General and social systems, (Rutgers University Press, New Brunswick, N.J., 1968).

[17] Bertalanffy L. von, Robots, men, and minds: psychology in the modern world, (G. Braziller, New York, 1967).

[18] Blauberg I.V. Sadovsky V.N. and Yudin E.G., Systems theory: philosophical and methodological problems, (Progress, Moscow, 1977).

[19] Burr W.R. et al. (Editors), Contemporary theories about the family, (Free Press, New York, 1979).

[20] Cannon W.B., The wisdom of the body, (W.W. Norton, New York, 1939).

[21] Capra F., The Tao of physics: an exploration of the parallels between modern physics and eastern mysticism, (Wildwood House, London, 1975).

[22] Carter E.A. and McGoldrick M. (Editors), The Family life cycle: a framework for family therapy, (Gardner Press, New York, 1980).

[23] Checkland P., Systems thinking systems practice, (Wiley, New York, 1981).

[24] Checkland P. and Scholes J., Soft systems methodology in action, (Wiley, New York, 1990).

[25] Churchman C.W., The design of inquiring systems: basic concepts of systems and organization, (Basic Books, New York, 1971).

[26] Churchman C.W., The systems approach, (Dell Pub. Co., New York, N.Y., revised and updated edition, 1979).

[27] Churchman C.W., The design of inquiring systems: basic concepts of systems and organization, (Basic Books, New York, 1971).

[28] Clemson B., Cybernetics: a new management tool, (Abacus Press, Tunbridge Wells, Kent, 1991).

[29] Davidson M., Uncommon sense: the life and thought of Ludwig von Bertalanffy father of general systems theory, (J.P. Tarcher, Distributed by Houghton Mifflin Co., Los Angeles, 1983).

[30] Fisher R. and Ury W., Getting to yes: negotiating agreement without giving in, (Houghton Mifflin, Boston, 1981).

[31] Flood R.L. and Carson E.R., Dealing with complexity: an introduction to the theory and application of systems science, (Plenum Press, New York, 2nd edition, 1993).

[32] Hubbard B.M., The evolutionary journey: a personal guide to a positive future, (Evolutionary Press, San Francisco, CA, 1982).

[33] Jantsch E., The self-organizing universe: scientific and human implications of the emerging paradigm of evolution, (Pergamon Press, New York, 1980).

[34] Jantson E. and Waddington C.H. (Editors), Evolution and consciousness: human systems in transition, (Addison-Wesley, Reading, Mass., 1976).

[35] King L.T., Problem solving in a project environment: a consulting process, (Wiley, New York, 1981).

[36] Koberg D. and Bagnall J., The Universal traveller: a companion for those on problem-solving journeys and a soft-systems guide book to the process of design, (Kaufmann, Los Altos, Calif., 1972).

[37] Koestler A., Janus: a summing up, (Random House, New York, 1978).

[38] Lerner D. (Editor), Parts and wholes, (Free Press of Glencoe, New York, 1963).

[39] MacIver R.M., Social causation, (Harper and Row, New York, 1964).

[40] Maltz M., Psycho-cybernetics, (Pocket Books, New York, 1969).

[41] Olsen M.E., The process of social organization: power in social systems, (Holt, Rinehart and Winston, New York, 2nd edition, 1978).

[42] Ouchi W.G., Theory Z: how American business can meet the Japanese challenge, (Addison-Wesley, Reading, Mass., 1981).

[43] Papanek V., Design for the real world: human ecology and social change, (Pantheon Books, New York, 1972).

[44] Pask G., An approach to cybernetics, (Hutchinson, London, 1961).

[45] Rapoport A., General system theory: essential concepts and applications, (Abacus Press, Cambridge, Mass., 1986).

[46] Robertshaw J.E. Mecca S.J. and Rerick M.N., Problem solving a systems approach, (Petrocelli Books, New York, 1978).

[47] Roe P.H.O. Soulis G.N. and Handa V.K., The discipline of design, (Allyn and Bacon, Boston, 1967).

[48] Salk J., Anatomy of reality: merging of intuition and reason, (Praeger, New York, 1985).

[49] Schumacher E.F., Small is beautiful: economics as if people mattered, (Harper Colophon Books, New York, 1975).

[50] Selye H., Stress without distress, (Hodder and Stoughton, London, 1974).

[51] Selye H., The stress of life, (McGraw-Hill, New York, 1956).

[52] Tubbs S.L., A systems approach to small group interaction, (Addison-Wesley Pub. Co., Reading, Mass, 1978).

[53] Van Gigch J.P., Applied general systems theory, (Harper & Row, New York, 2nd edition, 1978).

[54] Vickers G., Freedom in a rocking boat: changing values in an unstable society, (Allen Lane, London, 1970).

[55] Vickers G., Human systems are different, (Harper & Row, London, New York, 1983).

[56] Weinberg G.M., An introduction to general systems thinking, (Wiley, New York, 1975).

[57] Weinberg G.M., Rethinking systems analysis and design, (Dorset House Pub. Co., New York, NY, 1988).

[58] Whitehead A.N., Process and reality: an essay in cosmology, (Cambridge University Press, Cambridge, 1929).

[59] Wiener N., The human use of human beings, (Houghton Mifflin, Boston, 1954).

SPORT SEARCH - A SPORTS COUNSELLING PROGRAM FOR STUDENTS

Deborah Hoare[1]

Abstract

Thinking back to my early high school years I can remember the annual athletics carnival. This was the one time of the year I bothered to run around the oval. Even though I didn't attend little athletics or even run regularly I was fairly confident of a victory in any race over 400 metres. Despite this success no one bothered to tell me that I could possibly be talented at this sport. Needless to say I plugged away at my netball hoping to succeed but never really having any chance to get there. If only Sport Search had been around for me then.

I also remember the kids that no one wanted to have on their team. They were always last to be chosen and their selection was greeted with derision from their team mates. I wonder what has happened to these class mates of mine? Have they by chance found a sport they can participate happily in or have they been turned off sport forever? These students are potentially those that can benefit most from Sport Search. A child with a coordination problem will often not be suited to the traditional sports that schools offer (e.g. team ball games). In many cases they are better suited to individual sports or non-traditional sports (e.g. archery, swimming) but never have the opportunity to try these. Sport Search has the potential to provide these students with a whole range of different sporting options.

1. INTRODUCTION

Sport Search is a counselling package aimed at students in lower high school (11 to 15 years of age). This age group was chosen as recent research has indicated a significant level of drop-out occurring. This is also the time at which students may make decisions about their sporting future and start to specialise in particular sports.

The program developed from a desire to provide counselling for students of all ability levels in their sporting choices. The emphasis of the package is on informed decision-making. The program is not aimed at raising expectations of performance, but will provide students with information that can assist them in selecting sports which they may enjoy and achieve some success in.

Research was undertaken to determine a battery of physical and physiological tasks that were appropriate for developing a profile of a student's performance. As a result of this research the program involves measurement of height, sitting height, weight, arm span, hand-eye coordination, upper body strength, vertical jump, agility, 40 metre sprint and aerobic fitness (shuttle run). These tests require minimal equipment and are simple to administer in the school setting. A manual and video are provided with the package to facilitate this.

---

[1] Australian Sports Commission, Belconnen ACT 2616

The students enter these results into a software program. They then identify their preferences for different types of sport (e.g., team or individual, contact or non-contact). Based on the student's physical profile and their response to these categories they are provided with:

- A profile of their performance on the physical tasks.
- A list of sports they appear most suited to.
- A description of any sport they are unfamiliar with.
- Information regarding how to access particular sports in their community.

It is vital that the teacher assumes a role in guiding and counselling students in making use of the information provided in Sport Search.

2. THE SOFTWARE

The challenge when developing the computer software for Sport Search was to create a package which was easy to use and yet gave the student access to all that they might need to know.

The early prototypes of the package were written in Clarion for DOS-based computers. This language is text-based, and the interface was therefore unlike the graphic environment which would finally be expected for the Macintosh and Windows.

It was agreed with the team who were to develop the final software, that Hypercard for the Macintosh should become the prototyping tool.

Working on the software for the project, it was relatively straightforward to modify each version of the program as the design view clarified. The basic calculations within the heart of the package relied on a matching of norm tables against the student's own results. These tables were prepared by sports scientists using statistical tools and exported as text files. A new file, with new norms could be imported quickly. Similarly, the file of descriptions of the sports was exported from a word-processor each time it had been updated, and simply imported to the Hypercard "stack".

A key element in such a package is visual design. It was always intended that the package would be marketed to schools as part of the AUSSIE SPORT program. This meant that a certain look and feel would be essential in the final version. However, the graphic design and video elements of the package were not confirmed until relatively late in the overall design process. Again, the flexibility of Hypercard allowed us to work with our first clumsy stick figure drawings and simple icons, confident that these could easily be replaced once the graphic designers came on stream.

Because the package relates to such physical concepts and because the target age group is youthful, it was always felt that visual elements implying movement and activity would be important. Early discussions included the possibility of using some form of video or animation in the software. The main difficulties in this area were always going to be the broad range of computers available in schools and the need to reach the maximum number of users. For this reason, it was agreed that the Macintosh version needed to be in black and white and capable of running on a Mac Classic. The decision to make a Windows version for the PC was only taken after deciding that the visual elements were too important to sacrifice. A DOS version was not going to offer these elements.

One of the major concerns from the early stages of the project had also been user response-time. Especially dealing with younger students, it was important not to keep them waiting. Coding the calculations for speed presented minor challenges, in so far as some sports had been negatively weighted on certain characteristics (for instance, being in the upper percentiles for height could be a disadvantage to a weight lifter).

There was also the further complication that in some states certain sports are not available, so it would be silly to list them (for instance, snow skiing in the Northern Territory). These exceptions made the working out and matching of scores for students a little less straightforward, and entailed some extra calculations. Early experiments with displaying a "progress bar" during calculations helped to diminish the sense of waiting, but it was obvious that even so, more than a few seconds was an unacceptable wait. The solution was to use external commands for Hypercard. In this case, it was sufficient to create them using a simple compiler which is available as a resource for Hypercard scripters. In their compiled form, the Hypercard scripts now took less than half the previous time.

The final and most exciting phase was the work with the graphic designers. The challenge with such artistic talent at our disposal was to combine function with purpose. The animations and design had a purpose to serve: helping the user to navigate through information. They should not distract from or intrude on the student's travel through the program. The results were all that we had hoped for, and after testing and refinement with trial groups of students, the graphic interface was complete. The final task was to complete a parallel version of the stack which would run under Windows.

3. CONCLUSION

By participating in Sport Search it is hoped that students may become involved with sports they are better suited to and are therefore less likely to drop out of. Those students not participating in sport because of a perceived lack of skill may have the opportunity to take up a sport which better suits their abilities.

Sport Search has the potential to provide all students, no matter what their size, shape or physical prowess, with more appropriate sporting choices, and hopefully a more positive and enjoyable sporting experience.

DRIVING A BALL WITH A GOLF CLUB

Maurice N. Brearley[1]

Abstract

An investigation is made to determine if the muscular effort exerted by a golfer during the brief period of contact between club and ball makes a significant contribution to the length of a golf drive. The analysis makes use of information obtained from multi-flash photography of a drive by the American golfer Robert Jones. It is found that the length contribution is negligible, partly because of the extreme brevity of the contact between club and ball, and partly because of the relatively small force exerted by the golfer towards the bottom of the downswing of the club. A similar conclusion applies to many other 'hard ball' sports such as cricket and baseball.

1. INTRODUCTION

The object of this note is to determine whether or not the force exerted by a player makes a significant contribution to ball speed once contact between ball and club begins. The discussion will be conducted in terms of the driving of a golf ball, but will apply to most other 'hard ball' sports such as cricket and baseball.

To obtain a quantitative answer, data connected with a drive by the American golfer Robert Jones will be used. This is made possible by analysis of a multi-flash photograph of the downswing, which was contained in a published article (Williams [1]).

The velocity of the club head just before contact with the ball is also known from high-speed photography of a Jones drive (Edgerton and Killian [2]). The velocity imparted to the ball can then be found by using Conservation of Momentum and Newton's Law of Impact for the club head and the ball.

The duration of the contact between club and ball is also known from Reference [2], and this enables a calculation to be made of the extra velocity imparted to the club head during the contact. It is then an easy matter to convert this change to an increase in the distance that would be travelled by the driven ball.

The actual carry of the drive (ignoring air resistance) can also be found if values are allotted to the masses of the ball and club head and to the coefficient of restitution between ball and club head. Because these quantities are not known precisely for the particular case considered, estimates of them are used to calculate the approximate carry of the photographed drive. This vagueness will later be seen not to affect the

---

1 85 Dandarriga Drive, Clifton Springs, Victoria, 3222

validity of the conclusion concerning the increase in the distance produced during contact between club and ball.

2. THE VELOCITY AND TANGENTIAL ACCELERATION OF THE CLUB HEAD

Figures 1 and 2 show the photograph and stick diagram of the golfer's arms and club in Reference 1.

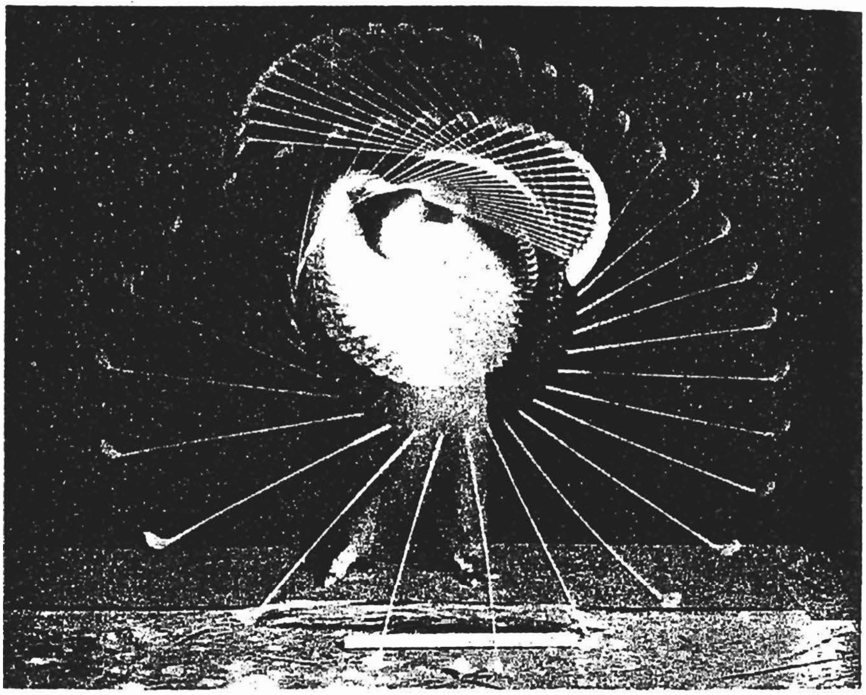


Figure 1: Downswing of Robert Jones' drive (multi-flash photograph at 1/100 second intervals; reprinted from Reference 1.)

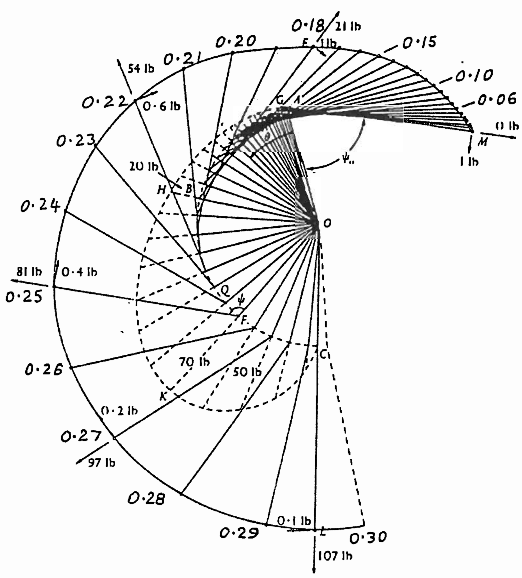


Figure 2: Stick diagram of arms and golf club corresponding to Figure 1 (1/100 second intervals; reprinted from Reference 1.)

In Reference 1, Figure 2 was made to match Figure 1 by postulating suitable kinematical relationships between the shaft of the golf club and the golfer's body. Figure 2 will now be used to determine the velocity, and hence the tangential acceleration, of the head of the golf club. To facilitate this, arcs have been drawn to link successive positions of the club head, and times in seconds from the start have been added.

The following notation will be used:

t = time from start of downswing,
s = arc length in mm of club head travel in Figure 2,
v = velocity of club head at time t,
V = value of v when club meets ball (at t = 0.295s),
= 50.3 ms^{-1} (Reference [2]).

Table 1 shows corresponding values of t and the arc lengths s measured on Figure 2.

Table 1

t (sec)	s (mm)	t (sec)	s (mm)	t (sec)	s (mm)
0	0	0.11	15	0.22	82
0.01	1	0.12	17	0.23	95
0.02	2	0.13	20	0.24	110.5
0.03	3	0.14	23	0.25	127
0.04	4	0.15	26.5	0.26	145
0.05	5	0.16	30.5	0.27	163
0.06	6	0.17	35	0.28	182
0.07	7.5	0.18	41	0.29	202
0.08	9	0.19	49	0.295	213
0.09	11	0.20	58.5		
0.10	13	0.21	69.5		

The data of Table 1 were plotted in Figure 3, and a smooth curve was drawn through them to produce the result designated in Figure 3 as Arc length.

To obtain the graph of club head velocity v versus time t during the downswing, the angles θ(s) of the slopes of the arc length curve were measured at various values of t. The results are shown in the first two columns in Table 2. To obtain from these the velocity v, use was made of the fact that at the final time t = 0.295 sec, where θ(s) = 63.7°,

$$v = V = 50.3\ \text{ms}^{-1}.$$

The velocities at other times were then found by using

$$v = 50.3 \tan \theta(s)/\tan 63.7^\circ\ \text{ms}^{-1}.$$

The results are shown in the third column of Table 2, using also the fact that v = 0 at the start of the downswing. The corresponding points were plotted in Figure 3, and the velocity curve was drawn to match them.

Table 2

t (sec)	θ(s) (deg)	v (ms^{-1})	θ(v) (deg)	f (ms^{-2})	t (sec)	θ(s) (deg)	v (ms^{-1})	θ(v) (deg)	f (ms^{-1})
0	0	0	14.0	62.5	0.19	39.2	20.3	58.3	405
0.02	5.7	2.5	14.0	62.5	0.195	42.1	22.5	60.5	442
0.04	5.7	2.5	11.7	51.8	0.20	45.5	25.3	61.5	460
0.06	7.6	3.3	8.8	38.7	0.205	47.3	26.9	60.3	438
0.08	9.2	4.0	7.4	32.5	0.21	49.6	29.2	59.4	423
0.10	11.0	4.8	11.7	51.8	0.22	52.5	32.4	56.0	371
0.12	14.0	6.2	16.8	75.5	0.24	58.0	39.8	49.7	295
0.14	17.7	7.9	26.9	127	0.26	61.1	45.0	41.6	222
0.16	23.0	10.6	39.7	208	0.28	63.0	48.8	30.0	144
0.18	35.6	17.8	52.5	326	0.295	63.7	50.3	18.8	85.1

The tangential acceleration f of the club head was found from measured values θ(v) of the angles of slope of the velocity curve. These values are shown in the fourth column of Table 2, and in the last column are the corresponding values of the acceleration given by the relation

$$f = 250 \tan \theta(v)\ \text{ms}^{-2}.$$

The coefficient 250 is the ratio of the scales of the vertical and horizontal axes of the velocity curve in Figure 3.

The calculated values of f were plotted in Figure 3, and the acceleration curve was then drawn through them.

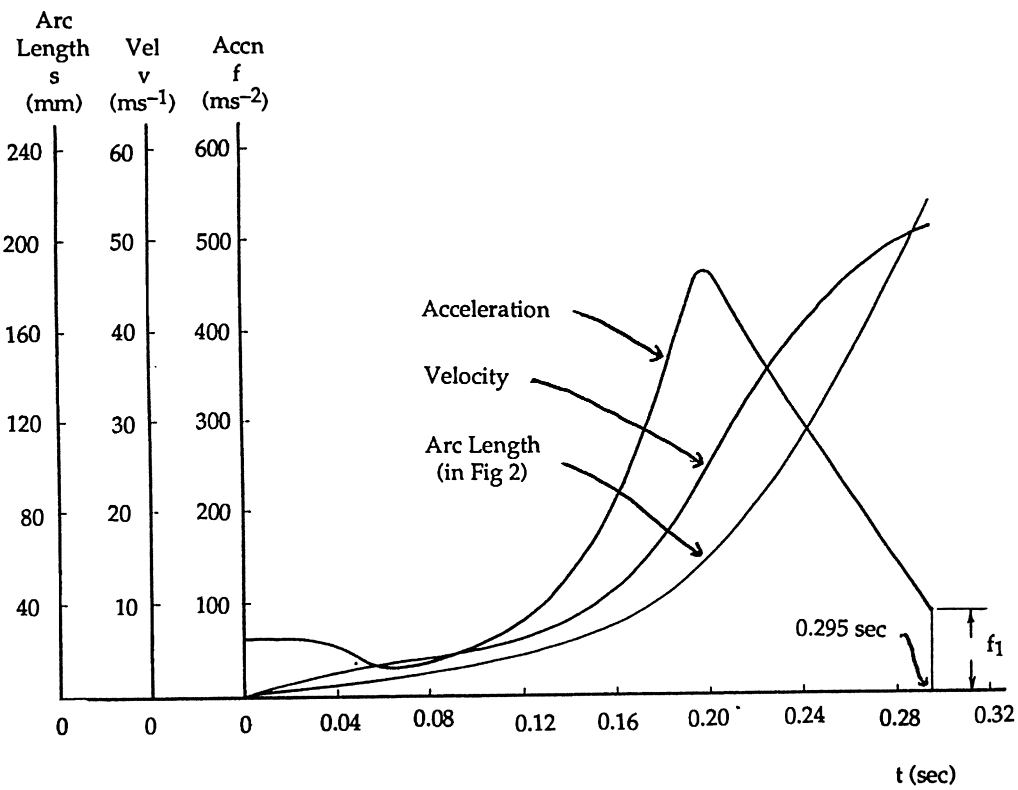


Figure 3: Club head arc length (in Figure 2), velocity and acceleration.

As a check on the accuracy of the acceleration curve in Figure 3, it may be used to calculate the velocity of the club head on impact with the ball (which is known Edgerton and Killian [2] to be $V = 50.3\ ms^{-1}$). By integration of the relation $dv/dt = f$ over the time interval (0, 0.295) of the downswing one obtains

$$V = \int_0^{0.295} f\, dt,$$

and the definite integral may be evaluated as the area under the $f - t$ curve in Figure 3. By counting squares on the graph paper on which the figure was originally drawn, and making allowance for the scales of the axes, the value of the integral was found to be $49.9\ ms^{-1}$, in excellent agreement with the known value $V = 50.3\ ms^{-1}$.

3. Distance Travelled by the Ball

The following notation will be used:

m = mass of golf ball,
M = mass of head of club,
u = velocity imparted to the ball,
e = coefficient of restitution between ball and club,
V_1 = velocity of club head after contact with the ball.

Conservation of momentum during contact gives

$$MV = MV_1 + mu,$$

and Newton's law of impact states that

$$u - V_1 = eV.$$

Eliminating V_1 between these two equations gives

$$u = (1 + e)\, MV/(M + m). \tag{1}$$

The distance travelled by the driven golf ball before striking the ground may be calculated from the formula

$$R = \left(u^2/g\right) \sin 2\alpha, \tag{2}$$

where α is the initial angle of elevation of the path of the ball (Bullen [3]). This formula ignores the effect of back spin as well as air resistance, but it will be seen later that this does not affect the conclusions reached.

Substituting for u from (1) into (2) gives

$$R = kV^2, \tag{3}$$

where

$$k = \left[(1+e)\ M/(M+m)\right]^2 (\sin 2\alpha)/g. \tag{4}$$

From Figure 1 it can be found that $\alpha \approx 12^\circ$, and experiments reveal that $e \approx 0.67$ in a hard hit drive (Cochran and Stobbs [4]). A typical golf ball has mass $m = 0.043$ kg. The mass of the head of the club used by Jones is not known, but it may be taken to have the mid-range value $M = 7$ oz $= 0.198$ kg. With all these values, equation (4) gives

$$k = 0.0781\ s^2m^{-1},$$

and equation (3) then yields

$$R = 0.0781 \times (50.3)^2 = 198 \text{ metres}.$$

This is a plausible value for the carry of a drive in the absence of air resistance, indicating that the process used to derive it was valid.

4. Contribution to Distance During Contact

The differential relation corresponding to equation (3) is

$$\delta R = 2kV\delta V, \tag{5}$$

and this can be used to find the contribution δR to the carry of the drive which is due to the extra velocity δV imparted to the ball during the brief period of time δt for which the club head is in contact with the ball.

From Table 2 and the acceleration graph in Figure 3 it can be seen that at time $t = 0.295$ sec, when the club head first contacts the ball, its acceleration is

$$f_1 = 85.1 \text{ ms}^{-2}.$$

From the shape of the graph at this time it is clear that the acceleration is decreasing, and so the extra velocity δV acquired by the club head during the time δt that it is in contact with the ball will be such that

$$\delta V < f_1 \, \delta t. \qquad (6)$$

High speed photography of the drive by Jones reveals [2] that

$$\delta t = 5 \times 10^{-4} \text{ s}.$$

Substitution from (6) in (5) shows that

$$\delta R < 2k \, V f_1 \, \delta t,$$

and on inserting the numerical values of k, V, f_1 and δt this becomes

$$\begin{aligned} \delta R &< 2 \times 0.0781 \times 50.3 \times 85.1 \times 5 \times 10^{-4} \text{ m}, \\ &< 0.33 \text{ m}. \end{aligned}$$

So the contribution to the length of a drive of carry 198 metres resulting from muscular action during contact between club and ball is less than 0.33 metre.

5. CONCLUSION

The analysis shows conclusively that the force exerted by the golfer on the club during the actual time of contact between the head of the club and the ball makes a negligibly small contribution to the length of the drive. Two factors prevent any significant input from the hands: the extreme brevity of the contact between club and ball, and the very small value of the tangential acceleration of the club head at the bottom of the swing.

In other 'hard ball' sports such as cricket, baseball, etc, the contact time between bat and ball will be longer than it is in golf, but not by enough to alter the conclusion that the effect of the hands is negligible during actual contact.

Consideration of the effect of the hands during contact with the ball in this paper does not include the part they play in making the stroke a good one. It is assumed that the stroke and follow-through are properly executed by the player.

ACKNOWLEDGMENTS

The writer is grateful to Mr B.M. Cameron for drawing his attention to the matter discussed in this paper, and to Associate Professor Neville de Mestre and Mr John Baker for helpful criticisms and comments on a previous version of it.

REFERENCES

[1] Williams D., The dynamics of the golf swing, *Quart. J. Mech. and Applied Math*, Vol XX, Pt 2, 247–264 (1967).

[2] Edgerton H.E. and Killian J.R., Flash: See the unseen by ultra-high-speed photography, (Branford, Boston, 1954).

[3] Bullen K.E., An introduction to the theory of mechanics, (Science Press, Sydney, 1971).

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

DIRECT GOLF PUTTING DYNAMICS AND STRATEGIES

John Scott[1] and Neville de Mestre[2]

Abstract

A mathematical model is developed for a golf ball rolling across a flat putting green. Experiments are performed to check the model. The model is then used to predict the range of initial putting speeds that propel a directly-aligned-and-hit putt into the hole.

1. INTRODUCTION

What is a good putting tactic in the game of golf? Is it to hit the ball with weight, aiming perhaps at a point past the hole? At the other extreme, is it to let the ball 'die' as it reaches the hole, effectively extending the size of the hole? This paper begins to explore these issues.

When a golf ball is putted on a flat green, and reaches the hole, its behaviour can be categorised into a number of separate actions. In chronological order they are:

(i) the ball drags on the green initially for a short interval of time until pure rolling commences,

(ii) the ball then rolls purely up to the hole,

(iii) the ball crosses the front (or near) portion of the hole's rim and, if it doesn't do this directly (ie. in the radial direction), its path may deviate,

(iv) the ball either rolls around the rim and drops in, rolls around the rim and goes out, or travels as a projectile across the top of the hole,

(v) the ball drops into the cup, or hits the back of the hole below the rim,

(vi) alternatively to (v), the balls strikes the back (or far) portion of the rim, pops up, and either falls into the hole or carries on past the hole on the far side of the green.

This simplified approach does not take account of the fact that during a normal golf day many players move around near the hole. Their combined actions tend to push the soil under the green into a little mound, bordering and just beyond the hole rim. Therefore, many putts are uphill near the hole, even on a supposedly flat green.

---

[1] Department of Mangement Systems, University of Waikato, New Zealand.
[2] School of Information Technology, Bond University, Gold Coast, Queensland, 4229

There are further complications if the hole is positioned on a sloping part of the green. An analysis of this scenario will be postponed until later; the mathematical analysis and resulting strategies for a flat green will be investigated first of all. However the authors are aware that the percentage of greens with slope near the hole is significant.

A further complication is the effect of wind on a golf putt. Fortunately the boundary layer characteristics of wind over the ground ensure that, except on days of extremely strong wind, the effect on a putted golf ball is small. Hence the model will ignore wind effects at this stage also.

Finally, the analysis in this paper will be developed for directly putted balls which travel in a line tht passes diretly over the centre of the hole. Obliquely putted balls will be considered in a later paper.

This paper will therefore explore a model to predict the behaviour of a golf ball putted on a flat green along a line which passes directly over the centre of the hole, when there is no wind effect on the putt.

The important geometrical factors in this problem are that the ball is considered to be a sphere with minimum diameter 4.267cm, and the hole is considered to be a circular cylinder of diameter 10.80cm and depth at least 10.16cm.

2. THE SIGNIFICANCE OF AIR DRAG

The speed of a putted golf ball depends on how hard the ball is hit, and this is determined by the distance from the hole that the putt is taken. For long putts an accepted strategy is to aim for a pseudo-hole with the same centre as the green hole but with a radius of approximately 1 metre. This should leave a final putt (hopefully) of less than 1 metre. For this final putt the ball should travel at speeds of the order of 2 to 4 ms^{-1} depending on the surface properties of the green.

If air drag was to be included, then one-dimensional motion is considered for a golf ball falling vertically through air. The governing equation is

$$m\ddot{y} = mg - k\dot{y}^2$$

where m denotes the mass of the golf ball (maximum 0.04593kg), y is the position coordinate measured vertically downwards, g denotes the acceleration due to gravity (9.81ms^{-2}), k denotes the air resistance coefficient, and a dot signifies differentiation with respect to the time (t).

From dimensional analysis and experiments in wind tunnels, the air drag on a golf ball is usually taken as $k\dot{y}^2$ where $k = \frac{1}{2}\rho A C_D$ Here ρ is the density of air (1.226kg m^{-3}), A is the area of cross-section of the ball (diameter 0.0427m), and C_D is taken as 0.45 since the fluid motion will not have reached the critical Reynolds' number at these putting speeds. Hence

$$k\dot{y}^2 \approx 0.000395\dot{y}^2,$$
$$mg \approx 0.49,$$

and so air drag will be less than 2% of the effect of weight for putting speeds less than 4ms⁻¹.

A series of experiments was carried out to confirm this. A golf ball was released from rest, next to a vertical 1-metre stick, and a video camera recorded its fall.

If weight was the only force acting then, from rest,

$$y = \tfrac{1}{2}gt^2$$

or alternatively

$$t = \sqrt{\frac{2y}{g}}.$$

Thus

$$t - t_0 = \sqrt{\frac{2y}{g}} - \sqrt{\frac{2y_0}{g}}$$

or

$$\sqrt{\frac{g}{2}}(t - t_0) = \sqrt{y} - \sqrt{y_0}$$

where y_0 is the position of the ball at some know time t_0. Typical experimental results are shown in the first two columns of Table 1, which gives the positions of the falling ball at each frame. Since the video image was recorded at 25 frames per second, there is a time interval 0.04 seconds between each frame. The ball was released at the 80cm mark on the stick to produce speeds around $4ms^{-1}$ near the bottom of its fall.

Table 1

Frame	Height (cm)	y(m)	$\sqrt{\frac{g}{2}}(t - t_0)$	$\sqrt{y} - \sqrt{y_0}$
5	65	0.15	0	0
6	57	0.23	0.089	0.092
7	48	0.32	0.177	0.178
8	37	0.43	0.266	0.268
9	25	0.55	0.354	0.354
10	11	0.69	0.443	0.443

If t_0 is chosen as the time at frame 5 to eliminate errors in the inaccuracies of release time, the final two columns give the calculated and measured values for a gravity-only assumption. Comparison of these columns indicates quite good agreement between the gravity-only (or drag-free) model and the measured values. Hence the assumption of no air drag for a 1-metre golf putt is confirmed, since the speeds attained in a 1-metre putt are of the same order as the speeds attained by the falling ball.

3. THE DYNAMICS OF A GOLF PUTT

Consider a ball of radius r struck by the putter and given an initial translational velocity $\dot{x}_0$ towards the hole with an initial rotational speed $\dot{\phi}_0$ about a horizontal axis perpendicular to the putting direction.

The ball will be considered as a solid sphere (ie. the dimples will be ignored in the geometrical model) and the putting surface will be considered as a horizontal plane with a constant coefficient of friction μ and a friction couple L on the golf ball, because of the grass being pushed down and partly springing back. In reality the frictional characteristics may be non-homogeneous because of the different directions that the green is mown. Of course there will also almost certainly be indentations on the green from many golfers walking on it, or from balls being pitched onto the green during numerous approach shots. Nevertheless, the chosen μ and L can be thought of as averages for the green for the particular dampness and cut of the grass at the time of putting.

A simplified model is considered in which the putt commences with its centre of mass G at the origin of the co-ordinate system, which is a horizontal distance C from the front edge of the hole. Gravity, the ground reaction and the friction couple are considered to be the only forces and moments acting after the putter strikes the ball.

When struck initially by the putter, the ball will slip on the surface if $\dot{x}_0 < r\dot{\phi}_0$, the ball will skid on the surface if $\dot{x}_0 > r\dot{\phi}_0$, or the ball will roll purely if $\dot{x}_0 = r\dot{\phi}_0$. Although the most likely situation is one of the first two, this will only occur for a very short interval until the ball adjusts its linear and angular speeds to rolling purely. Therefore the model to be considered in this paper is one in which the ball rolls purely immediately after being struck by the putter.

In this situation the frictional force may not necessarily be that of limiting fraction (μN), where N is the normal component of the ground reaction, but will be denoted simply by F where $0 \leq F \leq \mu N$.

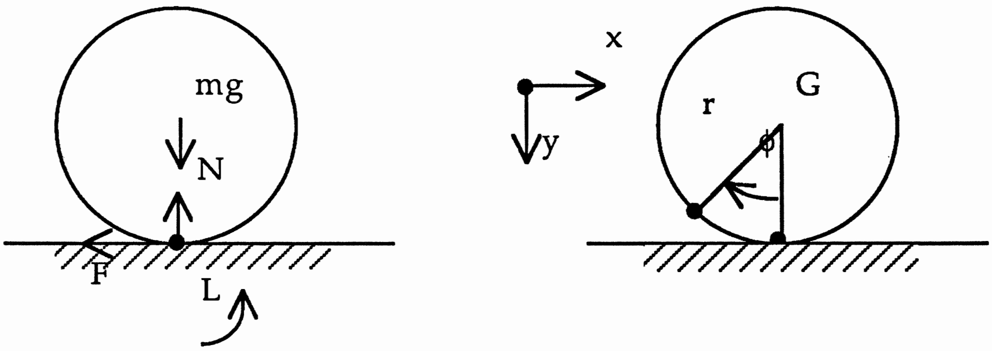


Figure 1

The co-ordinate system has the centre of mass G at the origin at time t = 0, 0x in the direction of initial motion, 0y vertically down, and ϕ is the angle turned through from the initial position (Figure 1). Application of Newton's laws of motion (Linear Momentum and Angular Momentum Principles) yield

$$\text{(x direction)} \qquad m\ddot{x} = -F \qquad (1)$$

$$\text{(y direction)} \qquad 0 = mg - N \qquad (2)$$

$$\text{(moments about G)} \qquad \frac{2}{5} mr^2 \ddot{\phi} = Fr - L \qquad (3)$$

where $\ddot{x}$ and $\ddot{\phi}$ denote the linear and angular accelerations respectively at time t.

Since the ball is assumed to roll purely, the kinematical condition

$$x = r\phi$$

holds. From this equation

$$\ddot{x} = r\ddot{\phi}$$

Hence, elimination of $\ddot{\phi}$ from equations (1) and (3) yields

$$L = \frac{7rF}{5} \qquad (4)$$

Thus, if the friction couple is ignored (L = 0), then the frictional component must also be zero. This means that $\ddot{x} = 0$, and hence $\dot{x} = \dot{x}_0$ for all t. That is, the golf ball, once putted, would roll purely forever, a property not observed experimentally or otherwise. Now air drag could be added to the model to slow the ball down, however it is clearly not the air which causes the ball to slow down but the grass on the green. The frictional couple and the frictional force must therefore both be included as an attempt to model the effect of the grass on the ball. Since the frictional force component should be zero when $\dot{\phi} = 0$, and the frictional force component should also be zero when $\dot{x} = 0$, a model for the friction component is chosen as

$$F = K\dot{\phi}^{\lambda} = Kr^{-\lambda}\dot{x}^{\lambda} \tag{5}$$

where K and λ are positive constants. By equation (4) the frictional couple L is now also modelled.

With this assumption, equation (1) yields

$$mr\,\ddot{\phi} = -K\dot{\phi}^{\lambda} \tag{6}$$

Integration of equation (6) with respect to t yields

$$\dot{\phi}^{1-\lambda} = \dot{\phi}_0^{1-\lambda} - \frac{(1-\lambda)K}{mr}\,t \tag{7}$$

while integration of equation (6) with respect to ϕ yields

$$\dot{\phi}^{2-\lambda} = \dot{\phi}_0^{2-\lambda} - \frac{(2-\lambda)K}{mr}\,\phi \tag{8}$$

Now if $\lambda < 1$ the ball comes to rest in a finite time t_F and at a finite distance x_F when $\dot{x} = r\dot{\phi} = 0$. Thus from equations (7) and (8)

$$t_F = \frac{mr\,\dot{\phi}_0^{1-\lambda}}{(1-\lambda)K} = \frac{mr^{\lambda}\,\dot{x}_0^{1-\lambda}}{(1-\lambda)K} \tag{9}$$

and

$$\begin{aligned} x_F &= r\phi_F \\ &= \frac{mr^2\,\dot{\phi}_0^{2-\lambda}}{(2-\lambda)K} = \frac{mr^{\lambda}\,\dot{x}_0^{2-\lambda}}{(2-\lambda)K} \end{aligned} \tag{10}$$

Elimination of K between equations (9) and (10) produces

$$\frac{x_F}{\dot{x}_0\,t_F} = \frac{1-\lambda}{2-\lambda},$$

and so

$$\lambda = \frac{\dot{x}_0\,t_F - 2x_F}{\dot{x}_0\,t_F - x_F} \tag{11}$$

This is clearly less than 1 and greater then 0, so the conditions of the model are satisfied. From equation (9) the value of K can be determined from

$$K = \frac{mr^{\lambda}\,\dot{x}_0^{1-\lambda}}{(1-\lambda)t_F} \tag{12}$$

once the value of λ has been established.

The two parameters λ and K are characteristics for a particular green on a particular day. They present a quantitative measure of the resistance and springiness of the grass to the rolling motion of the golf ball, which is governed by the cut and state of dryness of the green.

When $\dot{\phi}$ is eliminated between equations (7) and (8) the resulting expression is

$$\left[\dot{\phi}_0^{1-\lambda} - \frac{(1-\lambda)K}{mr}t\right]^{\frac{1}{1-\lambda}} = \left[\dot{\phi}_0^{2-\lambda} - \frac{(2-\lambda)K}{mr}\phi\right]^{\frac{1}{2-\lambda}}$$

or, since $\dot{x}_0 = r\dot{\phi}_0$,

$$\left[\frac{\dot{x}_0^{1-\lambda}}{r^{1-\lambda}} - \frac{(1-\lambda)K}{mr}t\right]^{\frac{1}{1-\lambda}} = \left[\frac{\dot{x}_0^{2-\lambda}}{r^{2-\lambda}} - \frac{(2-\lambda)K}{mr^2}x\right]^{\frac{1}{2-\lambda}}$$

Thus

$$x = \frac{mr^2}{(2-\lambda)K}\left[\frac{\dot{x}_0^{2-\lambda}}{r^{2-\lambda}} - \left\{\frac{\dot{x}_0^{1-\lambda}}{r^{1-\lambda}} - \frac{(1-\lambda)K}{mr}t\right\}^{\frac{2-\lambda}{1-\lambda}}\right]$$

$$= \frac{mr^2}{(2-\lambda)K}\left[\frac{\dot{x}_0^{2-\lambda}}{r^{2-\lambda}} - \frac{\dot{x}_0^{2-\lambda}}{r^{2-\lambda}}\left\{1 - \frac{(1-\lambda)K}{mr^\lambda\dot{x}_0^{1-\lambda}}t\right\}^{\frac{2-\lambda}{1-\lambda}}\right]$$

$$= \frac{mr^\lambda\dot{x}_0^{2-\lambda}}{(2-\lambda)K}\left[1 - \left\{1 - \frac{(1-\lambda)K}{mr^\lambda\dot{x}_0^{1-\lambda}}t\right\}^{\frac{2-\lambda}{1-\lambda}}\right]$$

From equations (9) and (10) this can be simplified to produce the expression

$$x = x_F\left[1 - \left\{1 - \frac{t}{t_F}\right\}^{(\dot{x}_0 t_F/x_F)}\right] \tag{13}$$

Experiments were carried out to measure $\dot{x}_0$, x_F and t_F. In addition, x and t were measured at short time intervals during the motion of the ball.

In the next section the experimental values are compared with the corresponding theoretical values given by equation (13).

4. EXPERIMENTS

Experiments were conducted on the chipping green at Robina Woods golf course using a video camera.

A number of putts were made using a golf club on two days separated by a six-month period. Each putt was performed on a level section of the green, and the ball's motion was recorded next to a horizontal 1-metre rule. Thus its position could be determined every time frame (0.04 sec) of the film.

Table 2 shows the time (t) in seconds, distance measured (x_m) in metres, and distance calculated (x_c) in metres using equation (13), for a typical experiment conducted in August 1993 and a typical experiment in February 1994.

Table 2

	August 1993		February 1994	
t (s)	x_m (m)	x_c (m)	x_m (m)	x_c (m)
0	0	0	0	0
0.2	0.21	0.23	0.23	0.27
0.4	0.37	0.42	0.43	0.50
0.6	0.50	0.57	0.58	0.67
0.8	0.61	0.68	0.71	0.80
1.0	0.69	0.76	0.82	0.88
1.2	0.77	0.81	0.89	0.93
1.4	0.82	0.83	0.94	0.94
1.52	0.83	0.83		

For the August 1993 experiment $t_F = 1.52$ secs, $x_F = 0.83$ metres, $\dot{x}_0 = 1.25\ \text{ms}^{-1}$ yielding $\lambda = 0.29$, $K = 0.016$ and

$$x_c = 0.83\left[1 - \left\{1 - \frac{t}{1.52}\right\}^{2.29}\right],$$

while for the February 1994 experiments $t_F = 1.40$ secs, $x_F = 0.94$ metres, $\dot{x}_0 = 1.5\ \text{ms}^{-1}$ yielding $\lambda = 0.19$, $K = 0.027$ and

$$x_c = 0.94\left[1 - \left\{1 - \frac{t}{1.40}\right\}^{2.23}\right]$$

The model appears to over-predict the intermediate distances travelled. However, when the motion of the ball is viewed frame-by-frame, it is seen that the ball does not roll purely on the green but skips along from bump to bump in very low trajectories. Every time it hits the ground it will loose energy because of the impact. A mathematical analysis of this behaviour would be extremely tedious, and would require detailed knowledge of the profile of the green in the direction of the path of the ball as well as knowledge of the coefficient of restitution at each point of impact.

There is one basic problem with the model so far, and that is that λ and K change as the initial putting speed $\dot{x}_0$ changes. This is unacceptable on logical grounds as one would naturally expect λ and K to depend on properties associated only with the green.

Notwithstanding all this, the current model based on equation (5) gives a reasonable approximation to the actual position of the ball, and can be used as a simple basis for further mathematical analysis.

5. Direct Putting Across the Hole

When the ball reaches the near part of the rim of the cup, that is, the semicircular segment closest to where the ball was struck, it will pass directly over the rim if it is travelling along an extended radius of the hole, or it will cling to the rim for a short time if it meets the rim of the cup obliquely. This paper will consider direct putting only.

From the model, the speed at any time during the putt is given by equations (8), (11) and (12) as

$$\dot{x} = \dot{x}_0 \left\{ 1 - \frac{x}{x_F} \right\}^{1/(2-\lambda)},$$

and so the linear and rotational speeds at the near edge of the hole are

$$\dot{x}_N = \dot{x}_0 \left\{ 1 - \frac{C}{x_F} \right\}^{1/(2-\lambda)}, \quad \dot{\phi}_N = \frac{\dot{x}_N}{r} \tag{14}$$

since C is the horizontal distance of the centre of the ball from the front of the cup at the beginning of the putt. Now $\dot{x}_N$ must be greater than zero at the front edge of the cup, or in other words $x_F > C$, otherwise the ball will not reach the cup. This places a lower limit on the initial putting speed $\dot{x}_0$.

After the ball passes over the near edge of the hole it is then behaving like a projectile with initial speed $\dot{x}_N$ and launching angle 0°. This projectile can be analysed using a gravitational force only, since drag is discounted as relatively unimportant at these speeds by the experiment in Section 2, and lift is also negligible because the rotation of the ball is much less than is required to produce a significant Magnus effect. Thus the equations of motion are

$$m\ddot{x} = 0$$
$$m\ddot{y} = mg$$
$$\frac{2}{5} mr^2 \ddot{\phi} = 0$$

yielding solutions

$$x = \dot{x}_N \tau + C \tag{15}$$

$$y = \frac{1}{2} g\tau^2$$
$$\dot{\phi} = \dot{\phi}_N \tag{16}$$

where τ is measured from the moment the ball leaves the near edge of the hole.

It will now become a successful putt if

(i) it doesn't reach the back portion of the hole

or (ii) it strikes the back portion of the hole below the rim

or (iii) it strikes the back edge of the hole and bounces back into the hole ($\dot{x} < 0$).

Unsuccessful putts will be those that strike the back edge of the hole and continue in the original direction of motion ($\dot{x} > 0$).

For successful putts the initial putting speed $\dot{x}_0$ is clearly smaller for cases (i) and (ii) than for case (iii). The significant value of $\dot{x}_0$ to be determined is the value which causes the ball to hit the back edge of the rim and rise vertically with no forward or backward speed. This will then determine the upper limit on the initial putting speed $\dot{x}_0$ for a successful putt; that is, when $\dot{x} = 0$ at the back rim after collision.

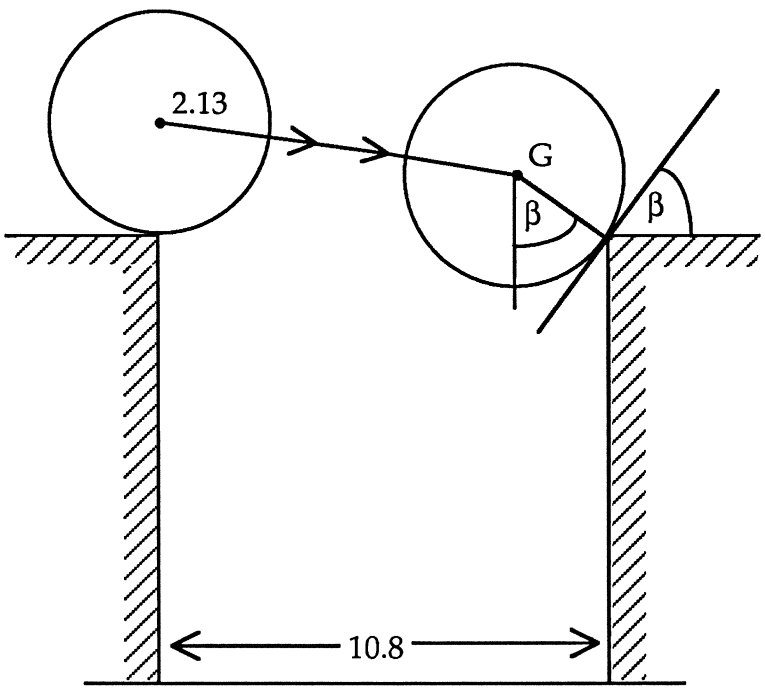


Figure 2

The edge of the cup's rim is not a sharp right angle but is actually rounded. Therefore, when the ball strikes the back edge, there will be a definite tangent at the connection point between the two surfaces (Figure 2). Let the tangent plane make an angle β with the green's surface. If $(\dot{x}_B, \dot{y}_B)$ denote the horizontal and vertical components of the ball's velocity just before striking the back edge, it is seen that the component in the tangent plane direction is $(\dot{x}_B \cos\beta - \dot{y}_B \sin\beta)$, and in the normal direction is $(\dot{x}_B \sin\beta + \dot{y}_B \cos\beta)$. Just after impact the tangent plane component remains $(\dot{x}_B \cos\beta - \dot{y}_B \sin\beta)$, but the normal direction component is reduced by a factor e and reversed in sign to $-e(\dot{x}_B \sin\beta + \dot{y}_B \cos\beta)$, according to Newton's law of impact. Thus the velocity component in the 0x-direction just after impact is

$$\begin{aligned} \dot{x} &= (\dot{x}_B \cos\beta - \dot{y}_B \sin\beta)\cos\beta - e(\dot{x}_B \sin\beta + \dot{y}_B \cos\beta)\sin\beta \\ &= \dot{x}_B(\cos^2\beta - e\sin^2\beta) - \dot{y}_B(1+e)\sin\beta\cos\beta \\ &= \dot{x}_N(1 - (1+e)\sin^2\beta) - g\tau(1+e)\sin\beta\cos\beta \end{aligned}$$

using equations (15) and (16). Therefore the condition for determining the upper limit of $\dot{x}_0$ is

$$0 = \dot{x}_N(1 - (1+e)\sin^2\beta) - g\tau(1+e)\sin\beta\cos\beta \tag{17}$$

The centre of mass G of the ball, when it strikes the back edge of the hole, is given by

$$\begin{aligned} x &= C + 10.80 - 2.13\sin\beta, \\ y &= 2.13\,(1 - \cos\beta). \end{aligned}$$

Substitution of these into equations (15) and (16) yield

$$\dot{x}_N\tau = 10.80 - 2.13\sin\beta, \tag{18}$$

$$\frac{1}{2}g\tau^2 = 2.13\,(1 - \cos\beta). \tag{19}$$

Equations (17), (18) and (19) have to be solved for the three unknowns $\dot{x}_N$, τ and β. Multiplication of equation (17) by τ, and elimination of $\dot{x}_N\tau$ and $g\tau^2$ using results (18) and (19), produces

$$(10.80 - 2.13\sin\beta)(1 - (1+e)\sin^2\beta) = 4.26\,(1+e)(\sin\beta\cos\beta - \sin\beta + \sin^3\beta),$$

which can be modified to a sixth-order polynomial in sin β. Once e is specified for the ball and the green impact, the appropriate solution for β can be obtained using a mathematical package such as MATHEMATICA. This can then be substituted into equations (18) and (19) to produce a value for $\dot{x}_N$, and finally the upper limit for $\dot{x}_0$ using equation (14). If the ball is putted initially with a speed greater than this limit, it will jump the hole and continue in its original direction on the other side of the cup.

6. FURTHER INVESTIGATION

Further analysis needs to be undertaken to determine the conditions that a ball rolls around the rim of the cup and pops out. This will involve an investigation similar to the one given in this paper, but for oblique incidence at the cup. The effect of an inclined green also needs to be taken into account.

Eventually the behaviour of the ball in rolling to the rim on an oblique putt, stopping, and then falling in needs to be compared with a ball rolling directly. This will then help to confirm or deny the strategy of striking the ball hard enough to make it stop just past the hole.

The authors acknowledge the assistance given by Terry Adcock, the professional at Robina Woods Golf Club, for permission to use the chipping green for the experiments, and to Steve Guttormsen from Audio-Visual Services at Bond University for videoing all the experimental runs.

TOWARDS A PREFERRED GOLF SWING AND ITS APPLICATION TO TEACHING GOLF

John Baker[1]

Abstract

This paper gives the rationale for a mathematical model of a *preferred* golf swing, and shows how the model can be used in golf teaching. The concept of preferred is described and arguments are given for and against the choice of one action over another. The mathematical model is then used to drive a computer program, which forms the basis of a golf teaching method called Coordination Tutoring. This teaching method uses a video camera and monitor to provide players with instantaneous information about their swing, together with an overlay of the preferred computer model which shows how they should be swinging. Experiences of teaching with this approach are described.

1. INTRODUCTION

This article will:

- compare the scientific and coaching literature on the topic of the dynamics of the golf swing,
- describe a mathematical model of a preferred golf swing,
- discuss the potential of the realisation of the computer program for the learning, and hence playing, of golf.

Having made this brief introduction, it remains only to welcome the reader to the minefield. For there are about as many good theories about the nature and composition of the golf swing as there are writers on the topic. Choices therefore have to be made for the development of the mathematical model. The term *preferred* is used to indicate that such choices have been made on which the model is based. Where possible the model will enable features of the swing to be parameterised, but the purpose of parameterisation will be to enable the model to cater for variation in human proportion and capability, rather than to cater for variation in the preferred way in which the swing should be made.

2. SWING BASICS

A key source for the development of the model has been the work of Williams [1] which was, as he acknowledges, made possible by the strobed photographic work of Edgerton and Killian [2]. Whatever shortcomings in Williams' analysis have been noted by others, the features that he uncovered have served as the basis for many

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

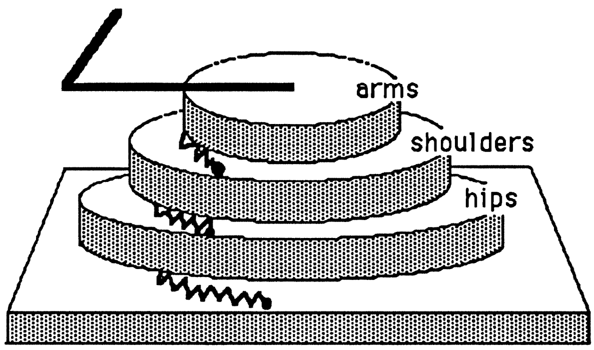


Figure 1: A mechanical model of a golf swing

later studies of the golf swing, and it is used as the basis of the mathematical model of the preferred golf swing. The other major source has been Cochran and Stobbs [3] who give mathematical details of some other parts of the golf swing. Indeed, there is no better model than that suggested by Cochran and Stobbs [3, page 51]. Their view is of a system of three disks that rotate around a common spindle and connected to each other and the ground by means of springs as in Figure 1.

The motion that this mechanical model supports is described by the graphs on the next page, which comprise much of the mathematical model developed in this study. A number of key features of a good golf swing are incorporated into the graphs of Figure 2. For example, in the downswing section, note how the hips initiate the downswing and the uncocking of the wrists is delayed as late as possible. These are features of a preferred swing, in the sense that, whatever else the player does, they should at least follow that form of the downswing. Further reasons for the shapes of the graphs are given in the rest of Section 2.

2.1. The Address Position

There is little mention of the address position in technical studies, even though it is clearly of great importance that the swing begin correctly. The essential features are:

1. The position of the feet relative to the ball and ball-to-target line.
2. The position of the body's centre of gravity.
3. The angles made by the player's limbs (legs, spine and arms) and the club.

All these aspects need to be considered when developing a model that assumes a correct stance.

2.1.1. Position of the feet

Cochran and Stobbs give the results of a statistical analysis of the stances of fourteen top professionals. The mean values they obtained are summarised on Figure 3.

In this study, the line of the toes is shown as exactly parallel to the ball-to-target line, though a range from 5.8° open to 3.5° closed was reported. In their depiction, the right foot is shown as being perpendicular to the toe line, whereas in Figure 3, both feet are shown at an angle to the toe line. The latter position is suggested by Jones [4, page 18]:

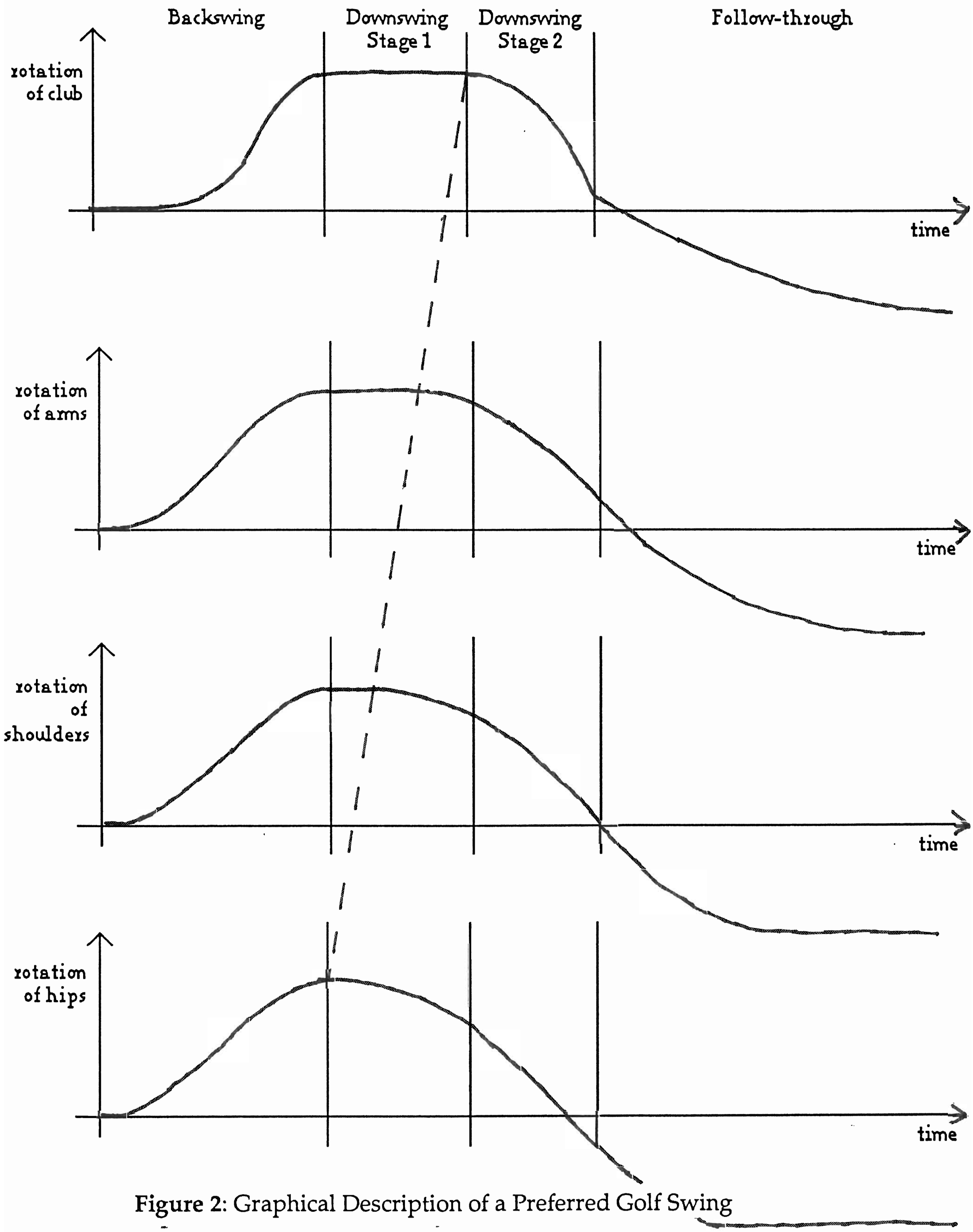


Figure 2: Graphical Description of a Preferred Golf Swing

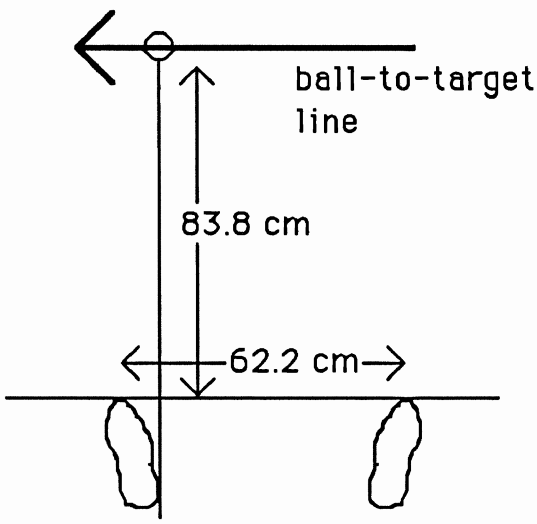


Figure 3: Mean foot positions of 14 top professionals

> This means that the toes of both feet should be turned slightly outward, that is, away from the ball.

Hogan [5, page 41] is emphatic in his disagreement with Jones:

> THERE IS ONE CORRECT BASIC STANCE: THE RIGHT FOOT IS AT RIGHT ANGLES TO THE LINE OF FLIGHT AND THE LEFT FOOT IS TURNED OUT A QUARTER OF A TURN TO THE LEFT.

In the model, the position of the feet can be altered to suit the player and/or the advice of the coach.

2.1.2. *Weight Distribution*

Most golf manuals are consistent in their advice on how the weight should be distributed at address position. For example, Jones [4] recommends that:

> If the posture of the player at address is natural and comfortable, his weight should be supported equally by his two feet; this should be so because the swing should rotate about the spine as an axis.

His remarks were confirmed by Richards et al [6], whose report on the weight transfer patterns during the golf swing included the trace of what they called the COVF (centre of vertical force). A feature of the results given by Richards is that while there is considerable left-right (lateral) weight transfer, there is very little front-back weight transfer. These findings have an important role in the development of the model, as it means that the direction of lateral movement can be restricted to the direction of the target line.

2.1.3. *Limb angles*

There are (at least) four limb angles and a club angle that will determine the limb positions at address; the inclinations of the shin, thigh, spine and left arm (which is usually straight at address). The need to specify all these angles for the model cannot

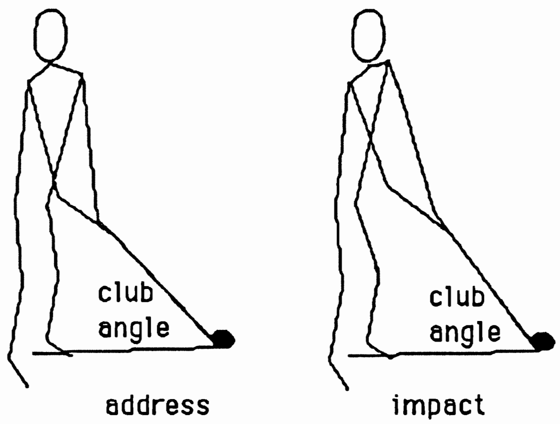


Figure 4: Club angle at address and impact

be avoided, nor can the conflicting information from the teaching manuals. For example, recorded values of club angle range from 41° – 54° for different players using a driver (see McIntyre and Snyder [7]) but the variation cannot be accounted for by the height of the player. Therefore the model caters for the club angle as a parameter.

While on the subject of club angle, it is worth noting here that the position of the hands at address is **lower** than their position at impact, with the result that the club angle at address is less than the club angle at impact. This is shown in Figure 4.

There are two reasons why the change occurs.

1. At address, the left shoulder is more to the front than it is at impact. The result being that, while the left arm is straight in both positions, the left shoulder is further from the ball at impact that it is at address. Thus, the angle between club and arms must be less at impact than at address, and hence the club angle is greater at impact than at address.
2. At impact, the weight and motion of the club generates a centrifugal acceleration of the clubhead away from the hands, in the line of the clubshaft. This acceleration will have the effect of decreasing the angle between the arms and the club, and hence of increasing the club angle.

This change needs to be part of the preferred swing for the model, as the angle of the swing plane will not be equal to the club angle at address but equal to the club angle at impact.

2.2. The Backswing

The purpose of the backswing is to get the club and player into the correct position at the top of the swing, ready for the real action. But despite its humble purpose, the backswing attracts considerable attention in the golf manuals, and there are conflicting views as to how it should be carried out. The major source of conflict is the question of inside–outside, between those who describe the downswing as being inside the backswing to those who describe it as being outside. The argument can be illustrated by a diagram of the two planes shown in Figure 5.

On Figure 5, Plane A is outside Plane B, as the points on Plane A are evidently further away from the player's feet before impact. So the question is:

> Is the backswing on Plane A and the downswing on Plane B, as the outside-to-inside writers propose, or is the backswing on Plane B and the downswing on Plane A, as the inside-to-outside writers propose?

Hogan [5, page 87] is quite clear that the downswing is inside the upswing. But Williams [8, page 51] is equally clear that the reverse is the case. A teaching manual by Adwick [9] provides evidence that makes it hard, if not impossible, to accept that Hogan could be right. Using his patented method, called X-ray tracings, Adwick provides consistent photographic records that can only be achieved by an inside to outside swinging action. For the purposes of the model, therefore, the preferred swing will be shown with the upswing inside the downswing, as proven by Adwick.

2.3. The Downswing

Many studies of the golf swing make the assumption that the downswing can be described in terms of a compound pendulum that includes a stop to restrict the club from jack-knifing onto the player's arms during the early stages of the downswing. This model was introduced by Williams [1]. The pendulum is shown in Figure 6.

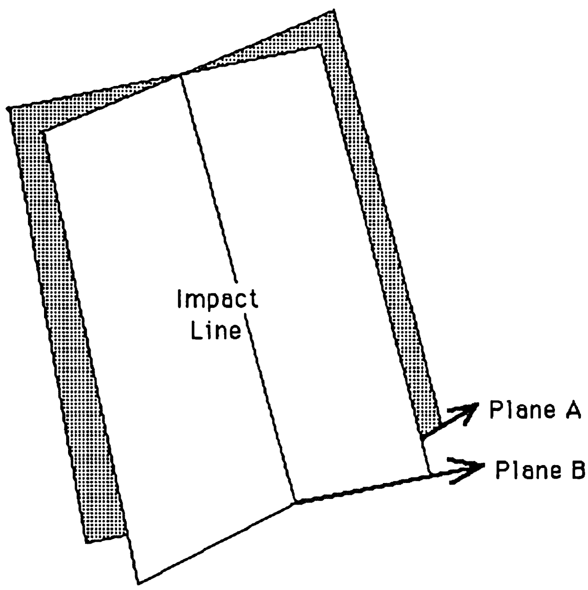


Figure 5:Two planes for the golf swing

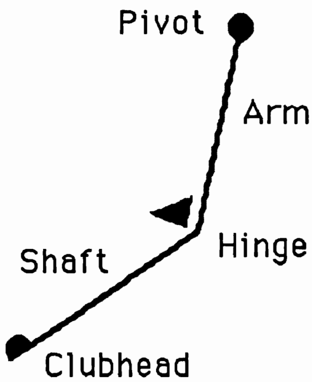


Figure 6: The golf swing as a compound pendulum

The upper pendulum, or 'arm', should not be confused with the player's left arm as it is by some writers, rather it is an imaginary line, pivoted at a point that is fixed in space during the downswing. The player can provide a couple about both the pivot and the hinge and the speed of the clubhead is obtained by means of a flail-like action of the shaft swinging away from the hinge. Williams assumes this couple to be small and positive; Vaughan [10], on the other hand, used an analysis of flash photographs based on digitisation and noted that a substantial initial positive couple about the hinge is followed by a negative couple that reaches -100 newton.metres at impact.

In either case, the resulting swing can be divided into two stages:

Stage 1. The arm rotates about the pivot, gradually gathering angular velocity, while the club is constrained to move by the action of the stop.

Stage 2. The centrifugal acceleration of the club is sufficient to cause it to fly away from the stop, and the arm and club move in compound pendulum motion.

In the following analysis, it is shown how two functions of time can be derived to form the mathematical model for Stages 1 and 2. These are:

$\theta(t)$ = the angle which the arm makes with a fixed direction, and equal to the sum of the arm, shoulder and hip turns represented by the graphs of Figure 2,

$\phi(t)$ = the angle that the club shaft makes with a fixed direction.

2.3.1. *First Stage*

During Stage 1, the distance of the club's centre of gravity from the pivot is constant, r, and so is the angle, β, that the club makes with the radius r, as in Figure 7.

Because the arm-club configuration is fixed during the first stage, the acceleration of the club's centre of gravity can be described in terms of θ and r alone in terms of the radial and tangential components of acceleration associated with circular motion. If these accelerations are resolved at right angles to the line of the club, they can be equated to the force at the centre of gravity which is produced by a couple of C about the hinge, this being the action of the stop on the motion of the system.

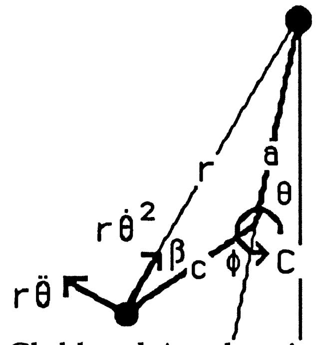


Figure 7: Clubhead Accelerations in Stage 1

Williams assumed that the couple is a non-zero constant, C_0, and hence that the equation of motion is:

$$\frac{C_0}{mc} = -r\,\dot{\theta}^2 \sin\beta \;-\; r\,\ddot{\theta}\cos\beta \qquad (1)$$

This is the equivalent of Equation 5 in Williams (1967). The solution can be obtained by setting

$$\ddot{\theta}=\frac{d}{d\theta}\left(\frac{1}{2}\dot{\theta}^2\right)$$

and solving the resulting first-order differential equation. For the given initial conditions,

$$\dot{\theta}^2 = \frac{C_0}{mcr\,\sin\beta}\left[\exp\{2\tan\beta(\theta_0 - \theta)\} - 1\right] \qquad (2)$$

where θ_0 is the angle of maximum arm rotation. In (2), let

$$\lambda^2 = \frac{C_0}{mc\,r\,\sin\beta}$$

and

$$\kappa = 2\tan\beta$$

Equation 2 can then be integrated to yield

$$\therefore \qquad \theta(t) = \theta_0 + \frac{2}{\kappa}\ln\left(\cos\left(\frac{\kappa\lambda.\,t}{2}\right)\right)$$

Remembering that we have assumed that

$$\phi(t) = \phi_0,\ \text{constant}$$

these formulae can be used to determine (in functional form) the shape of the graphs of Figure 2 between the start of the downswing and the end of Stage 1. These functions can then be used as the basis of a computer program in which a stick figure is made to replicate the actions of the golf swing.

2.3.2. *Second Stage*

The flail action now takes over, and the original assumption that ϕ is constant is abandoned. Following the lead of Williams [1] once more, we can reduce the complexity of the analysis by assuming that, for the second stage, $\dot{\theta}$ becomes constant, ω_0. This assumption is based on observations of the photograph of Jones in Edgerton and Killian [2], where the hands do appear to be equally separated in each exposure of Stage 2. Equation 1 now becomes:

$$\frac{C}{mc} = -a \sin\phi \cdot \omega_0^2 + c\,\ddot{\phi} \qquad (3)$$

Williams also makes the assumption that the couple C reduces exponentially from C_0 to close to zero at impact time:

$$C = C_0 \exp(\phi_0 - \phi)$$

Equation 3 now becomes:

$$\ddot{\phi} = \kappa' \sin\phi + \lambda' \exp(\phi_0 - \phi) \qquad (4)$$

where

$$\kappa' = \frac{a\omega_0^2}{c}$$

and

$$\lambda' = \frac{C_0}{mc^2}$$

This can be integrated with respect to ϕ, to obtain an equation of the form:

$$\dot{\phi}^2 = L + M \cos\phi + N \exp(-\phi)$$

where

$$L = \omega_0^2 + 2\lambda' - M \cos\phi_0$$

$$M = -2\kappa'$$

$$N = -2\lambda' \exp[\phi_0]$$

Unfortunately, this equation for $\dot{\phi}$ in terms of ϕ has no explicit integral that will enable ϕ to be expressed as a function of t. However, the equation can be solved numerically, and this approach is used in the model.

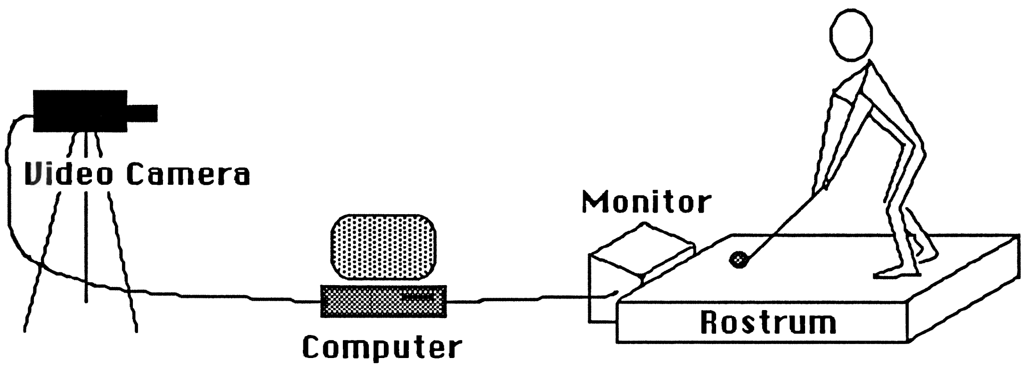


Figure 8: Basic Requirements of the Preferred Embodiment.

A key aspect of the above is the assumption that ω_0 is a constant for Stage 2. This assumption and those for Stage 1, which Vaughan [10] showed may not match the reality of all golf swings, nevertheless do match with good practice as described in golf coaching manuals and will be used in the preferred swing.

2.4. The Follow-through

Without exception, descriptions in the scientific literature of the mechanics of the golf swing stop very soon after the moment of impact. One reason for this omission is given by Daish [11] as:

> In the same way, in a drive at golf, the clubhead and ball remain in contact over a distance of less than two centimetres at the bottom of the swing. Thus, any follow through in the stroke is unable to affect the ball in any way—it will have lost contact before the follow-through has really begun.

Because the follow-through has no mechanical influence on the flight of the ball, it has not been a subject of study. The omission is, however, contrasted by the importance that golf manuals attach to the follow-through. One unmentioned feature of the follow-through is the way in which the right arm takes over from the left — to some degree, the follow-through is a mirror image of the backswing. The graphs of Figure 2 are based on this principle.

3. Application of the Model

The application of the model swing concept lies in a realisation of the system known as Coordination Tutoring in the field of golf. The hardware requirement for this system are depicted in Figure 8.

A player practises a particular golf shot. As they play, a video camera pictures the actions and sends the image to a computer. The computer has been programmed to generate the preferred swing as a wire-frame figure that has the same proportions as the player. The program is based on the mathematical model as described above and the computer hardware is able to superimpose the wire-frame image on the image of the player. The combined image is fed to a video monitor, which is cradled on a rostrum on which the player stands. Seeing their image combined with the preferred swing positions gives the player immediate feedback, or biofeedback, on their actions and enables them to learn how to correct mistakes that they make, by striving to emulate the positions adopted by the model.

3.1. Coordination Tutoring as a Teaching System

The system overcomes some of the difficulties that a golf coach often faces. Either by means of words, by demonstration or by showing the pupil their errors on video, the coach looks for some way of describing to the player how to improve. The player then has to carry that description onto the practice ground and attempts to put the description into action. But because of the nature of golf, the player cannot see their attempts to make the correction.

Coordination Tutoring provides a novel mechanism for the coach to describe what is required to the pupil. For example:

> "At the top of the backswing, the club should be parallel to the ground, and pointing toward the target."

This description is now achieved by displaying the model figure at the top of the backswing. No words are needed, the picture says it all.

The player now has to work on matching the position shown by the model. As they swing through the backswing, they can pause and immediately see whether they have swung too far, or not enough, and whether their action was on the suggested swing plane. No feedback from the coach is required, as the system itself provides instant information on the extent to which they have matched up to the preferred position.

3.2. Educational Effectiveness

The system described above has been used in its developmental stage with a small number of subjects. All have been high-handicap or beginning players and have had little prior coaching on the system. However, the initial results suggest that the system is capable of achieving the following:

1. The player can learn to recognise the difference between a swaying action and one based on rotating the body.

2. The player can learn to adopt a preferred address position, with a reasonably straight back and correctly bent knees.

3. The player can get the feel of the extent of the swing, and can determine the amount of body rotation of which they are capable.

4. The implication of the swing plane concept for the golfer can be easily comprehended, and the player can adopt a swing that conforms to a swing plane that suits their capability.

5. The two-stage quality of the downswing can be learned, in combination with the preferred unwinding mechanism (hips-shoulders-arms-club).

What has been remarkable has been the speed with which the subjects have been able to learn from the system. Most subjects have shown considerable progress with no more than 1 hour of practice on the system. By considerable progress is meant that a person who was showing a number of typical weak spots (e.g. extremely bent left arm at the top of the backswing because of insufficient body rotation) had completely corrected the fault. A study is in progress to quantify the potential of the system to help players correct faults.

REFERENCES

[1] Williams D., The dynamics of the golf swing, *Quarterly Journal of Mechanics and Applied Mathematics*, 20 (2), 247-264, 1967

[2] Edgerton H and Killian J. R., FLASH: Seeing the unseen by ultral high-speed photography, (Branford, Boston, 1954).

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

[4] Jones R. T., Bobby Jones on the basic golf swing, (Doubleday and Company, New York 1969).

[5] Hogan B with Wind W., The modern fundamentals of golf, (Nicholas Kaye, New York 1957).

[6] Richards J., Farrell M., Kent, J., and Kraft R., Weight transfer patterns during the golf swing, *Research Quarterly for Exercise and Sport*, 56 (4), 361 - 365, 1985.

[7] McIntyre N. and Snyder R., Compare your swing with 20 pros, *Golf Digest*, 15 (3), 48 - 51, 1964.

[8] Williams D. The Science of the Golf Swing, (Pelham Books, London, 1969).

[9] Adwick K., X-ray way to master golf, (Pelham Books, London, 1970).

[10] Vaughan C. L., A three-dimensional analysis of the forces and torques applied by a golfer during the downswing, in *Proceedings of the Seventh International Conference on Biomechanics*, (University Park Press, Baltimore, 1981).

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

GOLF SCORES, EXPLORATORY DATA ANALYSIS AND THE POISSON MODEL

Heather J. Whitford [1]

Abstract

It has often been recorded (see for example Singer and Willett [1]) that the teaching of statistics is enriched by the use of 'real' data sets rather than data that has been artificially constructed to demonstrate some statistical property. This paper looks at the final two rounds of some Australian golf tournaments and uses the relatively new tools from exploratory data analysis to summarize the data. The Poisson distribution is then used to model the scores of golfers who have qualified for the final two rounds.

1. Introduction

Exploratory data analysis (EDA) gives the user a range of tools that allows for the investigation of data in order to *discover* what 'story' is being told by the data. Specifically, EDA consists of a variety of techniques which allows for the summary of data while easily identifying outlying or extreme values. The beauty of the subject lies in the extreme simplicity of the calculations that are easily done using pencil and paper or by using a computer package. EDA should be seen as a precursor to more sophisticated analyses that make assumptions about the nature of the data used. While this paper does not describe all the techniques developed by Tukey [2], an extensive discussion of some which are used in a teacher education course is contained in Whitford [3].

The data used in the first section of this paper are the scores from the final two rounds of the 1994 Microsoft Australian Masters Championship held at the Huntingdale Golf Course in Victoria. Investigation will centre on the distribution of the scores, the shape of this distribution, the value of a 'typical' score, the spread of the scores and the identification of outliers.

The second section of this paper uses scores from the final two rounds of the last three Australian Open Championships and the last three Australian Masters Championships. Summary statistics, this time of parametric form, will enable the fitting of the Poisson distribution to the scores as proposed by Mosteller and Youtz [4].

[1] Flinders University of South Australia, Sturt Rd., Bedford Park, S.A., 5042

2. STEM AND LEAF PLOT

The stem and leaf plot while similar to the histogram in that it gives the shape of the distribution of the data also retains the original data in readable form in the plot. It consists of two parts- a stem that contains all but the last significant digit and the leaf that gives this digit. The scores are given here (Figure 1) in a back-to-back plot that enables easy comparison.

Saturday scores (n=67)		Sunday scores (n=67)
8	6	8899
11100	7	000000111
333333333322222	7	22222223333333
5555555544444	7	444444444445555555
7777777766666666	7	6666667777777
9999998888	7	89999
00000	8	001
	8	3
6	8	

Figure 1: Stem and leaf plot for Saturday and Sunday 1994 Masters, stem = tens, leaf = units.

The stem and leaf plot reveals that there were more low scores on Sunday with the highest score for the two days being shot on Saturday. The distribution of both sets of scores appears to be roughly symmetrical. While this plot is useful for a first look at smaller data sets some of the 'detail' can be filtered out by looking at summary statistics (Table 1) and a boxplot (Figure 2).

3. THE FIVE NUMBER SUMMARY

Table 1 contains summary statistics that have been calculated from the stem and leaf plot. The depth gives the number of values that have been counted in from either end to find the corresponding statistic. In the five number summary the median, the 25th and 75th percentiles and the minimum and maximum scores are used. The mids are the arithmetic means of the high and low values of the number summary. Their calculation can be used for a discussion of symmetry: increasing mids show that the data is positively skewed, and mids that remain roughly constant indicate symmetric data.

Table 1

Five number summaries for 1994 Masters

	Sat			Sun		
Depth	Low	High	Mids	Low	High	Mids
34	75		75	74		74
17.5	73	78	75.5	72	76	74
1	68	86	77	68	83	75.5

The five number summary for each data set show the properties identified in the discussion of the stem and leaf-the minimum, maximum and the impression of symmetry. In addition, the 'typical' score is obtained. This score is one less for the Sunday round reinforcing the impression already obtained that overall the scoring was slightly better on the Sunday.

4. BOX PLOT

The box plot results from using the information of Table 1. The median is identified by + and the box outlines the middle 50% of the scores. The horizontal lines or 'whiskers' go to what would be considered to be the minimum and maximum values if the data were 'well behaved'. Any unusual scores outside these whiskers are plotted individually and so identified as having the potential for causing problems in parametric analysis. Clearly, in sporting data where unusual scores may be of the greatest interest the boxplot can play a valuable role. In this example the slightly better overall performance on the Sunday is reflected by the lesser range of the middle 50% of Sunday's scores and the shorter left hand whisker. There are no unusually low scores but on each day there was one golfer who performed poorly compared with the rest of the field.

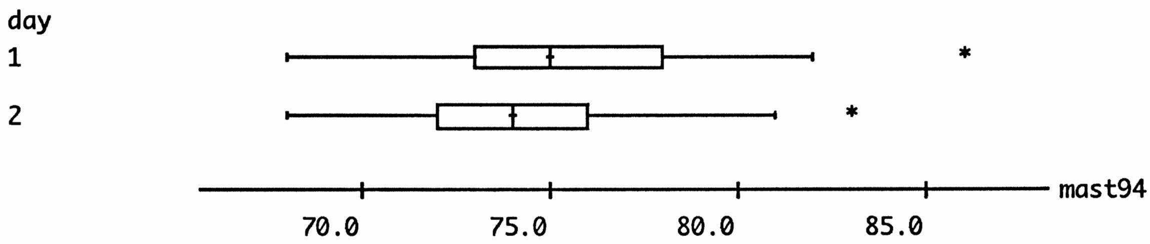


Figure 2: Boxplots of Saturday and Sunday scores of the 1994 Masters
1 = Saturday, 2 = Sunday.

5. OTHER INVESTIGATIONS

In analysing this data set the two sets of scores were treated as independent whereas this is clearly not the case. Students set the exercise of investigating these data sets may choose to use the pairing by forming differences or by using a two dimensional approach (eg. a scatter plot and median trace). In investigating data to see what *story* it can tell there should be no intention to constrain the investigation. Imaginative approaches can provide invaluable insights. In this case all investigations show that the data is reasonably well behaved and is suitable for use in exploring the Poisson distribution.

6. THE POISSON MODEL

While the previous sections have focussed on the statistical description of the distribution of scores from professional golf tournaments, another aspect for analysis is the fitting of a model to these scores. Mosteller and Youtz [4] examined many scores from the final two days of tournaments in America and fitted a model to describe the pattern of golf scores returned in a day's play.

Their approach assumed that a golfer's score, X, satisfied the relation

$$X = \text{base} + Y .$$

The base is regarded as the score for a perfect round while Y is the excess strokes above this fixed base. Y is a random variable which Mosteller and Youtz showed could be modelled by the Poisson distribution with parameter λ.

$$\text{ie. } P(Y = k) = \frac{e^{-\lambda}\lambda^{k}}{k!}, \; k = 0,1,2,....$$

The mean and variance of this distribution are λ, so if the model is to represent the situation the mean of the scores used should be approximately equal to (base + λ).

In using this model there are of course aspects of the problem that were simplified or ignored: (1) the scores on the final two days of a tournament were treated as independent whereas for some tournaments there was significant correlation (2) the effect of differing pools of players and the different courses was ignored and (3) there could be some scores *less than* the base-these were not considered.

Mosteller and Youtz used data from the final rounds of 33 tournaments. All scores were adjusted to relate to a par of 72. Since the weather can be a significant factor in scoring, the results from 10 tournaments where the weather was ideal were analysed and the model of base 64 with mean excess strokes (λ) of 8.1 proved to be a good fit. In the tournaments where the weather was less favourable the base was 62 with λ=10.4. This second model, although appearing to be a reasonable fit, was rejected by the chi-square goodness of fit test.

In this paper two tournaments will be considered- the Australian Open and the Australian Masters Championship. The Sydney office of the PGA kindly supplied the scores for the last two rounds of the Open 1991-1993 and the Masters 1992-1994. The Open was played on three different courses while the Masters is always played at Huntingdale. All Opens were played on courses where the par was 72. The par of Huntingdale is 73 so all Huntingdale scores were adjusted by subtracting one. Summary statistics are given in Table 2.

Table 2

Summary statistics for the last two days of Opens and Masters

		Open		Masters(Adj.)	
		Mean	Variance	Mean	Variance
1991	Sat.	73.8	7.5	-	-
	Sun.	74.4	8.2	-	-
1992	Sat.	75.0	12.3	72.1	10.0
	Sun	74.2	9	73.6	10.2
1993	Sat.	72.0	9.9	73.3	9.2
	Sun.	72.5	7.4	71.8	12.9
1994	Sat.	-	-	74.6	11.0
	Sun.	-	-	73.2	10.4
Comb.		73.6	10.0	73.1	11.3

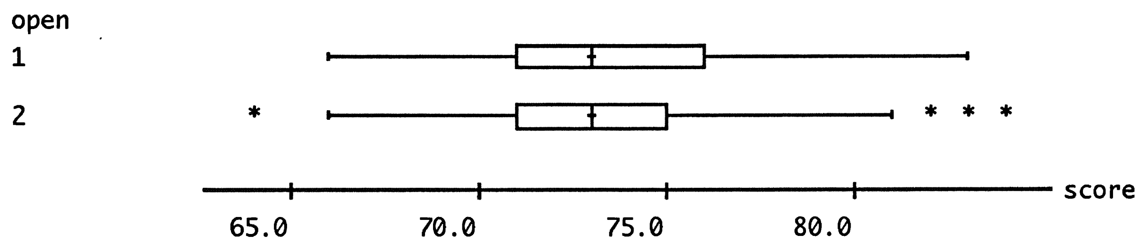
Strong winds were recorded on the Sunday of both the 1992 Open and 1992 Masters. The summary statistics did not reveal that the wind had any effect on scoring and a comparison of the Sunday boxplot with the corresponding Saturday plot showed no unusual scores. All rounds were retained for analysis.

Clearly there are many options for the choice of λ for each tournament and of these choices there will be several that will provide an adequate fit of the model. For the Open a base of 63 and a λ of 10.3 produced a χ^2 statistic of 9.48 on 12 degrees of freedom while for the Masters a base of 62 and a λ of 11.1 produced a good fit (χ^2 = 11.28, df = 13). Table 3 shows the observed and expected frequency for these models. Figure 3 gives the boxplot of the original data together with a boxplot of the same number of scores randomly selected from the corresponding Poisson model.

Table 3

Observed frequencies and expected frequencies using the Poisson model

	Open (n=380)	Base = 63 λ =10.3		Masters (n=388)	Base = 62 λ =11.1
Score	O	E	Score	O	E
-	-	-	≤66	8	5.5
≤67	7	8.5	67	5	8.2
68	12	12.3	68	16	15.2
69	13	21.2	69	25	24.2
70	28	31.2	70	30	33.5
71	37	40.2	71	50	41.3
72	48	46.0	72	37	45.9
73	47	47.3	73	45	46.3
74	50	44.3	74	46	42.8
75	38	38.0	75	32	36.6
76	33	30.1	76	29	29.0
77	21	22.2	77	25	21.5
78	22	15.2	78	19	14.9
79	10	9.8	79	12	9.7
80	4	5.9	80	3	6.0
≥81	10	7.8	≥81	6	7.4



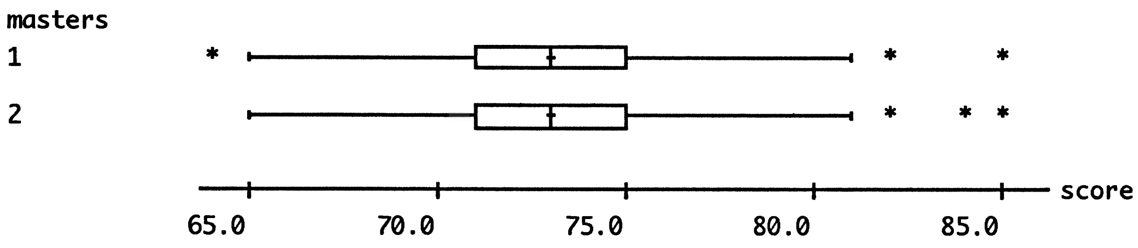


Figure 3: Boxplots of actual and simulated golf scores. 1 = actual scores, 2 = randomly selected scores.

The boxplots show that there is some similarity between the data and the model. In the case of the Open the data had more numbers in the middle range of scores than would be expected under the Poisson model and it was here that there was the greatest contribution to the chi-square statistic. Thus the main difference between the two plots is in the range of the middle 50% of scores. In the case of the Masters the contribution to the chi-square statistic was fairly uniform across the scores so resulting in boxplots that show very little differences.

7. DISCUSSION

The model proposed by Mosteller and Youtz had a base of 64 and a λ of 8.1. This does not provide a good fit for either of the data sets or the combined set (Open and Masters). The models show differences between the two countries and one could speculate on the effects of differing types of courses, depth of players, the data being gathered over a longer period and other factors. It is well known that Australia has produced golfers capable of leading any field but it could be that the larger variance occurs because our fields are not as strong. Similarly it is interesting to speculate why the data that was obtained at a single course produced a larger variance. Students with an interest in golf will find many questions arising from this type of analysis that can be the basis for further investigation.

REFERENCES

[1] Singer J.D. and Willett J.B., Improving the teaching of applied statistics: Putting the data back into data analysis, *American Statistician,* Vol 44, No. 3, 223-226 (1990).

[2] Tukey J.W., Exploratory Data Analysis, (Addison-Wesley, Mass., 1977)

[3] Whitford H.J., Introducing exploratory data analysis-the origins and structure of a mathematics unit for pre-service teachers, The Teacher Education Mathematics Project (J. Baxter ed.) 107-119 (1988).

[4] Mosteller F. and Youtz C., Where eagles fly, *Chance: new directions for statistics and computing,,* Vol 6, No.2, 37-42 (1993).

THE AUSTRALIAN GOLF HANDICAPPING SYSTEM

Anthony K. Whitford[1]

Abstract

Since 1992 the Australian Golf Union has directed clubs to use the Calculated Course Rating (CCR) system for determining the daily "degree of difficulty" of golf courses. This rating is the basis on which male golfers' handicaps are adjusted after each competition round, and is set at the 15th percentile of net scores returned. This study is based on fourteen competition rounds at the Glenelg Golf Club in Adelaide, where the 15% formula is felt to be not working properly. The study looks at alternatives to the 15th percentile and also seeks a CCR formula which takes into account the handicaps of the competitors.

1. BACKGROUND

In 1992 the Australian Golf Union (AGU) introduced a new method of determining the day to day "degree of difficulty" of golf courses, called the Calculated Course Rating (CCR) system. This system replaced the Daily Course Rating (DCR) system, and is now the basis on which male golfers' handicaps are reviewed after every competition round. In practice the CCR is a whole number within a few strokes of the par of the course, as was the DCR. However, whereas the DCR had to take into account such factors as tee placements, condition of the course and the weather, the CCR is determined statistically and is therefore far easier to administer. In all but a few special cases (e.g. small fields, low handicap fields) the CCR is simply equal to the net score returned by the 15th percentile of the competition field. For example, if there are 178 starters in a Stroke competition, one computes 15% of 178 (= 26.7), and the 27th ranked net score becomes the CCR for the day. Exactly the same method is used for Par and Stableford competitions, except that the scores are converted to equivalent Stroke scores (e.g. 3 up = 3 under par, 34 points = 2 over par).

2. STUDY AT GLENELG GOLF CLUB

For the purposes of this paper I have kept records of 14 competition rounds at the Glenelg Golf Club over the period February-May this year. Glenelg is recognised as a major club in Adelaide and for the past few years has been ranked by the Australian Golf Digest magazine amongst the top 50 courses in Australia. The course has a par of 72 and an Australian Course Rating of 72. The following table shows the date, field size, number of cards returned and CCR for the 14 rounds:

[1] University of South Australia, Holbrooks Rd, Underdale SA 5032

Table 1

Data from 14 competition rounds at Glenelg, Feb–May 1994

Date	5/2	12/2	26/2	5/3	12/3	19/3	9/4	23/4	27/4	30/4	4/5	7/5	11/5	14/5
Field	192	199	184	193	186	175	208	200	131	202	127	193	108	198
Cards	151	180	171	158	157	158	165	186	112	180	109	159	92	179
CCR	74	74	73	76	74	73	76	75	73	74	74	76	75	75

The sum of the fields is 2496 and the sum of the cards is 2157, so that over these 14 rounds the return of cards averaged 86.4%, a point we shall refer to later. Of most significance, however, is the fact that the CCR averaged 74.4 and was always above the par of 72, despite there being no rain and virtually no wind, and the course being in excellent playing condition throughout this period. It is also worth noting that under the previous system the DCR was often rated at par or below. The feeling at Glenelg is that the 15% formula consistently gives a CCR value one or two strokes too high.

3. ALTERNATIVE PERCENTILES

If the 15th percentile yields a CCR too high, which percentile consistently yields an acceptable value? The recording method used in the survey makes this easy to find. For each of the 14 competition rounds all net scores were recorded relative to the CCR of the day, so that 0 was used for a net score equal to the CCR, -1, -2, etc for net scores below the CCR, and 1, 2, etc for net scores above the CCR. The following table shows the percentiles (rounded to one decimal place) at which ranked net scores changed from -3 to -2, -2 to ƒ1, -1 to 0, and 0 to +1 respectively:

Table 2

Percentiles at integer change points for 14 rounds

Round	–2	–1	0	+1
1	5.7	7.8	11.5	17.2
2	5.0	9.0	14.1	21.1
3	4.3	9.2	11.4	19.6
4	4.7	7.8	11.4	17.1
5	4.3	11.3	14.0	20.4
6	3.4	6.3	12.6	18.9
7	2.4	5.8	9.1	18.3
8	4.5	10.5	14.5	20.5
9	6.1	8.4	10.7	16.8
10	5.4	10.9	14.4	22.8
11	4.7	6.3	15.0	26.8
12	7.8	9.3	14.5	17.1
13	4.6	7.4	14.8	24.1
14	8.6	10.1	14.1	19.2

It can be seen that the 11th percentile gives a CCR one lower in all rounds except 7 and 9, and two lower in round 5. The 10th percentile gives a CCR at least one lower in all rounds except 7, and two lower in rounds 5, 8, 10 and 14. It also reduces the average CCR to 73.2. The 9th percentile reduces the average to 73.0, the 8th reduces it to 72.9, and so on. It is obviously hard to know which is the best option, but I think 9% may be about right.

4. HANDICAP VARIATIONS

An apparent weakness with the 15% formula is that it does not take into account the handicap profile of the competition field. The AGU does acknowledge the fact that lower handicap golfers play to their handicap more often, by stipulating a 20% formula for fields in which at least 80% of the players have handicaps under 13, and conversely a 10% formula when at least 51% of the players have handicaps of 20 or more. However for the vast majority of fields the 15% formula must be used.

Another way in which handicap differences are acknowledged by the AGU is in the following table, which is used for handicap adjustments:

Table 3

Handicap adjustments to CCR

(Above CCR)		(Below CCR)	
Category	Handicap	Buffer	Subtract
A	up to 4	0	0.1
B	5 – 12	1	0.2
C	13 – 19	2	0.3
D	20 – 27	3	0.4
E	over 27	4	0.5

Note that for net scores above the CCR by more than the corresponding buffer, 0.1 is added to the player's handicap, while for net scores below the CCR the amount shown is subtracted for each stroke below.

5. AN ALTERNATIVE CCR FORMULA

The above five categories provide a suitable basis for developing a formula which reflects the handicap profile of the field. The following table is based on the 14 rounds surveyed, and shows the number of cards returned in each category, together with the numbers and frequencies (%) of net scores at least equal to the CCR (designated PTH for "played to handicap"), at least one under the CCR (PTH-1) and at least two under the CCR (PTH-2) respectively.

Table 4

Hdcp Cat Cards	A 150	B 754	C 705	D 480	E 68	All players 2157
PTH Freq	75 43.2	224 25.7	118 14.5	54 9.7	8 10.2	479 19.2
PTH–1 Freq	55 31.7	151 17.3	72 8.8	27 4.9	4 5.1	309 12.4
PTH–2 Freq	38 21.9	101 11.6	46 5.6	17 3.1	1 1.3	203 8.1

Note that in computing the frequencies, a factor of 0.864 was used throughout to correct for non-returned cards (see section 2).

Notice also that although in each of the 14 rounds surveyed the 15th percentile was used to determine the CCR, it turns out that 19.2% of starters actually played to their handicap. If we wanted to produce a handicap-based formula which on average produced 15% we would need to interpolate between the frequencies of PTH and PTH-1. Similarly to get a handicap-based formula which produces 9%, which we accepted earlier as perhaps the best option (see section 3), we need to interpolate between the frequencies of PTH-1 and PTH-2. These frequencies are illustrated in the following graph:

By interpolation the relevant frequencies (rounded to the nearest whole number) turn out to be A:24%, B:13%, C:6%, D:3% and E:2%.

Thus if in a field of N players there are a in category A, b in B, c in C, d in D and e in E the formula

$$P = (24a + 13b + 6c + 3d + 2e) / N \qquad (1)$$

gives the percentile to be used to find the CCR.

6. Testing the Formula

In practice a record would have to be kept of how many starters there are in each handicap category, preferably as players enter the competition because of the prevalence of non-returned cards. To illustrate the effect of formula (1), suppose we go back to the 14 rounds surveyed, treating the cards returned as if they were the whole field (for large fields this method gives about the same result anyway). The following table gives the numbers in each category together with the percentile as calculated from formula (1):

Table 5

Round	A	B	C	D	E	Percentile
1	10	63	44	33	1	9.4
2	14	61	64	37	4	9.1
3	13	67	54	34	3	9.4
4	11	58	52	34	3	9.1
5	12	56	56	32	1	9.2
6	10	61	51	32	4	9.1
7	12	57	53	39	4	8.9
8	13	72	57	39	5	9.2
9	7	26	40	29	10	7.6
10	11	66	57	40	6	8.9
11	3	24	41	32	9	6.8
12	14	62	49	30	5	9.6
13	4	18	32	26	12	6.8
14	15	65	55	40	4	9.3

The three lowest percentiles (rounds 9, 11, 13) coincide with the three Wednesday competitions, when there was a greater proportion of long handicappers in the field. The other eleven rounds were all in Saturday competitions. If we compare these percentiles with Table 2, we can see that the CCR would be lowered by one or two in each case.

7. CONCLUSION

This study used only 14 rounds and one golf course, so should not be taken as an indication of general practice. While it seems to me that the 9th percentile works better at Glenelg than the 15th percentile (see section 3), the same may not apply at other clubs. Similarly the frequencies incorporated in formula (1) need to be refined by a lot more analysis, preferably at a variety of clubs, before being taken too seriously. Also, they are based on the "guess" of 9%, one of several alternatives considered in section 3. Of the two ideas my view at this stage is that getting the correct percentile is more important than seeking a handicap-based formula.

LOW-ANGLE GOLF TRAJECTORIES

Ian Collings[1] and Neville de Mestre[2]

Abstract

The solution for two-dimensional motion of a driven golf ball using the full Runge-Kutta numerical technique is compared with a low-trajectory analytical approximation. The drag and lift coefficients are given in the form Kv^{λ}, with K and λ constants determined from a least-squares analysis of experimental data.

The low-trajectory approximation yields range predictions that are closer to the measured ranges than the Runge-Kutta predictions.

This has applications for simple computer algorithms to generate trajectory characteristics for golf balls driven into nets at driving ranges.

1. INTRODUCTION

The lack of space for golf courses in some countries has led to the construction of multi-storied driving ranges, particularly in Japan and Korea. At these ranges, players drive golf balls into a net which impedes the normal trajectory. There have been attempts to give accurate feedback to players on what would have been the characteristics of each drive, and the feedback required by the average player is the carry of the ball (plus the run) and the amount of deviation to the left or right due to a hook or slice. This information seems to be provided using crude inaccurate algorithms or, in many cases, is not provided at all.

The crudest of these algorithms is based on the club-head speed just before impact with the golf ball. An assumption then has to be made about the coefficient of restitution between the club-head and the ball, from which the initial velocity of the ball from the tee is deduced. One difficulty with this approach is that the contact between the club-head and the ball is assumed to be direct, and hence the amount of hook or slice deviation cannot be calculated.

Other algorithms are being developed based on photo-electric eyes just in front of the tee to give initial velocity values directly (Boland [1]). The advantage of this technique is that three components of the velocity vector can be determinedand used. It would require more sophisticated video equipment to measure the initial spin rates of the ball about horizontal and vertical axes.

It is the aim of this paper to survey experiments associated with the trajectories of golf balls driven from a tee at low angles. These experiments have been carried out in wind tunnels or on the driving range. The shortcomings of all models used so far to

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

estimate the carry will be discussed, and a better model will be presented. This model will be solved numerically to establish trajectory characteristics.

However the computational requirements are relatively heavy and not suitable for calculations at a commercial driving range where almost instant feedback is required.

Only one approximate analytical solution of the differential equations for a golf trajectory has ever been produced (Tait [2], [3], [4]), but it is limited to two-dimensional flight and to initial angles of elevation less than 20°. In this paper an improved low-angle approximation is made to the governing equations of the trajectory, which dramatically reduces the computational time, but still produces a reasonably accurate range. Other characteristics of the trajectory can also be easily determined, leading to a better understanding of the physics of driven golf balls in flight.

2. THE TWO-DIMENSIONAL MODEL

The simplest model for 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 can be considered as two-dimensional. This is only possible if the ball spins about a horizontal axis perpendicular to the initial vertical plane of motion, for then the lift force due to the Magnus effect remains in this plane.

The components of the two-dimensional Principle of Linear Momentum equation are

$$m\ddot{x} = -D\cos\psi - L\sin\psi, \tag{1}$$

$$m\ddot{y} = L\cos\psi - D\sin\psi - mg, \tag{2}$$

where D denotes the drag force, L denotes the lift force, m (= 0.0459kg) denotes the mass of the golf ball, and g (= 9.81ms^{-2}) the acceleration due to gravity (see Figure 1).

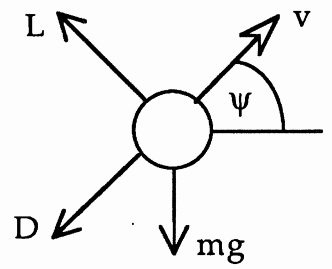


Figure 1: The forces of gravity, drag and lift in relation to the velocity-vector.

Here a dot denotes differentiation with respect to time (t), the 0x axis is horizontal, the 0y axis is vertically upwards. The linear (or translational) speed (v) of the ball makes an angle (ψ) with the horizontal axis as shown. Hence the horizontal and vertical velocity components are given by

$$\dot{x} = v\cos\psi \tag{3}$$

$$\dot{y} = v \sin \psi \tag{4}$$

The initial conditions for the flight of the golf ball are

$$x = 0, y = 0, v = v_0, \psi = \psi_0 \qquad \text{(when } t = 0\text{)} \tag{5}$$

Both Daish [5] and Bearman and Harvey [6] propose that the drag and lift forces be modelled by

$$D = \frac{1}{2} \rho S v^2 C_D, \tag{6}$$

$$L = \frac{1}{2} \rho S v^2 C_L, \tag{7}$$

where C_D and C_L denote the drag coefficient and the lift coefficient respectively, ρ denotes the density of air (1.226kg m^{-3} at sea level), and S(= 0.00143m^{-2}) denotes the cross-sectional area of the golf ball.

The drag and lift coefficients for a golf ball can be shown by dimensional analysis to be functions of the spin parameter (spin rate × radius/linear speed) and the Reynolds number (air density × linear speed × diameter/air viscosity). Thus they change as the linear and rotational speeds of the ball change during flight. A number of experiments have been carried out in wind tunnels to determine their values at different linear and rotational speeds.

Maccoll [7] considered a spinning smooth sphere supported by a spindle. Davies [8] 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 [6] utilized 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 [9] 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 [6], with greater experimental accuracy being claimed.

Bearman and Harvey [6] integrated equations (1) to (7) using a numerical step-by-step process that introduced the appropriate C_D and C_L values from their wind-tunnel data at each step of the computation. One of the difficulties with this approach is that the spin rate is not included in the two-dimensional formulation, and an assumption has to be made about its initial value and how it changes along the trajectory. Daish [5] estimated that a driven golf ball would still be rotating at more than 80% of its initial rotational speed just before it hits the ground. Bearman and Harvey [6] tried various rates of spin decay in their trajectory calculations, and came to the conclusion that these made little change to the ranges predicted. This suggests that a reasonable assumption would be to take the rotational speed of the ball as constant during its flight.

With this assumption, Bearman and Harvey [6] tested their numerical model against the ranges obtained by the Uniroyal Ltd driving machine. Reasonable agreement was obtained, indicating that the model based on equations (1) to (7) is soundly based.

3. Modified Two-Dimensional Computations

To solve equations (1) to (7) numerically requires a Runge-Kutta process which is extremely time consuming on the computer when reasonable accuracy is expected. Adding the further complication of referring to wind-tunnel data for C_D and C_L at each time-step makes the process even more computer-time consuming, and not suitable for driving-range applications.

The latter complication is removed by fitting curves of the form Kv^λ (K, λ constants), using a least-squares method, to the Bearman and Harvey data [6] for a specified spin rate. Since Bearman and Harvey published five calculated range values for five different initial speeds at an angle of elevation 10° and a spin rate 3500 rpm, it was decided to check the modified model against these. The expressions

$$C_D = 0.83\, v^{-0.27}$$
$$C_L = 2.53\, v^{-0.67}$$

were obtained by least-squares for this spin rate. Note that these are different from the values C_D = constant or proportional to v^{-1}, C_L proportional to v^{-1} assumed by many previous investigators (Tait [2], [3] and [4], Davies [8], Hart and Croft [10], Williams [11]). Although Daish [5] suggested a model along the above lines, no values of K or λ were given.

Therefore, for the purposes of this paper, the equations to be solved are (3) and (4) together with

$$\ddot{x} = -0.0158v^{1.73}\cos\psi - 0.0482v^{1.33}\sin\psi \qquad (8)$$

$$\ddot{y} = 0.0482v^{1.33}\cos\psi - 0.0158v^{1.73}\sin\psi - 9.81 \qquad (9)$$

with

$$x = 0,\ y = 0,\ \dot{x} = 0.985v_0,\ \dot{y} = 0.174v_0 \qquad \text{(when } t = 0) \qquad (10)$$

for the five different initial speeds v_0 = 21, 34, 46, 58 and 67 ms^{-1} respectively used by Bearman and Harvey.

Application of the complicated and relatively lengthy Runge-Kutta process can be removed by making a low-angle approximation, since the initial driving angle is small. Tait [4] made the assumption that for low-angle trajectories the equations could be simplified by making $\cos\psi \approx 1$ and $\sin\psi \approx \psi$. One of the main physical problems with Tait's approximation is that the x-component of the ball's position becomes independent of the initial angle of projection. It was pointed out by de Mestre [12] that this unrealistic physical characteristic can be removed, and the errors between the low-trajectory approximation and the Runge-Kutta solution made much smaller, if $\dot{y} << \dot{x}$ is used as the low-trajectory approximation instead. This approximation is true except for approximately the last 10% of the flight.

Therefore, with x replacing $\sqrt{\dot{x}^2 + \dot{y}^2}$ everywhere in equations (8) and (9), the low-trajectory approximations become

$$\ddot{x} = -0.0158(\dot{x})^{1.73} \tag{11}$$

$$\ddot{y} = 0.0482(\dot{x})^{1.33} - 0.0158\dot{y}(\dot{x})^{0.73} - 9.81 \tag{12}$$

These essentially indicate that the $\dot{x}$-component of velocity is affected by the drag, while the $\dot{y}$-component is affected by gravity, drag and lift.

For the initial conditions (10) the solutions of (11) and (12) are

$$x = 233.5\, v_0^{0.27}\left[1 - \left\{1 + 0.0114\, v_0^{0.73}\, t\right\}^{-0.37}\right] \tag{13}$$

$$\begin{aligned} y = {} & 3720\, v_0^{-0.13}\left[\left\{1 + 0.0114\, v_0^{0.73}\, t\right\}^{0.18} - 1\right] \\ & + 15905\, v_0^{-1.46}\left[1 - \left\{1 + 0.0114\, v_0^{0.73}\, t\right\}^{2}\right] \\ & + \left\{41.2\, v_0^{0.27} - 1791\, v_0^{-0.13} + 86006\, v_0^{-1.46}\right\}\left[1 - \left\{1 + 0.0114\, v_0^{0.73}\, t\right\}^{-0.37}\right] \end{aligned} \tag{14}$$

The time of flight is then obtained by determining t from equation (14) with y = 0. Table 1 compares the predicted ranges using Bearman and Harvey's full Runge-Kutta calculations with the predicted ranges using the low-trajectory approximation results given by equations (13) and (14). Five different initial speeds are considered at 10° angle of elevation and 3500 rpm.

Table 1

v_0	Range (m) [Bearman and Harvey]	Range (m) [Equns (13) and (14)]	Time of flight (s) [Equn (14) with y=0]
21	18	21	1.1
34	55	61	2.2
46	104	113	3.6
58	159	167	5.0
67	195	206	6.0

The low-trajectory approximations appear, at first sight, to over-predict the range. However Bearman and Harvey [6] tested their calculations against measured values obtained from a golf-driving machine. They found that the measured values were in general greater than their calculated values for conventional golf balls. This adds credence to the use of the low-trajectory approximations as good estimates of driving-range values.

4. FURTHER INVESTIGATION

Now that the viability and accuracy of the modified two-dimensional low-trajectory approximation has been established, further investigations need to be undertaken in this area before a useful program can be developed to yield driving-range information.

Firstly, the appropriate spin rate w should be deduced for a given initial speed of the ball. It is well-known that the harder the ball is driven from the tee then, the more it spins. Bearman and Harvey [6] show a table indicating that, on average, amateurs drive off at about $56.7 ms^{-1}$ with a spin rate of 2450 rpm, while professionals drive off at $68.1 ms^{-1}$ with a spin rate of 3450 rpm. A simple linear relationship of the form

$$w = 87.7\, v_0 - 2524$$

will be adequate to cover the values given, and any others by interpolation or extrapolation.

Secondly, the problem of determining the amount of hook or slice cannot be resolved until a three-dimensional model is considered. The extended versions of equations (1) and (2) are

$$\begin{aligned} m\ddot{x} &= -D\cos\psi\cos\phi - L\sin\psi\cos\phi + H\sin\phi, \\ m\ddot{y} &= -D\sin\psi + L\cos\psi - mg, \\ m\ddot{z} &= -D\cos\psi\sin\phi - L\sin\psi\sin\phi - H\cos\phi, \end{aligned}$$

where H is the sideways component of the Magnus force, ϕ is the azimuthal angle between the vertical plane containing the lift and drag and the 0xy-plane of initial aiming, and z is the component of drift to the right or left depending on its sign. Now w can be determined from v_0 as just indicated, but separating it into two components that determine L and H is more difficult. The separation values will depend on how much the club face cuts across the ball during impact. Once a suitable physical assumption is made about this, expressions for C_D, C_L and C_H can be obtained and low-trajectory approximations can be made to the governing differential equations. They can then be solved to produce relatively simple closed-form expressions for x, y, z as functions of t. From these the range $\sqrt{x^2 + z^2}$ and the drift z can be calculated for each drive, using y = 0 to obtain the time of flight.

It is hoped to give details of the low-angle three-dimensional analysis in a following paper.

REFERENCES

[1] Boland D., Private communication.

[2] Tait P.G., Some points in the physics of golf. Part I, *Nature*, 42, 420–423 (1890).

[3] Tait P.G., Some points in the physics of golf. Part II, *Nature*, 44, 497–498 (1891).

[4] Tait P.G., Some points in the physics of golf. Part III, *Nature*, 48, 202–205 (1893).

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

[6] Bearman P.W. and Harvey J.K., Golf ball aerodynamics, *Aeronaut. Quart.*, 27, 112–122 (1976).

[7] Maccoll J., Aerodynamics of a spinning sphere, *J. Roy. Aeronaut. Soc.*, 32, 777–798 (1928).

[8] Davies J.M., The aerodynamics of golf balls, *J. Appl. Phys.*, 20, 821–828 (1949).

[9] Aoyama S., A modern method for the measurement of aerodynamic lift and drag on golf balls, in Science and Golf, Proc. of the First World Scientific Congress of Golf (ed. A. Cochran), (E.& F.N. Spon, London, 1990).

[10] Hart D. and Croft T., Modelling with projectiles, (Ellis-Horwood, Chichester, U.K., 1988).

[11] Williams D., Drag force on a golf ball in flight and its practical significance, *Quart. J. Mech. Appl. Math.*, 12, 387–392 (1959).

[12] de Mestre N., Inflight golf drive analysis, Report 3-099B, (School of Information Technology, Bond University, 1993).