Proceedings of the
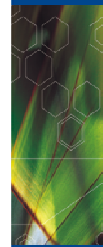## Seventh Australasian Conference

on

MATHEMATICS & COMPUTERS
IN SPORT

7M&CS

Edited by:
R Hugh Morton & S Ganesalingam

30 August – 1 September 2004

Massey University

Te Kunenga
ki Pūrehuroa

Proceeding of the Seventh Australasian Conference

on

MATHEMATICS AND COMPUTERS IN SPORT

Massey University
Palmerston North
New Zealand

Edited by: R Hugh Morton
              Institute of Food, Nutrition & Human Health
              &
              S Ganesalingam
              Institute of Information Science and Technology

7M&CS

30 August – 1 September 2004

Proceedings of the Seventh Australasian Conference
on
Mathematics and Computers in Sport

CONTENTS

Invited Speakers:

Contributed Papers:

Conference Director's Report

Welcome to the Seventh Australasian Conference on Mathematics and Computers in Sport (7M&CS). For the first time the conference is hosted in New Zealand, and I am particularly pleased to welcome you to this country, to Palmerston North, and to Massey University in particular. Our venue on the Massey Campus is the attractive NZ Institute of Rugby, established in 1999. It is a most suitable venue for a small intimate conference like ours. At these conferences, it is always a pleasure to renew contacts with colleagues who have attended and contributed to our conferences in the past, but as always there will be some new faces. It is great having you here too, and we look forward to seeing you again at our future gatherings. Although most delegates are from Australia and New Zealand, I welcome also those from the USA, UK and elsewhere. In the local dialect "Kia Ora".

Being hosted in New Zealand and at the Institute of Rugby, many of our papers have a rugby theme, with computer analyses of performance and results being a main strand. Professor Mike Hughes, Director of the Centre for Performance Analysis, University of Wales Institute, Cardiff, is a leading exponent of notational analysis and as our overseas keynote speaker, will share some of his knowledge with us. He is in frequent demand as a conference speaker around the world, and we are fortunate to have him visit. Our local keynote speaker is Professor Keith Davids, Dean of the School of Physical Education at the University of Otago, Dunedin. Keith is an author of a number of books on visual perception and action in sport, and the performance of interceptive actions and motor development. He will explain to us the role that dynamical systems plays in analysing and understanding movement systems. In addition there are several papers on both cricket and tennis, together with a range of other most interesting applications of mathematics, statistics and computing in sport.

As Conference Director there are many people who have assisted me along the way; in particular Grant Musgrave, the webmaster at http://7mcs.massey.ac.nz, and Christine Ramsay who managed the finances. There are also colleagues and friends at Massey University, too numerous to mention. To you all, a big thankyou. All papers in these *Proceedings* have been peer refereed, and a big thankyou to all those who freely gave their services.

I sincerely hope you enjoy attending this conference, spending time with like-minded colleagues, listening to their presentations, and reading these *Proceedings*.

Until 8M&CS…………..

R Hugh Morton
August 2004.

Programme

Sunday 29 August 2004

3:00-6:00 pm	Registration and Coffee
7:00 pm	Dinner at *"The Bath House"*, Broadway Avenue

Monday 30 August 2004

8:30-9:00 am	Registration and Coffee	
9:00-9:05 am	Conference Organiser	Hugh Morton
9:05-9:10 am	Opening Welcome	Vice-Chancellor Prof Judith Kinnear
9.10-9:30 am	Opening Address	Farah Palmer, Captain, NZ Black Ferns
9:30-10:30 am	Invited Lecture	Keith Davids and Chris Button School of Physical Education Otago University: "Variability and constraints in dynamical movement systems"
10:30-11:00 am	Morning Tea	
11:00-12:30 pm	Contributed Papers (3)	Norton: "Is there a difference in the predictability of men's and woman's basketball matches?" Newton & Pollard: "Service neutral scoring strategies in tennis" Pollard & Noble: "Some attractive properties of the 16-point tiebreak game in tennis"
12:30-2:00 pm	Lunch	
2:00-3:30 pm	Contributed Papers (4)	Pollard & Noble: "The benefits of a new game scoring system in tennis: the 50-40 game" Pollard & Noble: "The effect of having correlated point outcomes in tennis" Pollard: "Can a tennis player increase the probability of winning a point when it is more important" Barnett et al: "Optimal use of tennis resources"
3:30-4:00 pm	Afternoon Tea	

4:00-5:00 pm	Contributed Papers (2)	Croucher: "A mathematical view of sports betting" Bailey & Clarke: "Deriving a profit from Australian Rules football: a statistical approach"
7:00 pm	Conference Dinner	*"Ambitions",* Queen Street.

Tuesday 31 August 2004

9:00-10:00 am	Invited Lecture	Mike Hughes Centre for Performance Analysis University of Wales: "Notational analysis - a mathematical perspective"
10:00-10:30 am	Contributed Papers (1)	Serralach & Martin: "The fallacy of the use of pure statistics in sport"
10:30-11:00 am	Morning Tea	
11:00-12:30 pm	Contributed Papers (3)	Bracewell: "Extraction, interpretation and utilisation of meaningful information about individual rugby performance using data mining" Hammond: "Using performance analysis technology to evaluate the instructional process in sport" Green: "Augmented reality user interfaces for biomechanical data overlays"
12:30-2:00 pm	Lunch	
2:00-3:30 pm	Contributed Papers (3)	Forbes & Clarke: "A seven-state markov process for modelling Australian Rules football" Bedford: "Predicting womens world cup handball match outcomes using optimised rating models" Bedford & Da Costa: "A ratings based analysis of Oceania's road to the World Cup"
3:30-4:00 pm	Afternoon Tea	

4:00-5:30 pm	Contributed Papers (3)	Yelas & Clarke: "Forecasting the 2003 Rugby World Cup" Hoare: "An investigation of rugby test results" Cameron: "Using performance or results to measure the quality of sports competitors"
7:00 pm	Dinner	*"The Elm",* Fitzherbert Avenue

Wednesday 1 September

8:30-9:00 am	Business Meeting: 7M&CS	
9:00-10:30 am	Contributed Papers (3)	Harman & de Mestre: "The mechanics and perceptions of judging an outfield catch" Norman & Clarke: "Dynamic programming in cricket: batting on a sticky wicket" Cohen: "A new statistic in cricket - the slog factor"
10:30-11:00 am	Morning Tea	
11:00-12:30 pm	Contributed Papers (3)	Lewis: "Steps towards fairer one-day cricketing measures of performance" Ovens: "If it rains, do you still have a sporting chance?" Gurram & Narayanan: "Comparison of the methods to reset targets for interrupted one-day cricket matches"
12:30-2:00 pm	Lunch	
2:00-3:30 pm	Contributed Papers (3)	Morton: "On optimal race pace" Cogill: "The mathematics of bicycling part 3: the somersault" de Mestre: "Bouncing a water polo ball"
3:30 pm -	Conference Close and Afternoon Tea	

Participants

Mr Michael Bailey	mbailey@swin.edu.au
Mr Tristan Barnett	tbarnett@swin.edu.au
Dr Anthony Bedford	anthony.bedford@ems.rmit.edu.au
Mrs Penny Bilton	pennybilton@xtra.co.nz
Dr Paul Bracewell	paul@offlode.com
Mr Michael Cameron	mpc9@waikato.ac.nz
Prof Stephen Clarke	sclarke@swin.edu.au
Dr John Cogill	j.cogill@unsw.edu.au
Dr Graeme Cohen	graeme.cohen@uts.edu.au
Prof John Croucher	jcrouche@gsm.mq.edu.au
Miss Monica Daley	mdaley@dsr.nsw.gov.au
Prof Keith Davids	kdavids@pooka.otago.ac.nz
Prof Neville de Mestre	neville_de_mestre@bond.edu.au
Mr Don Forbes	dgforbes@swin.edu.au
Dr Ganes Ganesalingam	S.Ganesalingam@massey.ac.nz
Mr Graeme Gee	graeme@offlode.com
Dr Richard Green	richard.green@canterbury.ac.nz
Ms Manohita Gurram	manohita@yahoo.co.in
Dr John Hammond	jhammond@scu.edu.au
Assoc Prof Chris Harman	harman@usq.edu.au
Dr Ray Hoare	ray@hrs.co.nz
Prof Mike Hughes	MHughes@uwic.ac.uk
Mr Tony Lewis	ajlewis@brookes.ac.uk
Dr Andy Martin	a.j.martin@massey.ac.nz
Mr Doug McClymont	dougmcclymont@cce.ac.nz
Assoc Prof Hugh Morton	H.Morton@massey.ac.nz
Dr Sundaram Narayanan	snarayanan@cs.wright.edu
Prof John Norman	j.norman@sheffield.ac.uk
Dr Pam Norton	pam.norton@sci.monash.edu.au
Mr Darren O'Shaughnessy	darren@championdata.com.au
Mr Matthew Ovens	matthew.ovens@sci.monash.edu.au
Prof Graham Pollard	gpollard@austarmetro.com.au

VARIABILITY AND CONSTRAINTS IN DYNAMICAL MOVEMENT SYSTEMS

Keith Davids and Chris Button
School of Physical Education, University of Otago,
46 Union St West, Dunedin, New Zealand

Abstract

Dynamical systems theory was developed by the French mathematician, Henri Poincare at the beginning of the last century. It has been successfully applied to the study of a range of natural phenomena including biological movement systems. This paper examines its applications to the study of movement behaviour in sport, particularly understanding the functional role that variability could play in helping performers to come up with different coordination solutions to adapt actions in complex, dynamic environments. It is concluded that sport provides a range of useful movement models with which to study the behaviour of dynamical systems, adequately capturing the rich range of interacting constraints that shape the emergence of successful movement solutions at different timescales of analysis.

1. Introduction

Increasingly a rich range of sports and physical activities are providing the experimental models to reveal theoretical insights into how processes of perception and action subserve movement coordination and control [1]. Traditionally, selection of models to investigate motor learning and control has been biased away from dynamic, whole-body, multijoint actions prevalent in sports and physical activities because of the prevalent belief that experimental rigor could be better maintained in the laboratory, compared to studying performance of natural tasks during functional behaviour. How can we make sense of such a fundamental and remarkable transition in the choice of movement models favoured by many movement scientists and psychologists? The distinction between experimental rigor and ecological validity is now being recognized as a false dichotomy, mainly due to the influence of a relatively new theoretical paradigm on movement coordination, dominated by ecological psychology [2] and dynamical systems theory [3]. Consequently, insights into movement behaviour from the perspective of dynamical systems theory are being provided by experimental models from activities such as playground swinging [4], pole-balancing [5], hula-hooping [6], sit ups [7], javelin and discus throwing [8], rowing [9,10], sailing[11], pistol shooting[12], ball hitting [13], basketball ball bouncing [14] and golf driving [15].

Undoubtedly, technological advances have also impacted on the choice of movement models for studying coordination and control, but in this paper we identify that a particularly prominent driving force for this transition has been the emergence of a conceptual framework provided by dynamical systems theory. From this conceptual framework, sports tasks and physical activities represent relevant movement models for theoretical elaboration because they exemplify the rich range of interacting constraints that movement systems need to satisfy during goal-directed behaviour. In this paper we introduce key ideas from dynamical systems theory and outline how nested interacting constraints shape coordination and control of multijoint actions and their acquisition. To achieve our aim we selectively focus on research in basketball

and cricket to exemplify how sport is replete with relevant movement models such as catching, batting, throwing and running towards targets in space.

2. Dynamical Systems Theory

Prominent ideas from dynamical systems theory and ecological psychology have led to a focus on how movements are coordinated with respect to complex and dynamical environments. Research has adopted a systems perspective and sought to characterize biological movement systems as complex, dynamical systems, revealing how the many interacting parts, or degrees of freedom, are coordinated and controlled during goal-directed movements [16]. It is well established that patterns emerge between parts of dynamical movement systems through processes of self-organization ubiquitous to physical and biological systems in nature [17]. Dynamical systems are able to exploit surrounding constraints to allow functional, self-sustaining patterns of behavior to emerge in specific contexts. Interest has focused on the phase transitions (movements of the micro components of a system into a different state of organization) that can emerge spontaneously in natural dynamical systems because they are relevant for understanding transitions between patterns of coordination in biological movement systems. The type of order that emerges is dependent on initial conditions (existing environmental conditions) and the constraints that shape a system's behavior. Dynamical systems theory suggests that the existence of a common optimal motor pattern, presented as an idealized image of a motor skill to which all learners should aspire, is a fallacy owing to the significant amount of variability often observed in human motor performance [18]. Movement variability has traditionally been viewed as dysfunctional and a reflection of 'noise' in the central nervous system. Dynamical systems theorists, however, suggest that movement variability is an intrinsic feature of skilled motor performance as it provides the flexibility required to adapt to complex, dynamic sport environments [17]. This notion has been supported by studies on national and international javelin and discus throwers [8], which reported more variability in the groups of international athletes, particularly during the last 200 ms before release. Variability in movement patterns permits flexible and adaptive motor system behaviour, encouraging free exploration of performance contexts. This paradox between stability and variability explains why skilled athletes are capable of both persistence and change in motor output during sport performance [17]. The radical implication of this theoretical approach is that a performer's ideas, perceptions, memories, plans or actions may be conceived of as emergent, self-organising, macroscopic patterns formed by the interaction of the molecular constituents of the neuro-musculoskeletal system [19].

2.1. *The Role of Task Constraints*

Research has shown that very simple changes in task constraints can provide powerful insights into the adaptive coordinative structures that emerge as individuals find functional coordination solutions. Sports and physical activities abound with a variety of unique task constraints, and it is for this reason that these movement models are revealing rich insights into the nature of movement variability and its functional role in achieving movement outcomes [20]. Analysis of coordination solutions after manipulating task constraints could provide particularly useful windows on the nature of variability and the role of specific intentions and perceptual information sources in constraining movements.

Recent studies of activities common to sport have exemplified how a process-oriented, time-continuous approach can be successfully transferred to the study of more complex movement models [7,8]. One of the major reasons for this cross over has been the theoretical emphasis on how coordination emerges under interacting personal and task constraints. Sports and physical activities are replete with movement models accentuating an extensive range of task constraints including clothing, equipment, rules, targets, boundaries, surfaces and presence of other individuals, in which interaction with neuroanatomical, intentional and environmental constraints shape the emergence of coordination and control [20]. Coordination has been defined as a relation captured in a perceptual-motor workspace formed across the individual and environmental backdrop of task constraints [21] and for this reason there is a clear need for more work investigating coordination in a diverse range of movement models emphasizing interaction with different task constraints. Compelling arguments have been proposed for empirical work to adequately reflect 'enriched action environments' based on a broad range of task constraints, to ensure that movement model selection does not occur to confirm *a priori* theorizing [8,21]. To exemplify our arguments, in the following section we show how the sports of basketball and cricket provide a rich backdrop for the study of coordination and control processes in many different movement models.

3. Movement Models in Basketball and Cricket: Windows on coordination and control processes

3.1 Basketball

As discussed earlier, spontaneous movement variability can play an important role in allowing performers to create different solutions to fit different situations. In a dynamic sport such as basketball, one can get a good impression of how skilled players utilise a range of coordination patterns to achieve individual and team goals. For example, different types of shooting techniques (e.g., jump shot, free-throw, lay-up and dunk) may emerge in a game as appropriate solutions depending on the player's position on the court and the relative movement of team mates and opponents at that point in time. Furthermore, it has been demonstrated that emergent processes can also characterise decision-making performance in basketball [22]. These researchers modelled a nonlinear system based on the relative positioning of an attacker with the ball and a marking defender near the basket. In basketball, the aim of the attacker is to 'destroy' the stability of this system. When a defender matches the movements of an opponent and remains in position between the attacker and the basket, the form or symmetry of the system remains stable. When an attacking player dribbles past an opponent, near the basket, a break in the symmetry of the system is created. For skilled players the preferred way to break system symmetry may be to make an unexpected movement or feign to move one way but in fact to move in a different direction. Even though the defender may try to stay between an attacker and the basket, principal movements will occur after the attacker has dribbled with the ball towards the basket. The constant fluctuations and nonlinear perturbations between the positioning of opposing players is a characteristic of dribbling tasks in many team ball games, and can be understood as a type of interpersonal coordination [22].

An interesting question also concerns how the amount and structure of movement variability changes as a person becomes increasingly skilled at a task. The free-throw shot in basketball is a precision throwing task that is seemingly associated with a high degree of movement consistency amongst experienced performers. To test this

observation, Button et al. [23] conducted a detailed, individual analysis of the throwing kinematics of 6 female players performing 30 free-throw shots. The players comprised a range of expertise levels from novice (Participant 1) to international (Participant 6). The data revealed an increasing amount of intertrial consistency from the elbow and wrist joints as skill level improved (see Figure 1). One might associate such characteristics of the data with a stable attractor shell that enhances the reproducibility of the movement system under high levels of pressure and fatigue, which are important demands in competitive basketball. However, particularly at ball release, there was evidence of angular joint covariation to adapt to subtle changes in key release parameters of the ball. Regardless of expertise level, all players showed an increase in movement variability indicating that this behavioural feature was a function of the task constraints rather than skill level. It could be that to satisfy the unique task constraints of the basketball free-throw, a bandwidth of trajectory variability is necessary that all the players tended to exploit. For example, the elbow position tended to be most variable during the middle phase of the throwing action with a decrease towards ball release. It has previously been shown that increased variability is associated with rapid, accelerative phases of a movement and projecting the ball in an arc towards the basket demands a forceful extension of the shooting arm. Interestingly, Button et al. [23] found that the co-ordination between the elbow and wrist joints became more variable towards the end of the action. Even participant 6 showed a distinct peak in variability at approximately 90% of the movement duration. One explanation may be that as a result of changes in key release parameters, players need to maintain a functional level of joint-space adaptation towards the end of the action. For example, it has been observed that skilled throwers compensate the values of release velocity and release angle against each other in order to achieve a consistent outcome. An implication of this work is that whilst sports performers may appear to use increasingly stereotypical movements with expertise, functional variability in trajectory space can still be exploited on a trial-to-trial basis.

Figure 1: Elbow-wrist angle-angle plots for 6 individual basketball players. NB: Participants 1 & 2 were Novices; Participants 3 & 4 were Club-level; Participants 5 & 6 were International players.

Participant 1

Wrist angle (degrees)

Participant 2

Participant 3

Participant 4

Elbow angle (degrees)

Participant 5

Participant 6
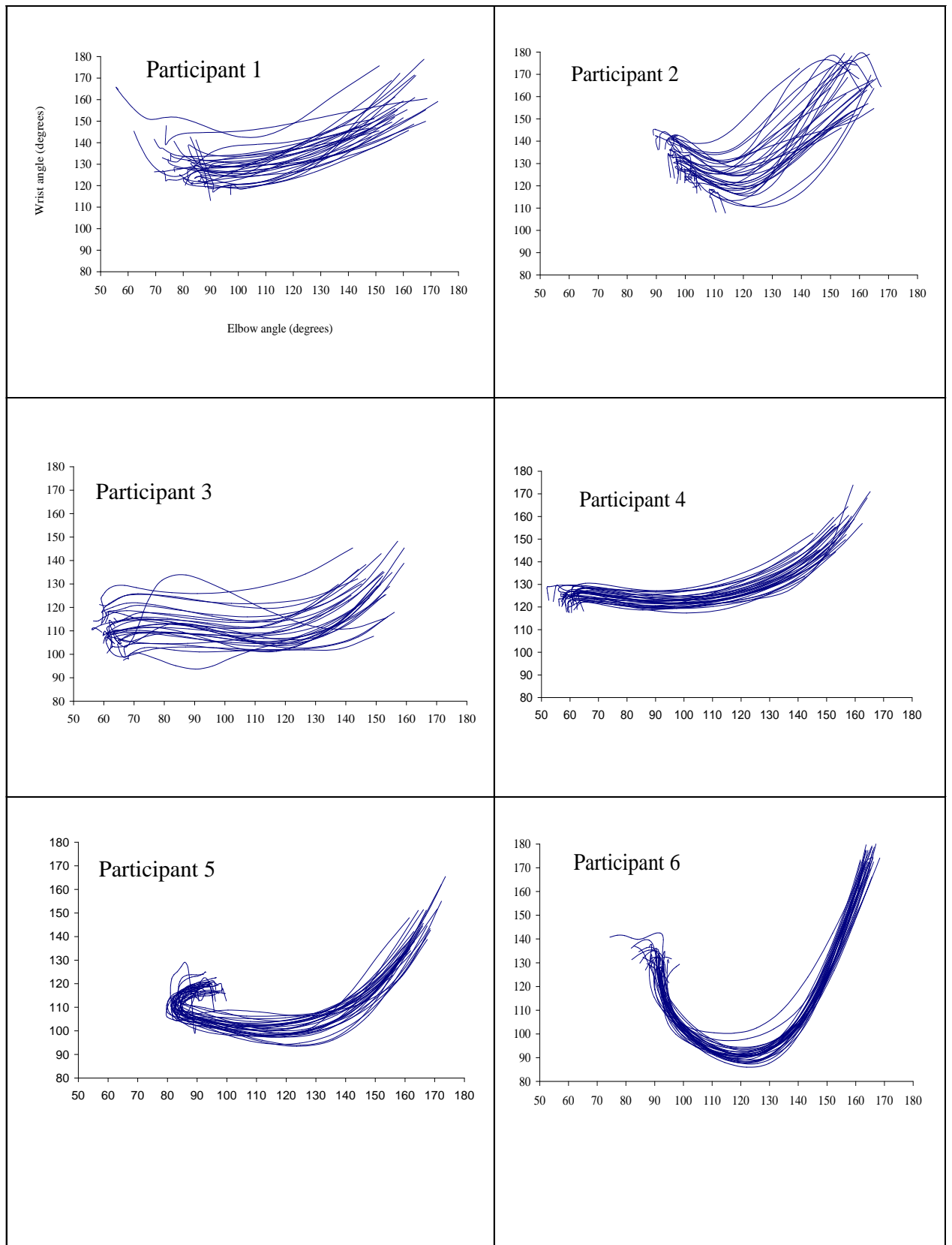
3.2 Dynamic Interceptive Actions in Cricket

Cricketers require skill in a variety of complex, multi-joint, interceptive actions related to batting, bowling and fielding. Interceptive actions involve controlled collisions during which perceptual information from the environment is picked up to guide the appropriate limb(s), or a striking implement, into the right place at the right time, while imparting an appropriate amount of force into a projectile. The task constraints of cricket are demanding, for example, batting requires the interception of a ball with a bat under severe time constraint (typically the time from ball delivery to contact ranges from around 1 second to 0.6 s). Skilled cricketers, facing fast bowling velocities of 160 km.h^{-1}, need to be able to perceptually discriminate the spatial trajectories in depth of balls to a precision of 0.5 degrees. Response timing precision in cricket batting has estimated margins of failure of around ± 2.5 milliseconds at the point of movement execution. Evidence for the role of perceptual variables in constraining interceptive actions has emerged in research on cricket batting and bowling. In this section we discuss recent research on the emergence of variable but functional coordination patterns under different task constraints. We show how slight variations in task constraints of performance lead to significant, functional adaptations by cricketers.

3.3 *Coordination modes emerge under practice task constraints*
Two-handed tasks such as slip-fielding and batting are often practiced with the use of ball projection machines and research from cricket batting is beginning to raise important questions on their role in fine-tuning processes of perception and action. Successful perception of relevant (i.e. constraining) information sources is dependent on the ecological constraints of performance that can deeply influence the movement behaviours that emerge [24]. Practice involves educating learners to pick up constraining perceptual variables rather than non-constraining (less relevant) variables in specific and relevant practice contexts. However, practice environments have traditionally been adapted to manage the information load on learners by decomposition of the movement model into micro-components of actions. This type of management strategy is prevalent in educational, training and practice contexts and can be achieved in many different ways. For example, in cricket, bowling machines allow accurate and stable projection of balls which purports to aid practice of batting skills in isolation from game contexts.

The problem is that it is well established that experienced performers use pre-ball flight information to constrain coordination modes, as revealed by studies of cricket batting [e.g., 25]. Furthermore, batting against a bowling machine induces highly specific coordination modes which are not similar to the movement patterns required when facing bowlers. Specificity of coordination was revealed by analyses of the forward defensive stroke when batting against a real bowler and a bowling machine. The forward defensive stroke has been broken down into two phases: (i) the back lift and stance; and (ii), the downswing to impact. To understand if the different task constraints of batting against a bowling machine or a bowler changes the nature of timing and coordination of the forward defensive, temporal organization of the shot played by English batsmen of high intermediate standard was examined from the moment of ball release (from the machine projection mouth or the bowler's hand) up to the point of ball-bat contact (velocity 26.76 m.s^{-1} under both conditions). Data generally showed significant differences in coordination and timing under these different ecological constraints.

Against the bowling machine, batters coupled the backswing to the moment of ball release (0.02±0.10 s), whereas against the bowler the backswing started later (0.12±0.04 s). Initiation of the front foot movement occurred later (0.16±0.04 s after ball release) in the bowling machine condition compared to facing the bowler (0.14±.0.03 s after ball release). Initiation of the downswing commenced earlier when facing the machine than when facing the bowler (0.32±0.04 s compared with 0.41±.03 s). There was a different ratio of backswing–downswing when batting against the projection machine (46%:54%) compared to the bowler (56%-44%). Timing of the placement of the front foot was delayed against the machine compared to the bowler (0.53±0.05 s compared with 0.55±.0.05 s). Peak bat height differed under the two constraints (bowling machine: 1.56±19.89 m and bowler: 1.72±10.36 m). Mean length of front foot stride was shorter against the machine (0.55±0.07 m) compared to the bowler (0.59±0.06 m). Correlation between initiation of backswing and front foot movement was much higher against the bowler (r=0.88) than the bowling machine (r=0.65). In summary, two different coordination patterns emerged under the different task constraints and performers were practicing two different actions. Practice constraints involving projectile machines should be restricted to learners needing to assemble basic, stable coordination patterns, or when coordination needs to be re-stabilised after absence due to illness or injury.

3.4 Cricket Bowling Run-Ups

As well as fielding and batting, many cricketers need to bowl the ball. Data shows that variability is an inherent characteristic of the run-up patterns of skilled cricketers. Cricket bowling requires the player to run towards a target area (the popping crease) to bowl a ball, placing the back foot before or across the front line of the popping crease to avoid bowling a 'no-ball'. Run-ups in sports like cricket, athletic jumping and gymnastics are examples of locomotor pointing or running to place the foot on a target in space, a movement model which has generated a lot of interest in the study of perception and action in goal-directed gait. Interesting questions concern the control mechanisms and the information sources used to regulate gait in the run-up phase of locomotor pointing tasks. de Rugy et al. [26] proposed a prospective control model to explain visually driven adaptations of basic locomotion and locomotor pointing performance. It was argued that, if information on current and required behaviour were optically available, then regulation of gait might be continuously based on the perception of the difference between them. To support their argument, a dynamical systems model of movement coordination was designed to show how a nervous system central pattern generator (CPG) generates stable gait patterns, which are instantaneously adapted during approach to a target through the modulation of a key gait parameter, step length. In the model an optic variable known as time-to-foot (TTF) was proposed to specify the time remaining before the current eye-foot axis of an approaching runner intersected a target in space. The variable was proposed to combine closely with a parameter representing the gain in the hip flexors-ankle extensors muscle complex, a neuroanatomical relationship that instantaneously provides the changes required during locomotor pointing. Computer simulation tests of the model showed that: (i) there was a marked decrease in toe-target distance variability occurring in the last few steps of a locomotor pointing task; and (ii), that the greater the adjustment needed during approach to place the foot on a target, the earlier visual regulation began. The conclusion was that a stable coordination pattern provided the basic stability needed for locomotion towards a target and that the gait parameter step length was used to adapt this pattern to variability in performance contexts through a continuous perception-action coupling strategy.

Behavioural support for the prospective control model of locomotor pointing has been obtained and a number of task vehicles in sports and physical activity readily lend themselves to testing its assumptions, including approach running in the athletic jumps, javelin throwing and gymnastic vaulting. For example, a study of the long jump run up found that locomotor pointing was a direct function of the optical flow generated by the performer and that the onset of stride length adjustment was a function of the amount of adjustment required [27]. Perhaps the major source of constraint in natural locomotor pointing tasks concerns the influence of the nested actions at the end of an approach run. That is, what athletes need to do after the run-up constrains the precise gait pattern adopted during the run-up. Some locomotor pointing tasks, like the horizontal jumps and gymnastics vaulting, require the generation of maximum velocity during the run up to hit the take-off board, others require a more controlled collision with a target area because of the need for further complex actions nested on the approach phase. Good examples of complex, nested task constraints in locomotor pointing include the javelin throw and cricket bowling, which involve a run up to a target area followed by re-orientation of the body into a new projectile delivery position. Under these task constraints, the earlier initiation of visual regulation is an advantage because it is necessary for adjustments to be spread evenly over more strides, causing less disruption to nested actions by horizontal velocity of the approach phase. Cricket ball deliveries are made up of 4 phases; run-up, bound, delivery stride and follow through. The aim of the run-up is to enable the bowler to move into the bound phase maintaining the velocity gained and positioning the body effectively for a successful 'link' to the delivery stride to be made. During the bound phase the bowler aims to jump forward and high enough to enable him or her to land in the correct position for the delivery stride to release the ball with the desired velocity. Clearly, the horizontal velocity generated in negotiating the bound phase needs to be managed carefully. The ideal final front foot placement in the delivery stride is one that cuts the popping crease, enabling the ball to be delivered as close as possible to the batter, reducing his or her potential response time. Failure to land with the heel behind the popping crease results in a "no-ball" being called by the umpire, with one run being added to the opponents' total score.

Analysis of run-ups of professional cricketers revealed support for the continuous perception-action coupling locomotor control model proposed by the simulations of de Rugy et al. [26]. Analyses of the run-ups of cricket bowlers showed that, due to the specific constraints of cricket bowling, the majority of the bowlers made adjustments early in the run-up, before making late adjustments just prior to the bound stride [28]. This aspect of the task was emphasised in the data by the high number of visually regulated run-ups in our study compared with those of long jumpers. Almost all of the run-ups of the cricket bowlers were regulated at some stage (91 out of 92) and these regulations were spread over the whole length of the run-ups.

Montagne et al. [27] showed that the amount of adjustment produced was a function of the point at which regulation was initiated. A linear relationship between stride number and amount of adjustment putatively showed that perception and action were closely coupled. In the study of cricket bowlers, however, few correlations were found between stride number and amount of adjustment. This is an interesting observation given that the inconsistent starting points of the bowlers, and initial high levels of variability, did not prevent them from achieving remarkably low levels of variability at the bound stride that are consistent with findings in previous studies of long

jumpers. To achieve such functional levels of footfall variability at the critical bound stride, the cricket bowlers were making adjustments based on need at a very early stage of the run-up, a finding in line with a key premise of de Rugy et al.'s model [26]: that regulation is continuous and based on perception of current and required behaviour. If a bowler can adjust a stride in relation to need, as predicted by the model, then there would be little requirement to continue making adjustments to the end of the run-up. Data show that the task constraints of cricket bowling benefited from a greater amount of adaptive visual control during the run-up, compared to the velocity-generation constraint which dominates the athletic jumps. When the inter-trial plots were followed with an intra-trial analysis, significant relationships were observed between amount of adjustment produced and the amount of adjustment needed at steps throughout the run-up.

The data illustrate nicely how nested task constraints shape the nature of control strategies implemented during locomotor pointing tasks, since the cricket bowlers began visually regulating very early in the run-up. Bowlers were making significant adjustments at steps over 20 m from the popping crease. Speed-accuracy trade-offs required that, in order to successfully arrive at the bound stride with the feet correctly oriented and at the optimum run-up velocity, it was essential that bowlers made on-line adjustments at early stages of their run-ups, at all stages throughout the run-up, as and when they were needed. Often the bowlers showed a period of no or very low levels of regulation after the initial adjustments have been made. This observation suggests that bowlers achieved stride length consistency that is within their perceptual tolerance limits enabling them to produce a stable 'rhythmic' approach. However, they were also capable of making late refinements in order to make very consistent take-off points at the bound step.

Interesting questions concern the informational constraints that are used to regulate gait during the cricket bowler's run up. Given that visual regulation seems to occur very early in the bowler's run up, compared to data revealed by analyses of the run up of long jumpers, it is possible that cricketers are able to use the position of the umpire to regulate the approach run from a distance of 20 m. Analysis of the run up of a professional bowler's run ups under three different visual informational constraints (normal match conditions, stumps only, and crease markings only) revealed differences in the ability to place the foot safely within the popping crease to prevent no-balls. The distance of each footfall placement from the crease was calculated, with standard deviation of each footfall around the mean distribution taken as a measure of inter-trial variability. Approximately similar levels of variability at front-foot contact (sd= 0.08 m, 0.08 m, 0.04 m) can be observed in all conditions, however the manipulation of information resulted in different patterns of footfall during the run-up and different mean distances of final footfall from the crease line (0.35 m, 0.21 m, 0.04 m) (see Figure 2). Without the presence of an umpire or stumps at the popping crease, the bowler seemed to regulate gait with reference to the popping crease line and the probability of no-balling was extensively increased. These findings, that need to be verified with group analyses, suggest that the presence of the umpire and/or stumps represents important informational constraints necessary for accurate footfall placement during the cricket bowler's run up.

Figure 2: Mean standard deviations of heel to popping crease distance in the run ups of a professional bowler under three different informational constraints. N.B.: Data were plotted over 12 trials in each condition.
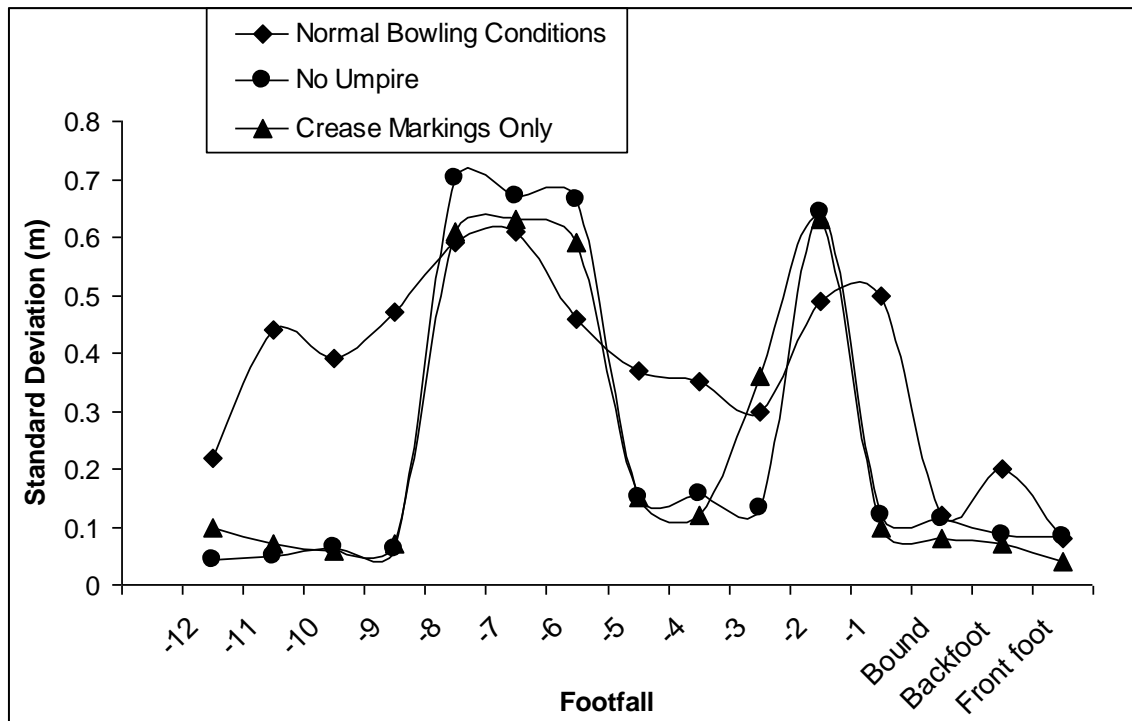
4. Conclusions and Implications for Future Research

Movement models from sports and physical activities provide a viable option to avoid reductionism in adopting a systemic perspective in studying processes of coordination and control. In this paper we outlined how movement models from the sport of cricket had shown that human movements need to be understood in relation to the interacting constraints of structural and functional neuroanatomical design, specific task goals and environmental contexts. We argued that the transition of complex task vehicles to the forefront of research on human movement behavior has coincided with advances in theorizing from a dynamical systems viewpoint. The emphasis on studying multijoint coordination patterns in movement systems as a function of specific task constraints has led movement scientists to investigate coordination processes in natural tasks in order to understand functional, emergent patterns of behaviour. Movement models from sports provide excellent vehicles for studying coordination within and between different sub-systems of human movement systems as well as coordination between individuals and the rich range of task constraints found in the environment [3]. This emphasis on how individual constraints interact with task constraints is signalling a fresh perspective on the role of variability in facilitating adaptation to dynamic task environments. From this standpoint, variability emerges under constraints, highlighting how athletes adapt movements to dynamic environments. Analysis of behaviour at the level of the performer-environment system is revealing coordination modes as an example of a natural, emergent phenomenon which can be functionally varied to suit the challenge of performing in dynamic contexts.

10

References

[1]   Davids, K., Button, C. & Bennett, S.J., "Acquiring Skill in Sport: A Constraints-Led Perspective", Human Kinetics, Champaign, Ill.(2005).

[2]   Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin, Boston,   (1979).

[3]   Jirsa, V.K. & Kelso, J.A.S. (Eds.), Coordination Dynamics: Issues and Trends, N.Y., Springer-Verlag, (2004).

[4]   Post, A.A., Peper, C.E. & Beek, P.J., "Effects of visual information and task constraints on intersegmental coordination in playground swinging." J. Mot. Behav., **35**, (2003), 64-78.

[5]   DeGuzman, G. C., "Using visual information in functional stabilization: Pole balancing example.", in V.K. Jirsa, & J.A.S. Kelso (Eds.), Coordination Dynamics: Issues and Trends, N.Y., Springer-Verlag, (2004), 91-101.

[6]   Balasubramaniam, R. & Turvey, M.T., "Coordination modes in the multisegmental dynamics of hula hooping." Biol. Cyb., **90**, (2004), 176-190.

[7]   Cordo, P.J. & Gurfinkel, V.S., "Motor coordination can be fully understood only by studying complex movements." Prog. Brain Res., **143**, (2004), 29-38.

[8]   Schöllhorn, W., "Coordination dynamics and its consequences on sports.", Int. J. Comp. Sci. Sp., **2**, (2003), 40-46.

[9]   Daffertshofer, A., Huys, R. & Beek, P.J., "Dynamical coupling between locomotion and respiration." Biol. Cyb., **90**, (2004), 157-164.

[10]   Shuttleworth, R., "Differential effects of varying attentional-focus through instruction in acquisition of movement coordination in rowing.", Communication to 31st Austr. Exp. Psychol. Conf., Dunedin, New Zealand, (2004), April 11-14.

[11]   Araújo, D., Davids, K., Serpa, S., "An ecological approach to expertise effects in decision-making in sailing.", Psych. Sp. Ex.: Under Review, (2004).

[12]   Scholz, J.P., Schoner, G. & Latash, M.L., "Identifying the control structure of multijoint coordination during shooting.", Exp. Brain Res., **135**, (2000), 382-404.

[13]   Scott, M.A., van der Kamp, J., Savelsbergh, G.J.P., Oudejans, R.R.D. & Davids, K., "Object Rotation Effects on the Timing of a Hitting Action.", Res. Quart. Ex. Sp, **75**, (2004), 130-137.

[14]   Broderick, M.P., Pavis, B. & Newell, K.M., "Assessment of the adiabatic transformability hypothesis in a ball-bouncing task.", Biol. Cyb., **82**, (2003), 433-442.

[15]   Knight, C.A., "Neuromotor issues in the learning and control of golf skill.", Res. Quart. Ex. Sp., **75**, (2004), 9-15.

[16]   Bernstein, N., "The co-ordination and regulations of movements.", Oxford, Pergamon Press, (1967).

[17]   Davids, K., Araújo, D., Shuttleworth, R. & Button, C., "Acquiring Skill in Sport: A Constraints-Led Perspective.", Int. J. Comp. Sci. Sp., **2**, (2003), 31-39.

[18]   Brisson, T.A. & Alain, C., "Should common optimal movement patterns be identified as the criterion to be achieved?", J. Mot. Behav., **28**, (1996), 211-223.

[19]   van Rooij, I., Bongers, R.M. & Haselager, W.F.G., "A non-representational approach to imagined action.", Cog. Sci., **26**, (2002), 345-375.

[20]   Davids, K., Bennett, S.J. & Newell, K.M., (Eds.), "Variability in the Movement System: A Multi-Disciplinary Perspective.", Champaign, Ill., Human Kinetics, (2005).

[21]   Newell, K.M., "Motor skill acquisition.", Ann. Rev. Psych., **42**, (1991), 213-237.

[22]   Araújo, D., Davids, K., Bennett, S., Button, C., & Chapman, G., "Emergence of Sport Skills under Constraints." In A.M. Williams and N.J. Hodges(Eds.), Skill Acquisition in Sport: Research, Theory and Practice, London, Routledge, (2004), 409-434.

[23]   Button, C., McLeod, M., Sanders, R. & Coleman, S., "Examining movement variability in the basketball free-throw action at different skill levels.", Res. Quart. Ex. Sp., **74**, (2003), 257-269.

[24]   Beek, P.J., Dessing, J.C., Peper, C.E. & Bullock, D., "Modelling the control of interceptive actions.", Phil. Trans. Roy. Soc. Lond. B, **358**, (2003), 1511-1523.

[25]   Renshaw, I., & Fairweather, M. M., "Cricket bowling deliveries and the discrimination ability of professional and amateur batters.", J. Sp. Sci., **18**, (2000), 951-957.

[26] de Rugy, A., Taga, G., Montagne, G., Beukers, M.J. and Laurent, G., "Perception-action coupling model for human locomotor pointing.", Biol. Cyb., **87**, (2002), 141-150.

[27]   Montagne, G., Cornus, S., Glize, D., Quaine, F., & Laurent, M., "A perception-action coupling type of control in long jumping.", J. Mot. Behav., **32**, (2000), 37-43.

[28]   Renshaw, I. & Davids, K., "Nested task constraints shape continuous perception-action coupling control during human locomotor pointing.", Neurosci. Lett., (2004), in press.

Notational analysis – a mathematical perspective.

Mike Hughes,
CPA, UWIC, Cyncoed, Cardiff, CF26 2XD.

Abstract

The role of feedback is central in the performance improvement process, and by inference, so is the need for accuracy and precision of such feedback. The provision of this accurate and precise feedback can only be facilitated if performance and practice is subjected to a vigorous process of analysis. Recent research has reformed our ideas on reliability, performance indicators and performance profiling in notational analysis – also statistical processes have come under close scrutiny, and have generally been found wanting. These are areas that will continue to develop to the good of the discipline and the confidence of the sports scientist, coach and athlete. If we consider the role of a performance analyst in its general sense in relation the to the data that the analyst is collecting, processing and analysing, then there a number of mathematical skills that will be required to facilitate the steps in the processes:- i) defining performance indicators, ii) establishing the reliability of the data collected, iii) ensuring that enough data have been collected to define stable performance profiles, iv) determining which are important, v) comparing sets of data, vi) modelling performances and vii) prediction. The mathematical and statistical techniques commonly used and required for these processes will be discussed and evaluated in this paper.

1. Introduction

Recent research has reformed our ideas on reliability, performance indicators and performance profiling in notational analysis – also statistical processes have come under close scrutiny, and have generally been found wanting. These are areas that will continue to develop to the good of the discipline and the confidence of the sports scientist, coach and athlete. If we consider the role of a notational analyst (Figure 1) in its general sense in relation to the data that the analyst is collecting, processing and analysing, then there a number of mathematical skills that will be required to facilitate the steps in the processes:-

1. defining performance indicators,

2. determining which are important,

3. establishing the reliability of the data collected,

4. ensuring that enough data have been collected to define stable performance profiles,

5. comparing sets of data,

6. modelling performances.

The recent advances made into the research and application of the mathematical and statistical techniques commonly used and required for these processes will be discussed and evaluated in this paper.

2. Mathematical and statistical processes in notational analysis

2.1 Performance indicators

A performance indicator is a selection, or combination, of action variables that aims to define some or all aspects of a performance. Analysts and coaches use performance

13

indicators to assess the performance of an individual, a team, or elements of a team. They are sometimes used in a comparative way, with opponents, other athletes or peer groups of athletes or teams, but often they are used in isolation as a measure of the performance of a team or individual alone.

Figure 1: A schematic chart of the steps required in moving from data gathering to producing a performance profile:
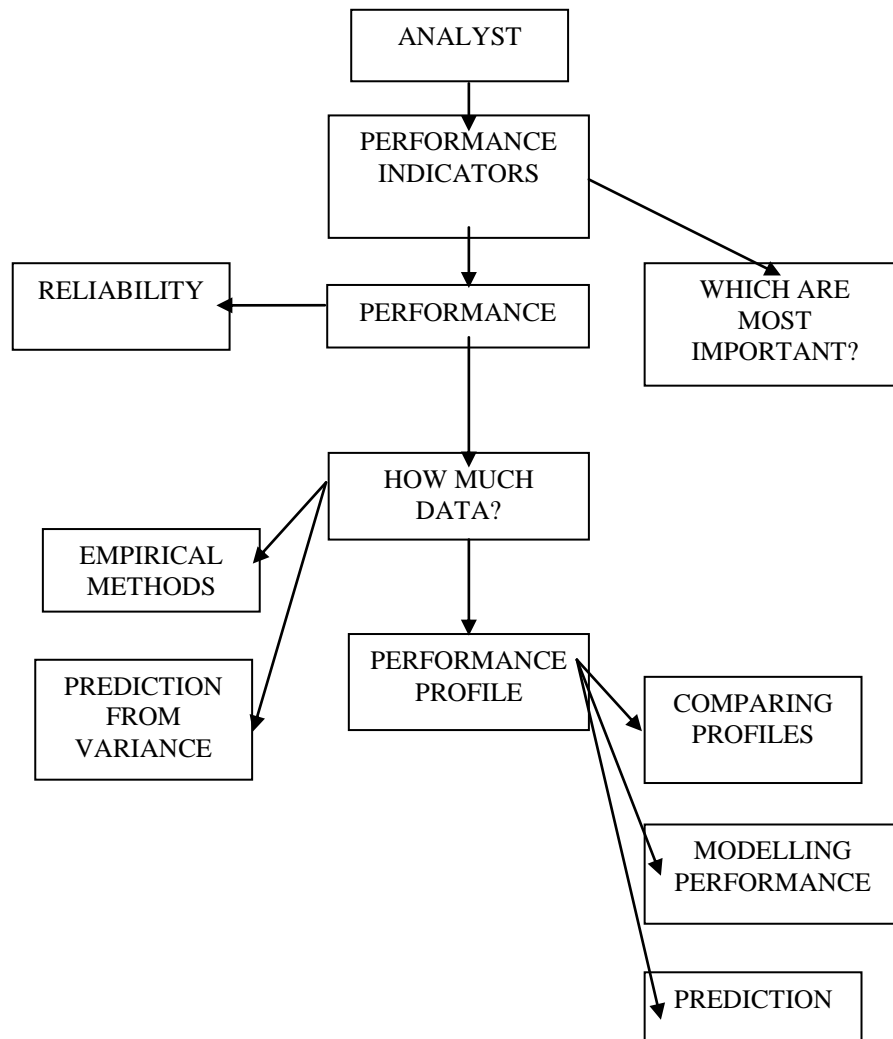
Through an analysis of game structures and the performance indicators used in recent research in performance analysis, Hughes and Bartlett [18] defined basic rules in the application of performance indicators to any sports. In every case, success is relative, either to your opposition, or to previous performances of the team or individual. To enable a full and objective interpretation of the data from the analysis of a performance, it is necessary to compare the collected data to aggregated data of a peer group of teams, or individuals, which compete at an appropriate standard. In addition any analysis of the distribution of actions across the playing surface must be normalised with respect to the total distribution of actions across the area.

Performance indicators, expressed as non-dimensional ratios, have the advantage of being independent of any units that are used; furthermore, they are implicitly independent of any one variable. They also allow, as in the example of bowling in cricket, an insight into differences between performers that can be obscure in the raw data. The use of non-

dimensional analysis is common in fluid dynamics, which offers empirical clues to the solution of multivariate problems that cannot be solved mathematically. Sport is even more complex, the result of interacting human behaviours; to apply simplistic analyses of raw sports data can be highly misleading. Current research [25, 30] is examining how normative profiles are established - how much data are required to reliably define a profile and how this varies with the different types and natures of the data involved in any analysis profile. This area is discussed in detail later.

Figure 2: A digital systems approach to the data sharing that the interactive commercial systems have enabled for performance analysts working with coaches and athletes [18]:
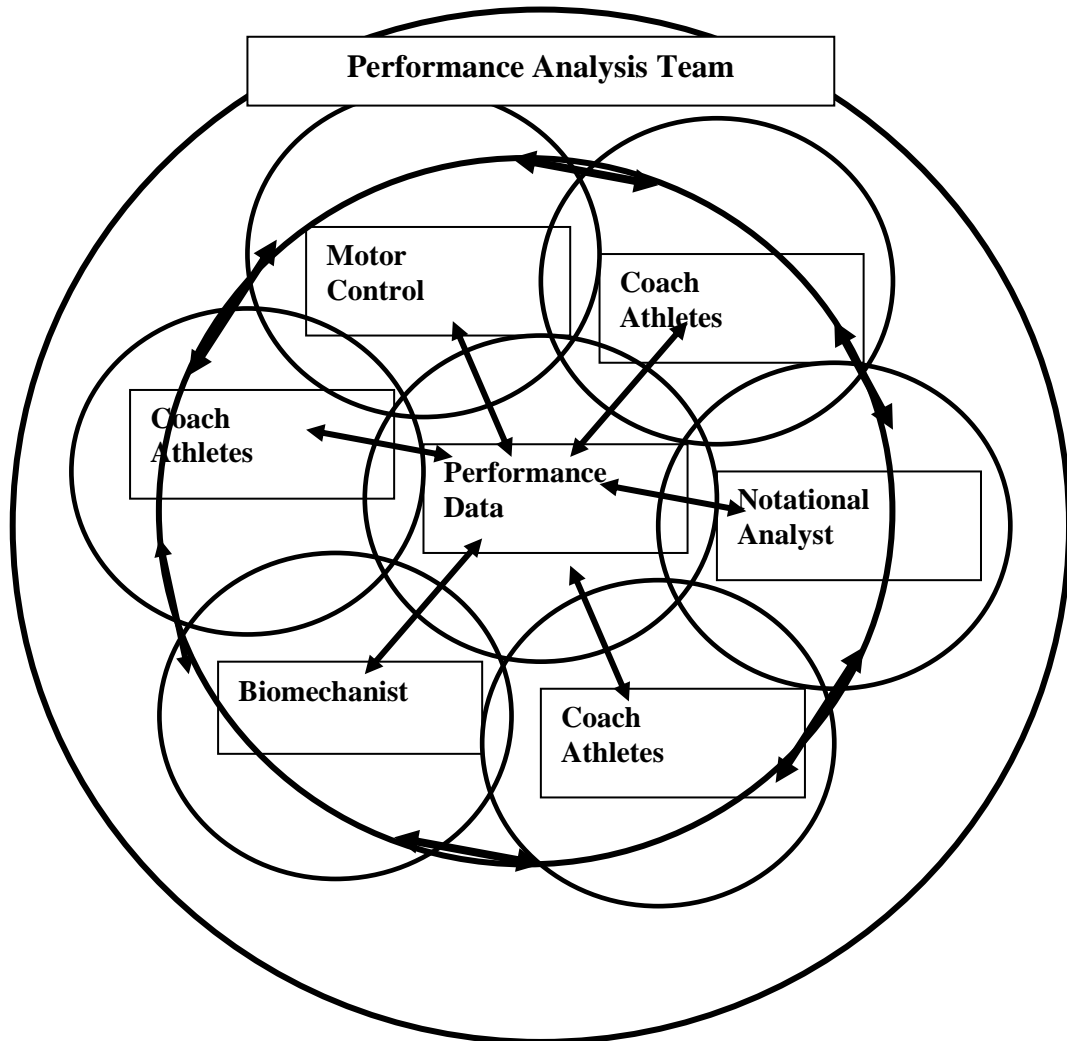
Many of the most important aspects of team performance cannot be 'teased out' by biomechanists or match analysts working alone – a combined research approach is needed (Figure 2.). This is particularly important for information processing – both in movement control and decision making. We should move rapidly to incorporate into such analyses qualitative biomechanical indicators that contribute to a successful movement. These should be identified interactively by biomechanists, notational analysts and coaches, sport-by-sport and movement-by-movement, and validated against detailed biomechanical measurements in controlled conditions. Biomechanists and notational analysts, along with experts in other sports science disciplines – particularly motor control, should also seek to agree on, and measure, those performance indicators that are

important from this perspective. This could help to clarify, for example, whether the movement variability, which has been measured in such skills as basketball shooting and cricket batting, is a function of the behaviour of the opponents or other team members or due to noise in measurements or the motor control apparatus.

For the different types of games considered, it has become clear that the classification of the different action variables being used as performance indicators follow rules that transcend the different sports. These are summarised in Table 1. Most of the research community in performance analysis have not followed these simple rules to date. The utility of performance analysis could be considerably enhanced if its practitioners agreed and implemented such rules in the future.

Table 1: Categorisation of the application of different performance indicators in games

Match Classification	Biomechanical	Technical/Tactical
Always compare with opponents' data and, where possible, with aggregated data from peer performances.	**Compare with previous performances and with team members, opponents and those of a similar standard. Consider normalising data when a maximum or overall value both exists and is important or when inter-individual or intra-individual across-time comparisons are to be made.**	**Always normalise the action variables with the total frequency of that action variable or, in some instances, the total frequency of all actions.**

2.2 Analysis of the relative importance of Performance Indicators

Defining performance indicators (PI's) is based on prior knowledge and research of the effect of relevant factors on relevant dependent variables. However, in many areas, defining performance indicators remains difficult to do accurately. This is due to the complexity of the behaviour of interest as well as the large number of relevant factors. Many factors are often difficult to define and measurements used may be of limited reliability and validity. To be able to comment appropriately (accurately and relevantly) on a given sport, it is necessary to know exactly which performance indicators are relevant to the results. Coaches and, to some extent, notational analysts have traditionally relied upon experience and their own ideas about their sport. This has led to a huge spectrum of performance indicators being used in some sports, particularly soccer. But as databases of multiple sporting events, such as the rugby or the soccer world cups or repeated grand slam events in tennis, are compiled by comprehensive and sophisticated computerised analyses, then comparative evaluation of PI's should be possible – using the outcomes of the tournaments as a correlating factor. Two recent pieces of research have given a starting point in the methodology to assess the relative importance of PI's.

Luhtanen et al. [37] analysed offensive and defensive variables of field players and goalkeepers in the EURO 2000 soccer tournament and related the results to the final team ranking in the tournament. All matches (n=31) of the EURO 1996 and 2000 were recorded using video and analysed by three trained observers with a computerised match analysis system. The written definitions of each event (pass, receiving, run with the ball,

shot, scoring trial, defending against scoring trial, interception and tackle) were applied in analysing the matches.

The quantitative (number of executions) and qualitative (percentage of successful executions) game performance variables were as follows: passes, receivings, runs with ball, scoring trials, interceptions, tackles, goals and goalkeeper's savings. The total playing times were recorded and the game performance results were standardised for 90 minutes playing time. Team ranking in each variable (quantitative and qualitative) was used as a new variable. The final ranking orders in the EURO 1996 and EURO 2000 tournaments were explained by calculating the rank correlation coefficients between team ranking in the tournament and ranking in the following variables: ranking of ball possession in distance, passes, receiving, runs with the ball, shots, interceptions, tackles and tackles. Selected quantitative and qualitative sum variables were calculated using ranking order of all obtained variables, only defensive variables and only offensive variables. The means and standard deviations of the game performance variables were calculated. Ranking order in each variable was constructed. Spearman's correlation coefficients (95 % CI) were calculated between all ranking variables describing game performance. In this way the relative importance of the variables were assessed. Although this research was orientated towards explaining the success of the top teams, and did not explore different combinations of performance indicators nor did they non-dimensionalise the variables, the methodology used would seem one way of ordering the huge spectrum of 'would-be' performance indicators that appear in the literature.

O'Donoghue at al. [54] and O'Donoghue and Williams [53] used a variety of statistical, predictive and modelling techniques in attempts to anticipate the results in the 2002 World Cup in soccer and the 2003 World Cup in rugby union. Some of their statistical techniques also point the way to future research being able to perhaps quantify objectively the relative importance of the performance indicators that determine success in a sport. Each factor as well as the match score was represented as a difference between the superior team and the inferior team. They used very simple performance variables, that were not measured in-match as in the Luhtanen et al. [37] paper, just the world ranking, the distance travelled to the tournament and the amount of rest between matches. Amongst the many techniques used in these two papers, O'Donoghue et al. [55] used multiple linear regression and binary linear regression. Using large databases to construct their predictive models, they created equations in which the relative size of the coefficients of the respective variables in these equations denote their relative importance. These methods have certain problems associated with them but, given the large databases of in-match analysis variables that are now available at many of the centres of research in performance analysis of sport throughout the world, this methodology would seem to point to an even better way to assess quantitatively the relative importance of performance indicators in any sport.

2.3 Reliability

It is vital that the reliability of a data gathering system is demonstrated clearly and in a way that is compatible with the intended analyses of the data. The data must be tested in the same way and to the same depth in which they will be processed in the subsequent analyses. In general, the work of Bland and Altman [6] has transformed the attitude of sport scientists to testing reliability; can similar techniques be applied to the non-parametric data that most notational analysis studies generate? There are also a number of questions that inherently re-occur in these forms of data-gathering.

Analysing 72 research papers recently published under the banner of notational analysis, Hughes et al. [23] found that 70% did not report any reliability study and a large proportion of the remaining used questionable processes given the recent ideas in

reliability testing in sports science [2]. In some cases the reliability studies were executed on summary data, and the system was then assumed to be reliable for all of the other types of more detailed data analyses that were produced. The most common form of data analysis in notation studies is to record frequencies of actions and their respective positions on the performance area, these are then presented as sums or totals in each respective area. What are the effects of cumulative errors nullifying each other, so that the overall totals appear less incorrect than they actually are?

Many research papers have used parametric tests in the past – these were found to be slightly less sensitive than the non-parametric tests, and they did not respond to large differences within the data. Further, the generally accepted tests for comparing sets of non-parametric data, Xi^2 analysis and Kruskal-Wallis, were found to be insensitive to relatively large differences between sets of the data – sets of data 25% apart were deemed to be not significantly different. In sport margins of 2% can mean the difference between winning and coming last (women's 400m final – Sydney, 2000). It would seem that a simple percentage calculation gives a simple indicator of the absolute differences of the data sets, and these are easily related back to the aims of the research. It was demonstrated that these tests can also lead to errors and confusion if the researchers involved were not specific enough with the definitions of how they defined their percentage differences. Hughes et al [25] suggested that the following conditions should be applied.

✓ The data should initially retain its sequentiality and be cross-checked item against item.

✓ Any data processing should be carefully examined as these processes mask original observation errors.

✓ The reliability test should be examined to the same depth of analysis as the subsequent data processing, rather than being performed on just some of the summary data.

✓ Careful definition of the variables involved in the percentage calculation is necessary to avoid confusion in the mind of the reader, and also to prevent any compromise of the reliability study.

✓ It is recommended that the calculation of % differences is based upon

$$(\Sigma(mod[V_1 - V_2]) / V_{mean}) * 100 \%,$$

(where V_1 and V_2 are variables, V_{mean} their mean, mod is short for modulus and Σ means 'sum of').

This is then used to calculate percentage error for each variable involved in the observation system, and these can be plotted against each variable, and each operator. All the variables should be carefully defined. This will give a powerful and immediate visual image of the reliability tests.

2.4 Establishing the stability of performance profiles

2.4.1 Empirical methods

It is an implicit assumption in notational analysis that in presenting a performance profile of a team or an individual that a 'normative profile' has been achieved. Inherently this implies that all the variables that are to be analysed and compared have all stabilised. Most researchers assume that this will have happened if they analyse enough performances. But how many is enough? In the literature there are large differences in sample sizes. Hughes Evans and Wells [25] trawled through some of the analyses in soccer shows the differences (Table 2). They suggested that there must be some way of

assessing how data within a study is stabilising. Many research papers in match analysis present data taken from an arbitrarily selected number of performances.

Table 2: Some examples of sample sizes for profiling in sport.

Research	Sport	N (matches for profile)
Reep & Benjamin [60]	Soccer	3,216
Eniseler et al., [9]	Soccer	4
Larsen et al., [33]	Soccer	4
Hughes et al., [27]	Soccer	8 (16 teams)
Tyryaky et al., [63]	Soccer	4 and 3 (2 groups)
Hughes [16]	Squash	12, 9 & 6 – 3 groups
Hughes & Knight [20]	Squash	400 rallies
Hughes & Williams [22]	Rugby Union	5
O'Donoghue [50]	Badminton	16, 17, 17, 16, 15
Hughes & Clarke [19]	Tennis	400 rallies
O'Donoghue & Ingram [52]	Tennis	1328<rallies<4300 (8 groups)

Research by Hughes, Wells and Matthews [28] attempted to 'validate' their performance profiles. To establish that a normative profile had been reached, the profiles of 8 matches were compared with those of 9 and 10 matches, using dependent t-tests, for each of the standards of players – no non-parametric software was available. The results showed that the elite players and the county players did establish a playing pattern that could be said to be reproduced reasonably consistently. The recreational players failed to show any real significance at a suitable level, therefore a normal playing pattern was not established. It must be understood that each squash match, even at the same level can differ from one another. A question that arose from this is, how much can they differ before they are suspected that the difference is caused by something other than chance? This was felt to be especially true for the recreational playing standards, where their differences were not caused by chance but by the fact that they have no fixed pattern of play. As players work their way up the different standards of play a set pattern was emerging. The elite players are at a level where a fixed pattern can be established. This method clearly demonstrated that those studies assuming that 5, 6 or 8 matches of performances were enough for a normative profile, without resorting to this sort of test, are clearly subject to possible flaws. It is vital that, in notational analysis, the same care is adopted in examining our validity and reliability techniques.

The nature of the data themselves will also effect how many performances are required – 5 matches may be enough to analyse passing in field hockey, would you need 10 to analyse crossing or perhaps 30 for shooting? The way in which the data is analysed also will effect the stabilisation of performance means – data that are analysed across a multi-cell representation of the playing area will require far more performances to stabilise than data that are analysed on overall performance descriptors (e.g. shots per match). It is misleading to test the latter and then go on to analyse the data in further detail.

Hughes Evans and Wells [28] suggested a practical way of examining whether a true performance profile has reached stable means, calculations of percentage variance from the 'end mean' were determined and then plotted appropriately. Percentage difference calculations were used to evaluate the reliability of system, for intra- and inter-operator observations, producing similar graphs to the Bland and Altman [6].

<ins>% difference = (No. of differing observations / Total no. of observations) x 100</ins>

Frequency data, from squash and badminton, were normalised to allow for true comparison of data from matches and games of differing lengths.

Normalised Match Frequency = (Match frequency / No. of rallies in match) x 100

Normalised Game Frequency = (Game frequency / No. of rallies in game) x 50

The cumulative means of each variable were examined over a series of matches/games.

   At the first point, the number of matches, 'N(E)', where the cumulative mean consistently lay within set 'limits of error' was recorded as the establishment of a normative template of play. These limits of error are a percentage deviation (+/- 1%; +/- 5%; +/- 10%) of the overall data mean about the overall mean.

Let n   =   the variable 'number of matches'
g   =   the variable ' number of games'
N(E)   =   value of n to reach limits of error
N(T)   =   total number of matches

       Cumulative mean = (Sum of the frequencies of 'n') / n

       Limits of error (10%) = Mean N(T) ± (Mean N(T) x 0.1)

       Limits of error (5%) = Mean N(T) ± (Mean N(T) x 0.05)

       Limits of error (1%) = Mean N(T) ± (Mean N(T) x 0.01)

For the working performance analyst the results provide an estimate of the minimum number of matches to profile an opponent's rally-end play. Whilst the results may be limited to badminton, men's singles and the individual, the methodology of using graphical plots of cumulative means in attempting to establish templates of performance has been served. From this study [28] the following conclusions were made:

✓ This method clearly demonstrated that those studies assuming that 4, 6 or 8 matches or performances were enough for a normative profile, without resorting to this sort of test, are clearly subject to possible flaws. The number of matches required for a normal profile of a subject population to be reached is dependent upon the nature of the data and, in particular, the nature of the performers.

✓ The number of matches required to establish a profile of elite women's movement was dependent on the nature of the data. For elite women, profiles were achieved within three to nine matches (under 10% error) depending on the variables being analysed.

✓ The main problem associated with any primary study aiming at establishing previously unrecorded 'normal' profiles remains reliability and accuracy. Any future studies that proclaim data as a performance profile should provide supportive evidence that the variable means are stabilising. A percentage error plot showing the mean

variation as each match/player is analysed is one such technique. This can be adapted to different sports when analysing profiles/templates of performance.

For the working performance analyst the results provide an estimate of the minimum number of matches to profile an opponent's rally-end play. Whilst the results may be limited, the methodology of using graphical plots of cumulative means in attempting to establish templates of performance has been served.

2.4.2 Confidence intervals

The procedure outlined above is expensive in terms of time when collecting data for the first time and is limited in its applicability in many cases due to fluctuations in factors such as team changes, maturation and the fact that some performances never stabilise. James et al. [30] suggested an alternative approach whereby the specific estimates of population means are calculated from the sample data through confidence limits (CL's). CL's represent upper and lower values between which the true (population) mean is likely to fall based on the observed values collected. Calculated CL's naturally change as more data is collected, typically resulting in the confidence interval (upper CL minus lower CL) decreasing. Confidence intervals (CI's) were therefore suggested to be more appropriate as performance guides compared to using mean values. Using a fixed value appears to be too constrained due to potential confounding variables that typically affect performance, making prescriptive targets untenable.

Figure 3: Example of the % difference plots:
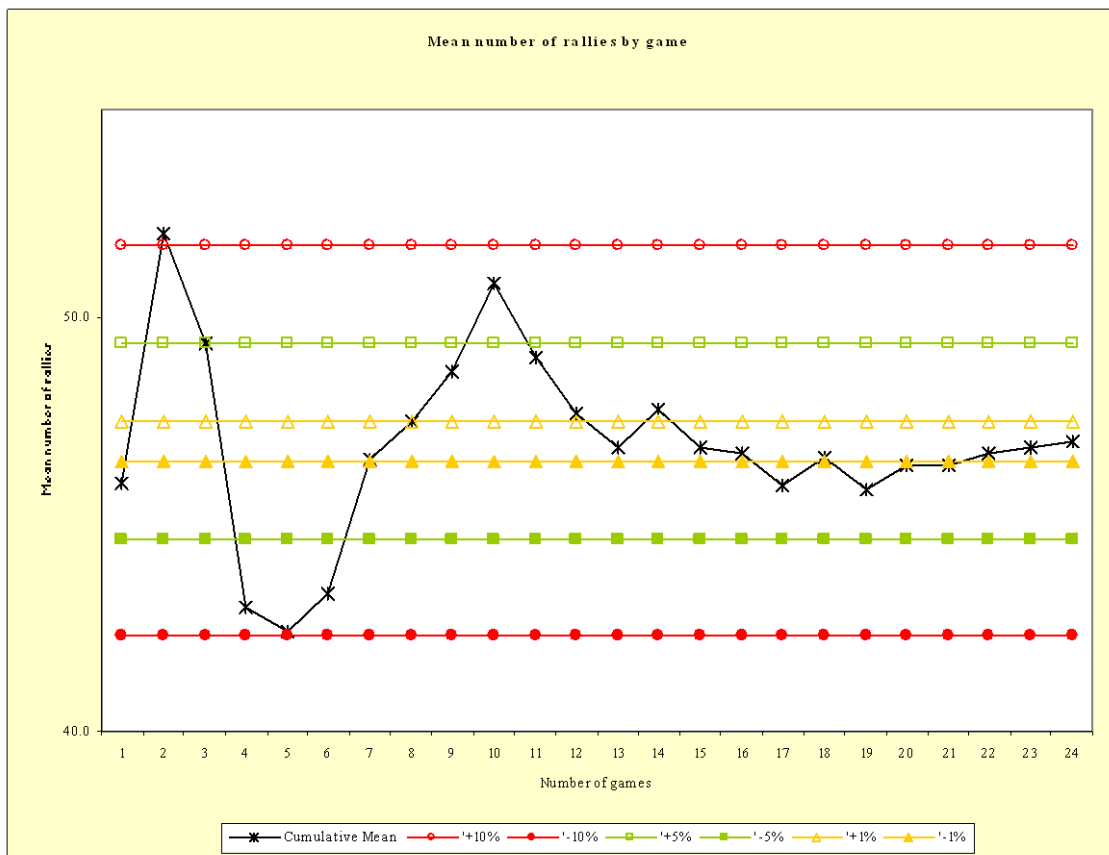
From a theoretical perspective, James et al. argued that the use of CI's can also add significance to the judgement of the predictive potential of a data set, i.e. whether enough data has been collected to allow a reasonable estimation. For their investigation a criterion was formulated to test the rate of change of the CI for stability. Initially 95% CI's were calculated for each performance indicator as soon as enough match data had been collected ($N = 2$) and each time more data was added the new CI was calculated. This meant that CI's could be constructed for each performance indicator after 2, 3 and…. N matches respectively. Behavioural frequencies fell outside the 95% CI more often for small data sets and less often as the data set increased. However, this was inevitable as any measure related to the mean of a data set becomes progressively more resistant to change as the data set increases.

The data in table 3 are an example of this type of presentation, it is a very useful way of assessing the stability of an analysts' data. I feel that the more this method is used then the more easily will analysts be able to interpret the CI's in relation to the experimental goals of their research or consultancy work.

Table 3: Mean profiles and 95% confidence limits for the positional clusters of prop, hooker and lock.

	Prop ($n = 3$)***			Hooker ($n = 1$)			Lock ($n = 4$)**		
	Mean	+CL	-CL	Mean	+CL	-CL	Mean	+CL	-CL
Successful Tackles	4.01	4.96	3.06	4.25	5.59	2.91	5.73	6.81	4.64
Unsuccessful Tackles	0.73	1.12	0.33	0.51	1.01	0	0.76	1.15	0.37
Successful Carries	4.25	5.32	3.18	3.51	5.37	1.66	2.20	2.81	1.60
Unsuccessful Carries	0.23	0.44	0.01	0.40	0.80	0	0.03	0.08	0
Successful Passes	1.76	2.46	1.06	0.87	1.68	0.06	1.23	1.76	0.69
Unsuccessful Passes	0.47	0.98	0	0	0	0	0.20	0.33	0.07
Handling Errors	0.33	0.56	0.09	0.27	0.69	0	0.68	0.94	0.41
Normal Penalties	0.68	0.97	0.39	0.40	0.79	0	1.12	1.47	0.77
Yellow Cards	0.05	0.11	0	0.08	0.24	0	0.09	0.20	0
Tries Scored	0.02	0.07	0	0.10	0.31	0	0	0	0
Successful Throw-ins	-	-	-	9.76	11.85	7.66	-	-	-
Unsuccessful Throw-ins	-	-	-	2.94	4.09	1.79	-	-	-
Successful Lineout Takes	-	-	-	-	-	-	4.05	4.83	3.28
Unsucc. Lineout Takes	-	-	-	-	-	-	0.43	0.71	0.15
Successful Restart Takes	-	-	-	-	-	-	0.65	0.91	0.39
Unsucc. Restart Takes	-	-	-	-	-	-	0.59	0.90	0.27

$*p < .05$ $**p < .01$ $*** p < .001$

2.5 Comparing sets of data

Hughes et al. [23] extended their survey of reliability methods used in recent research to also examine the types of statistical processes used in the analyses of data in a selected number of research papers in notational analysis. The subsequent data analyses (see Table 2) used a multiplicity of techniques but there were a large number of studies that did not present any statistics to compare sets of data. Those labelled 'not specific' did cite probability values, but did not mention which statistical process had been used. In a number of studies, parametric techniques were used with data that were non-parametric, although, in some cases, the means of the data sets appeared ordinal, they were often means of nominal data and therefore the use of a parametric test put the conclusions at risk.

Table 4: An analysis of the different statistical processes used in subsequent data analyses in some randomly selected performance analysis research papers.

Statistical processes for data analysis	No.	%
Chi-square	21	29
None	19	26
Not specific	12	17
t-test	8	11
ANOVA	5	7
Factor analysis	2	3
ANCOVA	1	1
Mann Whitney	1	1
Hotelling T^2 test	1	1
Wilcoxon	1	1
Bivariate analysis	1	1
Total	**72**	**100**

There are many similarities in the nature of the data generated by experiments in performance analysis of sport. Parametric statistical techniques are often applied to notational analysis data, either through ignorance, or lack of availability of the relevant software. Correlation was the most common technique used in confirming reliability, when a technique was used. To test the sensitivity of correlation, and other techniques, more tests were applied to some experimental data, with the intent of assessing the threat to interpretation of reliability using these types of method. More correctly [67] a Xi^2 analysis should be applied to data of this nature. As a means of comparison, these analyses are presented in Table 5.

Table 5: Correlation and Xi^2 analysis applied to the intra-operator data from Table 8.

	L	S	I	G	O
R	1.000	1.000	0.996	0.999	0.998
P value	0.000	0.000	0.000	0.000	0.000
Xi^2	0.069	0.005	0.383	0.165	0.289
P value	1.000	1.000	0.996	0.999	0.998

It is disturbing that, given the % error calculations on the different variables above, that these tests indicate little sensitivity to the differences in the sets of data. It seems puzzling that these tests are so insensitive to differences as large as 5%. At what size of differences do they indicate a significant difference? To examine further the sensitivity of these tests to differences in data sets, the reliability test data were manipulated with increments of 10% changes in values. When these differences were calculated all in the same direction, then the correlation coefficient, and the equivalent for Xi^2 analysis, is always perfect (r=1). Consequently the increments were alternately added and subtracted from the items

in descending order in the column of data. The correlation data dropped below the P>0.05 (r_{crit}= 0.811) level when differences of 40% were added and subtracted to the original data, that is data sets over 30% different would be indicated to be the same (P<0.05) using correlation. The Xi^2 analysis data only showed significant differences (P<0.05) when between 20% and 30% incremental differences had been added and subtracted in the data.

Table 6: Kruskal-Wallis and Anova applied to the different variables for 5 different sets of data.

	Kruskal-Wallis		Anova	
Variables	H-value	P-value	F-value	P-value
Tackle	2.89	0.576	0.79	0.577
Pass	6.65	0.156	4.68	0.061
Kick	4.34	0.361	1.23	0.405
Ruck	6.03	0.197	1.45	0.342
Lineout	1.50	0.827	0.25	0.898

A number of the research papers examined used Anova, which is not the correct method for non-parametric data. The textbooks [66] explain that this is a threat to the validity of the testing process but by how much? There is little in the literature to help with this difficulty. Because of this a comparison of the Kruskal-Wallis analysis and Anova is presented in Table 6.

The surprising fact in the analysis is that neither test indicates that there are any significant differences at the P<0.05 level. Both tests indicate that the biggest differences are in the variable the pass, while the tackle and the ruck had the largest differences (9.4% and 10.3% respectively) when calculated by % measurements. This is because these tests measure the overall variance about the median and mean – and may therefore miss one relatively large difference measurement if all the others are close to the mean or median. It is interesting to manipulate the data, as before, to examine what levels of errors will cause these tests to indicate that significant differences are present (Table 7).

The F value is a ratio of between-group variances to those within the groups. Consequently, although the means of the data are moving further apart, because the variation within each group varies as much, more in some cases, the F values do not approach any closer to the critical values. Consequently using these tests does not appear to be practical for comparison studies. Another common problem in notation studies is with the amounts of data collected for the reliability study – have enough repetitions been completed? How many are required? Perhaps the small amounts of data here are contributory towards the poor performance of these tests.

Table 7: Manipulation of some sample data to test the sensitivity of Kruskal-Wallis and Anova for comparing multiple sets of non-parametric data.

	Kruskal-Wallis		Anova	
% changes	H-value	P-value	F-value	P-value
5%	2.89	0.576	2.84	0.118
10%	6.65	0.156	0.86	0.551
15%	4.34	0.361	1.57	0.297
20%	6.03	0.197	0.97	0.503

Although Atkinson and Neville [2] have produced a definitive summary of reliability techniques in sports medicine, the only attempt that has been attempted to make

recommendations in the use of techniques in performance analysis for solving some of the common problems associated with these types of data was that by Nevill et al [50]. This is written very much from a statistician's point of view particularly extolling the virtues of the high power software GLIM. They concluded that investigating categorical differences in discrete data using traditional parametric tests of significance (e.g. ANOVA, based on the continuous symmetric normal distribution) is inappropriate. More appropriate statistical methods were promoted based on two key discrete probability distributions, the Poisson and binomial distributions. In the opinion of Nevill et al. [50] the first approach is based on the classic x^2 test of significance (both the goodness-of-fit test and the test of independence). The second approach adopts a more contemporary method based on log-linear and logit models using the statistical software GLIM. Although investigating relatively simple one-way and two-way comparisons in categorical data, both approaches result in very similar conclusions, which is comforting to the 'old guard' notational analysts who have been using x^2 for the last few decades. However, as soon as more complex models and higher-order comparisons are required, the approach based on log-linear models is shown to be more effective. Indeed, when investigating factors and categorical differences associated with binomial or binary response variables, such as the proportion of winners when attempting decisive shots in squash or the proportion of goals scored from all shots in association football, logit models become the only realistic methods available. Nevill et al. concluded that with the help of such log-linear and logit models, greater insight into the underlying differences or mechanisms in sport performance can be achieved.

2.6 Modelling performances.

"The modelling of competitive sport is an informative analytic technique because it directs the attention of the modeller to the critical aspects of data which delineate successful performance. The modeller searches for an underlying signature of sport performance which is a reliable predictor of future sport behaviour. Stochastic models have not yet, to our knowledge, been used further to investigate sport at the behavioural level of analysis. However, the modelling procedure is readily applicable to other sports and could lead to useful and interesting results."

Franks and McGarry [11]

Some exciting trends are to be found in modelling performances and match play, using a variety of techniques, many examples can be found in the Journals now available in these disciplines, the International Journal of Performance Analysis of Sport (electronic - eIJPAS) and the International Journal of Computers in Sport Science (electronic – eIJCSS). The simplest, and traditional, form is using empirical methods of producing enough performance data to define a performance profile at that particular level. Some researchers are extending the use of these forms of data bases to attempt to predict performances; stochastic probabilities, neural networks and fuzzy logic have been used, singly or in combinations, to produce the outputs. McGarry and Perl [45] presented a good overview of models in sports contest which embraces most of these techniques. So far results have been a little disappointing in practical terms. It does seem to have potential – perhaps if we added a dash of f'eng shui?

Early research of modelling in sport includes Mosteller [47], who set out guidelines when he developed a predictive model, and these ideas are eminently practical and many researchers in the area use these, or modifications of these to delimit their models.

Alexander et al [1] used the mathematical theory of probability in the game of squash. Mathematical modelling can describe the main features of the game of squash and can reveal strategic patterns to the player. Squash is an example of a Markov chain mathematical structure, a similar technique was used later by McGarry and Franks [39]. Alexander et al. went on to recommend practical strategies at 8-8 in a game, in calling to finish on 9 or 10, the choice being that of the player receiving serve, based on these

relative probabilities of winning rallies. This is one of the few practical outcomes from this form of modelling

   Other attempts to model team games [32] theoretically have tended to founder upon the complexity of the numbers of variables involved and, at that time, did not base their predictions upon sound databases. The advent of computer notation systems have enabled the creation of large databases of sports performances in different sports, these in turn have helped the development of a number of different techniques in modelling performance in sport. These will be discussed under the following generic headings:-

➤   Empirical Modelling

➤   Stochastic Modelling

➤   Dynamic systems

➤   Statistical techniques

➤   Artificial Intelligence

•   Expert Systems

•   Neural Networks

2.6.1 Empirical models

Hughes [15] established a considerable database on different standards of squash players. He examined and compared the differences in patterns of play between recreational players, country players and nationally ranked players, using the computerised notational analysis system he had developed. The method involved the digitization of all the shots and court positions, and these were entered via the QWERTY keyboard. Hughes [16]) was able then to define models of tactical patterns of play, and inherently technical ability, at different playing levels in squash. Although racket developments have affected the lengths of the rallies, and there have been a number of rule changes in the scoring systems in squash, these tactical models still apply to the game today. This study was replicated, with a far more thorough methodology, for the women's game by Hughes et al. [28].

Fuller [12] developed and designed a Netball Analysis System and focused on game modelling from a data base of 28 matches in the 1987 World Netball Championships. There were three main components to the research - to develop a notation and analysis system, to record performance, and to investigate the prescience of performance patterns that would distinguish winners from losers. The system could record how each tactical entry started; the player involved and the court area through which the ball travelled; the reason for each ending; and an optional comment. The software produced the data according to shooting analysis; centre pass analysis; loss of possession; player profiles; and circle feeding. Fuller's intention of modelling play was to determine the routes that winning, drawing and losing teams took and to identify significantly different patterns. From the results Fuller was able to differentiate between the performances of winning and losing teams. The differences were both technical and tactical.

   The research was an attempt to model winning performance in elite netball and more research needed in terms of the qualitative aspects i.e. how are more shooting opportunities created. the model should be used to monitor performance over a series of matches not on one-off performances.

Treadwell, Lyons, Potter [65] suggested that match analysis in rugby union and other field games has centred on game modelling and that their research was concerned with using the data to predict game content's of rugby union matches. They found that clear physiological rhythms and strategic patterns emerged. They also found that at elite level it was possible to identify key "windows" i.e. vital "moments of chronological expectancy where strategic expediency needs to be imposed." It appears that international matches and successful teams generate distinctive rhythms of play which can exhibit a team fingerprint or heartbeat".

Empirical models enable both the academic and consultant sports scientist to make conclusions about patterns of play of sample populations of athletes within their sport. This, in turn, gives the academic the potential to examine the development and structures of different sports with respect to different levels of play, rule changes, introduction of professionalism, etc. The consultant analyst can utilise these models to compare performances of peer athletes or teams with whom the analyst is working.

2.6.2 Dynamic systems

Modelling human behaviour is implicitly a very complex mathematical exercise, which is multi-dimensional, and these dimensions will depend upon 2 or 3 spatial dimensions together with time. But the outcomes of successful analyses offer huge rewards, as Kelso [31] pointed out:-

If we study a system only in the linear range of its operation where change is smooth, it's difficult if not impossible to determine which variables are essential and which are not.

Most scientists know about nonlinearity and usually try to avoid it.

Here we exploit qualitative change, a nonlinear instability, to identify collective variables, the implication being that because these variables change abruptly, it is likely that they are also the key variables when the system operates in the linear range.

Recent research exploring mathematical models of human behaviour led to populist theories exemplified by 'Catastrophe Theory' and 'Chaos Theory'.

2.6.2.1 Chaos Theory

The ability to predict performance is inherent in the process of effective planning., but is very difficult to do accurately. Errors in statistical methods of predicting are often attributed to forecasting error but chaos theory suggests that those errors imply that performances follow natural trends and are better explained by non-linear rather than the more traditional linear mathematics.

"Chaos theory is the science that discovers order in nature's seeming randomness."

In more recent times scientists have discovered that certain systems within nature have chaotic dynamics and have an infinite variety of unpredictable forms but through a systematic process of self-organisation. The disorder of nature produces orderly patterns such as snowflakes. Other examples of non-linear chaotic systems are: weather, national economies, fibrillating hearts. It is possible to mathematically equate the beating of the heart that will provide values for the process over time. These solutions can be mapped through an "attractor" graph, which shows a chaotic system's solutions converge towards a specific path. A small change to the input will vary the pattern. Although this variation

appears to be chaotic and random it is a reflection of a high order of complex events. This provoked interest in whether this pattern of chaos and self-organisation could also be evident in human situations.

In the management of human organisations chaos theory points towards the need for managers to create an unstable environment for effective learning and hence new strategic directions to evolve. There are certain key points on the behaviour of dynamic systems and their applicability to human situations.

1. Chaos is a fundamental property of non-linear feedback systems. All human behaviour are non-linear because one action always leads to a subsequent one and people tend to over or under react. Therefore in any situation involving human interaction there is a possibility of chaotic behaviour as well as stable or unstable behaviour. The key question is which state leads to successful performance. Success will lie at the border between a state of stable equilibrium (ossification and team work) and an unstable state of equilibrium (disintegration and individual performance), that is in a non-equilibrium state between the two.

2. Chaos is a form of instability where the long-term future is not known. When irregular patterns of behaviour operate away from equilibrium they will be highly sensitive to tiny changes and will completely alter the behaviour. Small changes leading to larger ones are commonplace occurrences in human situations. Stacey [62] cited the example of VHS and Betamax.

3. Chaos has boundaries around its instability. Chaos is disorder and randomness at one level and qualitative pattern at another. When the future unfolds it often repeats itself but never in exactly the same way. "Chaos is an inseparable intertwining of order and disorder."

4. Unpredictable new order can emerge from chaos.

The applications of chaos theory to analysis of sports are so far limited to using the theories as a source of metaphors [38] - the idea of reducing some of the apparent chaos and linear dynamic models of sports such as soccer to predictive and stable differential equations is attractive but would seem some way in the future. An interesting fact is the similarity between the equations of Chaos Theory, those of the simplest model for Critical Incident Technique (Simple Harmonic Motion – see below) and the equations that Kelso [31] uses to underpin his models in his ideas of motor learning and the self-organisation of brain and behaviour. It is tempting to think that there might be parallels between these theories and that there are possible links between these models, just as there are links between these different aspects of performance. Developing the skills and imagination to tease these links out of these fascinating areas will make very rewarding research in the future?

2.6.2.2 Critical Incident Technique

At the First World Congress of Notational Analysis of Sport (1991), Downey talked of rhythms in badminton rackets, athletes in "co-operation" playing rhythmic rallies, until there was a dislocation of the rhythm (a good shot or conversely a poor shot) – a 'perturbation' – sometimes resulting in a rally end situation (a 'critical incident'), sometimes not. A good defensive recovery can result in the re-establishment of the rhythm. This was the first time that most of us had considered different sports as an example of a multidimensional, periodic dynamic system.

The term ´critical incident´ was first coined by Flanagan [10] in a study designed to identify why student pilots were exhausted at flight school.

"The critical incident technique outlines procedures for collecting observed incidents having special significance and meeting systematically defined criteria."

The critical incident technique is a powerful research tool but as with other forms of notating behaviour there are limitations inherent in the technique. Flanagan admitted that "Critical incidents represent only raw data and do not automatically provide solutions to problems". Another limitation of the technique is the total dependence on the reporters' opinions and this subjective element inherent in the use of the technique is often stated as a disadvantage. But Flanagan also pointed to the advantages of such a technique:

"The critical incident technique, rather than collecting opinions, hunches and estimates, obtains a record of specific behaviours from those in the best position to make the necessary observations and evaluations."

These opinions sound very much like the debates that have surrounded notational analysis within the halls of sports science over the last decade. Research in sport has addressed some of these issues and notation and movement analysis systems have been developed that can overcome some of these disadvantages. Since Downey's suggestions in 1991, some researchers have investigated the possibilities that analysing 'perturbations' and 'critical incidents' offer. McGarry and Franks [40, 41, 42, 43 and 44] applied further research to tennis and squash. They derived that every sporting situation contains unique rhythmical patterns of play. This behavioural characteristic is said to be the stable state or dynamic equilibrium. The research suggested that there are moments of play where the cycle is broken and a change in flow occurs. Such a moment of play is called a 'perturbation'. It occurs when either a poorly executed skill or a touch of excellence forces a disturbance in the stability of the game. For example, in a game of rugby this could be a bad pass or immediate change of pace. From this situation the game could unfold one of two ways; either the flow of play could be re-established through defensive excellence or an attacking error, or it could result in loss of possession or a try that would end the flow. When the perturbation results in a loss of possession or a try, it was then defined as a 'Critical Incident', sometimes the perturbation may be 'smoothed out', by good defensive play or poor attacking play, and not lead to a critical incident. A more in-depth approach is essential in order to derive a system for analyzing the existence of perturbations. In order to do so, entire phases of play must be analysed and notated accordingly.

Applying McGarry and Franks' [40] work on perturbations in squash, Hughes et al. [29] attempted to confirm and define the existence of perturbations in soccer. Using twenty English league matches, the study found that perturbations could be consistently classified and identified, but also that it was possible to generate specific profiles of variables that identify winning and losing traits. After further analyses of the 1996 European Championship matches (N=31), Hughes et al. attempted to create a profile for nations that had played more than five matches. Although supporting English League traits for successful and unsuccessful teams, there was insufficient data for the development of a comprehensive normative profile. Consequently, although failing to accurately predict performance it introduced the method of using perturbations to construct a prediction model. By identifying 12 common attacking and defending perturbations that exist in English football leading to scoring opportunities, Hughes, Dawkins and Langridge [24] had obtained variables that could underpin many studies involving perturbations. These twelve causes were shown to occur consistently, covering all possible eventualities and had a high reliability. Although Hughes et al. [29] had classified perturbations; the method prevented the generation of a stable and accurate

performance profile. In match play, teams may alter tactics and style according to the game state; for instance a team falling behind may revert to a certain style of play to create goal-scoring chances and therefore skew any data away from an overall profile.

In some instances, a perturbation may not result in a shot, owing to high defensive skill or a lack of attacking skill. Developing earlier work on British league football, Hughes et al. [24] analysed how the international teams stabilise or 'smooth out' the disruption. Analysing the European Championships in 1996, attempts were made to identify perturbations that did not lead to a shot on goal. Hughes et al. refined the classifications to 3 types of causes; actions by the player in possession, actions by the receiver and interceptions. Inaccuracy of pass accounted for 62% of the player in possession variables and interception by the defence accounted for the vast majority of defensive actions (68%). Actions of the receiver (12%) were dominated by a loss of control; however these possessions have great importance because of the increased proximity to the shot (critical incident). Conclusions therefore focussed on improvements in technical skill of players, however with patterns varying from team to team, combining data provides little benefit for coaches and highlights the need for analysing an individual teams 'signature'.

Squash is potentially an ideal sport for analyzing perturbations, and as such has received considerable attention from researchers. It is of a very intense nature and is confined to a small space. The rhythms of the game are easy to see, and the rallies are of a length (mean number of shots at elite level = 14 [21]) that enables these rhythms and their disruption i.e. perturbations and critical incidents. In defining perturbations, it may help to understand the reasons for this occurrence, consider the simplest model of a dynamic oscillating system. What are the main parameters and do the equivalent variables in sports contests conform to the same behaviour?

Simple Harmonic Motion
Simple Harmonic Motion is the simplest model of periodic motion and has a simple elegance to it.

Figure 4: A schematic example of Simple Harmonic Motion (SHM) from Nave [48].
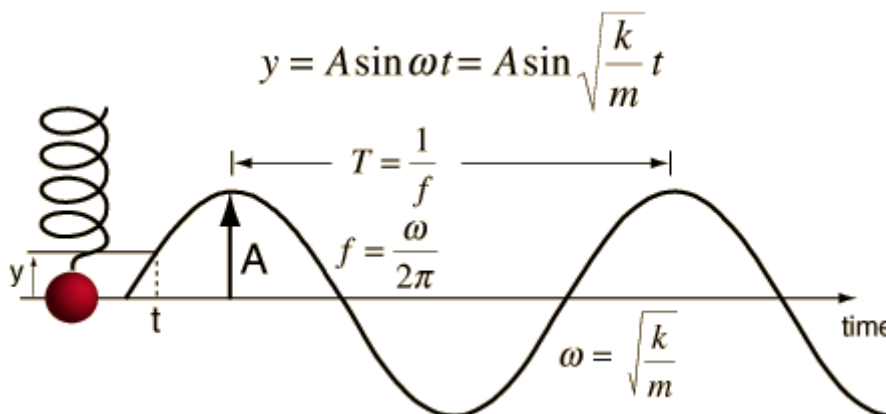
Simple harmonic motion refers to the periodic sinusoidal oscillation of an object or quantity. Take, for example, a mass suspended on a spring (see Fig. 4). When displaced the mass will move up and down in Simple Harmonic Motion, under the forces of the tension in the spring and gravity. This motion is not dissimilar in rhythm to a squash player's movement around the 'T' (the centre of the court) during a rally. Simple harmonic motion is executed by any quantity obeying the differential equation:

Displacement:-

$$Y = A \sin(wt).$$

This ordinary differential equation has an irregular singularity at ∞ .

Hence the velocity, (dx/dt):-

$$V = (dx/dt) = wA \cos(wt),$$

and the acceleration, (d^2x/dt^2):-

$$a = (d^2x/dt^2) = -w^2A \sin(wt) = -w^2 \, y$$

where the two constants, A and w, are determined from the dynamic conditions.

Figure 5: Distance Time Graph for squash players [55].
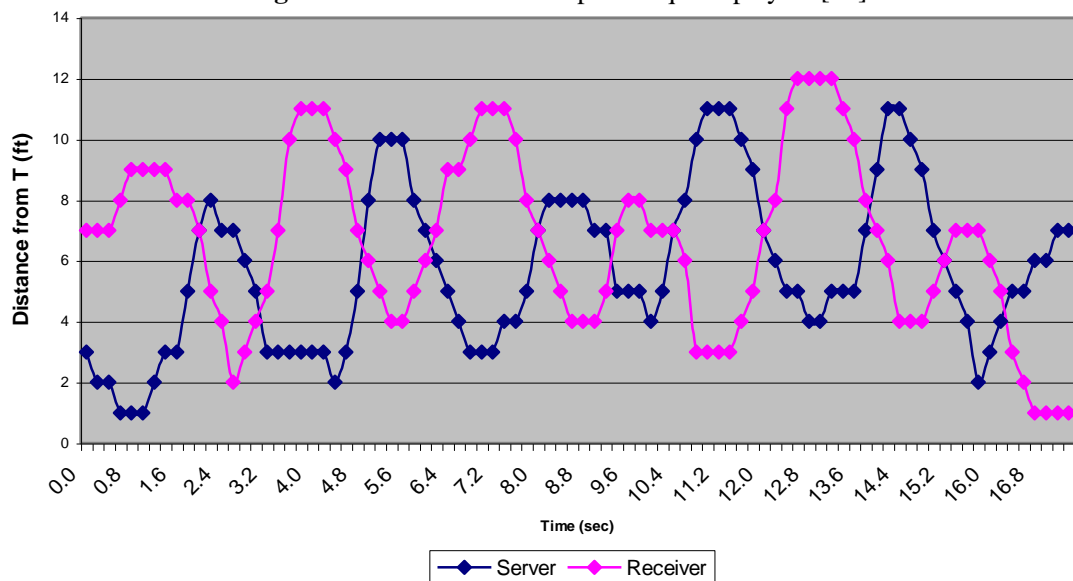
If the basic principles of Simple Harmonic Motion (SHM) are used as a model for the movement of a squash player in the game – with the 'T', the central focus of the court, as the origin of movement, then we can examine some simple concepts. This is the simplest form of oscillating motion and is therefore a good place to start. There are immediate limitations in these ideas – that the 'T' is neither the geometrical centre of the court, and the player does not really follow the velocity patterns of classic SHM. A squash player will return to the 'T' and stop, retaining their balance to be able to move in any direction, while in SHM the oscillating particle attains maximum velocity at the central point. It is surprising then how good the actual motion (Figure 5.) matches the SHM model.

By examining the 'stability' of these constants A (the amplitude) and w (the frequency of the movement) or T (the wavelength), it is hoped that we will get a reasonable fit to the model, and then be able to identify quantitatively, where the variations become large enough to be classed as perturbations. Squash Coaches often talk about players imposing their own rhythms on the game – can this be measured by identifying specific values of w, the frequency of play, for specific players and how does this vary against different opponents? This could be repeated in different individual and team sports.

Many physical systems undergoing small displacements, including any objects obeying Hooke's Law, exhibit simple harmonic motion. This equation arises, for example, in the analysis of the flow of current in an electronic CL circuit (which contains a capacitor and an inductor). If a damping force such as Friction is present, an additional term β (dx/dt) must be added to the differential equation and motion dies out over time. Perhaps more complex models such as Damped SHM can be pursued in further research.

Critical incident and/or perturbation analysis seems to offer a way of making sense of all the masses of data that is available from an analysis of a team sport such as soccer – even a 'simple' sport such as squash will have 4000 – 6000 bits of data per match. The resultant data output can be so overwhelming so as to leave coaches and sports scientists struggling to see significant patterns amongst the thousands and thousands of bits of data. This method appears to direct analysts to those important aspects of the data that shape winning and losing models.

Figure 6: Example of a perturbation that was 'smoothed out' at time = 2.6:08:
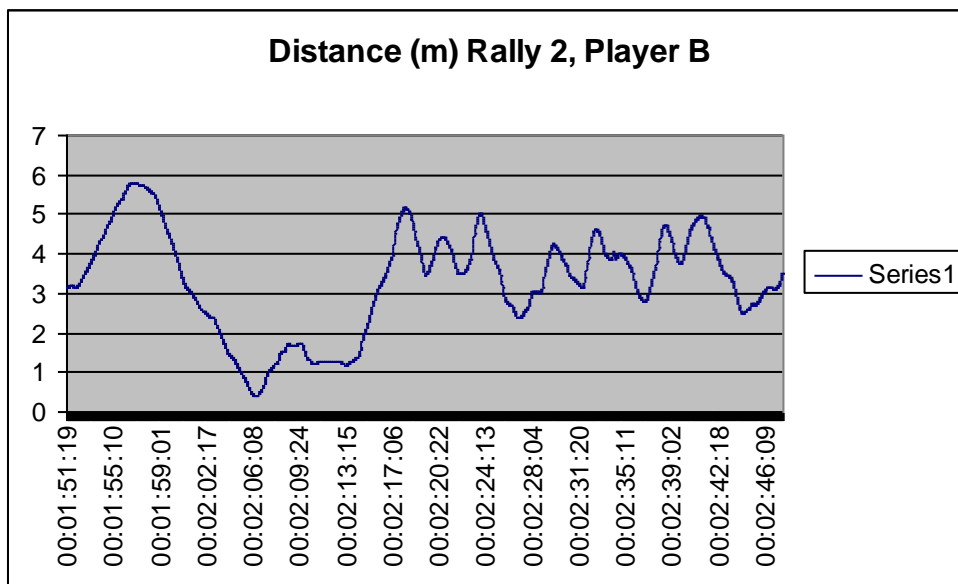
2.6.3 Statistical techniques

Every match is a contradiction, being at once both highly predictable and highly unpredictable.

Morris [46]

Historically, the prediction of sports performance has been a concept usually reserved for those associated with the betting culture. However, in reality, each and every person involved within sport will subconsciously process information to predict sports performance. Performance prediction could be described as the ability to draw conclusions upon the outcome of future performance based upon the combined interaction of previously gathered information, knowledge or data. For players and coaches, predictions are often made about forthcoming opponents based upon previous encounters and known traits. Therefore, is it not reasonable to assume that with valid and reliable information, using the correct techniques, the accurate prediction of performance should be possible? From a sports science perspective, the most common approaches to performance predication use large amounts of data and apply statistical techniques [7]. Human predictions however, are entirely derived from one's underpinning knowledge and

subjective bias, although the 'experts' are able to accommodate a greater understanding and opportunity for the element of chance and uncertainty, unlike computerised models of performance predication that are entirely statistically driven, such as Multiple Linear Regression.

In an evaluation of human and computer-based prediction models for the 2003 Rugby World Cup (Table 8), O'Donoghue and Williams [53] identified that the best human predictor performed better than any computer-based model, although the mean score of all the human predictions fell below each of the computer-based models that were used, unquestionably a result of the subjectivity within human prediction. Interestingly, in a similar study during the 2002 Soccer World Cup [55], although the best computer-based models outperformed the human predictors once again, their overall effectiveness in predicting results was far inferior compared to the 2003 Rugby World Cup. Ironically, only the human-based focus group was able to predict 4 of the 8 quarter-finalists and no method predicted more than 1 of the semi-finalists, whereas *all* of the computer-based models for the 2003 Rugby World Cup predicted 7 of the 8 quarter-finalists and 3 of the 4 semi-finalists. However, this is understandable considering the inherent differences between soccer and rugby union. Very few upsets occur within rugby union and with only one drawn match during World Cup rugby from 1987 [54], results generally go to form. Research by Garganta and Gonçalves [13] led to the notion that among team sports, soccer presents one of the lowest success rates in the ratio of goals scored to the number of attacking actions performed, subsequently increasing the likelihood of drawn matches and upsets. This considered, unlike rugby union in which the number of points scored is far greater, the accurate prediction of soccer matches is far more difficult, a notion shared by O'Donoghue and Williams [53] and demonstrated by O'Donoghue et al. [54].

Table 8: Marks awarded for each prediction.

Method	Marks Awarded					
	Group Matches (40)	Quarter-finals (4)	Semi-finals (2)	3rd Place Play Off (1)	Final (1)	Total (48)
Human-based methods						
Best Individual Human	39.00	4.00	2.00	1.00	0.00	46.00
Mean Human Prediction	35.63	3.45	1.17	0.26	0.31	40.66
Expert Focus Group	37.00	4.00	1.00	0.00	1.00	43.00
Computer-based methods						
Multiple linear regression (satisfies assumptions)	38.00	4.00	1.00	0.00	0.00	43.00
Multiple linear regression (violates assumptions)	38.00	4.00	1.00	0.00	1.00	44.00
Binary logistic regression	37.00	4.00	1.00	0.00	1.00	43.00
Neural networks with numeric input	37.00	4.00	1.00	0.00	1.00	43.00
Neural network with binary input – 4 middle layer nodes	34.00	3.00	1.00	0.00	1.00	39.00
Neural network with binary input – 8 middle layer nodes	34.00	4.00	1.00	0.00	1.00	40.00
Neural network with binary input – 16 middle layer nodes	36.00	4.00	1.00	0.00	1.00	42.00
Neural network with binary input – 32 middle layer nodes	37.00	4.00	1.00	0.00	1.00	43.00
Simulation Program	38.50	4.00	1.00	0.00	1.00	44.50

Making suggestions upon the types of data that should be used is difficult and ultimately reliant upon what is actually available. Historically, research in soccer, and similar team sports, that has attempted to identify the characteristics of a successful team has used game related performance indicators rather than factors such as distance travelled

(Hughes et al., [27], Yamanaka et al., [67]). By using process orientated data, such pass completion, shots on target and entries into the attacking third for example, a more accurate picture of a team's abilities would be created which directly relates to the dynamic processes involved in soccer. Although large databases of such information are available, the validity of the data in terms of defining successful performance is questionable and unsubstantiated. In order for performance prediction to move forward, not only within soccer, it is imperative that issues such as this are addressed, along with the continued development of valid and reliable methods of performance prediction.

By using MLR, a number of conditions must be accepted in considering its use as a prediction tool. The method is not based upon any 'artificial learning' process in order to generate predictions and predicts each game on its own merit without consideration for other factors. The simulation package used by O'Donoghue et al. [54] favoured Brazil to win the 2002 WC, rather than France who were the strongest team in the tournament. The model took into consideration the probabilities of qualifying from the group stages and then progressing through the knock-outs against different ranked opposition and predicted that Brazil had a greater chance of winning the fixture against France, should they have actually qualified from the group stages, because of the events preceding the potential tie. However, MLR would simply identify that France had both a superior rank and far less travelling distance than Brazil and would predict France to win the tie. This simplistic approach of MLR and indeed other algorithmic methods such as binary logistic regression is their fundamental drawback, although the results of these research papers quoted proved that even the most simplistic approach can be relatively effective.

2.6.4 Artificial intelligence

Bartlett [4] presented a broad overview of Artificial Intelligence (AI) encompassing speech recognition, natural language processing, computer vision (which would include on-line motion analysis systems that automatically track and assign markers, e.g. EVA real time, Vicon), and decision making. He suggested that the intelligent core of AI included:

➢ Expert systems :
  o rule-based,
  o fuzzy,
  o frame-based, and
  o their uses.

➢ Artificial neural networks (ANNs):
  ▪ biological and artificial neural networks,
  ▪ the Perceptron,
  ▪ multi-layer neural networks,
  ▪ recurrent neural networks (we won't consider these),
  ▪ self-organising neural networks, and
  ▪ their uses.

2.6.4.1 Expert Systems

Rule-based Expert Systems
These are, effectively, a database combined with a knowledge base, 'reasoning' and a user interface [49]. They can encompass a knowledge base that contains specific knowledge for 'domain', e.g. diagnosis of sports technique 'errors'. The 'reasoning' comes from rules that can be relations, recommendations, directives, strategies, or heuristics. They can include logic operations, just as a computer programme will contain conditional operators. The inference engine 'reasons' by linking the rules with the facts,

and finally, the explanation facilities explain how a conclusion was reached and why a specific fact is needed.

Uncertainty in Expert Systems
In most applications, knowledge rules based on simple logic, as in the mixed technique example, do not apply. This 'uncertainty' can be managed by (Bayesian) probability theory, e.g.

- IF 'counter-rotation' is high
- THEN 'technique' is mixed (p=0.8).

This example was chosen to illustrate that much information is vague – 'high' in the above example has in recent research varied from 10 to 20 to as much as 30 to 40 degrees. This type of information is classed as 'fuzzy'.

Fuzzy Expert Systems
The difference between 'crisp' and 'fuzzy' knowledge is shown for fast bowling in Figure 7.This type of representation deals with fuzzy knowledge, which may be shown by linear or non-linear (e.g. sigmoidal) functions. Qualitative statements ('Hedges'), such as 'a little' or 'very' can be used to modify the fuzzy shapes. Domain knowledge can be interpreted by fuzzy rules, similar to those for rule-based systems but using fuzzy logic not simple logic. They offer great potential in quantifying the vagueness of much knowledge.

Frame-based Expert Systems
In these expert systems, knowledge is represented as 'frames', with 'attributes', rather as in many conventional databases. They 'structure' knowledge, for example a frame-based system might contain a player's name, weight, height, age, position in last game, goals scored, passes made, tackles made, infringements and so on. By using 'object-oriented programming' frames communicate with each other by rules. Frames can belong to 'classes', e.g. forwards, mid-fielders, defenders, keepers.

Expert Systems - Advantages
•Separate knowledge from processing, unlike conventional programs.
•Provide an explanation facility.
•Can deal with incomplete and vague data.
•Can model fuzzy human decision-making.
•Are good for diagnosis.
•'Shells' for development of expert systems are widely available (e.g. add-ons to MATLAB).

Expert Systems - Disadvantages
•Need to acquire knowledge from experts; this is a major problem.
•Very domain-specific; the fast bowling system could not be used for javelin throwing.
•Opaque relationships between rules.
•In general, do not have an ability to 'learn'.
•Have to manage conflicts between rules.
•Ineffective rules searching – trawl through all rules in each cycle.
•Relatively complex programs.

Expert Systems in Performance Analysis
Barlett [4] reported that a search of Medline with 'expert systems' yielded 480 references, but when these were redefined toward 'sport' or 'exercise' the results were reduced to 9 strikes, and none of these was on sport. Expert Systems in gait analysis [8, 14, 62, 65] offer a great deal more than any other areas of performance analysis. They have been

shown to be relatively successful and can identify abnormalities that human observations missed [34].

The development and application in gait analysis of such advanced systems can be explained by a number of contributory factors [4]. Gait analysis is complex but there has been a strong developmental motivation – patient health. Clinicians are expensive and invariably very well paid. There are very confined expert domains and the analyses are laboratory-based so automatic systems are commonplace. There are vast amounts of data and there are many experts in this field.

Figure 7: A comparison of 'crisp' and 'fuzzy' knowledge for fast bowling [3].
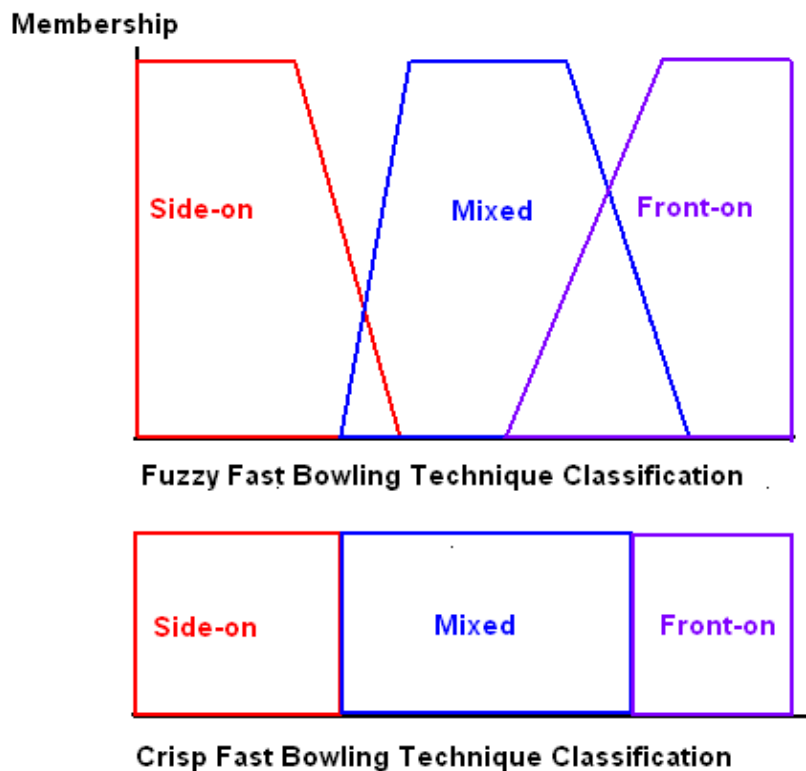
The lack of applications in Performance Analysis of sport can be attributed to the facts that in sport the analyses are even more complex than gait analysis. There has been weak developmental motivation, as in sport performance coaches and sport scientists are not expensive – they are poorly paid generally in comparison to the health profession. There are further difficulties because the analyses are often field-based, prohibiting, to date, automatic tracking. The great number and variety of sports, create a large amount of various specialists – technique analysts, notational analysts, thus in turn making a broad expert domain. There are not a lot of data for technique analysis expert systems and comparatively fewer experts in each specific field than gait analysis.

It is still surprising that these types of expert systems have not been taken up and applied in the area of coaching. The type of diagnostic logic needed in the coaching of technique lends itself perfectly to the structure of rule-based expert systems. Perhaps this is an area that we can see developing hand in hand with coach development and education? In terms

36

of modelling sports performance and for research purposes, other AI tools appear more promising.

2.6.4.2 Neural networks

The future of research within performance prediction undoubtedly lies in the development of artificial intelligence systems such as neural networking. In simplistic terms, an artificial neural network is a computer simulated mathematical model of the neurons within the human brain which is able to learn from experience [49]. It was suggested by O'Donoghue and Williams [53] that neural networks are commonly used to analyse complex information and could ultimately be used as a tool for predicting soccer matches using the complex game related data identified previously. However, it must be accepted that predicting the outcome of sports performance consistently and accurately is unlikely to be made possible even with the most sophisticated of computerised systems, after all, it is the unpredictability of sport that continues to capture the imagination of millions of people across the world.

The term neural network refers to a specific type of computational model, whose origins are found in the properties of neurons (i.e., nerve cells) and their interactions. Simply put, neurons receive inputs and send outputs in relay fashion, that is to say that the outputs serve as inputs to other neurons in the network. There are two main tasks that neural networks are used for, and these two tasks lend the specific characteristics to the two types of networks that service these tasks. The first task that neural networks can be used for is to recognise patterns, such as pictures, situations, processes, or any other types of structured objects. For this task, a special type of neural network is used which is called a Kohonen Feature Map (KFM). The characteristics of a KFM are that these networks are able to learn by forming clusters in response to a set of training patterns. These learning features allow the network to later identify other arbitrary patterns by associating these patterns to one of the already learned clusters within the network.

Figure 8: Neural Network Architecture: Fully Interconnected Three Layer Network.

37
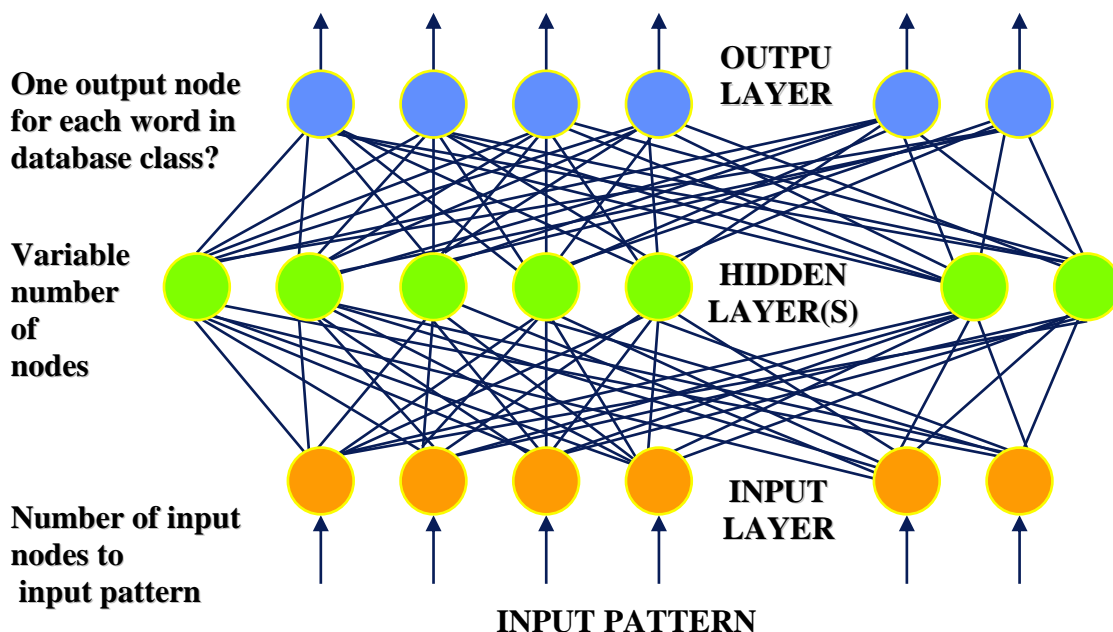
The second task that neural networks can be used for is in the selection of actions, in particular those actions that are identified as being optimal for given situations. One application of these tasks is the ability to make behavioural decisions in the context of sports games. For this task of selecting actions, a different type of network is used that often makes use of the attributes of "feed forward" and "back propagation".
The 'training' or 'learning' process is as follows:

❑ ANNs are capable of learning through experience by adjusting network weightings (w_i).

❑ For example, $Z = sum(x_i.w_i)$ for i = 1 to n; $Y = sign(Z)$, where Y = +1 if $Z \geq \theta$ or -1 if $Z < \theta$, where θ is a threshold value, normally 0-1; x_i are inputs and Y the output.

❑ Set all weights and thresholds randomly to number in range -2.4/N to 2.4/N; N=no. of neurons in network.

❑ Calculate outputs of hidden neurons (z_j):

$$z_j = sigmoid((sum(w_{ij}.x_i) - \theta_j).$$

❑ Ditto for output layer:

$$y_k = sigmoid((sum(w_{jk}.z_j) - \theta_k).$$

❑ Compare outputs with known values (y_{dk});
$e_k = y_{dk} - y_k.$

❑ Update weights in back-propagation network by propagating output errors backward from output to middle to input layers:

$$w_{jk} = w_{jk} + \Delta w_{jk} \ (= \alpha.z_j.\delta_k),$$

where δ_k is related to e_k;

similarly: $w_{ij} = w_{ij} + \Delta w_{jk} \ (= \alpha.x_i.\delta_j).$

❑ Repeat until errors acceptably small.

Pattern learning with Kohonen Feature Maps

Each neuron of a KFM contains an information structure that stands for the representation of a pattern. For example, this information structure can consist of a pair of coordinates. The pattern, then, is one of a two-dimensional position, which is represented by its x- and y-coordinate. As sketched in Figure 9 (left graphic), a learning impulse, which itself represents a two-dimensional pattern (i.e., position), can be applied to the network. The result of the stimulus is to bring about local changes within the network. This process of change is usually controlled by a set of rules and parameters that determine how the network adapts to a specific learning situation. Of course, different stimulus patterns will bring about different changes in the network. Once the learning process is finished, the neural network, as sketched in Figure 9 (right graphic), forms clusters that represent the structure, or distribution, of the inputs that have been "learned". In the second phase, the network can be used to recognise other inputs, or patterns – i.e. the network is given the pattern and then indicates the cluster of

neurons that the pattern belongs to. From a higher point of view, the clusters can each be thought of as representing a certain set of properties or attributes, and the new inputs are assigned to these properties in order for them to be recognised or identified.

In Figure 10 (right graphics) an example of a neural network trained on data from squash contests is presented. The dots stand for types of processes, or inputs, and the diameters of the dots indicate the frequency, or strength, of the types of inputs that occurred during the game. Once a sports contest is represented in terms of network and process information, as in Figure 10, then a lot of game analysis concerning the frequencies, distributions, and the success of actions, as well as tactical structures and the success of action sequences can be undertaken.

ANN - advantages
❖ Learn by experience; in the case of self-organising ANNs, without a teacher!
❖ Are good for classification, clustering and prediction tasks.
❖ Can be adapted for inexact or incomplete data through 'fuzzy' ANNs.
❖ Are widely available, e.g. the MATLAB Neural Network Toolbox, and relatively simple programs.
❖ Seem to mimic brain processes.
❖ Provide link to dynamic systems theory as non-linear program representations of non-linear biological systems.

Figure 9: Learning step (left) and information clusters on a Kohonen Feature Map (right) [45].
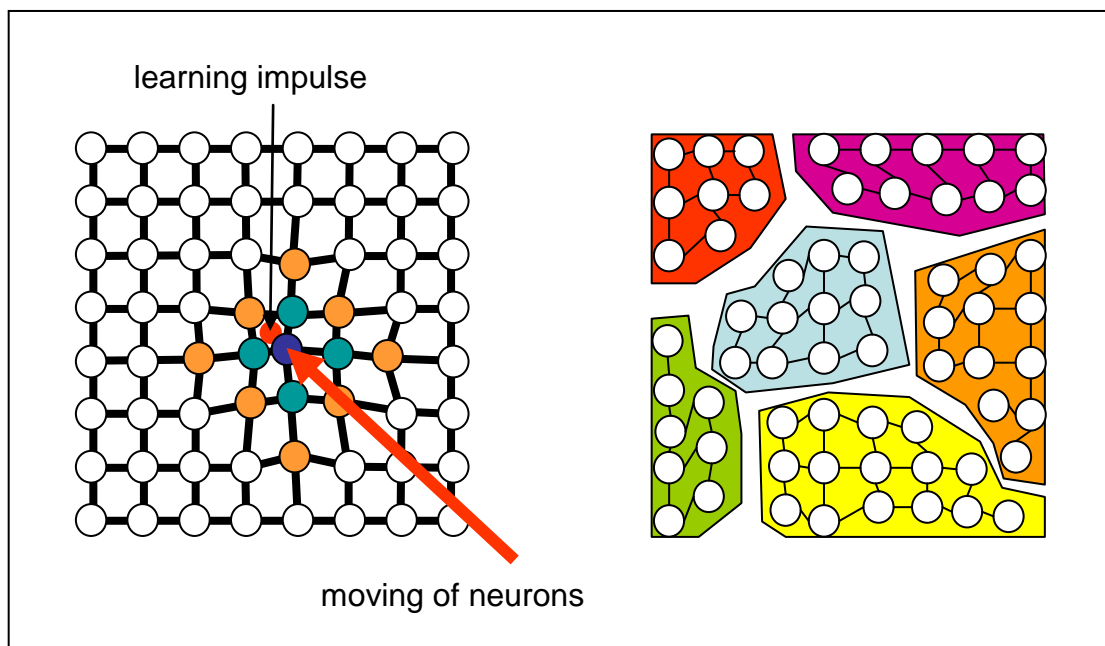
ANN - disadvantages
▪ They are opaque 'black boxes' with no explanation of the reasoning process.
▪ The 'rules' within the non-linear network are not well understood; the non-linear characteristics may prohibit simple and understandable rules.
▪ To validate their output, they need test cases for which output is known.
▪ They often do not work well for inputs outside the range used for learning.
▪ Back propagation is very slow, although widely used for pattern recognition.

- Kohonen SOMs need lots of learning data and aren't dynamic.
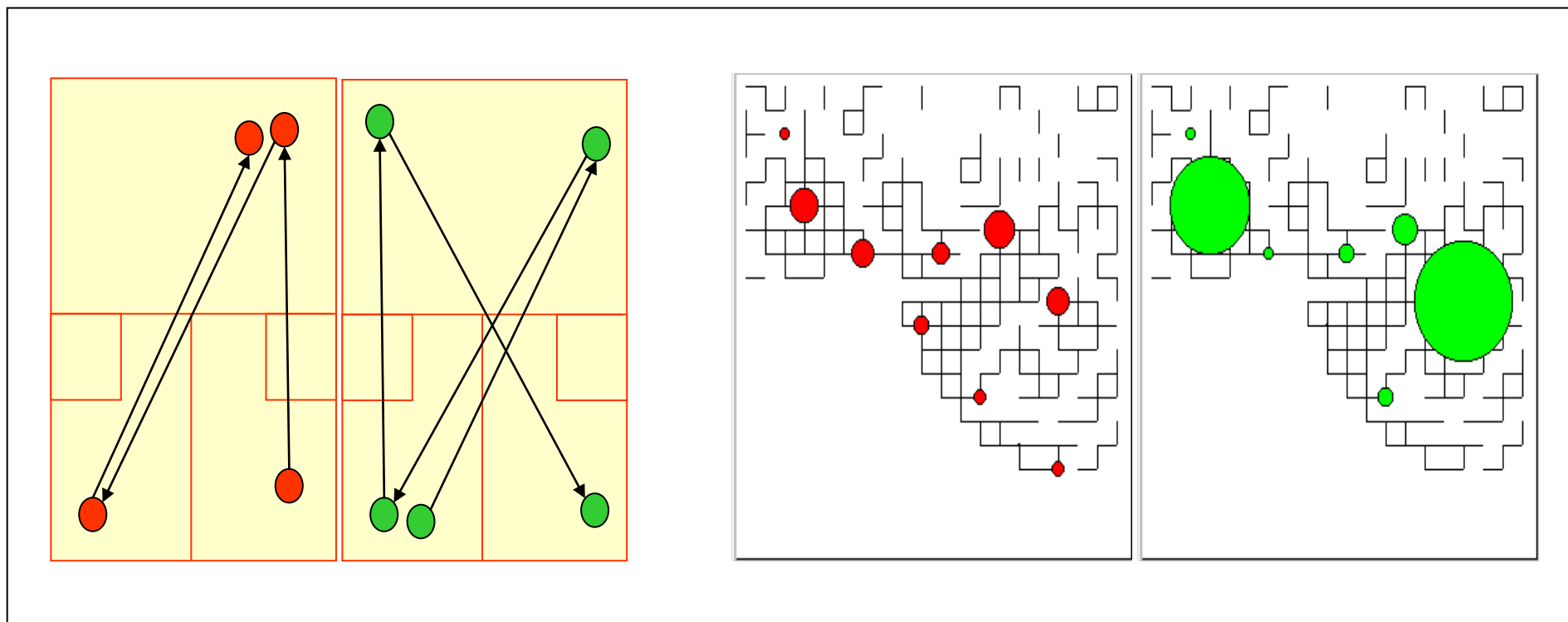
Figure 10: Squash processes on the court (left) and process clusters on a squash network (right) [45].

Summary - ANNs in Performance Analysis

ANNs have been more widely used than expert systems in performance analysis [56, 57, 58, 5, 60, 35, 36, 68, 69, 45]. In technique analysis, Kohonen self-organising maps have been claimed to reveal the 'wood' rather than the 'trees'. Whilst some of the research is difficult to replicate, the potential uses do seem to be very good – particularly for dynamically controlled networks – their prospective applications in notational analysis have been recognised if not extensively realised. Most intriguing is that they represent a possible and an important link to non-linear dynamics and support ecological motor control views of movement variability as omnipresent. Finally the potential for multi-layer ANNs to optimise decision-making processes [46] has not yet been realised or even attempted.

2.643 Summary of modelling and prediction

In an evaluation of human and computer-based prediction models, including neural networks, for the 2003 Rugby World Cup, O'Donoghue and Williams [53] identified that the best human predictor performed better than any computer-based model, although the mean score of all the human predictions fell below each of the computer-based models that were used, unquestionably a result of the subjectivity within human prediction. Interestingly, in a similar study during the 2002 Soccer World Cup [54], although the best computer-based models outperformed the human predictors once again, their overall effectiveness in predicting results was far inferior compared to the 2003 Rugby World Cup. Ironically, only the human-based focus group was able to predict 4 of the 8 quarter-finalists and no method predicted more than 1 of the semi-finalists, whereas *all* of the computer-based models for the 2003 Rugby World Cup predicted 7 of the 8 quarter-finalists and 3 of the 4 semi-finalists. However, this is understandable considering the inherent differences between soccer and rugby union. Very few upsets occur within rugby union and with only one drawn match during World Cup rugby from 1987 [54], results generally go to form. Research by Garganta and Gonçalves [13] led to the notion that among team sports, soccer presents one of the lowest success rates in the ratio of goals scored to the number of attacking actions performed, subsequently increasing the likelihood of drawn matches and upsets. This considered, unlike rugby union in which the number of points scored is far greater, the accurate prediction of soccer matches is far more difficult, a notion shared by O'Donoghue and Williams [53] and demonstrated by O'Donoghue et al. [54].

There are a range of statistical methods that can be used to predict performance, such as Multiple Linear Regression, Discriminant Function Analysis, Binary Logistic Regression, forms of artificial intelligence such as Artificial Neural Networks, or indeed any combination of these. The analysis of sports behaviours using stochastic processes would seem to represent the most complete type of model for sports contests to date. This observation is unsurprising given the widespread uses of statistics based on probabilities in many different fields (e.g., actuarial tables, insurance statistics, stock market forecasts, sports betting, etc.). Indeed, the behaviours in sports contests would seem to be good candidates for this type of analysis, not least since the strategies for sports contests are often designed on the basis of future expectations (e.g. the use of strategies based on "percentage play"). This said, some limitations of this type of analysis (for squash contests) were reported, indicating that any such type of system description on the basis of probability is not as straightforward as might otherwise be imagined. These results led in time to the consideration of a new type of description for sports contests. However, the data required to pursue the description of sports contests as a dynamical system remains sparse at the time of writing.

  Finally, we briefly considered the use of neural networks for recognizing structures, or processes, within sports contests. Each of these system descriptions, while incomplete, may assist in our understanding of the behaviours that form sports contests. Furthermore, these descriptions for sports contests need not be exclusive of each other, and a hybrid type of description (or model) may be appropriate in the future, a suggestion that remains only a point of conjecture at this time. For these reasons, further research on sports contests using various types of system descriptions is warranted. The use of simulation procedures in the modelling of sports contests will continue to be a useful technique in the pursuit of these objectives.

3. Conclusions

It has been suggested that the processes necessary for a Notational Analyst working either as a consultant or an academic researcher are as follows:-

1. defining performance indicators,

2. determining which are important,

3. establishing the reliability of the data collected,

4. ensuring that enough data have been collected to define stable performance profiles (performance profiling),

5. comparing sets of data,

6. modelling performances.

It was concluded that recent research has demonstrated:-

✓ clear methods for determining which performance indicators are relevant and which are more important,
✓ simple absolute measures of reliability need to be used together with accepted non-parametric measures of variance,
✓ that if performance profiles of teams or individual athletes are being applied, then some measures of confidence in the stability of these profiles need to expressed,
✓ the comparison of sets of data in notational analysis needs to be considered carefully, as the data are usually non-parametric and conform most likely to Poisson and Binomial distributions. The use of x^2 seems to be a simple answer, however, as soon as more complex models and higher-order comparisons are required, the approach based on log-linear models is shown to be more effective,
✓ the sensitivity of these x^2 and log-linear models to the small differences in performance, that differentiate between winning and losing at the elite level, is open to question,
✓ that there are many techniques used to model sport, some of these are providing the greatest challenges to notational analysts and mathematicians alike.

It is clear that the working notational analyst must have a broad set of mathematical and statistical skills and be prepared to maintain and extend those skills just as the research in this area develops the knowledge base. Further research is urgently needed in some of these areas:-

❖ How do we define performance indicators in a general and generic sense? At the moment they are arbitrarily defined for each sport depending upon the subjective opinions of the analyst and/or coach.
❖ The statistical methods that we use are improving, but more work needs to be done on making the more sophisticated systems more transparent, in terms of how they relate to the experimental aims of the comparisons, and also the basic practical demand of them being easier to apply.
❖ The sensitivity of the tests needs to be examined – how can we determine the significant differences in performance when the increments of comparison are very small?
❖ More research in modelling in performance analysis is vital as we extend our knowledge and databases into those exciting areas of prediction.

Using the word prediction in the same phrase as sport almost certainly creates a form of an oxymoron, because of the inherent nature of sport. Nevertheless, working towards the extended aims of modelling, and therefore forecasting, must be the most exciting of the ways to further develop performance analysis.

References

[1]   D. Alexander, K. McClements and J. Simmons, "Calculating to win". *New Scientist.* 10 December (1988), 30-33.
[2]   G. Atkinson and A. Nevill,"Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine". *Sports Med*., **26** (1998), 217-238

[3]   R. M. Bartlett, "The science and medicine of cricket: a overview and update". *Journal of Sports Sciences*, **21** (2003), 733-752.

[4]   R. M. Bartlett, **"**Artificial intelligence in performance analysis". In P. O'Donoghue and M. Hughes (eds.) *Notational Analysis of Sport VI*, (2004) (In press).

[5]   H. U. Bauer and W. I. Schöllhorn, "Self-organizing maps for the analysis of complex movement patterns" *Neural Processing Letters*, **5** (1997), 193-199.

[6]   J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement". *The Lancet*, **1** (1986), 307-310.

[7]   A. G. Buchner, W. Dubitzky, A. Schuster, P. Lopes, P. G. O'Donoghue, J. G. Hughes, D. A. Bell, K. Adamson, J. A. White, J. M. C. C. Anderson, and M. D. Mulvenna, "Corporate evidential decision making in performance prediction domains". *13th Conference of Uncertainty in Artifical Intelligence (UAI'97),* Providence, Rhode Island, (1997).

[8]   J. M. Dziezanowski, J. R. Bourne, R. Shiavi, H. S. H. Sandall, and D. Guy, "GAITSPERT: An expert system for the evaluation of abnormal human locomotion arising from stroke", *IEEE Transactions on Biomedical Engineering*, **32** (1985), 935-942.

[9]   N. Eniseler, B. Doðan, S. Aydin, S. V. Üstün, and Y. Taþkiran, "Area and time analysis of duels in matches of the Turkish football team in elimination stages of the European Cup". In (ed. M. Hughes) *Notational Analysis of Sport III.* Cardiff: CPA, UWIC (2000), pp. 121 – 132.

[10]   J. C. Flanagan, "The Critical Incident Technique". *Psychological Bulletin*, **51** (1954), 327-358.

[11]   I. M. Franks and T. McGarry, "The Science of Match Analysis", in *Science and Soccer* (ed T. Reilly), London:E. & F.N. Spon, (1996).

[12]   N. Fuller, "*Game modelling in netball through computerised performance analysis".* Master´s Dissertation, Sheffield University, (1994).

[13]   J. Garganta & G. Goncalves, "Comparison of successful attacking play in male and female Portuguese national soccer teams". In M.D. Hughes (ed) ***Notational Analysis of Sport - I & II***, Cardiff : UWIC (1996), pp. 79 - 85.

[14]   D. E. Hirsch, S. R. Simon, T. Bylander, M. A. Weintraub, and P. Szolovits, "Using causal reasoning in gait analysis", *Applied Artificial Intelligence*, **3** (1989), 253-272.

[15]   M. Hughes, "A comparison of the patterns of play of squash". In *International Ergonomics '85* (eds I.D.Brown, R.Goldsmith, K.Coombes & M.A.Sinclair), London: Taylor & Francis, (1985), pp. 139 -141.

[16]   M. Hughes, "A review of patterns of play in squash". In *Sports Science* (eds J.Watkins, T.Reilly & L.Burwitz), London: E. & F.N. Spon, (1986).

[17]   M. Hughes, "Performance analysis – a 2004 perspective". *eIJPAS*, **4** (1), (2004), (In press).

[18]   M. Hughes and R. Bartlett, "The use of performance indicators in performance analysis". *Journal of Sports Science,* 20 (2002), 739 – 754.

[19]   M. Hughes and S. Clarke, "Surface effect on patterns of play of elite tennis players". In *Science and Racket Sports* (edited by T. Reilly, M. Hughes, and A. Lees), London: E&FN Spon., (1995), pp. 272-278.

[20]   M. Hughes and P. Knight, "Playing patterns of elite squash players, using English and point-per-rally scoring". In T. Reilly, M. Hughes & A. Lees (Eds), *Science and Racket Sports*. E & FN Spon: London, (1995), pp. 257-259.

[21]   M. Hughes and C. Robertson, "Using computerised notational analyisis to create a template for elite squash and its subsequent use in designing hand notation systems for player development". In A. Lees, I. Maynard, M. Hughes and T. Reilly (Eds), *Science and Racket Sports II*. E & FN Spon: London, (1998), pp. 227-234.

[22]   M. Hughes & D. Williams, "The development and application of a computerised Rugby Union notation system". *Journal of Sports Sciences,* **6** (1988)**,** 254-255.

[23]   M. Hughes, S-M. Cooper and A. Nevill, "Analysis procedures for non-parametric data from performance analysis". *eIJPAS,* 2 (2002), 6 – 20.

[24]   M. Hughes, N. Dawkins and C. Langridge, "Perturbations not leading to shots in soccer". In M.Hughes (ed.) *Notational Analysis of Sport III,* Cardiff: CPA, UWIC, (2000), pp. 108 -116.

[25]   M. Hughes, S. Evans and J. Wells, "Establishing normative profiles in performance analysis". *eIJPAS,* **1** (2001)**,** 4 – 27.

[26]   M. Hughes, I. M. Franks & P. Nagelkerke, "A video-system fo the quantitative motion analysis of athletes in competitive sport". *J.H.M.S.*, **17** (1989), 212-227.

[27]   M. Hughes, K. Robertson, & A. Nicholson, "An Analysis of 1984 World Cup of Association Football". In T.Reilly, A.Lees, K.Davids & W.Murphy (Eds.), *Science and Football.* London: E. & F. Spon, (1988).

[28]   M. Hughes, J. Wells, and C. Matthews, "Performance profiles at recreational, county and elite levels of women's squash". *Journal of Human Movement Studies*, **39** (2000), 85-104.

[29]   M. Hughes, N. Dawkins, R. David and J. Mills, "The perturbation effect and goal opportunities in soccer". *Journal of Sports Science,* **16** (1997), 20.

[30]   N. James, S. D. Mellalieu and N. M. P. Jones, "The development of position-specific performance indicators in professional rugby union". (Manuscript submitted to *Journal of Sports Sciences*), (2004).

[31]   S. Kelso, *"Dynamic Patterns".* Champaign, Illinois: Human Kinetics, (1999).

[32]   S. P. Ladany and R. E. Machol (Eds.), *"Optimal Strategies in Sports".* Amsterdam. North Holland, (1977).

[33]   O. Larsen, H. Zoglowek and K. Rafoss, "An analysis of team performance for the Norwegian women soccer team in the Olympics in Atlanta 1996". In (ed. M. Hughes) *Notational Analysis of Sport III.* Cardiff: CPA, UWIC, (2000), pp. 158 – 166.

[34]   A. C. Latham, and R. M. Bartlett, "The use of artificial intelligence in the analysis of sports performance: A review of applications in human gait analysis and future directions for sports biomechanics", *Journal of Sports Sciences*, **13** (1995), 229-237.

[35]   A. Lees and G. Barton, "A characterisation of technique in the soccer kick using a Kohonen neural network analysis". In *Proceedings of the 5th World Congress of Science and Football.* (In press), (2004).

[36]   A. Lees, G. Barton and L. Kershaw, "The use of Kohonen neural network analysis to qualitatively characterize technique in soccer kicking". *Journal of Sports Sciences*, **21** (2003), 243-244.

[37]   P. Luhtanen, A. Belinskij, M. Häyrinen and T. Vänttinen, "A computer aided team analysis of the euro 2000 in soccer. *EIJPAS,* **1** (2002), 74 – 83.

[38]   K. Lyons and M. Hughes, "The science of chaos". *New Scientist,* July, (1996).

[39]   T. McGarry and I. M. Franks, "A stochastic approach to predicting competition squash match-play". *Journal of Sports Sciences,* **12** (1994)**,** 573-584.

[40]   T. McGarry and I. M. Franks, "Modelling competitive squash performance from quantitative analysis"**.** *Human Performance,* **8** (1995), 113-129.

[41]   T. McGarry and I. M. Franks, "Analysing championship squash match play: In search of a system description". In S. Haake (ed.) *The Engineering of Sport.* Rotterdam: Balkema, (1996a), pp. 263-269.

[42]   T. McGarry and I. M. Franks, "In search of invariance in championship squash." In M.D. Hughes (ed) *Notational Analysis of Sport - I & II,* Cardiff : UWIC, (1996b), pp. 264 – 271.

[43]   T. McGarry and I.M. Franks, "Development, application and limitation of a stochastic Markov model in explaining championship squash performance." *Research Quarterly for Exercise and Sport,* **67** (1996c)**,** 406-415.

[44]   T. McGarry and I. M. Franks, "In search of invariant athletic behaviour in competitive sport systems: An example from championship squash match-play." *Journal of Sports Sciences,* **14** (1996d), 445-456.

[45]   T. McGarry and J. Perl, "Models of sports contests: Markov processes, dynamical systems and neural networks". In *Notational Analysis of Sport: Systems for Better Coaching and Performance in Sport* (edited by M.D. Hughes and I.M. Franks), London: Routledge, (2004), pp. 227-242.

[46]   D. Morris, *"The Soccer Tribe".* London: Jonathan Cape Ltd, (1981).

[47]   F. Mosteller, "A resistant analysis of 1971 and 1972 Professional Football", in *Sports, Games & Play* (ed J. H. Goldstein), New Jersey: Lawrence Erlbaum Associates, (1979), pp. 371-401.

[48]   C. R. Nave, http://hyperphysics.phy-astr.gsu.edu/hbase/shm.html. (2004), (July).

[49]   M. Negnevitsky, *"Artificial Intelligence: A Guide to Intelligent Systems".* Harlow: Addison-Wesley, (2004).

[50]   A. M. Nevill, G. Atkinson, M. Hughes and S-M. Cooper, "Statistical methods for analysing sport performance and notational analysis data". *Journal of Sports Sciences*, **20** (2002), 829 - 844.

[51]   P. O'Donoghue, "Notational analysis of rallies in European club championship badminton". In M. Hughes and F. Tavares (eds.) ***Notational Analysis of Sport IV,*** Porto: Faculty of Sports Sciences and Education, University of Porto, Portugal. (2001), pp. 225 - 228.

[52]   P. O'Donoghue and W. Ingram, "Notational analysis of the cause of elite tennis players approaching the net in men's and ladies' singles at grand slam tournaments". In M. Hughes and F. Tavares (eds.) *Notational Analysis of Sport IV*, Porto, Portugal, (2001), pp. 248 - 256.

[53]   P. G. O'Donoghue and J. Williams, "An evaluation of Human and Computer-based predictions of the 2003 Rugby Union World Cup", *International Journal of Computer Science and Sport-e.,* (2004) (in press).

[54]   P. G.O'Donoghue, W. Dubitzky, P. Lopes, D. Berrar, K. Lagan, D. Hassan, A. Bairner and P. Darby, "An Evaluation of Quantitative and Qualitative Methods of Predicting the 2002 FIFA World Cup", *Proceedings of the World Congress of Science and Football V,* Lisbon, Portugal, (2003), pp. 44-45.

[55]   T. Peakman, "Perturbations in squash". Undergraduate dissertation. BSc Sports Coaching, UWIC, Cardiff, (2001).

[56]   J. Perl, "Artificial neural networks in sports: New concepts and approaches". *International Journal of Performance Analysis in Sport*, **1** (2001), 106-121.

[57]   J. Perl, "Game analysis and control by means of continuously learning networks". *International Journal of Performance Analysis in Sport,* **2** (2002a)*,* 21-35.

[58]   J. Perl, "Adaptation, antagonism and system dynamics". In G. Ghent, D. Kluka & D. Jones (eds.). *Perspectives. The multidisciplinary series of physical education and sport science,* **4***,* Oxford: Meyer & Meyer Sport, (2002b), pp. 105-125.

[59]   C. Reep and B. Benjamin, "Skill and chance in association football". *Journal of the Royal Statistical Society, Series A,* **131** (1968), 581-585.

[60]   W. I. Schöllhorn and H. U. Bauer, "Identifying individual movement styles in high performance sports by means of self-organizing Kohonen maps". In *Proceedings of the XVI ISBS 98 Konstanz* (edited by H.J. Riehle and M. Vieten), Konstanz: ISBS, (1998).

[61]   E. H. Shortcliff, "*Computer-Based Medical Consultation: MYCIN"*. New York: Elsevier, (1976).

[62]   R. D. Stacey, "Strategy as Order Emerging from Chaos"**.** *Long Range Planning.* **26** (1993)**.**

[63]   G. Tiryaki, S. Cicek, A. T. Erdogan, F. Kalay, A. T. Atalay, and F. Tuncel, "The analysis of the offensive patterns of the Switzerland soccer team in the World Cup, 1994". In M. Hughes (ed) *Notational Analysis of Sport - I & II*, Cardiff: UWIC, (1997), pp. 91 – 98.

[64]   K. Tracy, E. Montague, R. Gabriel and B. Kent, "Computer-assisted diagnosis of orthopaedic gait disorders". *Physical Therapy*, **59** (1979), 268-277.

[65]   P. Treadwell, K. Lyons and G. Potter, "The Predictive Potential of Match Analysis Systems for Rugby Union Football". In *Science and Football II* (eds T.Reilly, J. Clarys and A. Stibbe). London: E. & F. Spon, (1991), pp. 221 – 229.

[66]   W. Vincent, "*Statistics in Kinesiology",* Champaign, Ill.: Human Kinetics, (1999).

[67]   K. Yamanaka, M. Hughes and M. Lott, "An analysis of playing patterns in the 1990 World Cup for association football". In T.Reilly (Ed.), *Science and Football II.* London: E. & F. Spon, (1993), pp. 206 – 214.

[68]   B. Yan, and M. Li, "Shot put technique using an ANN AMT model". In *Proceedings of the XVIII International Symposium on Biomechanics in Sports, Volume 2* (edited by Y. Hong and D.P. Johns), Hong Kong: The Chinese University, (2000), pp. 580-584.

[69]   B. Yan and Y. Wu, "The ANN-based analysis model of the sports techniques". In *Proceedings of the XVIII International Symposium on Biomechanics in Sports, Volume 2* (edited by Y. Hong and D.P. Johns), Hong Kong: The Chinese University, (2000), pp. 585-589.

DERIVING A PROFIT FROM AUSTRALIAN RULES FOOTBALL: A STATISTICAL APPROACH

Michael J Bailey & Stephen R Clarke
School of Mathematical Sciences
Swinburne University
PO Box 218 Hawthorn, Vic, 3122
Australia

Abstract

Using match information gathered from 100 seasons of Australian Rules Football played prior to 1997, multiple linear regression model was used to identify and weight numerical features that could independently explain statistically significant proportions of variation associated with the outcome of matches. Prediction models constructed at both a team and player level were applied to matches played between 1997 and 2003 with results compared against an existing benchmark for AFL prediction and bookmakers' prices. Although bookmakers have appeared to improve in their price setting processes over the past seven years, it is still possible to derive an annual profit, with statistically significant improvement coming through the use of data derived at an individual player level.

1. Introduction

Commencing in 1897 with eight teams from Victoria only, the Victorian Football League (VFL) has grown into a national competition under the Australian Football League (AFL), and now boasts 16 teams from five Australian states.

Applying mathematical models to sport is not a new concept. In the past, attempts at modelling sporting outcomes were always hampered by the lack of quality data and the time taken to accurately enter the data for computer modelling. Although some work has been published on modelling AFL football, no attempt has been made to utilise all past matches to establish a prediction process. Clarke [1] uses only the names of the competing teams and the venue to predict winners, margins and chance of winning via an exponential smoothing process to develop team ratings and ground advantages. The program was originally optimised in 1980 using only one year's past data, and re-optimised in 1986 using 6 years of data. Even so, the program has proved it can match it with the media experts in picking winners and margins [2, 3]. Using similar data, Stefani [4] achieves similar success using a least squares procedure to produce team ratings and long term averages for home advantages. Both these approaches report around 67% accuracy in predicting winners.

With the growth of the Internet has come a rapid increase in the amount of readily accessible data from which to explore sporting outcomes. Use of this data should improve modelling accuracy. Bailey [5] uses team playing statistics such as turnovers and forward thrusts along with bookmaker's prices to predict match outcomes. This paper continues this approach by developing models using all previous match results, and investigates the additional benefits of incorporating individual player statistics in the prediction process.

The development of the Internet has been accompanied by a growth in sports betting. Several government controlled and private bookmakers take bets over the Internet on a range of outcomes in football. While the operator's percentage margin in traditional betting on horseracing is anywhere between 12 and 18%, it can be as low as 5% in Head to Head betting on the AFL. Contrary to popular perception, price setting by bookmakers is not an exact science, but rather a combination of experience, knowledge and intuition. Several Internet sites compare differing bookmaker prices, and by choosing the best available it is possible to operate in a market with a margin as low as 2 to 3%. This increases the chance that mathematical models can produce a long-term profit by performing better than the public in estimating probabilities.

While most AFL media experts predict winners, profitable betting requires predicting probabilities. For example, if for a \$1 bet Essendon starts favourite at \$1.50 against Fremantle \$2.40, and a prediction model gives Essendon a 57% chance of winning, betting on Essendon has an expected return of 57% of \$1.50 = 86 cents. However betting on Fremantle has an expected return of 43% of \$2.40 = \$1.03 for a profit or advantage of 3%. In order to maximise growth of wealth, a betting strategy should incorporate three specific

features in determining the size of the wager, namely bank size, size of advantage and probability of winning. Kelly [6] developed a betting strategy designed to maximise the growth of wealth by maximising the expected log of wealth. This formula can be effectively simplified to proportion of wealth bet = expected profit divided by maximum profit. So in the above example, we should bet $3/1.40 = 2.3\%$ of our stake. While Kelly is designed for a growing bank, there are some difficulties in using it to assess alternative betting strategies. For example, final profit depends on the order of bets. However for a wagering strategy such as the Kelly's to be successful, the modelling process must be able to produce a consistent profit whilst operating with a fixed bank size. Thus to determine feasibility, a fixed bank size of $1000 has been chosen with bet size determined by only two features, size of the perceived advantage and probability of winning.

As a benchmark to compare the more complicated methods developed here, we use that of Clarke[1]. The predicted winners of this fully automated program have been published in various media outlets, including newspaper, radio, TV and the WWW almost continually since 1981. For many years margins have also been published. Studies have shown this program to consistently predict as many winners as the best of the expert tipsters, and to outperform them in predicting margins. In recent years, when analysis of various betting strategies showed it might assist punters to exploit market inefficiencies, the complete output including estimated probabilities of winning, has been distributed via subscription. Records of predicted winners, margins and probabilities for several years past were available.

The structure of this paper is as follows. The next section offers some background into the factors that are important in predicting the outcome of AFL matches. In Section 3, using only information prior to 1997, a multivariate model is constructed to effectively weight the relative importance of contributing variables. In Section 4 this regression equation is applied to matches played between 1997 and 2003 to assess the predictive capacity of data gathered at both a team and individual level, with results compared against the benchmark program for the same period. All analysis was conducted using SAS version 8.2 (SAS Institute Inc, Cary, NC, USA).

2. Background

Because each match is played between two teams competing at a single venue, by convention, each game is assigned to have a home team and an away team. Data on home team, away team, venue and final scores for each of the 12462 games played prior to 2004 were obtained from the WWW. By considering the margin of games as being the home team score minus the away team score, the match result or margin of victory is well approximated by a normal distribution with a mean of eight points and a standard deviation of 40 points. (Figure 1)

Figure 1: Histograms of results (Home team score minus Away team score)
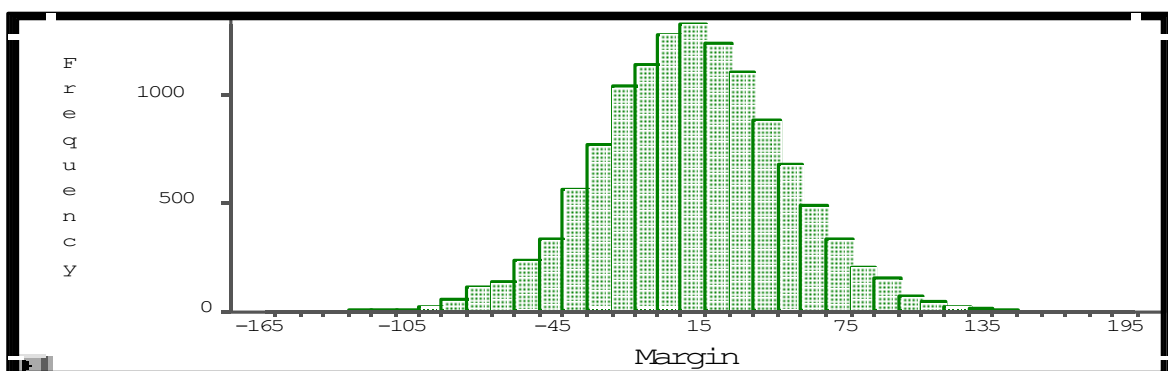
Compliance with normality enables the use of multiple linear regression to weight the contributing effects of home ground advantage, team quality and current form to produce a prediction equation. Predicted margins can be divided by standard errors and compared with the standard normal distribution to determine the winning probability of competing teams.

2.1 Home ground advantage

The concept of home ground advantage (HA) has long been recognised as a contributing factor for sporting success. One of the earliest studies conducted by Schwartz and Barsky [7] found differing but significant HA in each of four American sports, namely baseball, football, ice hockey and basketball. This led them to hypothesised that HA can be attributed to a combination of ground familiarisation, travel fatigue and home crowd support. By looking at pairs of matches from the 1980s, Stefani and Clarke [4] found evidence for individual home advantages. Clarke [2] has a detailed analysis of HA in the AFL for the period 1980-1998. He found significant evidence for an interstate advantage. Although the draw is not balanced for opponent or ground, Clarke also showed there is little difference in the HA produced by a regression model allowing for team ability and the simpler process of taking the average winning margin of the nominal home side.

Figure 2 shows the total score and common HA (as measured by average winning margin of the nominal home side) in the AFL in five-year periods. Although the overall score for the matches has risen consistently, the home ground advantage has remained reasonably constant with a mean figure of eight points. In the past 20 years, it would appear that home ground advantage has increased slightly, although this could primarily be attributed to the increase in matches played between teams from differing states.

Figure 2 Home ground advantage and total game score stratified in five-year periods
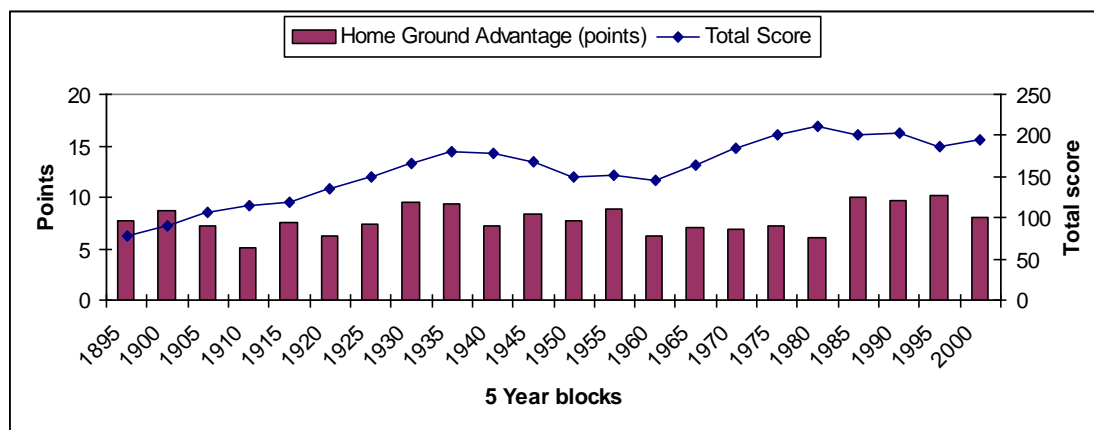
Because information on crowd numbers and more specifically crowd passion is not readily available, it is difficult to quantify the effect due to crowd support, but it is possible to differentiate the effects of travel fatigue and ground familiarisation.

2.2 Travel fatigue

By measuring the distance travelled by the opposing team it is possible to tease out the negative effects due to travel. Prior to 1982, all AFL matches were played within Victoria. Since then teams have established home bases in New South Wales, Western Australia, Queensland and South Australia, creating the need for interstate travel. The most simplistic approach to quantifying the effects due to travel is to introduce a binomial variable to identify interstate travel. Overall, 13% of matches played have been between interstate opponents, although currently, approximately half of all matches are played between teams from differing states.

On average, the home ground advantage when opponents travel from interstate is almost double the advantage experienced when teams are from the same state (14.1 points vs 7.1 points $p<0.0001$). The debilitating effects of interstate flights can be further quantified using a cut-off of 1500km (approximately 2hrs travel). Teams travelling for longer than two hours are disadvantaged by an additional goal, with their opponents enjoying a HA of 16.5 points compared with 10.9 points when travel is less than two hours. ($p<0.0001$).

2.3 Ground familiarisation

The more often a side plays at a particular venue, the more familiar they become with the surroundings. In the 107 years of football, 35 different grounds have been used to host matches, although only 20 venues have been used on more than 100 occasions. By considering the difference in the number of times the two competing teams have played at the chosen venue, it is possible to numerically quantify the effects of familiarisation. Given that some teams have been using the same home venue for hundreds of matches, there is obviously a limit to just how familiar a side can become with a particular venue. To allow for this, the upper limit for experience gained at a particular venue was set at 100 matches. By stratifying the difference in matches played at the chosen venue into 3 categories, we can observe a clear statistically significant difference ($p<0.0001$).

Table 1: Advantage of familiarisation at a given venue

Difference in Experience	N	Average Result
Less than 10 matches difference	2848	3.8 points
Between 10 and 50 matches difference	2942	7.1 points
Greater than 50 matches difference	6672	10.2 points

Although by definition every AFL game played has a home team and an away team, 12.4% of all matches played have been played on a shared or neutral venue. When this is the case, the home advantage is approximately halved (4.3 vs 8.5 $p<0.0001$). It could be hypothesised that when games are played at a neutral venue, the HA effects due to travel and familiarisation are removed. This suggests that the HA experienced at neutral venues (4.3 points) acts as a surrogate marker of crowd support. In reality, differences in ground familiarisation can still be shown to exist at neutral venues, indicating that ground familiarisation is a more specific measure of neutrality, thus alleviating the need for a separate variable to adjust for matches played at neutral venues.

2.4 Measures of Performance

Various measures of team performance were considered. A simple way to gauge the performance of the competing teams is to take a moving average of past results, where the result of the game is given by the margin of victory, with past losses recorded as negative results. An alternative approach is to give more weight to more recent performances by exponentially smoothing past results. To produce the final predicted result, the away teams predicted margin is subtracted from the home teams predicted margin. Both moving averages and exponentially smoothed predictors with different smoothing constants were compared.

To measure the quality of each prediction approach, two performance measurements are considered namely the percentage of winners predicted by each approach and the absolute difference between the predicted and actual margins.(AAE)

From Figure 3 it can seen that when considering all past matches, exponentially smoothed predictors produce a lower margin of error and a higher percentage of winners that do arithmetic averages. With the exception of Average Ever, a clear relationship exists between the models that produce the lowest errors and the models that produce the highest number of winners.

Figure 3: Comparison of performance predictors for all matches played ($n=12,462$)
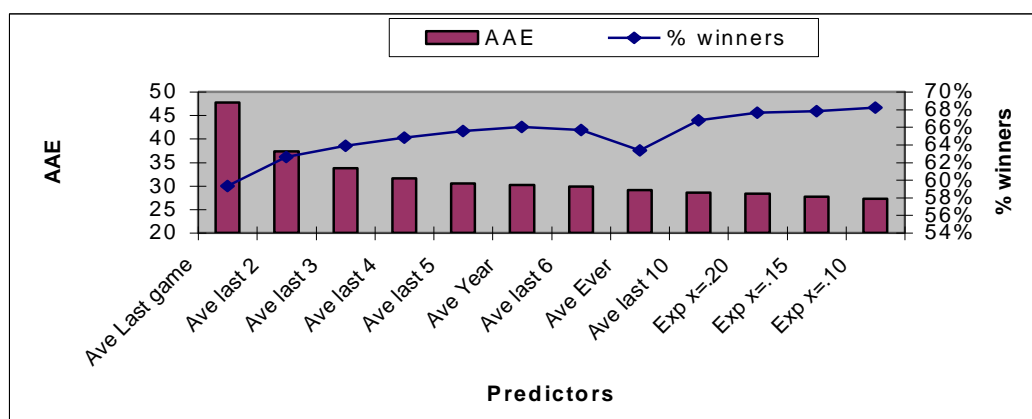
Figure 4: Average Absolute Error for exponentially smoothed predictor ($\alpha=.1$)
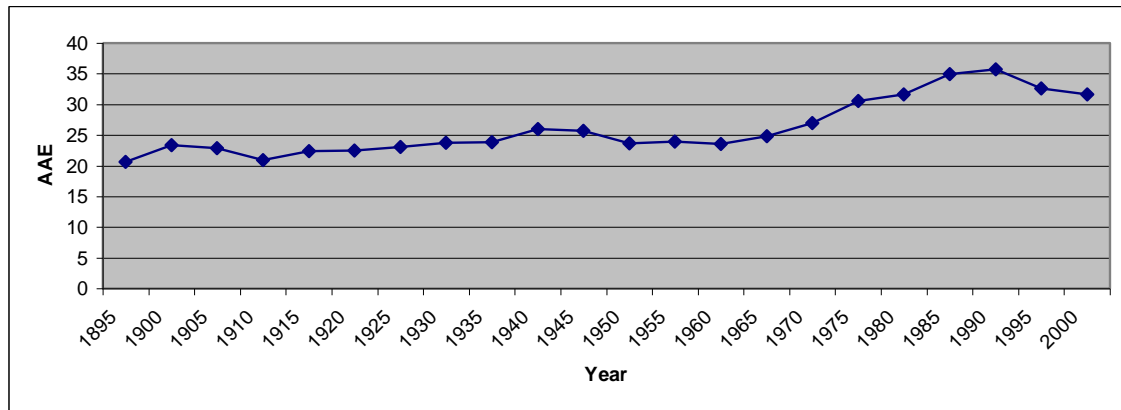
Figure 4 shows the AAE for an exponentially smoothed predictor ($\alpha=0.1$) over the history of football. The steady increase in AAE suggests that match results are becoming less predictable. This decrease in predicability, specifically over the past 30 years, could well be attributed to the increased number of teams in the competition, salary cap and draft constraints as well as additional travel requirements.

2.5   Bookmakers' prices

Bookmakers are quite efficient at predicting the winner of AFL football games, with the designated bookmaker favourite winning two thirds of matches. Because of the dynamic nature of fixed price betting markets, the bookmaker's greatest vulnerability occurs when initially setting prices. Bookmakers will traditionally post an opening market for each AFL match approximately 5-6 days prior to commencement of each game. In accordance with supply and demand, by the start of each match the bookmaker price will reflect the opinion of the general public, or more specifically, those in the general public who have placed the largest wagers on the game. Bookmaker prices for all matches from 1997 were collected from a leading bookmaker, Centrebet, on the Friday morning prior to the commencement of each round of matches. Figure 5 shows that the majority of prices offered on the home team are below $2.00, (mean $1.92, median $1.68) reflecting the bookmakers perception of a home ground advantage.

Figure 5: Histogram of prices offered by Centrebet for home team (1997-2003)
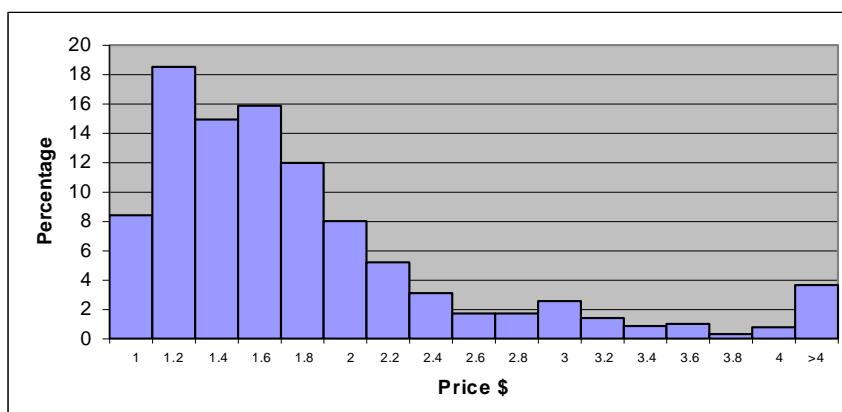
3. Method

A multiple regression model was constructed to predict the margin or difference in points scored between the home and away teams. Variables included in the multiple regression were home ground advantage, interstate travel, ground familiarisation, team quality and current form, with all variables being statistically significant with a *p*-value <0.0001. We have found that using such a stringent significance level creates more robust predictors. The final model developed for *Pt*, the predicted difference between *home* and *away* teams derived at a team level was

$$Pt = A + BI + C(Ft_h - Ft_a) + D(Qt_h - Qt_a) + E(A2t_h - A2t_a)\dots\dots\dots\dots\dots\dots(1)$$

Where:
- A, the intercept, is the home ground advantage.
- I = Interstate travel {0,1(<2hrs) or 2(>2hrs)}.
- Ft = Number of matches team has played at the venue (max=100).
- Qt = Exponentially smoothed predictor of team performance.
- $A2t$ = Average team result for last 2 games

and the subscripts *h* and *a* indicate the home and away teams.

Although exponential smoothing produced a better predictor of team quality (Figure 1.3), it was interesting to note from the multiple regression that the average for the past two games also proved to be an independent predictor of outcome, thus providing a more accurate reflection of current form.

The final form of the model, and the values of *A*, *B*, *C*, *D*, and *E* (suppressed for commercial reasons) were developed using data from all matches played prior to 1997 (11167 games).

One flaw of a team performance model that is based solely on past scores is an inability to adjust for the loss of key players from within the team. Whether through injury or suspension, the loss of key players can severely reduce a team's probability of winning. Conversely, if quality players return to the team, the probability of success may increase.

One approach to compensate for individual players is to derive prediction variables at a player rather than team level. For each match that an individual player competes in, there will be a match result. (team score – opposition score) By smoothing past results for each player it is possible to derive a predicted result at an individual player level. By averaging the predicted result for the 18 players who are named in the starting line-up for the team, it is possible to derive a team prediction that compensates for changes within the team. Starting line-ups for each team are initially named on the Thursday night prior to the weekend matches, and must be finalised 24 hours prior to the commencement of the game. Although last minute changes have been known to occur, the public is generally aware when key players are unlikely to play.

Because data at an individual player level was only available for matches played from 1997 onwards, it is impossible to derive separate parameter estimates for prediction variables using a sample of matches played prior to 1997. Since we wish to use the post 1996 data as a holdout sample for testing, an alternative approach was adopted. The form of the team model was used, with the parameter estimates A, B, C, D and E derived from the team model the same, whilst the values of the prediction variables were calculated using individual player statistics. Ground familiarisation, overall quality and current form are all calculated at an individual player level and then averaged to give a team rating. This means that the overall weighting of the five variables included in the multivariate model does not change, just that the predictor variables used become more representative of the actual players on the field. Thus we obtain the following model for *Pi*, the predicted difference between home and away teams derived at an individual player level.

$$Pi = A + BI + C(Fi_h - Fi_a) + D(Qi_h - Qi_a) + E(A2i_h - A2i_a)\dots\dots\dots\dots\dots\dots(2)$$

Where:
- Fi = Average number of matches played at venue for the 18 starting players.
- Qi = Average exponentially smoothed predictor for the starting players.
- $A2i$ = Average result for last 2 games for the starting 18 players

Parameter estimates were applied to 1286 matches[1] played after 1996 with results compared against a benchmark of Clarke [1] for the same period.

Goodness of fit was assessed by three criteria, namely the Average Absolute Error (AAE) between the predicted and actual results, the percentage of winners successfully predicted and the potential Return on Investment (ROI) that could be derived from fixed price bookmakers.

4. Results and discussion

Table 2 compares the three models. It can be seen that the model derived at an individual level produced the lowest AAE, the highest percentage of winners and the greatest ROI. Both the individual and team models produced average profits that were significantly greater than zero ($p<0.0001$).

Table 2: Model comparison for AAE, % winners and ROI.

Model	AAE	%win	Number of bets	Total Outlay ($1000s)	Profit ($1000s)	Ave. Bet Size	Ave. Profit per bet (std error)	ROI %
Individual	29.8	66.7	981	159.1	24.0	$163	$25 ($6)*	15.1
Team	30.2	65.8	1049	203.3	20.4	$193	$19 ($6)*	10.1
Benchmark	30.5	64.6	923	157.0	2.0	$170	$2 ($7)	1.3

*Significantly greater than zero p<0.0001

With an AAE of 29.8 points per game, the individual model was significantly more accurate in predicting the margin of victory than both the team model (30.2, $p=0.025$) and the benchmark model (30.5 $p=0.001$). The difference between the team model and the benchmark model was bordering on significance ($p=0.06$). With a percentage of successfully predicted winners of 66.7%, the individual model was significantly better in predicting winners than the benchmark (64.6% $p=0.02$) but did not achieve statistical significance in comparison to the team model (65.8% $p=0.21$). Although also predicting more winners, the team model was not significantly better performed than the benchmark (65.8% vs 64.6% $p=0.20$). When considering profit derived from betting on all situations in which there was perceived advantage, there was no significant difference between the team and individual models, although both models were significantly better performed than the benchmark.

Table 3: Statistical comparison between models

Outcome	A	B	Mean Difference (std error)	P-value*
AAE	Individual	Team	0.38 (0.16)	0.025
AAE	Individual	Benchmark	0.72 (0.25)	0.001
AAE	Team	Benchmark	0.34 (0.26)	0.06
% winners	Individual	Team	0.90%	0.21[#]
% winners	Individual	Benchmark	2.10%	0.02[#]
% winners	Team	Benchmark	1.20%	0.20[#]
Ave. Profit	Individual	Team	$3 ($3)	0.19
Ave. Profit	Individual	Benchmark	$17 ($5)	0.0003
Ave. Profit	Team	Benchmark	$14 ($5)	0.0007

*Wilcoxon Sign Rank test
[#]McNemar Test for paired proportions

---

[1] Individual player and bookmaker information was unavailable for the 1997 finals series

Figure 6: AAE from 1997 to 2003
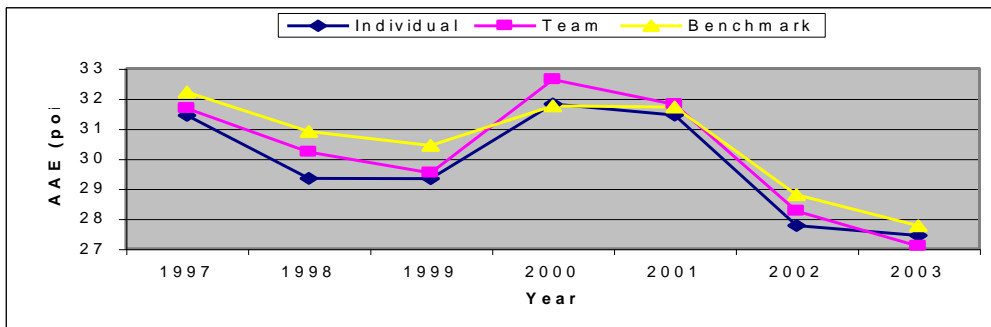
From Figures 6 and 7, we can see that over the past couple of years, the AAE has gone down and the percentage of predicted winners has gone up. Unfortunately, Figure 8 shows this has not equated to an increase in ROI suggesting that the bookmaker, Centrebet, has improved with their price setting process. Although variability exists from year to year, both the individual and team models were able to produce a positive ROI for all seven years.

Figure 7: Winning Percentage from 1997 to 2003
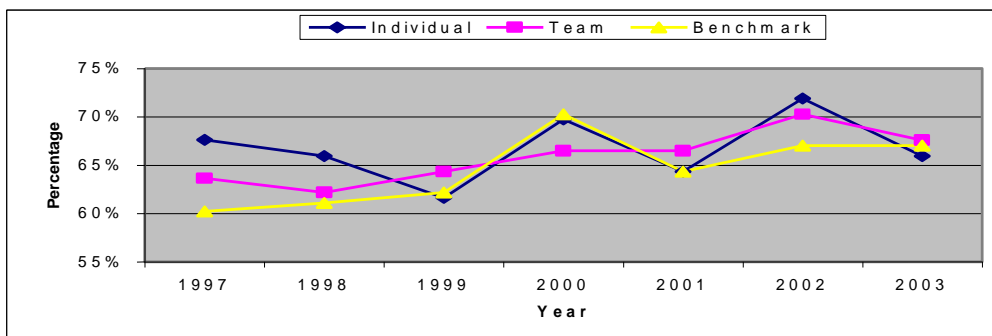
Figure 8: Return on Investment (ROI) from 1997 to 2003
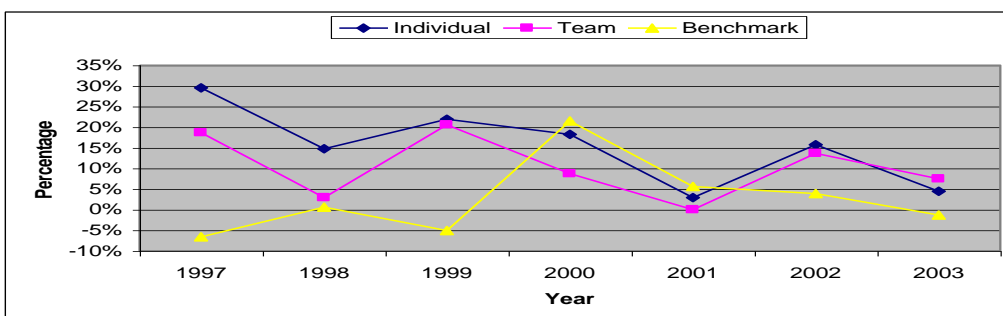
5. Conclusion

The use of multiple linear regression to identify and weight highly significant predictors of outcomes can clearly aid in the prediction of AFL matches. The use of data derived at an individual level can further benefit the prediction process.

Clearly the AFL betting market is to some extent inefficient. Working with a starting bank of $1000, and making bets based on a fixed bank size of $1000, it is possible to see from Table 2 that a significant profit can be derived. Although the individual model can been seen to produce the greatest profits, this model is dependent upon team selection and cannot be utilised until the Friday prior to the weekends rounds of matches. Because both the team and benchmark models are based solely on past team scores, predictions for these models can be produced immediately after the last round has finished and can be available to use when

markets are posted early in the week. Because bookmakers are most vulnerable when prices are initially posted, it is realistic to assume that greater profit could be derived for both of these models by placing bets earlier in the week.

In addition there are many betting strategies that can be employed. Variations such as betting only when the advantage is at least some pre determined figure, betting only on some rounds in the season, betting only on favourites, are some of the strategies employed by punters. These all have the possibility of increasing returns or reducing risk. In addition, this paper has only investigated head to head betting. Evidence suggests that mathematical models are relatively better than the media experts (and thus possibly the bookmakers and the public) at selecting more complicated outcomes such as margins and final ladder order. Inefficiencies in these markets may also be open for profit by using statistically derived prediction approaches.

6. References

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

[2]   Clarke, S.R., Home advantage in the AFL. *Journal of Sports Sciences*, 2004. (under review)

[3]   Yelas, S., Analysis of Swinburne computer performance in identifying inefficiencies in the AFL betting market, Unpublished MSc Thesis, 2003, Mathematical sciences, Swinburne University: Melbourne.

[4]   Stefani, R.T. and S.R. Clarke, Predictions and home advantage for Australian rules football. *Journal of Applied Statistics*, 1992. **19**(2): p. 251-261.

[5]   Bailey, M., Identifying arbitrage opportunities in AFL betting markets through mathematical modelling, in *Proceedings of the Fifth Australian conference on Mathematics and Computers in Sport*, G. Cohen and T. Langtry, Editors. 2000, University of Technology Sydney: Sydney. p. 37-42.

[6]   Kelly, J., A New Interpretation of Information Rate. *Bell System Technical Journal*, 1956. **35**: p. 917-926.

[7]   Schwartz, B. and S.E. Barsky, The home advantage. *Social Forces*, 1977. **55**: p. 641-661.

OPTIMAL USE OF TENNIS RESOURCES

Tristan Barnett, Alan Brown and Stephen R Clarke
School of Mathematical Sciences
Swinburne University, PO Box 218
Hawthorn, Victoria 3122, Australia

Abstract

This paper demonstrates how tennis players can optimize their chance of winning a match by using strategies to utilize their energy resources. This can be achieved by either increasing effort on certain points, games and sets in a match, or by increasing and decreasing effort about an overall mean. The results show that increasing effort on any point in a game before deuce is reached has the same effect on the player's chances of winning the game. By increasing effort on the important points and decreasing effort on the unimportant points in a game, players can increase their chances of winning a game. For the better player, this gain is a result of the variability about the mean and also the importance of points. The results obtained in tennis are used to investigate problems related to warfare.

1. Introduction

The scoring structure of tennis is designed in such a way, that to win, a player must reach a winning score rather than be ahead at a predetermined time. For a 5 set match, the player that first reaches 3 sets wins the match. To win a classical or advantage set, a player must win at least 6 games and be ahead by at least 2 games. And finally to win a game, a player must win at least 4 points and be ahead by at least 2 points. As a consequence of this scoring system, it is possible for a player to win the match by winning well under half of the points played (Ferris [2]). This raises questions about whether some points, games and sets are more important than others, and whether players should distribute their energy resources accordingly?

Tennis commentators often state that the best players can raise their level when faced with break points, e.g 30-40, 15-40. A counter argument is that if these players raised their level earlier in the game, then they could avoid being down break points altogether. Morris [3], O'Donoghue [4] and Pollard [5] show that expending additional physical and mental effort on the important points in a game whilst relaxing on the unimportant points increases the chances of winning a game. In particular Morris [3] states "*If he increased p from 0.60 to 0.61 on half his service points, and decreased from 0.60 to 0.59 on the unimportant half, he would increase his winning percentage by 0.0075 from 0.7357 to 0.7432*". We demonstrate that this increase in the chances of winning the game is produced as a result of the variability about the mean p, as well as the importance of points to winning a game.

Sections 2 and 3 establish some important results. This is achieved by calculating the conditional probabilities of players winning a best of 3 set match for a constant probability of winning a set. Then calculations are produced for the conditional probabilities of players winning the match by increasing probability, or increasing and decreasing probability on certain sets within the match. Definitions and properties of importance, time-importance and weighted-importance are produced in section 3. Section 4 uses the established results to solve problems on utilizing energy resources. Sets within a match are considered first. Optimal strategies are determined for players increasing, or increasing and decreasing effort on certain sets within a match. Similar problems are solved for points in a game and games in a set. By considering a tennis match comprised of games and sets, a problem is formulated to determine on which one game a player should increase their effort to optimize their chances of winning a match. In section 5, the results obtained in tennis are used to investigate problems related to warfare. The paper concludes in section 6 with a summary and a general discussion.

2. Probabilities on winning a match

A best of 3 set match is a contest where the first player to win 2 sets wins the match. Analyzing this system is non trivial despite its relatively simple structure, because it is not certain that the third set will be played.

Consider the situation where a player has a constant probability p of winning a set. What is the probability of this player winning a 3 set match? The player can win the match by either winning in straight sets with

probability p^2, losing the first set and winning the last two sets with probability $(1-p)p^2$ or winning the first and last set and losing the second set with probability $p(1-p)p$. Summing these, the chance of the player to win the match is given by $p^2(3-2p)$.

An alternative method, described in Barnett and Clarke [1], uses recursion to calculate the probability of a player winning the match from any position. This method allows more flexibility and easier calculation when the probability of winning a set is not constant.

If $P(e,f)$ is the probability that the referred player wins the match when the match score in sets is (e,f) ($e=$ referred player's score, $f=$opponent's score), the recurrence formula becomes:

$$P(e,f) = pP(e+1,f) + (1-p)P(e,f+1)$$

The boundary values are $P(2,0)=P(2,1)=1$, $P(0,2)=P(1,2)=0$.

Table 1 represents the conditional probabilities of a player winning a match. The match score at (0,0) agrees with our former calculation of $p^2(3-2p)$.

Table 1: The conditional probabilities $P(e,f)$ of players winning a match.

		opponent score in sets		
		0	1	2
	0	$p^2(3-2p)$	p^2	0
player score in sets	1	$p(2-p)$	p	0
	2	1	1	

Similar calculations are used to calculate the probabilities of reaching score lines within a match. If $N(e,f|k,l)$ is the probability of reaching a match score (e,f) in a match from match score (k,l), the recurrence formulas become:

$$N(e,f|k,l) = pN(e-1,f|k,l) \qquad \text{for } e = 2, \text{ or } f = 0$$
$$N(e,f|k,l) = (1-p)N(e,f-1|k,l) \qquad \text{for } e = 0, \text{ or } f = 2$$
$$N(e,f|k,l) = pN(e-1,f|k,l)+(1-p)N(e,f-1|k,l) \qquad \text{for } e = 1, f = 1$$

The boundary values are $N(e,f|k,l) = 1$ if $e = k$ and $f = l$.

Table 2 lists the probabilities of reaching various score lines in a match given $k=0$, $l=0$. The probability of playing the first and second set is always one, but the probability of playing the third set is $2p(1-p)<1$.

Table 2: The probabilities $N(e,f|0,0)$ of reaching various score lines in a match.

		opponent score in sets		
		0	1	2
	0	1	$1-p$	$(1-p)^2$
player score in sets	1	p	$2p(1-p)$	$2p(1-p)^2$
	2	p^2	$2p^2(1-p)$	

Now suppose a player increases his effort for one set at a match score (e,f), so as to change his probability of winning this set from p to $p+\varepsilon$, where $p+\varepsilon < 1$. This is equivalent to the opponent decreasing his effort at a match score (e,f) so as to change his probability of winning this set from $1-p$ to $1-p-\varepsilon$, since an increase of the probability of winning to one player is a decrease to the other player. If the increase in effort is applied at (0,0), the chance for the player to win the match becomes $(p+\varepsilon)p(2-p) + (1-p-\varepsilon)p^2 = p^2(3-2p) + \varepsilon 2p(1-p)$. The same result is obtained if an increase in effort is applied at (1,1). Similarly the chances of a player to win the match when an increase in effort is applied at one of (1,0) or (0,1) is $p^2(3-2p) + \varepsilon p(1-p)$. Conditional on the match score reaching (1,0), the chance for a player to win the match when an increase in effort is applied at (1,0) or (1,1) is $p(2-p)+ \varepsilon(1-p)$; and conditional on the match score reaching (0,1), the chance for a player to win the match when an increase in effort is applied at (0,1) or (1,1) is $p^2+ \varepsilon p$. Table 3 gives the increase in probability when effort is applied throughout the match.

Table 3: The increase in probability when effort is applied throughout the match.

Current match score	Match score at which an increase is applied	Increase in probability of winning match
(0,0)	(0,0)	$\varepsilon 2p(1-p)$
	(1,0)	$\varepsilon p(1-p)$
	(0,1)	$\varepsilon p(1-p)$
	(1,1)	$\varepsilon 2p(1-p)$
(1,0)	(1,0)	$\varepsilon(1-p)$
	(1,1)	$\varepsilon(1-p)$
(0,1)	(0,1)	εp
	(1,1)	εp
(1,1)	(1,1)	ε

The first set played begins with the match score at (0,0). The third set is played only if the match score reaches (1,1). The second set played occurs with the match score at either (1,0) or (0,1). The chance of a player winning the match when one increase in effort is applied on the first, second or third set played is equal to $p^2(3-2p) + \varepsilon 2p(1-p)$.

Now suppose a player adopts a strategy of increasing his effort on the first, second or third set played by ε, and decreases p on the first, second or third set played (but a different set played from that of the increase) by ε, where $0 < p+\varepsilon < 1$. Calculations show the chance of the player winning the match for this situation is equal to $p^2 (3-2p)+ \varepsilon^2(2p-1)$.

3. Importance, Time-Importance and Weighted-Importance

3.1 Notation for points, games, sets and matches

Tennis is a game where the structure of the scoring system is hierarchical. It becomes convenient in the following sections to repeat the use of a symbol at various levels, and to distinguish between the levels by means of accents. For example, let p be the probability of a player winning a point, p' be the probability of a player winning a game, p'' be the probability of a player winning a set, and p''' be the probability of a player winning a match.

Subscripts are used where it is necessary to distinguish between a player and his opponent, e.g. p_A and p_B are the probabilities of players A and B winning a point on serve respectively.

3.2 Importance and Time-Importance

Morris [3] defines the importance of a point to winning the game $I(a,b)$ as the probability a player wins the game given that he wins the point, minus the probability that he wins the game given he loses the point. If $P(a,b)$ is the probability a player wins the game from game score (a,b), this can be represented by $I(a,b) = P(a+1,b) - P(a,b+1)$. It follows that the importance of a set to winning the match is represented by: $I''(e,f) = P''(e+1,f) - P''(e,f+1)$. These formulas apply to both players, since every point in a game or set in a match is equally important for both players (Morris [3]). Table 4 lists the importance of a set to winning the match and shows clearly the set (1,1) has the highest importance to winning a match. If $p''> 1/2$, then (0,1) is more important than (1,0).

Table 4: The importance of a set to winning the match.

		opponent score in sets	
		0	1
player score in sets	0	$2 p'' (1- p'')$	p''
	1	$1- p''$	1

Morris [3] defines time-importance using the following formula when g=0,h=0:

$$T(a,b/g,h) = I(a,b)E(a,b/g,h)$$

where $T(a,b/g,h)$ is the time-importance of point (a,b) in a game from game score (g,h) for a player, and $E(a,b/g,h)$ is the expected number of times the point (a,b) is played in a game from game score (g,h). With this definition, Morris [3] considers advantage server to be the same score as $(a=3,b=2)$, since these points are logically equivalent. Similarly he considers advantage receiver to be the same score as $(a=2,b=3)$.

3.3   Weighted-Importance

We define weighted-importance using the following formula:

$$W(a,b/g,h) = I(a,b)N(a,b/g,h)$$

where $W(a,b/g,h)$ is the weighted-importance of point (a,b) to winning the game for a player from game score (g,h), $N(a,b/g,h)$ is the probability of reaching point (a,b) in a game for a player from game score (g,h). It follows that $W''(e,f/k,l) = I''(e,f)N''(e,f/k,l)$ is the weighted-importance of set (e,f) to winning the match from match score (k,l). It also follows that $W''(e,f/k,l) = T''(e,f/k,l)$ for all e,f,k,l but $W(a,b/g,h) \neq T(a,b/g,h)$ for all a,b,g,h.

Table 5 represents the weighted-importance of sets in a match from match score $(k=0,l=0)$ and notice that $W''(0,0|0,0)$, $W''(1,1|0,0)$ and $W''(1,0|0,0)+W''(0,1|0,0)$ all equal $2\,p''\,(1-p'')$.

Table 5: The weighted-importance of sets in a match from (0,0).

		opponent score in sets	
		0	1
player score in sets	0	$2\,p''\,(1-p'')$	$p''\,(1-p'')$
	1	$p''\,(1-p'')$	$2\,p''\,(1-p'')$

Morris [3] states the following formula for time-importance when g=0,h=0:

$$\Sigma_{(a,b)}T(a,b/g,h) = \frac{dP(g,h)}{dp}$$

We introduce the corresponding formula on weighted-importance:

$$\Sigma_{(e,f)} W''\,(e,f/k,l) = \frac{dP''(k,l)}{dp''}$$

Morris [3] states the following property about time-importance. Suppose a server, who ordinarily has probability p of winning a point on his serve, decides that he will try harder every time the point (a,b) occurs. If by doing so he is able to raise his probability of winning from p to $p+\varepsilon$, ($\varepsilon > 0$ but small) for that point alone, then he raises his probability of winning the game from $P(g=0,h=0)$ to $P(g=0,h=0) + \varepsilon T(a,b|g=0,h=0)$.

Likewise it can be shown that for weighted-importance of sets to winning a match, has the following property:

Property 1: Suppose a player, who ordinarily has probability p'' of winning a set, decides that he will try harder every time the set (e,f) occurs. If by doing so he is able to raise his probability of winning from p'' to $p'' + \varepsilon$, ($p''+ \varepsilon < 1$) for that set alone, then he raises his probability of winning the match from $P''(k,l)$ to $P''(k,l) +\varepsilon W''(e,f|k,l)$.

Similar properties can be obtained for points in a game (Property 2), games in a set (Property 3) and games in a match (Property 4).

Because $N''(e,f|k=e,l=f)=1$, Property 5 arises as a special case of Property 1.

Property 5: Suppose a player, who ordinarily has probability p'' of winning a set, decides that he will try harder every time the set (e,f) occurs. If by doing so he is able to raise his probability of winning from p'' to $p''+\varepsilon$, $(p''+\varepsilon<1)$ for that set alone, then he raises his probability of winning the match from $P''(e,f)$ to $P''(e,f)+\varepsilon I''(e,f)$.

Morris [3] states the importance of winning a point to winning the match $I''(a,b:c,d:e,f)$ can be obtained from the importance of a point to winning the game $I(a,b)$, the importance of a game to winning the set $I'(c,d)$ and the importance of a set to winning the match $I''(e,f)$ as follows:

$$I''(a,b:c,d:e,f) = I(a,b)I'(c,d)I''(e,f) \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(1)$$

Let $N'(c,d|i,j)$ = the probability of reaching a set score (c,d) in a set from set score (i,j), $N''(a,b:c,d:e,f|g,h:i,j:k,l)$ = the probability of reaching a match score $(a,b:c,d:e,f)$ in a match from match score $(g,h:i,j:k,l)$. The following equation is obtained as a result of independence:

$$N''(a,b:c,d:e,f|0,0:0,0:0,0)=N(a,b|0,0)\ N'(c,d|0,0)\ N''(e,f|0,0) \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(2)$$

Let $W'(c,d|i,j)$ = the weighted-importance of game (c,d) in winning the set from set score (i,j), $W''(a,b:c,d:e,f|g,h:i,j:k,l)$ = the weighted-importance of point $(a,b:c,d:e,f)$ in winning the match from match score $(g,h:i,j:k,l)$. The following equation can be obtained from equations 1 and 2 and verifies the multiplication result for the importance of points in a match holds for weighted-importance only if $(g=0,h=0:i=0,j=0:k=0,l=0)$.

$$W''(a,b:c,d:e,f|0,0:0,0:0,0)=W(a,b|0,0)W'(c,d|0,0)W''(e,f|0,0)$$

Also the weighted-importance for any point in the match from any score line within the match is represented by equation 3 and the weighted-importance for any game in the match $W''(c,d:e,f|i,j:k,l)$ is represented by equation 4.

$$W''(a,b:c,d:e,f|g,h:i,j:k,l)=I''(a,b:c,d:e,f)N''(a,b:c,d:e,f|g,h:i,j:k,l) \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(3)$$

$$W''(c,d:e,f|i,j:k,l)=I''(c,d:e,f)N''(c,d:e,f|i,j:k,l) \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(4)$$

4. Interpretation of results

4.1   Sets in a match

Suppose a player can apply an increased effort in a match on any set played so as to increase p'' to $p''+\varepsilon$, $p''+\varepsilon<1$. On which set, should the player apply the increase to optimize their chances of winning the match?

From Table 3, applying an increased effort at $(0,0)$ or $(1,1)$ results in an increased chance of $\varepsilon 2 p''(1-p'')$. Applying an increase in effort at $(1,0)$ or $(0,1)$ results in an increased chance of only $\varepsilon p''(1-p'')$. However $(1,0)$ and $(0,1)$ constitute the second set played and the sum of these increases is equivalent to $\varepsilon 2 p''(1-p'')$. Also conditional on the match score reaching $(1,0)$, increasing effort on $(1,0)$ or $(1,1)$ results in an increase of $\varepsilon(1-p'')$ and conditional on the match score reaching $(0,1)$, increasing effort on $(0,1)$ or $(1,1)$ results in an increase of $\varepsilon p''$. Therefore an increase in effort could be applied at either the first, second or third set played to optimize the chances for the player to win the match. The same conclusion can be obtained by looking at the weighted-importance of sets in a match as a result of Property 1.

Suppose a player can apply M increases of effort in a match, $0 < M \leq 3$, on any set/s played by increasing p'' to $p''+\varepsilon$, $p''+\varepsilon < 1$. On which set/s, should the player apply increases to optimize their chances of winning the match?

Since it is optimal to apply an increase on any set played, an optimal strategy is to apply the M increases on every set played throughout the course of the match until there are no increases remaining.

The chance of a player winning a match based on a constant probability is given by $p''^2(3-2\,p'')$. When an increase and a corresponding decrease in effort is applied in any order to the first, second or third set played, there is an additional term in the chances of winning the match of $\varepsilon^2(2\,p''-1)$. When $p''=\frac{1}{2}$, $2p''-1 = 0$, and there is no change in the chances for either player to win the match. When $p'' > \frac{1}{2}$, the chance for the player to win the match increases by $\varepsilon^2(2\,p''-1)$ and therefore the opponent's chances to win the match decrease by $\varepsilon^2(2\,p''-1)$. This implies that it is an advantage for the better player to vary his effort whilst maintaining his mean probability of winning a set. It follows by symmetry that the weaker player is disadvantaged by varying his effort.

The weighted-importance of the first set is $2\,p''\,(1-p'')$, which is the same as the weighted-importance of the second set. Since these sets are always played in a 3 set match, the chances of a player winning the match when an increased effort is applied on the first set and a corresponding decreased effort on the second set, is the same as when a decrease in effort is applied on the first set and an increase in effort on the second set. In this situation, the increase or decrease in probability of winning the match for a player is caused by the variation about the mean probability of winning a set. However this is not the case for the third set played which has the highest importance in the match. This set is only played a proportion of the time, and the better player could further increase his chance of winning the match by increasing their effort on the third set played and a proportion of the time on the second set played.

For example, if a player has a probability of winning a set given by $p'' = 0.6$, then the probability of this player winning the match is 0.648. If a decrease in probability by $\varepsilon = 0.1$ occurs on the first set and an increase in probability by $\varepsilon = 0.1$ occurs on the third set, then the probability of this player winning the match becomes 0.650. However, since the third set is only played a proportion of the time, additional increase in effort can also be applied on the second set with probability z, where z is found by solving the equation: $0.5(0.7z+0.6(1-z))+0.5(1-(0.7z+0.6(1-z)))+z=1$, i.e. $z=0.5$, in which case the probability of this player winning the match now becomes 0.675. Out of the $0.675-0.648 = 0.027$ increase in probability of winning the match for this player, $(0.675-0.650)/0.027 = 92.59\%$ is contributed by the fact of the third set being more important than the other sets. Similar calculations show that the opponent with a probability of 0.4 of winning a set also gains an advantage by decreasing effort on the first set and increasing effort on the third set and a proportion of the time on the second set. But the increase gained of 0.025 is less that what their opponent achieves since $p'' > \frac{1}{2}$.

4.2 Points in a game

Suppose a player can apply an increased effort in a game on any one point played by increasing p to $p+\varepsilon$, $p+\varepsilon < 1$. On which point, should the player apply the increase to optimize their chances of winning the game?

Property 2 about weighted-importance of points in a game is used to solve this problem. Recurrence formulas for a game are entered in a spreadsheet and $p = 0.61$ is used to represent the men's average percentage of points won on serve. Table 6 represents the weighted-importance of points in a game for $p = 0.61$. The sum of the group of points for the nth point played (represented by the diagonals) are equal to 0.257 for $n \leq 6$ and 0.122 for $n = 7,8$. It can be verified that $W(a,b/g,h)$ for all g,h give similar results. Therefore a player can increase their effort on any point in the game, providing this increase is applied before deuce is reached, and they have optimized the usage of their one available increase.

Table 6: The weighted-importance of points in a game from (0,0) and p=0.61

		opponent score in points				
		0	15	30	40	Ad
player score in points	0	0.257	0.134	0.057	0.016	
	15	0.123	0.154	0.124	0.063	
	30	0.046	0.107	0.154	0.157	
	40	0.010	0.040	0.100	0.122	0.075
	Ad				0.048	

Suppose a player with p=0.61 can apply M increases in a game, on any point/s played by increasing p to $p+\varepsilon$, $p+\varepsilon < 1$. On which point/s, should the player apply an increase in effort to optimize their chances of winning the game?

It is optimal to apply an increase on any point in a game, providing the increase is applied before deuce is reached. It can also be shown that the sum of the weighted-importance of points for the $n = 7+2m$ points played is equal to the sum of the weighted-importance of points for the $n = 8+2m$ points played and less than the sum of the weighted-importance of points for the $n = 9+2m$ points played for all $m \geq 0$. Therefore an optimal strategy for a player is to apply an increase in effort on every point in the game until either the game is finished or there are no increases remaining.

Since the sum of the weighted-importance of points for $n \leq 6$ are all equal, then the chances of a player winning the game are the same irrespective of which points played an increase and decrease in effort is applied, providing $n \leq 6$. However, similar to sets in a match, a player can gain a significant advantage by increasing effort at $n = 5$ or 6 due to the fact that $n = 5$ or 6 only occurs a proportion of the time. It can be shown that for a player on serve with p=0.61, 30-40 is the most important point in a game, followed by 30-30 and deuce. The least important point is 40-0. Therefore a player can gain a significant advantage by increasing effort on the important points in a game and decreasing effort on the unimportant points.

4.3 Games in a set

Suppose a player can apply an increased effort in a set on any one game played by increasing p' to $p'+\varepsilon$, $p'+\varepsilon < 1$. On which game should a player apply the increase to optimize their chances of winning the set?

Property 3 about weighted-importance of games in a set is used to solve this problem. Once again, recurrence formulas are entered into spreadsheets to produce the weighted-importance of games in a set conditional on the set score, but now two sheets are required for each player serving. The situation is analyzed using values of $p_A = 0.62$ and $p_B = 0.60$ to represent the men's average percentages of points won on serve. As a result of this analysis, the following set of rules are produced that can be used in order of precedence when making decisions about an increased effort by a player when player A has a stronger serve than player B.

1. If A or B is serving and the set score is $(e=4+n,f=5+n)$ or $(e=5+n,f=4+n)$, $n \geq 0$ then apply an increase.
2. If A is serving and the set score is equal then apply an increase.
3. If B is serving and the set score is equal then apply no increase.
4. If A is serving and player A is ahead then apply an increase.
5. If A is serving and player A is behind then apply no increase.
6. If B is serving and player A is ahead then apply no increase.
7. If B is serving and player A is behind then apply an increase.

As a result of rules 2 and 3, at the start of the match it would be correct for a player to apply an increase in effort if player A is serving, but incorrect to apply an increase in effort if player B is serving. This is because player A is given a higher chance of winning a point on serve compared to player B. Now suppose player B starts serving and the set score progresses with A score always

represented first: (0,0), (0,1), (1,1), (1,2), (2,2), (2,3), (3,3), (3,4), (4,4), (4,5), (4,6), then an increase in effort would not be applied by a player until (4,5).

Suppose a player has M increases available in a set and this player has 1 increase remaining when the set score reaches (5,3) (A score = 5, B score =3), player B serving. It can be shown that player A should conserve energy to serve out the set if the set score reaches (5,4) rather than expend energy to break serve and win the set.

4.4   Games in a match

Although it might be correct for a player to increase effort on a particular game within a set, it might be incorrect to increase effort on the same game within a match as a result of the extra level of hierarchy. Once again decisions about increasing effort on games in a match can be obtained as a result of Property 4. The weighted-importance for games in a match is represented by equation 4. $N''(c,d{:}e,f|i,j{:}k,l)$ can be calculated from spreadsheets on the conditional probabilities of winning a set and the probabilities of reaching score lines in a set. For example if player A is serving then $N''(c{=}0,d{=}0{:}e{=}1,f{=}1|i{=}1,j{=}0{:}k{=}0,l{=}0) = P'_A(i{=}1,j{=}0)(1{-} P'_A(i{=}0,j{=}0))+(1{-} P'_A(i{=}1,j{=}0))\, P'_A(i{=}0,j{=}0)$.

Suppose a player has M increases available in a match and has one increase remaining at the following match score ($i{=}0,j{=}3{:}k{=}1,l{=}0$) with player B serving. Let $p_A{=}0.62$ and $p_B{=}0.60$. Since $W''(c{=}0,d{=}0{:}e{=}1,f{=}1|i{=}0,j{=}3{:}k{=}1,l{=}0) > W''(c{=}0,d{=}3{:}e{=}1,f{=}0|i{=}0,j{=}3{:}k{=}1,l{=}0)$, it would be incorrect to apply an increase at this state of the match, but it would have been correct if it had been in the final set since player B is serving and player A is behind in the set. A player ahead on sets, but behind in the current set, may be better off to save energy to try and win the next set, rather than expend additional energy in the current set.

5. Applications to warfare

This paper is a result of a problem organized by the Defence Science of Technology Organization at the 2003 Mathematics in Industry Study Group (MISG). The following analogue between tennis and warfare is proposed by representing a tennis match as a war conflict, where the levels of hierarchies in a tennis match consisting of points, games, sets and match now become skirmishes, battles, campaigns and war. However in warfare, there are costs associated in applying an increase on points, games or sets. This could involve the cost of firing an extra missile.

In the model it is assumed there are a large number of increases in effort available for use, and if the allocated M increases run out, the supply can always be replenished. There is a reward r for winning the overall war and a cost c for applying an increase at a particular skirmish. Ultimately the hope is to win the war by applying M increases, to maximize $r{-}Mc$. There might be a good chance of winning the war by applying an increase on every skirmish, but overall the war might be a financial loss because of the high costs associated with the large number of increases. Clearly there is a trade off between the value of winning the war and the number of increases in effort that are applied.

In order to see what can be learnt from this model of warfare, it is convenient to return to the tennis contest. For each point of the match, a decision must be made on whether it is worth applying an increase, to maximize the expected payout of the match.

Firstly, consider sets within a match, where g = the cost of applying an increased effort to a set in the match.

Let $EX''(e,f){=}[\, p''P''(e{+}1,f){+}(1{-} p'')P''(e,f{+}1)]r$ and $EX''_I(e,f){=}[(\, p''{+}\varepsilon)P''(e{+}1,f){+}(1{-} p''{-}\varepsilon)P''(e,f{+}1)]r{-}g$, where $EX''(e,f)$ = expected payout at (e,f) in a match with no increase and $EX''_I(e,f)$ = expected payout at (e,f) in a match with an increase. If $EX''_I(e,f)$ - $EX''(e,f) > 0$ then an increase should be applied at (e,f). $EX''_I(e,f)$-$EX''(e,f)$ simplifies to $\varepsilon[P''(e{+}1,f){-} P''(e,f{+}1)]r{-}g$, which is equivalent to $\varepsilon I''(e,f)r{-}g$. This implies that an increase should be applied at (e,f) if $\varepsilon I''(e,f)r{-}g > 0$, or equivalently if

$$\Gamma''(e,f) > \frac{g}{r\varepsilon}$$

Similar definitions can be obtained for applying an increase for points in a game and games in a set and an increase should be applied on points in the match for which

$$\Gamma''(a,b:c,d:e,f) > \frac{c}{r\varepsilon}$$

6. Conclusions

This paper has demonstrated that strategies do exist in tennis as a result of the scoring structure. It has been shown that a player can increase their effort on any point in a game before deuce, and they have optimized the usage of this one available increase. It has also been shown that an increased chance of a player winning a game by varying effort on the first, second, third or fourth points played, for $p > $ ½, is due to the variation about the mean p. However, since the fifth or sixth points played only occur a proportion of the time, the better player can obtain an even greater advantage by increasing effort on the most important points and decreasing effort on the least important points in a game. By considering a tennis match comprised of different levels of hierarchies, it has been demonstrated how a player with one increase in effort, could determine whether this increase in effort should be applied at a particular game in the match. A model is formulated from the results obtained in a tennis match, to show how extra resources can be utilized in warfare.

There are many other problems that arise from the problems considered in this paper. Examples include the effect on the chances of players winning the match (if any) of momentum or morale, depleting available capability through the effort to win the point, or other psychological effects. The generality of the formulae and properties obtained throughout this paper, make it feasible to analyze other scoring structures that may be more applicable to warfare.

Acknowledgements

7. References

[1]   T. Barnett and S.R. Clarke, "Using Microsoft Excel to Model a Tennis Match", *6th Conference on Mathematics and Computers in Sport*, G. Cohen (editor), Bond University, Queensland, Australia (2002), 63-68.

[2]   T. Ferris, submitted, "Emergence: An Illustration of the Concept for Education of Young Students", INCOSE 2003, Washington, July.

[3]   C. Morris, "The Most Important Points in Tennis", *Optimal Strategies in Sports*, (1977), 131-140.

[4]   P. O'Donoghue, "The Most Important Points in Grand Slam Singles Tennis", *Research Quarterly for Exercise and Sport No 2*, **72** (2001), 125-131.

[5]   G. Pollard and K. Noble, "The Effect of a Variation to the Assumption that the Probability of Winning a Point in Tennis is Constant", *6th Conference on Mathematics and Computers in Sport*, G. Cohen (editor), Bond University, Queensland, Australia (2002), 63-68.

PREDICTING WOMEN'S WORLD CUP HANDBALL MATCH OUTCOMES USING OPTIMISED RATINGS MODELS

Anthony Bedford
School of Mathematical and Geospatial Sciences
RMIT University
Bundoora East, Australia

Abstract

This paper examines three differing ratings models in order to determine final rankings and round-based predictions for women's handball. The 2001 World Cup results were used to optimise the three models post tournament. This involved maximising the percentage of correctly predicted matches, then minimising the error whilst varying each model's weighting constant. These models were then used to predict match results for the 2003 World Cup. A modified Elo model, a smoothing model and a Pythagorean Projection model were analysed. All three returned high success rates in 2001 and proved useful in determining the tournament outcome in 2003.

1. Introduction

Handball is one of the world's oldest sports. Differing ancient forms of handball have been played by various European nations for at least 2000 years. The Ancient Greeks played a form known as 'Urania'. The Romans played a form known as 'Harpaston' around 150 A.D. Later forms of handball included the German game 'Fangballspiel' meaning 'catch the ball' dating from around the 11th century. The French also played a form of handball in the 1400's.

The formal rules were not laid out until the mid-1800's in Holland. In 1919, a German sports teacher, Karl Schelenz, improved the rules and is credited with establishing the modern game. The first field handball championship was held in Germany in 1938. The governing body of handball is the International Handball Federation (IHF) (www.ihk.info) which was founded in 1946.

Handball enjoys high popularity in Europe, Asia and North Africa. The two premier tournaments are the Olympic games and the World Cup. There have been nineteen women's World Cups contested, of which eighteen have been hosted in Europe. The other tournament was held in Asia. Handball is played around the rest of the world at various levels. With the exception of North America, all continents were represented at the 2003 women's tournament. The IHF currently consists of 150 member nations.

The women's World Cup held in Italy in 2001 and in Croatia in 2003 began with four groups of six. The first stage of the championships involved each team playing every other opponent in the same group once. In 2001, the bottom two teams from each group were automatically eliminated from the championships. The remaining 16 played elimination matches to determine the best eight, with the losers automatically eliminated. The remaining eight then continually played elimination matches with the losers playing placement matches to determine third to eighth rankings.

In 2003, the bottom three nations from each group were eliminated after the first stage, leaving twelve teams. The second stage (known as the main round) involved the remaining teams forming two groups of six. These teams played matches only with opponents gathered from different preliminary rounds. Points from the matches played with opponents from the same preliminary round were added to end up with rankings. The semi-finals had the first team in each group play the second team in the other group, with the winners playing off in the final.

It is these two tournaments that were the focus of this paper. The aim was to predict the outcomes of women's World Cup matches for both 2001 and 2003. Three ratings methods were fitted to the 2001 results and optimised retrospectively in the hope of high success in 2003. Further, the ratings models were compared to the final standings.

The structure of this paper is as follows: In Section 2 the three ratings methods are detailed, including the optimisation procedures, and an example for each is given. In Section 3, the three models are compared using the 2001 World Cup data. In Section 4, I examine the predictions made for the 2003 World Cup based on the Section 2 results, and then conclude in Section 5.

2. The Ratings Models

The three models used in this paper were a smoothing model, a modified Elo model and a modified Pythagorean Predictor. All of the models were optimised (using different methods to be detailed) with RiskOptimizer for Microsoft Excel. The stopping rule for the optimisations was either convergence to a feasible solution or the completion of 5000 iterations.

2.1 A modified smoothing model

The smoothing model had the following form:

$$R_t^{team} = R_{t-1}^{team} + K(Observed\,diff - Expected\,diff)\ldots\ldots\ldots\ldots\ldots\ldots(1)$$
where
$Expected\ diff = R_{t-1}^{team} - R_{t-1}^{opponent}$.

This model is similar to the one used in Bedford and Clarke [3] for tennis. It allows for a team's rating to decrease even in the event they win, as the predictions are margin based. The ratings of each nation were set at 26 commencing round 1 of the 2001 World Cup. This was chosen as 26 is the rounded mean goals scored per match for women's handball. The smoothing constant, K, was optimised in two stages. The first was to maximize the number of correct predictions by varying K, and the second was to minimise the average absolute error subject to no decrease in the number of correct predictions. The optimal smoothing constant reached was $K = 0.37$. The distribution of differences fitted the logistic distribution best (using BestFit for Excel), with $p - $ value of 0.8447.

2.2 A modified Elo model

The second model used was a variation of the famous Elo [3] model used for chess. The model is given by

$$Rate_t^{team} = Rate_{t-1}^{team} + W\left(Observed^{team} - Expected^{team}\right)\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

where

$Rate_t^{team}$ = the new rating for a nation(after game);

$Rate_{t-1}^{team}$ = the old rating of the nation(pre-game);

W is the weighting, set here at 20. It is then adjusted based on the goal difference. For the goal difference, $gd = team\ goals - opponent\ goals,$

$$W = \begin{cases} 20 & if \quad gd = 0,1 \\ 25 & if \quad gd = 2 \\ 20 + 20\left(0.25 + \dfrac{gd-2}{32}\right) & if \quad gd \geq 3 \end{cases} ; Observed^{team} = \begin{cases} 1 & for \quad a \quad win \\ 0.5 & for \quad a \quad draw \\ 0 & for \quad a \quad loss \end{cases}$$

$$Expected^{team} = \frac{1}{10^{\left(\frac{-diff}{85}\right)} + 1}, diff = \left(R_{t-1}^{team} - R_{t-1}^{opponent}\right)\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

A series of optimisations were undertaken to arrive at the final value of *W* as given above. The optimisation was a two-stage process. The first stage involved maximising the percentage correct by varying *W*. The second stage involved minimising the total error by varying *W* subject to no reduction in the percentage correct. Each nation started with a rating of 1000 in 2001.

2.3 Pythagorean Projection

Designed for baseball, the Pythagorean Projection was founded by the prolific sports author Bill James [4]. Also known as the Pythagorean win theorem or Pythagorean winning percentage, it is a simple formula used to evaluate the expected win percentage of a team based on cumulative scores for and against. He designed it primarily for use at the end of the season. It is evaluated by using a team's total runs scored and total runs against to determine how many wins should have occurred. This model has also been applied to hockey, basketball and NFL football. Incidentally the formula has nothing to do with Pythagoras' theorem. The name was given as the formula reminded Bill James of the famous theorem due to the squaring of the variables. The generic formula is as follows:

$$Winning\% = \frac{\left(Runs\,Scored\right)^{k}}{\left(Runs\,Scored\right)^{k} + \left(Runs\,Allowed\right)^{k}} \quad \dots\dots\dots\dots\dots\dots(4)$$

This is used in this paper to project the number of wins and losses a team should have achieved. At the end of the tournament, (4) was also used to determine under and over achievers (detailed in (8)). It has been shown that the value of $k = 2$ is not the most accurate measure of prediction. Davenport [2] found a more accurate way of determining the exponent *k* using regression. He determined

$$k = 1.50\log\left(Total\,Runs\,per\,game\right) + 0.45 \dots\dots\dots\dots\dots\dots(5)$$

This improves the accuracy for baseball, but not for other sports due to the different range of scoring and margins. Handball scores are typically in the twenties or thirties. So I optimised (3) for handball (maximizing % correct, then minimizing error) yielding $k = 8.03$. The final equation used to calculate the Pythagorean Projection for women's handball was

$$Winning\% = \frac{\left(Goals\,Scored\right)^{8.03}}{\left(Goals\,Scored\right)^{8.03} + \left(Goals\,Allowed\right)^{8.03}} \quad \dots\dots\dots\dots\dots\dots(6)$$

The approach used by Davenport [2] to determine *k* on a game-by-game basis, as in (5), was fitted to the 2001 World Cup data. No improvement to predictions was achieved. Therefore the fixed $k = 8.03$ (i.e. model (6)) was used for each match.
The team with the highest Pythagorean Projection (winning percentage) was deemed the predicted winner. The error measured for each nation was calculated by:

$$\varepsilon = \left|\left(wins + 0.5 \times draws\right) - \left(Winning\% \times games\,played\right)\right| \dots\dots\dots\dots\dots\dots(7)$$

The under / over achievement was calculated by:

$$+/_{-} = \left(\frac{wins + 0.5 \times draws}{matches\,played}\right) - Winning\% \quad \dots\dots\dots\dots\dots\dots(8)$$

An over achievement is reflected by a positive value, and an under achievement by a negative value.

2.4 The models in practise

Table 1 illustrates the different models in operation. The match cited is the 2001 World Cup quarterfinal between Denmark and Austria. The result was Denmark defeating Austria by 27 goals to

26. All three models predicted a victory to Denmark. The Elo model rewarded Denmark with a rating rise and reduced Austria's rating. The Pythagorean projection reduced both nations ratings, impacting more on Austria due to their lower overall goal difference. The smoother penalizes Denmark for not winning by enough goals and rewarded Austria for their valiant defeat.

Table 1: Example of the three ratings models

Method		Denmark	Austria
Elo	$Rate_{t-1}^{team}$	1053	1033
	diff	20	-20
	P(Win)	0.63	0.37
	New Rate	1060	1025
Pythagorean Projection	Win %	0.96	0.75
	New Win %	0.94	0.71
Smoother	$Rate_{t-1}^{team}$	35.3	31.9
	Exp margin	+3.4	-3.4
	P(Win)	0.62	0.38
	New Rate	34.4	32.8

3. Comparison of the Models for 2001

In this section, I examine the post-tournament rankings and ratings of the nations using the three models. Furthermore, I examine how the predictions went. It must be noted that the predictions were optimised retrospectively on the entire tournament. There were 24 nations that competed in the 2001 World Cup coming from three continents, Europe, Asia and Africa.

Table 2 : 2001 Comparison of methods using post-tournament ratings

	Observed Results						Smoother		Elo		Pythagorean		
	Nation	W	D	L	P	GD		Rating		Rating		Rating	+/-
1	RUS	9	0	0	18	58	RUS	38.14	RUS	1091	NOR	0.94	-0.06
2	NOR	8	0	1	16	87	YUG	36.04	NOR	1067	YUG	0.91	-0.18
3	YUG	6	1	2	13	77	NOR	35.00	YUG	1062	RUS	0.89	0.11
4	DEN	7	0	2	14	48	DEN	32.98	FRA	1053	DEN	0.84	-0.06
5	FRA	6	0	3	12	39	KOR	31.67	DEN	1035	SLO	0.81	-0.15
6	HUN	6	0	3	12	21	FRA	30.92	HUN	1028	FRA	0.80	-0.13
7	AUT	5	1	3	11	20	HUN	29.86	AUT	1027	HUN	0.67	0.00
8	SWE	6	0	3	12	4	AUT	29.77	SLO	1012	AUT	0.65	-0.04
9	SLO	4	0	2	8	28	SLO	28.76	SWE	1009	KOR	0.61	-0.27
10	ESP	3	0	3	6	1	SWE	28.14	ESP	1000	SWE	0.54	0.13
11	CHN	3	0	3	6	-13	ANG	27.32	BRA	998	ESP	0.51	-0.01
12	BRA	3	0	3	6	-23	NED	26.93	CHN	996	ANG	0.47	-0.14
13	KOR	2	0	4	4	9	ESP	26.17	KOR	996	NED	0.42	-0.17
14	ANG	2	0	4	4	-2	ITA	26.05	ANG	992	ROM	0.41	-0.21
15	ITA	2	0	4	4	-18	CHN	25.83	NED	985	CHN	0.34	0.16
16	NED	1	1	4	3	-6	ROM	25.23	ITA	982	ITA	0.26	0.08
17	ROM	1	0	4	2	-6	BRA	23.10	ROM	978	BRA	0.23	0.27
18	TUN	1	0	4	2	-32	UKR	21.68	UKR	971	UKR	0.18	0.02
19	UKR	0	2	3	2	-25	MKD	19.52	TUN	969	JPN	0.10	0.10
20	JPN	1	0	4	2	-36	JPN	19.35	JPN	965	TUN	0.10	0.10
21	MKD	0	1	4	1	-35	CGO	17.42	MKD	956	MKD	0.10	0.00
22	URU	0	0	5	0	-65	TUN	17.02	URU	946	CGO	0.02	-0.02
23	CGO	0	0	5	0	-56	URU	14.96	CGO	945	URU	0.01	-0.01
24	GRL	0	0	5	0	-75	GRL	14.52	GRL	943	GRL	0.00	0.00

Table 2 is divided horizontally into three sections based on the actual results. The first section contains the eight teams that made it through to the quarterfinals, semi-finals, finals and placement matches (to determine $3^{rd} - 8^{th}$). The next eight competed in the eighth finals, where the winners advanced to the finals and the losers were eliminated. The bottom eight do not qualify for the finals. The last three grouped columns

represent the three models, with the final single column (+ / -) stating the under or over achievement based on the Pythagorean Projection.

The smoother and Elo methods had Russia the highest ranked team at the end of the tournament. The Pythagorean projection had Norway top due to their superior goal difference.

South Korea's result is one of interest. The smoothing model had South Korea ranked 5th after their better than expected four goal loss to Norway in the eighths. The Elo model does not reward defeat and had South Korea in their actual final place at 13th. The Pythagorean projection had them 9th based on their impressive goal difference.

The Elo method is closest in ranking terms to the final places. For the Pythagorean Projection, the under and over achievement column shows that South Korea was the biggest underachiever (27% below expected wins) and Brazil the highest overachiever (27% above expected wins).

The predictive abilities of each of the methods were fairly close, as detailed in Table 3. The predictions commence for round 2 matches. All the methods delivered good success ratios.

Table 3 : 2001 percentage correctly identified and associated mean absolute error for each method.

Round	Pythagorean Projection		Smoother		Elo	
	% correct	$\overline{|\varepsilon|}$	% correct	$\overline{|\varepsilon|}$	% correct	$\overline{|\varepsilon|}$
2	91.67	0.1419	91.67	7.2854	91.67	0.3991
3	83.33	0.2051	83.33	2.6748	91.67	0.3266
4	66.67	0.3549	75.00	5.2127	83.33	0.3676
5	91.67	0.4473	83.33	2.9174	100.00	0.3243
Round 16	87.50	0.6671	87.50	5.1638	87.50	0.2916
Quarters	100.00	0.5196	100.00	3.7056	75.00	0.4225
Semi Finals / Placement	50.00	0.7386	25.00	6.4997	75.00	0.4697
Finals / Placement	75.00	0.9226	75.00	3.2783	50.00	0.5089
Mean	82.35	0.4027	80.88	4.5931	86.76	0.3669

The Pythagorean projection method, the smoothing method and the Elo method predicted 56, 55 and 59 winners out of 68 matches respectively. None of the methods successfully identified the World Cup final result in which Russia defeated Norway.

4. Predictions for the 2003 World Cup

Based on the fitted models used on the 2001 World Cup data, the 2003 World Cup was predicted. New nations to the Cup included Australia, Serbia & Montenegro, Ivory Coast, Croatia, Czech Republic, Argentina and Germany. As detailed in section 1, the tournament varied after the first stage, with only twelve nations qualifying for the second or main round. I shall look at the predictions of each of the models then compare the final standings.
For all three ratings methods, the percentage correctly predicted decreased from 2001 predictions, with the mean percentage correct rate using the Elo model dropping nearly 12%. This is given in Table 4.

Table 4 : 2003 predictions and mean absolute error for each method.

Round	Pythagorean Projection		Smoother		Elo	
	% correct	$\|\overline{\varepsilon}\|$	% correct	$\|\overline{\varepsilon}\|$	% correct	$\|\overline{\varepsilon}\|$
1	-	-	83.33	8.4158	75.00	0.4118
2	75.00	0.1302	75.00	8.0677	75.00	0.3879
3	91.67	0.2880	91.67	7.1718	83.33	0.4070
4	83.33	0.3618	75.00	7.7954	91.67	0.3913
5	91.67	0.4567	91.67	3.7479	75.00	0.4558
Finals Round 1	50.00	0.6271	50.00	4.4882	33.33	0.5234
Finals Round 2	50.00	0.6842	66.67	1.8729	66.67	0.3383
Finals Round 3	83.33	0.9075	50.00	7.8109	33.33	0.5068
Semi Finals / Placement	66.67	0.5062	100.00	1.9545	66.67	0.4055
Final / Placement	100.00	0.4722	50.00	5.9881	100.00	0.4183
Mean	78.87	0.4312	77.11	6.3284	75.00	0.4206

The post tournament ratings in Table 5 show clear differences in the models. The Elo method again posts the champion as top. The margin dependent smoother has Austria 1ˢᵗ, although they fell well short in reality, coming 11ᵗʰ.

Table 5 : 2003 Comparison of methods using post-tournament ratings

	Observed Results						Smoother		Elo		Pythagorean		
	Nation	W	D	L	P	GD		Rating		Rating		Rating	+/-
1	FRA	9	0	1	18	57	AUT	38.08	FRA	1146	NOR	0.863	-0.11
2	HUN	7	1	2	15	56	FRA	35.62	RUS	1128	FRA	0.861	0.04
3	KOR	7	0	3	14	55	HUN	34.49	NOR	1125	RUS	0.841	-0.15
4	UKR	6	1	3	13	29	KOR	34.09	HUN	1102	HUN	0.817	-0.07
5	ESP	6	1	2	13	28	RUS	34.07	SCG	1063	KOR	0.815	-0.12
6	NOR	7	1	2	15	57	ESP	33.25	KOR	1056	AUT	0.810	-0.31
7	RUS	5	1	2	11	42	NOR	32.19	ESP	1053	ROM	0.796	-0.23
8	SCG	5	0	3	10	18	ROM	31.83	SLO	1051	CRO	0.772	-0.37
9	SLO	5	0	3	10	8	SCG	31.36	DEN	1047	ESP	0.723	0.00
10	ROM	4	1	3	9	38	UKR	31.02	AUT	1043	UKR	0.702	-0.05
11	AUT	4	0	4	8	42	SLO	30.59	CRO	1027	SCG	0.641	-0.02
12	GER	3	1	4	7	12	CRO	30.25	GER	1024	GER	0.611	-0.17
13	DEN	2	1	2	5	-6	ANG	29.47	CZE	1020	SLO	0.570	0.06
14	CRO	2	0	3	4	20	CZE	27.11	UKR	1006	CZE	0.516	-0.12
15	CZE	2	0	3	4	1	CHN	26.75	ROM	1000	ANG	0.483	-0.28
16	JPN	2	0	3	4	-20	GER	26.62	CHN	965	DEN	0.398	0.10
17	ANG	1	0	4	2	-1	BRA	24.38	ANG	963	TUN	0.277	-0.08
18	TUN	1	0	4	2	-15	DEN	24.02	BRA	962	CHN	0.268	-0.07
19	CHN	1	0	4	2	-18	TUN	21.99	JPN	945	BRA	0.259	-0.06
20	BRA	1	0	4	2	-19	CIV	20.84	TUN	926	JPN	0.245	0.15
21	CIV	0	0	5	0	-49	JPN	18.94	CIV	917	CIV	0.057	-0.06
22	ARG	0	0	5	0	-97	AUS	12.08	AUS	901	ARG	0.001	0.00
23	AUS	0	0	5	0	-104	ARG	9.32	ARG	887	AUS	0.001	0.00
24	URU	0	0	5	0	-134	URU	6.13	URU	863	URU	0.000	0.00

The Pythagorean Projection and Elo method both predicted France's victory in the final. Whilst Elo's method had the lowest correct %, it was the only method to have France ranked as best team. The host nation Croatia was the biggest under-achiever (+/- = -0.37). All methods had them placed in the top 12. However, they failed to make the finals.

4.1 Modifications and ratings for each of the models for 2003

Each of the models was carried over from 2001 without any modification to the formulae. However, decisions needed to be made regarding the starting rating of each nation. For the Elo and smoothing model, the ratings were carried over from 2001. For nations that were new to the World Cup, ratings were assigned based on the mean ratings of all the other teams from the same continent, and then

standardised. For the Elo model, the scores were standardised so that the sum of ratings totalled 24000. So for the Elo and smoother, the ratings were carried over from 2001. As both these models are a measure of a team's ability over time, this seemed the obvious decision, as opposed to resetting the ratings. Further there are tournaments held in-between the World Cups that could potentially improve the accuracy of the ratings, and this is planned in the future.

4.2   Elo model

The Elo model was surprisingly poor compared to the other models. To see if an error in judgement was made in retaining the ratings from 2001, they were all reset at 26. There was a marginal increase in the % correct (noting that round 1 results were eliminated) so this could not be blamed. An optimization was then conducted to maximize the % correct whilst varying W. Once completed, the errors were minimized again changing W subject to no decrease in the % correct. The optimal value was $W = 45$ giving 75.9% correct, again a marginal improvement. I surmised the method was biased to winning nations.

The tracking of the ratings for 2003 is illustrated in Figure 1. The pre-match ratings each round show clear directional movements based on whether the nation won or lost.

Figure 1: Pre-round Elo ratings 2003
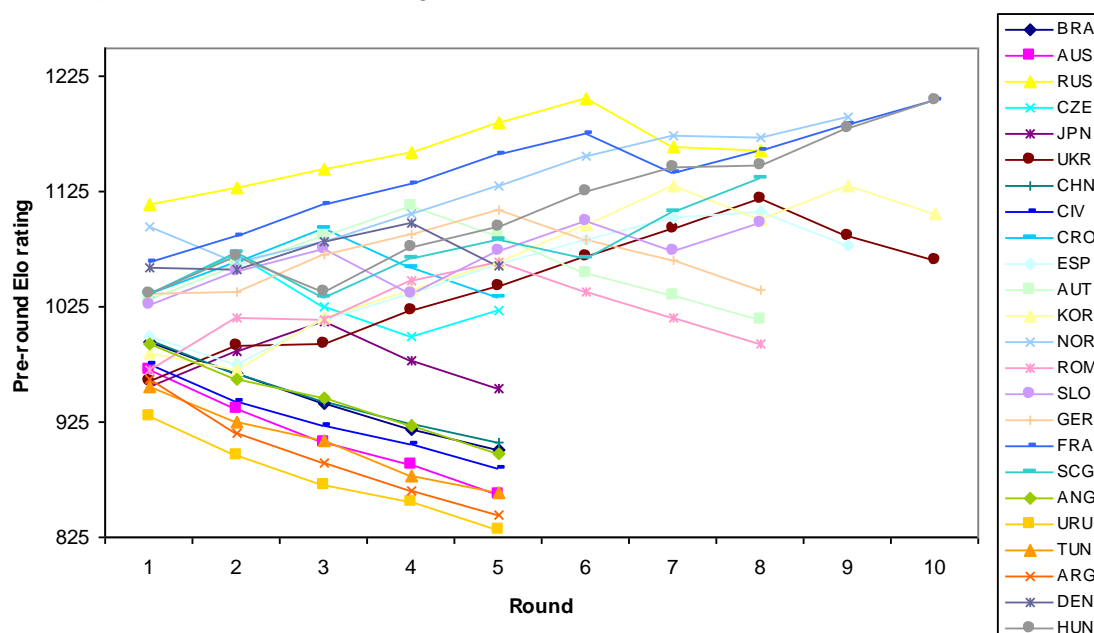
4.3   Pythagorean Projection

As this model is based on cumulative goals for and against, the more matches played, the more the Projection would approach a constant value. So I decided not to carry over goals scored and conceded. This also avoided problems with what quantities to give new nations. For interest, carrying over the cumulative goals scored/conceded from 2001 would have proved disastrous, with only 68.67% correct. Figure 2 illustrates the ratings changes in the Pythagorean model.

Figure 2: Pre-round Pythagorean Projection 2003
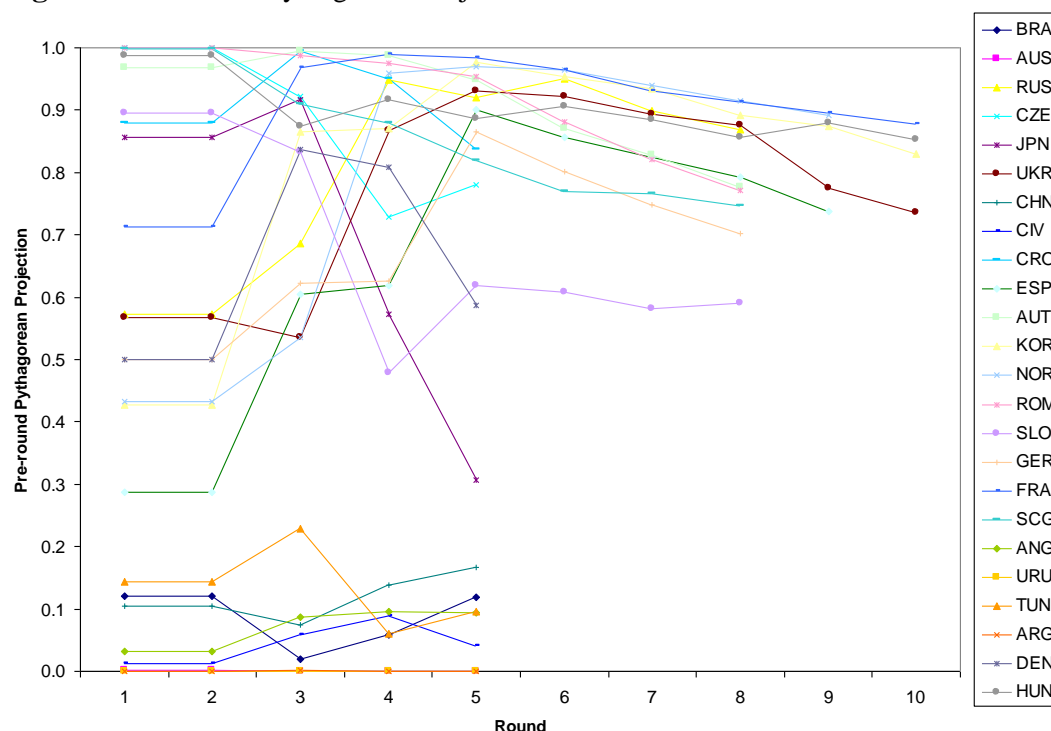
4.4

Smoother results

The smoother model would have been slightly improved if round based optimisation was used. Table 6 details the optimal value of K (denoted $K*$). After each round, K was optimised for the prior round to increase the % correct for the round just completed. So after round 2 matches, the K used in round 1 was optimised in the hope of increasing the percentage correct in round 3.

Table 6 : Round based optimised smoother results

Round	% correct fixed K	% correct optimised K	$K*$
1	83.33	83.33	0.0127
2	75.00	91.67	0.3720
3	91.67	91.67	0.3720
4	75.00	83.33	0.3720
5	91.67	91.67	0.3720
Round 1 finals	50.00	50.00	0.6323
Round 2 finals	50.00	83.33	0.3720
Round 3 finals	66.67	50.00	0.3720
Semi Finals /Placement	100.00	100.00	0.3720
Finals/ Placement	50.00	50.00	0.3720
Mean	77.11	81.93	

There are clear improvements in the predictions using the stage-based optimization.

Figure 3 illustrates the ratings for the 2003 tournament, with most highly rated teams at round 5 continuing on to the finals.

Figure 3 : Smoother ratings 2003
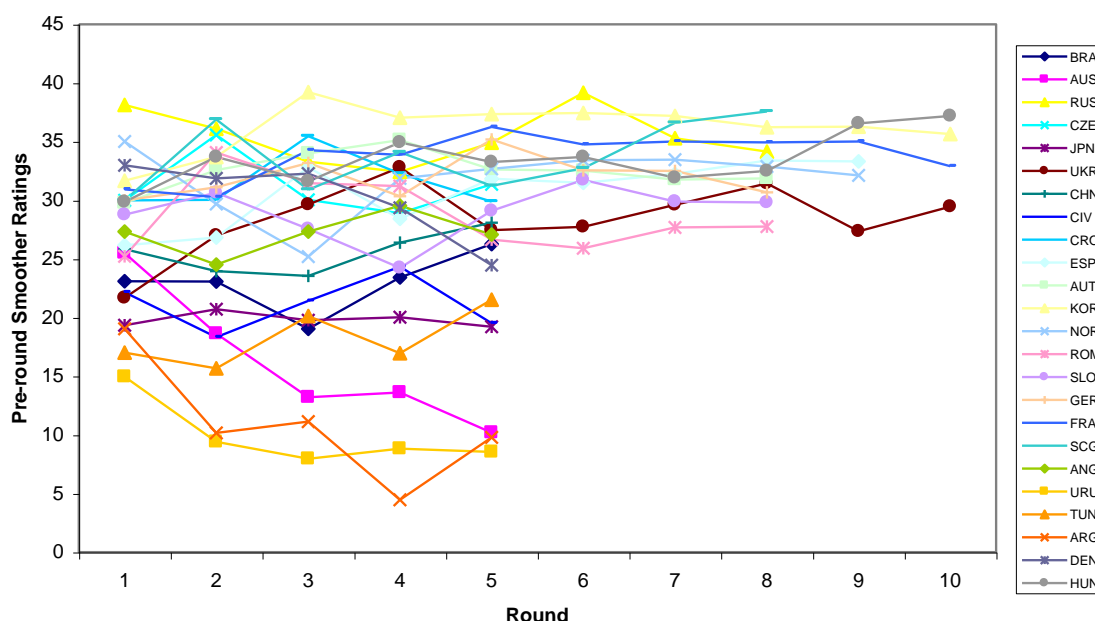
5. Conclusions

Women's handball may well be one of the best sports to predict, judging by the success rates of the three models introduced here. The Elo model performed well on the 2001 data (86.76%), however performed poorly as a predictor for the 2003 data (75%). Both the smoother and Pythagorean Projection methods showed 3-4% reductions in success from 2001 to 2003. However, they seem valuable predictors of outcome in women's World Cup handball.

Each of the three methods has merit. The Elo method is best for post-tournament rankings. It seems the fact that it doesn't allow for losing nations to increase rating somewhat hamstrings its predictive usefulness. The smoother seems best for full tournament prediction, and the ratings from 2001 certainly assisted in the success in 2003. The Pythagorean Projection was the best predictor, but is probably only useful in tournament predictions such as those analysed here. A longitudinal rating might not update as swiftly as the smoother.

Future work is planned on extending the ratings to other tournaments, with men's handball the obvious starter. With results dating back to the 1930's a larger analysis will be considered.

6. References

[1]   A. Bedford and S.R. Clarke, "A Comparison of the ATP Ratings with a Smoothing Method for Match Prediction", in *MathSport* **5**, G. Cohen and T. Langtry (Eds), University of Technology, Sydney, Australia (2000), 43-51.

[2]   C. Davenport, "Revisiting the Pythagorean theorem."
       http://www.baseballprospectus.com/news/19990630davenport.html

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

[4]   B. James, "The Bill James Baseball Abstract 1982". Ballantine Books, New York: (1982).

A RATINGS BASED ANALYSIS OF OCEANIA'S ROAD TO THE WORLD CUP

Anthony Bedford and Cliff Da Costa
School of Mathematical and Geospatial Sciences
RMIT University
Bundoora East, Australia

Abstract

The nations of the Oceania Football Confederation (OFC) have a very poor record at football's highest level. With only two World Cup appearances since 1966, are the OFC teams poor footballers, or are there other forces at play? Are they merely victims of poor scheduling by FIFA? What if the route to the finals was via Asia? In this paper these questions are answered retrospectively using an established statistical ratings method as a means of predicting the expected results for each nation. Analysis is also conducted on how the OFC nations might have performed if scheduling circumstances were different. Then the focus shifts back to Oceania, where successes are evaluated for the suggested alternative route to qualify, and a comparison of all systems is made to see if, as claimed, the OFC is somehow disadvantaged. We conclude our system of qualification via Asia is clearly fairer than the current system, and our prior tournament models show the potential benefit to both Asian and OFC nations.

1. Introduction

No sport is globally more popular and arguably more passionate than football. Elation, disappointment, blame, frustration, and even war have resulted from the outcomes of international matches. Although it could be argued that football is not the most popular sporting code in Oceania, the world's newest confederation has recently experienced more disappointments than any of the other football confederations. In the last four World Cup campaigns, the OFC champions, Australia, missed qualifying for the World Cup in what could have been considered as unfair circumstances. Australia did not lose one qualifying match in their 1998 campaign, yet still missed out on a World Cup final spot due to the away-goal rule. In both their 1994 and 2002 qualifying campaigns, they lost only the last game played: out of 10, and 8, matches respectively. Much criticism has been levelled at FIFA for their scheduling of opponents for the OFC champions. In the last 12 years, Australia has played solely against weak nations within Oceania (with the exception of New Zealand) before playing in an eliminator for a World Cup final place. Historically this decider has been a home and away eliminator, with recent opponents coming from either UEFA (Union des Associations Européennes de Football) or CONMEBOL (Confederación Sudamericana de Fútbol). These opponents are seasoned, experienced nations, given a second chance at qualifying after losing to highly ranked opponents in their localised qualifying round robin campaign. So the OFC champions end up with *no* top-class experience before the eliminator, whereas their opponents end up with *volumes* of high-classed experience before the eliminator. Essentially we have a homogeneously weak tournament champion playing a heterogeneously strong confederations 'worst of the best' nation.

This potential imbalance is the motivation for the retrospective ratings based schedule analysis undertaken in this paper. The aim here is to assess the OFC qualifying tournament using an impartial ratings method in order to determine any inadequacies in either the OFC nations' performances or the scheduling. In a sense, the objective is to see whether the claims of a poor qualification system are justified, or if the OFC champions should not have expected to qualify based on ratings. Finally, an analysis is completed on how the OFC champions would have performed if they had to qualify via the AFC (Asian Football Confederation). Based on these results, we suggest the better method of the two considered options, that is, qualification via the AFC or FIFA's existing system.

The structure of this paper is as follows: In Section 2, we look back at the World Cup places available for each conference since 1966, and detail the probabilities of making the World Cup finals from the OFC based on a simple equal chance model. In Section 3, the Elo [4] model used throughout the paper is introduced then implemented for the purposes of predicting what 'should have happened' based on expected results. In Section 4, a retrospective evaluation of the commonly proposed alternative – what if the OFC champion qualified via Asia? In Section 5 the expected results of the OFC nations are compared. Finally, we state the conclusions and suggest further research.

2. The Preliminary Competitions – Oceania and Beyond

The 'Road to the World Cup' began after the success of the first World Cup held in Uruguay in 1930. The initial preliminary tournaments involved competing nations formed loosely into groups based on geographical proximity. The first North and Central American nations (Mexico, USA and Cuba) vied for a World Cup final place in 1950. Asian teams first contested for a final place in 1954 as did the South American nations. African nations did not get underway properly until 1962. In that year, the World Cup preliminaries took the first form of what we have today – groupings by confederations. The last confederation formed was Oceania. Australia was the first OFC nation to compete for a World Cup place from the yet to be formalised conference, appearing against North Korea in 1966. Held in North Korea, Australia was soundly beaten 6-1 and 3-1 by the eventual World Cup quarterfinalists. Oceania did not have its own homogenous qualification process until the 1970 qualifiers, the year when New Zealand joined. Ironically, the two neighbours never met as Israel and Zimbabwe were grouped into the stages of the tournament. They later played in a combined OFC/AFC league in 1974.

2.1 The Oceania Football Conference

The full detailed history of the Oceania Football Conference can be found at www.oceaniafootball.com. The current level of membership in the OFC and year joined is as follows: Full Members: Australia (1966), Fiji (1966), New Zealand (1966), Papua New Guinea (1966), Samoa (1986), Solomon Islands (1988), Vanuatu (1988), Tahiti (1990), Cook Islands (1994), Tonga (1994), American Samoa (1998). Provisional Members: New Caledonia. Associate Members: Niue, Northern Mariana Islands, and Palau. Australia and New Zealand have each once made the World Cup finals, Australia for the World Cup of 1974 held in Germany and New Zealand for the World Cup of 1982 held in Spain. The best result from either of these teams at the finals was a 0-0 draw by Australia against Chile in their second group match. All other matches resulted in losses for the OFC representatives.

2.2 The Confederations

The FIFA world is divided into six confederations (or zones): the African (CAF), the Asian (AFC), the European (UEFA), the North / Central Americans and Caribbean (CONCACAF), the Oceania (OFC) and the South American (CONMEBOL). The youngest of these confederations is Oceania, formed solely of Oceanic teams in a preliminary competition for the first time for the 1994 World Cup. Previously the OFC teams (solely Australia and New Zealand) played against Asian teams to qualify, twice within the preferred league system (without the threat of elimination) or by knock out. Membership for each of the confederations can be found at www.fifa.com.

The systems for qualification have changed as more nations have attempted to qualify for the World Cup. Typically there are two types of matches – round robin matches and eliminators. Ideally all matches are played on a home and away basis, with round robin groups usually playing each other twice and on some occasions more (such as South America). The eliminators, ideally, are also home and away.

Each zone is allocated a certain number of places for the finals. Some preliminary competitions commence 2 years prior to the World Cup finals. This is typically due to either a long round robin style tournament as seen in South America, or a series of early eliminators for nations that are lowly ranked prior to round robin matches. Politics plays an important role in the allocation of the places, and much recent publicity concerning Oceania has been about the on again off again allocation of a direct qualification place to the World Cup final.

Table 1 contains the number of places available for each confederation by the World Cup final year. Although the World Cup began in 1930, our concern is with 1966 onward, the years in which the OFC became part of the World Cup.

Table 1: World Cup places for each conference by year.

Conference

World Cup Year	AFC	CAF	CON CACAF	CON MEBOL	OFC	UEFA	Total
2006	4.5	5	3.5	4.5	0.5	14 [H]	32
2002	4.5 [H H]	5	3	4.5	0.5	15.5 [C]	32
1998	3.5	5	3	5 [C]	0.5	15 [H]	32
1994	2	3	2.5 [H]	3.5	0.5	13 [C]	28
1990	2	2	2	3.5 [C]	0.5	14 [H]	28
1986	2	2	2 [H]	4	0.5	13.5 [C]	24
1982	1.5	2	2	4 [C]	0.5	14 [H]	24
1978	0.5	1	1	3 [H]	0.5	10 [C]	16
1974	0.5	1	1	4 [C]	0.5	9 [H]	16
1970	0.3	1.3	2 [H]	3	0.3	9 [HC]	16
1966	0.5	0	1	4 [C]	0.5	10 [H]	16

[H] = includes a host nation; [C] = includes the World Cup champion except for 2006.

What is noticeable is that the OFC has never had an increase in the number of places available and has never hosted a World Cup. The AFC and CAF have had steady increases as the push to develop football in these confederations continues. In fact, all the confederations have increased over the last 40 years with the exception of the OFC.

2.3 Equally likely model for Oceania

Table 2 shows that over the years the probability of an OFC nation making the World Cup based on a 50-50 chance has been typically around 6-8%. For the next World Cup (2006), the OFC nations have the lowest probability of all confederations of making the World Cup.

Table 2: Equally likely model for OFC

World Cup Year	Probability of a team making the World Cup finals	Round Robin matches	Elim. matches	Total	Result
1966	0.5000	0	2	2	DNQ
1970	0.0833	4	5	9	DNQ
1974	0.0625	6	5	11	Australia Qual.
1978	0.0667	12	0	12	DNQ
1982	0.0750	14	1	15	New Zealand Qual.
1986	0.125	6	2	8	DNQ
1990	0.0833	4	4	8	DNQ*
1994	0.0417	4	6	10	DNQ
1998	0.0208/0.0833[+]	6	4	10	DNQ
2002	0.05	4	4	8	DNQ
2006	0.0167/0.0833[+]	9	4	13	TBD

*No Oceania team made the eliminators to contest for a World Cup spot.
[+]OFC was divided into low ranked and high ranked nations, hence two probabilities

A further case against FIFA's OFC qualification system is that it is the only system whereby elimination decides a winner of a conference. The UEFA conference uses eliminators to break up the lowest of the best seconds to decide the final three places. Also the worst AFC and CONCACAF meet via eliminators. However, they have previously played home and away games.

3. Elo ratings in World Cup prediction

The task faced in deciding how to evaluate past performances commenced with the current source of ratings adopted by football's governing body, FIFA. The FIFA world ratings/rankings seemed the obvious place to start, and the hunt was on to find archived rankings. The results were both a help and a hindrance – helpful in that the FIFA ratings are judiciously archived at www.fifa.com - but a hindrance in that they only date back to 1993. Dyte and Clarke [3] have shown that these ratings can be used to predict results with a certain amount of success. As these ratings changed form, and furthermore did not date back far enough, the move to an Elo [4] model was chosen.

The task of obtaining past results for the preliminary competitions was made simple, as FIFA (www.fifa.com) has results of all the preliminary matches since the conception of the World Cup. An Elo model was found already in operation: www.eloratings.org. Unfortunately details are scarce regarding the statistical assumptions and predictive power of the model. However, as the aim of this paper is to assess the fate and 'what if' of the OFC preliminaries, the choice was made to adopt and modify the Elo rating model available.

The archived ratings at the beginning of tournaments were used throughout this analysis. They were not designed as a predictive measure. However, in this work they are used to predict match outcomes. Interestingly, similar ratings are used by other websites, in differing forms, to predict many different sports and games. One modification needed in future is to abolish the arbitrary value of 100 added to teams playing a home match. Research on the topic of home ground advantage are numerous, with a selection involving football represented by [2] and [5].

The Elo ratings model used here is as follows:

$$Rate_t^{team} = Rate_{t-1}^{team} + W\left(Observed^{team} - Expected^{team}\right) \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

where

$Rate_t^{team}$ = the new rating for a nation(after game); $Rate_{t-1}^{team}$ = the old rating of the nation(pre-game);

W is the weighting, based on the type of the match, which is set at 40 for World Cup qualifiers, and then adjusted based on the goal difference. So if the goal difference, gd = team goals – opponent goals,

$$W = \begin{cases} 40 & if \quad gd = 0,1 \\ 60 & if \quad gd = 2 \\ 40 + 40\left(\dfrac{3+gd}{8}\right) & if \quad gd \geq 3 \end{cases}$$

$$Observed^{team} = \begin{cases} 1 & for \quad a \quad win \\ 0.5 & for \quad a \quad draw \\ 0 & for \quad a \quad loss \end{cases}$$

$$Expected^{team} = \frac{1}{10^{\left(\frac{-diff}{400}\right)} + 1}, diff = \left(R_{t-1}^{team} - R_{t-1}^{opponent}\right) + 1_{\{team=away\}}(-100) + 1_{\{team=home\}}(100)$$

This model is based on the logistic distribution, and the factor W varies according to the strength of the match and the goal margins. Some matches (such as Olympic qualifiers) are not counted due to differences in the teams fielded. The derivation for W is based on the tournament importance that is then adjusted based on the outcome of the match. As this paper is solely interested in qualifiers, W is set at 40. Unlike Bedford and Clarke's [1] model of tennis players, the ratings of a team can only decrease if they draw or lose. As the ratings are result and not margin based the possibility of a team losing points if they do not win by an expected margin cannot occur.

One could argue the change in ratings does not award enough points for teams that draw against higher rated opponents. For example, Paraguay (rated 1767 and ranked 20th) drew 0 – 0 with Brazil (rated 1963 and ranked 4th) in Paraguay on March 31, 2004 in a World Cup qualifier. Whilst this is clearly a good result for Paraguay the revised ratings sees only 5 points going to Paraguay and 5 points deducted from Brazil. To illustrate,

Table 3: Brazil versus Paraguay example

	Paraguay	Brazil
$Rate_{t-1}^{team}$	1767	1963

	40	40
W		
diff	1767-1963+100= -96	1963-1767-100= 96
Expected	0.3653	0.6347
Paraguay Win	1767+25 = 1792	1963 - 25 = 1938
Draw	1767+5 = 1772	1963 - 5 = 1958
Brazil Win	1767-15 = 1752	1963 + 15 = 1978

As can be seen in Table 3, Brazil had an expected win probability of 0.6347 and Paraguay of 0.3653. If Brazil had won, their rating would have increased slightly. Even if Brazil won 6-0, the rating would only increase slightly because of the victory, and not the score, determining a rise in ratings.

To compare FIFA and Elo ratings, the ratings of all 204 nations at the beginning of 2004 were paired, and the ratings were highly correlated ($r = 0.941$, $p = 0.000$).

3.1 Elo ratings to predict goal difference

The Elo ratings used were not devised to predict winners. However, through retrospective analysis the pre-match Elo ratings were quite accurate in determining the margin, or goal difference, of the match. It has also been successful in determining recent results. Predicting the margin is useful in that the outcomes of a football match contain three potential results: win, draw or loss. Further the margin could be used to construct table positions in tournaments based not only on number of wins-draws-losses but also goal difference. Whilst the ratings are based on a logistic distribution function, and an associated probability of victory is simple to evaluate, deciding what probability range constitutes a draw makes the derivation a little clouded. Further, the ratings do not predict the score of a team. Clearly, lowly ranked teams are still capable of scoring as much as highly ranked teams if playing against a weaker opponent. So for this analysis, the difference in adjusted ratings pre-match proved a good predictor of the outcome of the match. The predicted margin is calculated as follows:

$$gd^{team} = \left(R_{t-1}^{team} - R_{t-1}^{opponent} + 1_{\{team=away\}}(-100) + 1_{\{team=home\}}(100) \right)/100 \quad\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

The margins were rounded to arrive at an actual goal difference for use in constructing tables. The predictions will be written in terms of the home or first team listed, so +1 is a one goal win to the home team and -2 a two goal away team win.

Some examples from the OFC championship finals in 2001 are given in Table 4.

Table 4: OFC Championship examples

Date	Home	Away	gd^{home}	Predicted Margin	Actual Margin
20-Jun-01	New Zealand 1549	Australia 1774	-1.2460	-1	-2
24-Jun-01	Australia 1787	New Zealand 1536	3.5084	+4	+3

3.2 Process used in applying the Elo predictive model

Two separate major analytical processes were undertaken in this work. The first process was the analysis of what actually happened in the OFC qualifiers, and how it compared to the Elo model. The second process required both the results from the first process and results from the AFC qualifiers. Then a schedule was designed for each World Cup campaign based on the actual AFC schedule with the inclusion of the OFC champion. Once the schedules were designed, every match was evaluated using the Elo model. Over 1000 matches were analysed. For both FIFA's and our schedules the venue had to be taken into account for the allocation of home ground advantage. At each qualifying stage, every nation's wins, losses and draws were tallied based on the predicted goal difference in each match. The results seen in subsequent tables are summaries of all the individual matches contested. Every nation competing had to be updated with new ratings at the commencement of each qualifying campaign, as other matches and tournaments usually occurred between World Cups. Microsoft Excel was used to compile the predictions, and the data amassed 17Mb by the time this study concluded.

3.3 Retrospective pre-tournament and updated stage analysis

Let us look at how the OFC nations should have performed based on the modified Elo model. The preliminary competitions are assessed based on the predicted margin of victory within each stage and tallied into tabular results. Two analyses are given, a pre-tournament prediction, and an updated round prediction. The pre-tournament prediction is an analysis of how the entire qualifying tournament should have turned out based on the FIFA schedule. The updated round predictions account for any errors in the early round predictions, and the teams that make it through each stage are used based on the actual results. The nations' ratings are updated based on the expected result from the last game, that is, if no error was to occur. For the updated round method, the ratings are updated at each stage, and no analysis is given for first rounds, as it is the same as pre-tournament.

The presentation in Table 5 gives the results of the two analytical methods undertaken and how the top OFC nation was predicted to perform. The following format is used for the top teams predicted: (wins-draws-losses:points Goal Difference). For example, Korea (4-1-3:13 +7) reads as 4 wins, 1 draw, 3 losses, 13 points and a goal difference of +7. All values represent the predicted totals at the completion of the stage. No modification is made to FIFA's scheduling process, with all venue locations factored into the calculations at each stage.

Table 5: Predictions based on Elo models

World Cup Year	Pre-tournament Predicted Qualifier/s	Updated Stage Predicted Qualifier/s	Actual Qualifier/s	OFC predicted finish	OFC actual finish
1966	KOR (2-0-0:6 +2)	KOR (2-0-0:6 +2)	KOR (2-0-0:6 +7)	AUS (2nd)	AUS (2nd)
1970	KOR (6-1-0:19 +15)	ISR (3-1-0:7 +5)	ISR (3-1-0:7 +7)	AUS (2nd)	AUS (2nd)
1974	AUS (10-0-0:30 +26)	AUS (9-1-0:28 +26)	AUS (5-5-1:20 +11)	AUS (1st)	AUS (1st)
1978	IRI (5-1-2:16 +9)	IRI (5-3-0:18 +12)	IRI (6-2-0:20 +9)	AUS (3rd)	AUS (4th)
1982	AUS (12-1-1:37 +34) KUW (3-2-2:11 +3)	NZL (10-0-2:30 +13) KUW (3-2-2:11 +3)	KUW (4-1-1:13 +2) NZL (9-5-1:32 +33)	NZL (1st)	NZL (2nd)
1986	SCO (2-0-0:6 +3)	SCO (2-0-0:6 +3)	SCO (1-1-0:4 +2)	AUS (2nd)	AUS (2nd)
1990	COL (1-1-0:4 +1)	COL (2-0-0:6 +4)	COL (1-1-0:4 +1)	AUS (2nd)	AUS (3rd)
1994	ARG (1-1-0:4 +2)	ARG (2-0-0:6 +4)	ARG (1-1-0:4 +1)	AUS (2nd)	AUS (2nd)
1998	AUS (8-0-0:24 +30)	AUS (8-0-0:24 +30)	IRI (0-2-0:2 0)	AUS (1st)	AUS (2nd)
2002	AUS (7-1-0:22 +43)	AUS (7-1-0:22 +44)	URU (1-0-1:3 +2)	AUS (1st)	AUS (2nd)

For World Cup 1966, the OFC competition was held at a neutral venue. Two head to head eliminators saw North Korea go through easily and they ended up quarter-finalists. For the 1970 World Cup, both the pre-tournament and updated stage selected the non-OFC teams to make it to the World Cup. Australia performed better than predicted, making it through to the final stage. The updated model picked the exact result. Australia clearly achieved better than expected with a 1st stage win. However it was still Israel as predicted and Australia did it tough against Zimbabwe in 'neutral' Maputo, Mozambique, a country neighbouring Zimbabwe.

In the 1974, Australia was predicted to win the tournament and did so, but not as easily as expected. The group tournament was held with all but one match in Australia. In the first eliminator, Australia

capitalised against Iran but not against South Korea. The neutral venue decider against South Korea saw Australia's predicted one goal margin get them through.

Coupled with 1982, the 1978 campaign was the fairest opportunity to gain experience and qualify without a preset eliminator. Australia was predicted pre-tournament to miss out by the barest of margins (1 goal). However they fell well short of this.

In 1982, Australia performed poorly, being unconvincing against Chinese Taipei and the clincher was New Zealand's excellent win on Australian soil. The updated ratings based on this first round result indicated that the winner should have come from Oceania, and it did, but not as easily as predicted. This is a common trend for both the times Oceanian teams have made the World Cup. New Zealand benefited from a 2nd against 3rd eliminator at a neutral venue where they then qualified.

In 1986, Australia performed well to go through undefeated to the playoff, winning away games in both New Zealand and Israel. Australia achieved better than expected against Scotland, holding them to a draw at home. Unfortunately it was still not enough for them to qualify.

The 1990 world cup campaign yielded the worst ever result for the OFC, with no nation making it to the playoff.

In 1994, a better than expected result was achieved with Australia doing well against Argentina, although just sneaking past Canada on penalties. Again, no wins at home for Australia against non-OFC opponents made the difference in the end in not qualifying.

The 1998 campaign was the most heart breaking for the OFC champions Australia. Their inability to defend the 2-0 half time lead in the return leg eliminator with Iran saw them miss out due to the away goal rule. All predictions had Australia through comfortably.

The last campaign in 2002 saw Australia again fail at the final hurdle, with a substantial away loss to Uruguay.

For all the OFC matches contested, the pre-tournament method yielded an overall 68.3% correct with an average absolute error of 1.79 goals. The error was high due to the blow-out results for the OFC in the last two World Cups. The performance of the ratings is summarized in Table 6. These results are solely based on every OFC World Cup qualifier played and no other matches.

Table 6: Summary of the Elo ratings as predictions for OFC

World Cup Year	Updated stage predictions		Correct World Cup Qualifier?		Correct Finalists	
	Correct	$\overline{|\varepsilon|}$	Pre	Updated	Pre	Updated
1966	100%	2.5	Yes	Yes	-	-
1970	58.3%	0.92	No	Yes	1 of 2	2 of 2
1974	41.2%	1.65	Yes	Yes	2 of 2	2 of 2
1978	69.2%	0.93	Yes	Yes	2 of 2	2 of 2
1982	57.6%	1.85	No	Yes	1 of 2	2 of 2
1986	71.4%	1.36	Yes	Yes	1 of 2	2 of 2
1990	66.7%	1.08	Yes	Yes	1 of 2	2 of 2
1994	77.8%	1.44	Yes	Yes	2 of 2	2 of 2
1998	62.5%	2.17	No	No	2 of 2	2 of 2
2002	87.5%	3.54	No	No	2 of 2	2 of 2

As can be seen the updated round predictions correctly identifies the World Cup finalists on each occasion. The pre-tournament results are also quite accurate, especially in recent times.

4. The Asian Way

Whilst other qualifying options are likely to be more feasible politically, such as qualifying via the CONCACAF, the focus is to now investigate how the OFC champion would have performed via an AFC qualifying tournament. The system of qualifying used here is based on what actually happened schedule-wise to reduce the error in differences in team ratings and to match it with reality. In the years where a 1st round was played in the AFC, the qualifiers from that round are taken. The OFC champion is then taken into the AFC with their post-championship rating.

A schedule was designed so that it varied little from FIFA's design. The aim was to not have any elimination matches, as that would simply replicate what already happens to the OFC champions. Further, the AFC has always been against losing a spot to the OFC. So the following criteria were devised:
(a) The AFC and OFC must not lose any of their original places to the World Cup finals when combined.
(b) All matches are round robin home and away with every nation playing every other twice.
(c) Eliminators are avoided.
(d) The home and away format is optimised to reduce sequential home or away matches.
(e) Group sizes are constrained to a maximum of six for practicality.
(f) In the event of tied points, goal difference is used to decide the winner. If they happen to be equal then the highest rating team qualifies.

The design of each tournament for the combined OFC/AFC playoffs was based on the finalists of both confederations top nation/s. As the AFC grew, the number of groups increased to two. In years 1986 and 1994, UEFA and CONCACAF nations were included due to FIFA scheduling. In those years, it was not possible for those nations to play against an alternative 'best of the worst' in an eliminator. This arose due to host nations taking places.

The presentation in Tables 7 and 8 follows the following format: Win-Loss-Draw: Points Goal Difference. For example, KOR 4-1-3: 13 +7 reads as 4 wins, 1 draw, 3 losses, 13 points and a goal difference of +7. All values represent the predicted totals at the completion of the stage. Again all venue locations are factored into the calculations at each point. The nations underlined move through to the World Cup finals.

For Table 7, as in reality, Israel would have qualified with Australia a credible third. In 1974 FIFA divided the AFC/OFC into 4 groups with a few classification matches that eliminated weaker Asian nations. In groups A1 and A2, first and second moved on to the next round, whilst only the top from B1 and B2 went through. The schedule was revised to allow the four group winners (plus second in A1 and A2) to play in a round robin finale.

In this instance, Australia would not have qualified with Israel going through to the World Cup. In 1974, South Korea eliminated Israel in the group final and Australia never played Israel, instead defeating Iran to go to the World Cup. Hence Australia would have missed out if qualifying via Asia under this fairer system.

Table 7: Elo predictions using our combined AFC/OFC schedule for World Cup 1970 and 1974.

World Cup Year	Group 1		Actual Qualifier
1970	Israel	6-4-0:22 +10	Israel
	Korea	6-1-3:19 + 5	
	Australia	5-2-3:17 + 5	
	Japan	5-2-3:17 +4	
	New Zealand	1-2-7: 5 -11	
	Zimbabwe	1-1-8: 4 -13	
1974	Israel	8-2-0:26 +22	Australia
	Australia	6-3-1:21 +14	
	Iran	5-2-3:17 + 6	
	Korea	4-2-4:14 + 2	
	Japan	1-2-7: 5 -14	
	Hong Kong	0-1-9: 1 -30	

In 1978 and 1982 the system was round robin via Asia and was the ideal scenario. Iran qualified in 1978 and Kuwait and New Zealand qualified in 1982.

In 1986 FIFA scheduled the OFC winner to play in an eliminator with the UEFA 'worst of the best qualifier' - avoiding the AFC completely. The AFC qualification process saw eight groups playing round robin, with the best of each playing an eliminator to find the best four, then again eliminators to find the best two who qualified. In our system we included Australia and Scotland in the final eight leaving a ten team round robin tournament, with the best three qualifying (as the OFC/UEFA contribute one place and the AFC two). Two groups were designed with the best six going into an Asian group and the top two qualifying. The determinant of the best six depends on firstly the highest points ratio (points/games played), then average goal difference. This was necessary as the eight groups contained different numbers of teams. The bottom two teams of the final eight played in a round robin with Australia and Scotland to decide the final place. This ensures two AFC places with the possibility of three. Scotland, Iraq and Japan would have gone to the World Cup finals (South Korea only missed on GD) using our system. In reality South Korea replaced Japan as seen in Table 8.

Table 8: Elo predictions using our combined AFC/OFC schedule for World Cup 1986, 1994 – 2002.

World Cup Year	Group 1		Group 2		Actual Qualifiers
1986	Iraq	10-0-0:30 +35			
	Japan	7-0-3:21 +12	Scotland	6-0-0:18 +24	Iraq
	South Korea	7-0-3:21 +10	Australia	4-0-2:12 +11	South Korea
	Hong Kong	3-1-6:10 -10	UAE	1-1-4:4 -14	Scotland
	Syria	1-2-7:5 -19	Bahrain	0-1-5:1 -21	
	Indonesia	0-1-9:1 -28			
1994			Argentina	8-0-0:24 +25	
	North Korea	3-2-1:11 +3	Australia	4-2-2:14 +5	Saudi Arabia
	Iraq	3-2-1:11 +3	Saudi Arabia	2-3-3:9 -4	South Korea
	Japan	1-2-3:5 -3	Iran	1-3-4:6 -11	Argentina
	South Korea	1-2-3:5 -3	Canada	0-2-6:2 -15	
1998 (1)	Australia	9-1-0:28 +24	Japan	8-1-1:25 +16	
	Iran	7-2-1:23 +12	South Korea	8-1-1:25 +15	Saudi Arabia
	Kuwait	4-2-4:14 0	UAE	5-3-2:18 +5	Iran
	Saudi Arabia	4-2-4:14 -1	Kazakhstan	2-1-7:7 -13	Japan
	China	1-3-6:6 -11	Tajikistan	2-0-8:6 -11	South Korea
	Qatar	0-1-9:1 -24	Uzbekistan	2-0-8:6 -12	
1998 (2)	Iran	7-3-0:24 +16	Australia	9-1-0:28 +26	
	Kuwait	5-3-2:18 +6	South Korea	6-2-2:20 +9	
	Saudi Arabia	5-3-2:18 +5	Japan	6-2-2:20 +8	
	China	4-3-3:15 +2	UAE	3-3-4:12 -3	
	Qatar	1-1-8:4 -14	Uzbekistan	1-2-8:5 -17	
	Tajikistan	1-1-8:4 -15	Kazakhstan	0-1-9:1 -23	
2002	Iran	8-1-1:25 +17			
	Saudi Arabia	7-1-2:22 +17	Australia	8-0-0:24 +32	Saudi Arabia
	China	7-1-2:22 +13	Iraq	3-2-3:11 -1	China
	Qatar	3-2-5:11 -3	Uzbekistan	3-2-3:11 -3	Ireland (d.Iran)
	Thailand	1-2-7:5 -16	Bahrain	1-2-5:5 -14	
	Oman	0-1-9:1 -28	UAE	1-2-5:5 -14	

In 1990 no OFC nation made it through the round robin stage to play off for a final place. Israel was grouped in the OFC, and lost in a two-way eliminator to Colombia. No analysis was conducted.

In 1994 Australia played the CONCACAF 'worst of the best' nation Canada, then the CONMEBOL 'worst of the best' nation Argentina. Both matches were eliminators. The AFC had a final round robin of six teams for two places (although not a true home and away). The addition of Australia, Canada and Argentina would have created a third place in which, potentially, all the Asian teams could miss out (if dealt with a single qualification group). To avoid this in our system, the nine teams were divided into two groups. In this way, a

guarantee of a place to the AFC is ensured if one group contains all Asian teams. Hence the four best Asian teams play off and the worst two mix it with Argentina, Canada and Australia. So all together there are three places available for nine teams. The top two teams from the Asian group go through and the top team of the other group goes through. North Korea, Iraq and Argentina qualify based on the ratings. In reality, North Korea bombed out dismally to finish bottom and Saudi Arabia, South Korea and Argentina qualified.

In 1998 Australia played the third best Asian team in an eliminator for the final World Cup spot. The Asian section had two groups of five with the top teams from each group automatically qualifying, then the second best playing for the right to play for an automatic spot whilst fourth had to play the OFC champion. Our chosen method was to let the OFC champion play in the second round AFC contest. Therefore the AFC finals were increased to two groups of six, with the top two qualifying. As the OFC champion could qualify via either group, both groups were analysed with and without Australia, labelled (1) and (2) in Table 8. To balance the groups Tajikistan was included as the 'worst of the best' Asian nation. Australia, Iran, Japan and South Korea go through in system (1), well clear of the other four opponents in their group. In reality, Australia was replaced by Saudi Arabia. For system (2), Australia, Iran, South Korea and Kuwait go through, with Saudi Arabia and Japan unlucky on goal difference. In reality, Kuwait was replaced by Saudi Arabia and Australia by Japan.

In 2002 the same system applied in the AFC as in 1998. However, it was the UEFA 'worst of the best' qualifying through instead of the OFC. Asia brought two and a half spots (due to Japan/Korea being hosts a spot was lost). The OFC qualified through the CONMEBOL. So the AFC must be guaranteed two spots and brings half a spot, as does the OFC. The consequences are that Ireland plays Uruguay in an eliminator instead of the OFC, remembering that these two teams are the 'worst of the best'. There are eleven teams, and to guarantee two Asian spots, the best six play off for two spots, with the worst four playing with Australia for the remaining one place. Iran and Saudi Arabia go through as the top two of group 1. Australia goes through easily. In reality, Iran missed out to Ireland and Saudi Arabia and China went through. It should be noted that if Australia were allocated to either of the two actual AFC groups they would qualify on top in both.

In summary, success would have come only in recent times for Australia, with clear cases for qualification in 1998 and 2002. Interestingly, it was likely that the 1974 team would have fallen three points short of qualifying. New Zealand was not affected as they qualified via Asia in 1982. Whilst this is all hypothetical, many believe this path still begs inquiry. The impression is that politics will get in the way of such a change occurring in the near future. However our models show that the AFC need not lose any spots and in return may capture an extra place to the World Cup.

5. Comparisons and Closing Comments

Some comparisons of all the expected and actual results for the two leading OFC nations are now given. Whilst much blame is heaped upon FIFA for the scheduling, one may also be critical of the record of these OFC nations. Figure 1 shows the points difference between the updated ratings model and actual results, with Australia being at or under expectation every year since 1986, and only playing above expectation twice in ten attempts. New Zealand has also not fared well since 1990. The asterisk denotes that the nation qualified for the World Cup. Interestingly, in the two times that Australia has played beyond expectation, they still did not qualify.

Figure 1: Points difference by World Cup year for Australia and New Zealand
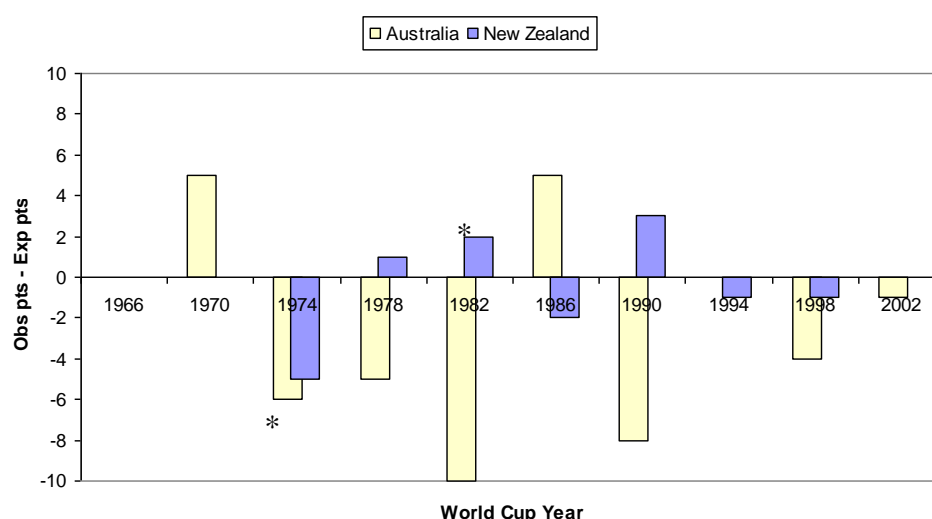
Figure 2 shows the expected placing via the FIFA draw, expected placing via the AFC draw and observed placing for Australia. The asterisk denotes when Australia qualified for the World Cup.

Figure 2: Final Placing by World Cup year for Australia
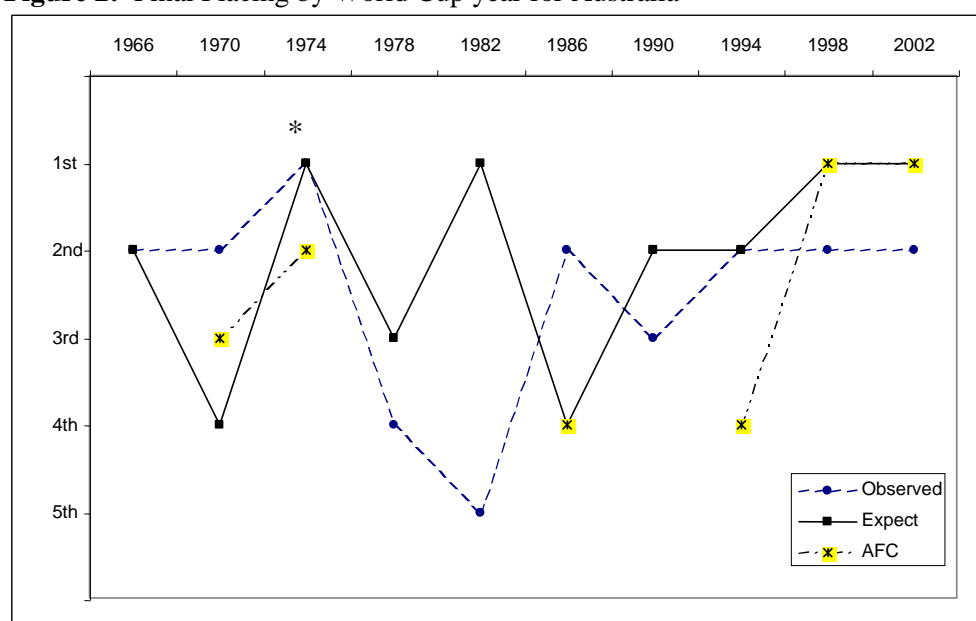
Clearly in the last two World Cup preliminaries Australia under-achieved. In 1998, Australia squandered a 3-1 aggregate lead with 45 minutes left on home soil to draw and miss out on away goals. In 2002, one could argue that Australia failed to capitalise at home. It could also be argued that the fairer AFC route would have seen Australia qualify in both 2002 and 1998, but based on the ratings it is hard to argue that scheduling is entirely to blame.

The problem may lie in something a little more simplistic – the eliminator record. The eliminators are what have been the greatest obstacle to OFC representation. The record below is damning:

Table 9: World Cup eliminators – non-OFC opponents (W-D-L GD)

Country	Home		Away		Neutral	
	Observed	Expected	Observed	Expected	Observed	Expected
Australia	1-5-0 +1	3-1-2 +5	0-2-4 -7	1-2-3 -5	1-0-2 -6	1-0-2 -1
New Zealand	-	-	-	-	1-0-0 +1	0-1-0 0

All the results in Table 9 are against non-OFC opponents. Australia's only win on home soil was the eliminator against Uruguay in 2002. Away, they have not won a game. This table certainly highlights the lack of international ability of the premier OFC nations.

Table 10 covers all the eliminators except final round matches.

Table 10: All non-final round World Cup preliminary tournament eliminators (W-D-L GD)

Country	Home		Away		Neutral	
	Observed	Expected	Observed	Expected	Observed	Expected
Australia	6-0-0 +16	6-0-0 +19	3-0-3 +2	4-2-0 +7	1-2-0 +2	3-0-0 +3
New Zealand	2-0-4 -1	2-1-3 +2	0-0-4 -12	0-0-4 -12	-	-

Australia's record when not facing the final hurdle is on target at home, but a little under expectation at neutral and away venues. New Zealand slightly underachieves at home.

Comparing these two tables, it is at the final hurdle that Australia underachieves. Australia also fails to score or win enough when at home.

5.1 Conclusions and further research

Certainly a path via Asia would have benefited Australia in their 1998 and 2002 World Cup campaigns. This system is more equitable in that it would allow the OFC champion a loss or two without it meaning the end of their road to the World Cup. This is a luxury afforded to most other qualifiers currently denied the OFC. However, it was also found that Australia should still have qualified under FIFA's system in these years, with the final hurdle identified as continually problematic. Australia's home record in crucial matches is poor, and one could say the FIFA schedule is not solely to blame for past failings.

Clearly, the OFC nations have a tough path to qualify for the World Cup. The dearth of victories in crucial matches may be indicative of the lack of exposure the top OFC nations have to other quality opponents. The statistical analysis suggests that Australia should have qualified in the last two campaigns. One could argue that the unexpected losses maybe attributed to the lack of quality opponents earlier. Further, FIFA's current system affords the OFC nations little scope for the occasional failure and crucial points in qualification.

Clearly, the route via the AFC presents a fairer system for the OFC nations. Interestingly both Australia and New Zealand benefited from the original 'unfair' system in 1974 and 1982 respectively. In 1974, Israel was knocked out by South Korea in an eliminator, however the Elo model determined Israel would have made it under the fairer round robin. In 1982, New Zealand actually qualified third in the round robin stage but played an eliminator to determine the final spot, in which New Zealand won.

Our analysis suggests FIFA abandon the unfair eliminator approach in the OFC for the round robin system designed in this paper. The draw is both fair and equitable to both the AFC and OFC nations and has the benefit of more matches played in the OFC against quality opponents. It also addresses the underlying political animosity between the OFC and AFC in that no places are lost to the AFC, and in fact they have the benefit of obtaining an extra spot from the OFC if good enough.

The Elo ratings method provided a useful approach to assessing past performances. However modifications are planned, including a more accurate measure of home and away ground advantages; travel distance factors; margin based ratings; and conference strength constants. The other potential route is via CONCACAF, and will be considered in the future. However the distances are substantially further than to the AFC and are consider far less practical than via Asia. Finally, a look at the perceived advantage in hosting the first game in eliminators is intended.

6. References

[1]   A. Bedford and S.R. Clarke, "A Comparison of the ATP Ratings with a Smoothing Method for Match Prediction", in *MathSport* **5**, G. Cohen and T. Langtry (Eds), University of Technology, Sydney, Australia (2000), 43-51.

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

[3]   D. Dyte and S.R. Clarke, "A ratings based Poisson model for World Cup soccer simulation", *JORS,* **51** (2001), 993-998.

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

[5]   A. M. Nevill and R. L. Holder, "Home advantage in sport: an overview of studies on the advantage of playing at home", *Sports Med.,* **28** (1999), 221-236.

Extraction, Interpretation and Utilisation of Meaningful Information about Individual Rugby Performance using Data Mining

Dr. Paul J. Bracewell
Offlode Ltd, PO Box 8024
Wellington, New Zealand

Abstract

Statistics are having an increased influence in the rugby-coaching environment. Many of the statistics used are exposed to changeable match constraints and conditions, reducing the practical significance of this data. Assuming that the events occurring on the rugby field are largely due to the manifestation of individual ability; then the mining of match data provides insight into individual performance. The importance of understanding the data is emphasised, placing the extraction of performance ratings in context. Consequently data mining techniques allow useful statistics to be extracted from match data. These summarise individual performance and negate the variability in match involvement. Standard data mining techniques are used to construct a stable measure of overall performance, providing the coach or domain expert with a simplified data set that can be explored using graphical techniques, such as control charts, to guide decision-making. This increases the power of the statistical tool available to coaches by enabling deficient or superior performances to be identified.

1. Introduction

Describing the performance of individual rugby players in any number of competitions motivates many conversations throughout New Zealand during winter. As part of the entertainment spectacle, the media provides a vast array of information to provide an objective overview of proceedings. Statistics are a natural by-product of competitive sport and this information is often used to determine the result of a contest (runs, goals, points, time). Hidden within the resultant match data are gems relating to an individual's performance in a match. Extracting this information can provide valuable objective insight into the relative capacity or ability of an individual.

Underlying the use of match statistics is the assumption that they are a reflection of an individual's ability. This assumption is used to create a match rating using dimension reduction capable data mining techniques. The goal of data mining is to extract useful, but previously unknown, information from typically massive collections of non-experimental, sometimes non-traditional, data [17]. However, it must be remembered that there is a story associated with the data, and it is this story that we wish to interpret. Consequently, domain knowledge is crucial when exploring data. Standard data mining techniques are used to highlight how easy it can be to extract powerful solutions.

Rugby provides an interesting and challenging application, as unsupervised data mining techniques must be used due to the lack of defined outcomes. Therefore a sound understanding of the nature of the data is crucial to ensure valid and ethical ratings are obtained. The ideology of the rating discussed in this paper is no different to any other sport rating system, such as cricket's Pricewaterhouse Coopers Rating (formerly known at various times as the Deloitte Rating, Coopers and Lybrand Rating) [2] and Baseball's PGP (Player Game Percentage) [1]. All are designed to indicate how well an individual has performed in a match situation. This enables assessment of both a competitor's ability and their worth to the team. However, this assumes that the occurrences on-field relate directly to the individual's ability. Whilst this discussion focuses on rugby, parallels can be drawn to other sports. Unlike cricket or baseball, rugby has few defined outcomes that represent successful contribution in a match context. That is, from a competitive team perspective, cricket batsmen have generally played well if they have scored runs, but a winger in rugby can have played extremely well, yet not scored any points. Rugby performance must therefore make use of successfully completed tasks such as the number of players beaten or the number of running metres gained. Instead of quantifying performance through structured tests, rugby performance must focus on unstructured rugby match participation, assuming that over time the nature of an individual's ability will be imposed on collated statistics. Each time an individual participates in a game of rugby this can be viewed as a sampling opportunity. The rugby match is then a forum for individuals to express their inherent ability. The vast array of information that can be obtained from match play can be summarised effectively using unsupervised dimension reduction techniques, provided caution is applied in the analyses.

In essence a tool is created to replicate the thought processes of an unemotional, non-forgetful, unbiased, highly-experienced human observer that can identify underlying structures, patterns and/or relationships of individual rugby player performance using match data. Importantly, due to the high noise present in the data, this system is only a tool to assist coaches in decision making, although it is suitable for use in the media.

2. Data Overview

Between 1996 and 2001 data was collected commercially from which a single match rating, summarising each individual's performance, is computed for each game. Over 130 raw measures are collected via notational analysis [6]. These measures cover most conceivable events on a rugby field, including variables such as the number of tackles made, number of turnover tackles, attacking metres gained, players beaten and kicking metres. Approximately ten hours of labour are required to convert the 80 minutes of action into useable data.

The rating described in this paper was launched early in 2000 during the Super 12 competition – a provincial competition involving teams from New Zealand (5), South Africa (4) and Australia (3) – as part of the fantasy game, '*Ultimate Rugby*'; a licensed New Zealand Rugby Football Union (NZRFU) product. Sky Television then used this rating as the basis for the Sky Man-of-the-Match competition during the 2000 All Black tour of France and Italy.

Details of the data are kept brief to protect commercial interests. For the analyses discussed in this paper, data from the 2000 Super 12 competition is used. Of the 137 variables available, 93 related to the individual. These were summarised manually leaving 19 for the initial analyses as shown in Table 1 [6]. A selection of these measures is shown for All Black and former Otago Highlanders halfback, Byron Kelleher for the 12 matches he played in the Super 12 Tournament of 2000.

It is impractical for a human observer to scan through many variables to understand individual performance. This information needs to be summarised in such a way that it empowers the interested parties. To do so it is important to understand the meaning of the raw data and extracted information. This enables valid and ethical solutions to be provided to the end user. Given the accessibility to data mining software and the apparent ease of use, the validity of analyses can often be neglected.

Table 1: Super 12, 2000 Match Statistics for Kelleher

Opposition	Breaks	Defence Beaten	Errors	Laybacks	Passes	Running Metres	Tries	Loose Ball	Tackle Assists	Tackles	Tackles Missed	Kicks in Play
Queensland Reds	10	2	0	3	54	107	0	2	1	1	0	2
Coastal Sharks	10	0	0	1	55	49	1	3	0	7	2	7
Wellington Hurricanes	9	1	3	0	46	48	0	1	0	7	1	5
Golden Cats	9	5	1	2	68	110	2	1	1	1	1	4
Canterbury Crusaders	9	0	0	0	53	31	0	1	2	4	0	3
ACT Brumbies	5	0	1	0	38	29	0	2	4	10	0	1
Auckland Blues	8	0	3	0	24	63	0	0	0	8	1	3
Waikato Chiefs	13	2	1	0	44	51	0	3	4	11	2	7
Western Stormers	13	3	0	1	54	50	0	2	0	11	0	4
Northern Bulls	14	4	3	1	53	80	1	1	2	10	3	2
NSW Waratahs	11	1	1	0	43	68	0	2	4	12	3	9
Canterbury Crusaders	9	0	2	1	61	43	1	2	2	15	1	5
TOTAL	120	18	15	9	593	729	5	20	20	97	14	52

3. Applicability of Data Mining to Create Sport Ratings

The impact statistics has in a rugby environment is readily identifiable in coaching publications such as those issued by the New Zealand Rugby Football Union (NZRFU)[21]. Objective match statistics provide an unbiased record of the game, albeit from a limited scope due to an information/resource trade-off [6].

The underlying assumption of sport statistics is that they provide an insight into performance. Performance is the perception of how well an individual played in a single match. Sports skill can be defined as any behaviour tending to improve performance in sport ([24] p.126). According to the Oxford Senior Dictionary [13], the combination of skills gives rise to ability. This definition fits nicely with rugby terminology and the continual reference to a player's skill set by coaching staff determining player value [8][23]. Sporting ability is the combination of skills that tend to improve the performance in sport [9]. It is possible to quantify skill, as outlined in the definition of ability. However, it must be noted that there is a difference between what is measured, and what is inferred from any measurement. This is an important distinction and a critical assumption that underwrites the development of any sport rating. Skill is inferred from performed tasks. Tasks are directly measurable (countable) and to complete the task successfully, the skill to perform such a task must exist. Mental and physiological tasks that contribute to the performance of a physical task, such as passing the ball, are not directly measurable from a game and are inferred from the performance on physical tasks. Thus, skill also comprises the set of mental, physical and physiological tasks that are required to increase the probability of success. To successfully perform a physical task, such as kicking the ball in a game situation, the other factors that comprise skill must also be successfully performed. Therefore, by measuring physical tasks, skill is inferred because the physical task cannot be completed unless the full set of tasks that comprise the skill (physiological, physical and psychological) are completed. Consequently, quantified skill is inferred from quantified tasks. The combined measurement of skill then leads to the quantification of ability. By measuring the performance on physical tasks one can infer the level of an individual's skill from the univariate data relating to physical tasks in our possession. The skill set of an individual to tackle, catch, kick, and evade defenders is evident within the data set. This provides the justification for calculating a value such as the rating described in this paper. Figure 1 is a schematic representation of the discussion presented, incorporating *m* measurable physical tasks that can be explained by *n* skills from which ability is inferred. A note of caution must be included at this point: It would be naive to assume that an individual's ability to play rugby could be completely described by a set of numerical measures. However, a large portion of an individual's talent is expressed, exemplified by the desire for such statistics in the media.

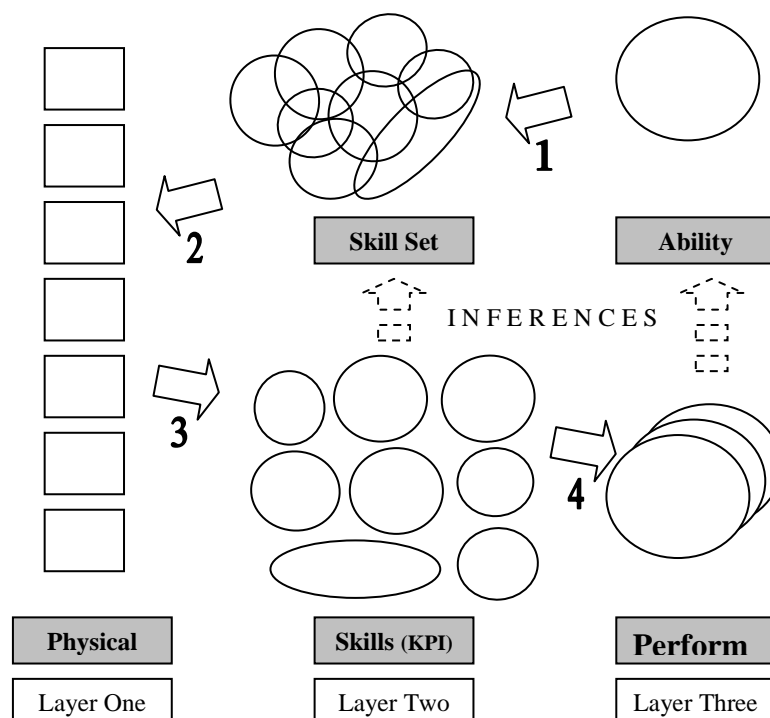**Figure 1:** Hypothesised Relationship between Performance and Ability

Figure 1 illustrates the comments described previously. It is assumed that ability determines the skill set of an individual (step 1). These skills then determine the physical tasks that are performed by the individual (step 2). The physical tasks are used to calculate the key performance indicators (KPI) for an individual (step 3). The collection of KPI is an estimate of an individual's skill set. The KPI are then used to calculate an estimate of ability. Ability cannot be estimated from only one performance. Consequently a series of performance measures needs to be examined. A point estimate of performance is calculated from the KPI (step 4) via the perfection method [5]. Finally this is used to estimate ability.

> Collection of match data referring to the performance by an individual on physical tasks is the primary starting point for quantifying performance. Data mining techniques are utilised to infer the structure and key components of performance from the typical multivariate profile that is expressed from many match situations by many relative individuals. Further, every time a physical task is performed and recorded from a match-type situation, the existence of some associated skill is supported. Consequently, the greater an individual's ability to play rugby translates directly to a greater relative capacity for the game. This leads to greater potential involvement in the game. Work-rate, or direct match involvement, is especially obvious in the statistics that are generated from a match. The more often a player is involved, the higher their work rate. Obviously, the higher the successful involvement on physical tasks, the greater the individual's inferred ability.

Fitness and mental skills are core components of rugby ability. Whilst not directly measurable from a rugby match, existence can be implied. When a skill cannot be replicated in a game environment due to either a lack of mental application or fitness, the absence of such data will suggest potential weakness. Specifically, what is the point of a player being able to wrestle the ball of any opposition player if that individual can never get to a situation where that skill is required due to lack of physiological attributes such as fitness and mobility? This example acknowledges that confounding variables such as fitness and mental attributes do not pose a problem. The mediating effects will influence task performance, which is manifested in the displayed skill-set recorded in a match situation.

From a statistical perspective, rugby brings about a special set of challenges because it is complex and chaotic. Circumstances change from game to game, even from phase to phase, due to numerous varying conditions; for example, weather, tactics, available personnel and standing in the competition. This exposure to match volatility must be considered when establishing a rating system. That is, one week an individual's team may be starved of possession and, consequently, will be required to defend for the majority of the game. Conversely, on another week the same team may have a wealth of possession such that attacking opportunities abound. Ultimately, to survive in the fiercely contested Super 12 competition, individuals must be multi-skilled; that is, competitors must be equally adept at attack, defence and any other positional responsibilities such as scrummaging, as required. Consequently, performance and ability are multifaceted – attacking skills are required when the team is in possession and defensive skills are called upon when the opposition has the ball. This expected overlap is useful when establishing a rating system. Bracewell [6] showed that assessment of each raw measure is vulnerable to match volatility. However, by considering the overall contribution of an individual to a team, a relatively robust performance measure is obtained. This assumes an individual's potential involvement is likely to remain constant and indicative of his or her worth to the team. In rugby union this is a relatively safe assumption due to the well-rounded skill-set required for competitors to reach the first class level.

The basic principles for obtaining an individual performance measure from match data are based on four key steps. Firstly, individual performance must be defined and then operationalised by listing all relevant physical tasks such as tackles, passes, and kicks. This allows match involvement to be quantified (see layer 1 of Figure 1). Quantification of match involvement enables performance measures representing core skill groupings to be calculated (see layer 2 of Figure 1). Finally, this allows overall performance to be established (see layer 3 of Figure 1). This template provides the necessary foundation to extract performance measures from match data.

Depending on the dimension reduction technique used, the skills (KPI) are likely to be uncorrelated in the second layer after step 3 (i.e. orthogonal rotation of factors). This is useful from a coaching and/or selection perspective as it creates a portfolio of measures that describes the profile of an individual.

4. Creating Meaningful Ratings

The extraction of an individual performance measure is similar to the approach adopted by Charles Spearman in his attempt to quantify human intelligence by examining the correlations between specific mental tests [19]. This lead to the development of factor analysis which is based on the premise that a number of highly linearly correlated variables is indicative of an underlying structure – comprised of factors – determining the behaviour of the measured characteristics [25]. Consequently, it was hypothesised that the data produced in a match was shaped by the individual's underlying ability in different facets of rugby. Due to the range of tasks undertaken in a rugby match, it is necessary to extract more than one factor. Each factor represents a core trait of performance, also known as a Key Performance Indicator (KPI). KPI are essentially a summary of the single physical task variables for each match, which is shown as the left layer in Figure 1. The second layer shown in this figure, representing the core skill-sets or KPI, are extracted using dimension reduction techniques as described next.

4.1   Dimension Reduction of Rugby Data

The data needed to be cleaned prior to commencing dimension reduction. Primarily, this involved identifying heterogeneous positional clusters from the 15 playing positions. Clustering of expert opinion was used to define positional clusters that needed to be identified due to the different demands placed on each of the different positions in a game situation. Eight clusters were identified, listed as follows: props, hooker, locks, loose forwards, halfback, first five eighth, midfield backs and outside backs.

For the amount of data generated in a rugby match to be useful to rugby observers, match statistics need to be summarised. Dimension reduction techniques are suitable for creating meaningful summaries of the data, provided a lower intrinsic dimensionality exists [11]. Three techniques (factor analysis, self-organising maps (SOM) and self-supervising feed-forward neural networks) were applied to physical task data.

4.2   Evaluating Key Performance Indicators

Factor analysis was the ideal starting point for analyses of player performance due to the interpretability of the latent factors using factor loadings. Five factors were required for each positional cluster to explain more than 60 percent of the variability within the data using principal component extraction and a varimax rotation. The factors were meaningful as they represented key aspects of rugby such as attack, defence, and kicking. A major advantage of using factor analysis is the ease of implementation as the structure is a linear equation. The relatively low number of adjustable parameters ensures that it is easy to update the system given changes in style of play. The use of continuous data is also advantageous for identifying small differences between comparable players. The factors enable strengths and weaknesses in an individual's skill-set to be determined.

Unfortunately, factor analysis lacks robustness, which means that the rating system may not be meaningful. This limitation is a consequence of the linear functions used to formulate the factors, making the ratings vulnerable to the amount of play in a match. This problem was highlighted during the All Blacks European Tour, 2000. Whilst there is some justification for strictly increasing ratings with respect to positive involvement, feedback provided by the Sky Television commentary team suggested that this is not desirable in a rating system for individual rugby players. Essentially, as the total number of sequences in a match increases so does the original rating developed for Ultimate Rugby and Sky Television [6]. In the test series against France, ratings were generally higher in the second test defeat than in the first test victory. This was due to the increased positive involvement of the individuals in the second test. This paradox highlights the problem of incorporating a linear structure into a model for individual performance. Obviously, re-scaling the performance measure by accounting for overall quantity of play is an option. However, other methods are pursued to establish relative capabilities.

Neural networks cater for non-linearity in the latent model structure [10]. In Kramer's [16] non-linear principal component analysis the need for target data in this five-layered feed-forward network is negated by replicating the original inputs as the target outputs [18]. The introduction of neural

networks as a potential modelling tool served as a reminder that typically sport generates relatively little data. To combat the vulnerability of neural networks to over-fitting the variables are summarised in an attempt to have five to ten observations per adjustable parameter as recommended by Berry and Linoff [3]. As a result the original 93 variables are reduced to just 19 (see Table 2) by combining similar variables, with minimal loss of information due to the small number of direct interactions with the ball a player has in any match. For example, it is more appropriate to consider total tackles made, rather than the distinct types of events such as turnover tackles with the left shoulder coming from the outside-in at chest height, due to the rarity of such events for a single player in a single match.

To further condense the variables, appropriate variables were then put into two major attribute categories, attack or defence for separate analysis. This action is supported by the results obtained from the two-dimensional SOM, and has huge implications on how performance is perceived and ability is quantified. Indeed, the apparent overlap between the methods validates the extraction of latent performance variables.

To quantify performance using the two main attributes (attack (7) and defence (7)), a five layer feed-forward self-supervising network, with a bottleneck layer representing the dimension reduction "output" was applied for the defence and attack categories for each positional cluster as suggested by Kramer [16]. Specifically, these five layers are an input layer, hidden mapping layer, bottleneck layer (which serves as the "output"), hidden de-mapping layer and a target layer. Bias was included on all layers except in the bottleneck "output" layer. To improve the generalisation of the back-propagation trained network, early stopping was employed during training and a hyperbolic tangent was utilised as the activation function in the mapping and de-mapping layers. The network architecture was established heuristically to minimise the average error in the test data set. Whilst the number of nodes in the mapping and de-mapping layers varied according to positional cluster, the middle bottleneck layer held two nodes in all instances. For example, the architecture – the number of nodes in each layer – for the midfield back cluster was 7,7,2,7,7 obtained from a sample size of 279 meaning that there were 1.29 observations per training parameter.

Table 2: Summary Attributes

Attributes		
Attack	*Defence*	*Other*
Breaks	Breakdown Impact	Infringements
Defence Beaten	Harassment	Kicks
Errors	Kicks Caught	Lineouts Lost
Laybacks	Loose ball Gained	Lineouts Won
Passes	Tackle Assists	Metres Kicked
Running Metres	Tackles	
Tries	Tackles Missed	

A number of benefits arise from the use of self-supervising feed-forward neural networks. Continuous data is obtained, allowing small differences in performance to be identified. Unlike factor analysis, the use of sigmoid activation functions reduces the impact of outliers. The 'S' shaped curves of sigmoid functions dampen the effect of extreme values reducing the impact of match involvement, which is apparent with a linear structure. As the network is functionally independent, relationships between variables can be non-linear. This enables the limiting relationships that are part of rugby performance to be better handled. Limiting relationships are best illustrated by the contrast between attack and defence, where an individual can only do one or the other at any time.

Ratings need to be transparent, or interpretable. This is necessary to ensure that the ratings are contextual, which leads to understanding, credibility and marketability. Whilst this was not a problem for factor analysis and SOM, neural networks are renowned for their lack of interpretability. As so few variables were involved, influence of individual variables on the dimension reduction output was established using graphical methods associated with sensitivity analysis [12]. This allowed the networks to be interpreted easily. Ensuring transparency in the case of the neural networks, commonly referred to as black boxes in the literature [22], was assisted by the development of a new statistical tool enabling significant variables to clearly identifiable. The half-moon statistic (HM) was

developed to promote transparency, understanding and interpretation of dimension reduction neural networks [6]. This novel heuristic, non-parametric multivariate method is a tool for determining the strength of a relationship between variables, whilst not requiring prior knowledge of the relationship between multiple input variables and a single output variable. This resolves the issues of interpretability, which is necessary to ensure a contextual rating system. This was achieved by comparing input-output pairs of observations with an enveloping circle to detect any changes from a null state of independence. Additionally, this method shows a great deal of promise as a diagnostic tool for use with multivariate data. The HM statistic provided output analogous to factor loadings enabling the important variables in a multiple input-output relationship to be established.

Ultimately the use of neural networks in quantifying ability is limited due to the vast amount of data required to justify the inclusion of adjustable parameters without over-fitting. However, a component-based system isolating key measures reduces greatly the number of adjustable parameters. Realistically, despite the effort to reduce the number of variables and consequently the associated number of adjustable parameters, the networks employed were still prone to over-fitting. Typically, there was one observation per adjustable parameter, rather than the recommended 5-10 observations. Not all models were contextually meaningful, invalidating the use of neural networks as a stand-alone method. However, for some positional clusters the results were contextual, indicating that the self-supervised feed-forward neural networks could be of potential use. Consequently, the contextual neural network components were retained and augmented with contextual factors from the factor analysis to produce two attribute indices relating to attack and defence. Positional specific indices were created for each cluster to highlight the core skills required in each position. For example, lineout-throwing accuracy was included for hookers, punting variables were considered for inside backs and scrummaging formed the basis for the prop specific attribute.

Of the three applicable techniques for dimension reduction, self-organising maps produce discrete output that is too coarse for use in the coaching environment. Due to the small sample sizes it is difficult to calculate the differences between and within individuals. Consequently this method is not discussed here. However, as a dimension reduction tool it successfully identified a contextual two-dimensional surface, with attacking performance represented by the first dimension and defensive performance on the other [7]. This validated the results from both the factor analysis and unsupervised neural network.

The data mining of match data provided three key performance indicators, or core skill groupings for each positional cluster. In order to create a contextual overall match rating, these indicators needed to be combined meaningfully.

4.3   Evaluating Overall Performance

The KPI are collapsed to a single performance measure via the perfection method [5]. This is an adaptation of quality control ideology and observations are compared with absolute perfection rather than the average to suit the circumstances presented in sport, producing a one-dimensional value that is reflective of an individual's performance in a single match. Sports people aim for perfection, not mediocrity, thus to apply control chart philosophy to sports data the focus needed to change. The control statistic is therefore defined as a Mahalanobis distance from "unattainable perfection" rather than the average as suggested by Hotelling [14]. The effect of comparing the KPI to perfection reduces the impact of dynamic match conditions, by considering the contribution to the team (combined impact of all KPI). These dynamic match conditions can also make the assessment of raw measures inappropriate in one-off situations, limiting between-player and within-player comparisons. Thus a contextual, transparent univariate measure for individual performance had been successfully created. Importantly, these ratings are approximately normally distributed meeting a major assumption necessary for statistical inference.

5. Monitoring Meaningful Ratings

Sport statistics for individuals seek to quantify sporting performance, the 'process' in this case. Quantification of sporting performance on a match-by-match basis is similar to the statistical area of quality control in that the nature of a process is quantified, such that any deviation from normal (expected level of

ability) is quickly identified [20]. Control charting procedures provide the ideal medium to allow coaches, selectors and other interested parties to quickly evaluate the performance of a given individual in a single match with respect to their expected ability based on past performances [5]. The use of a performance rating and the underlying key performance indicators requires the construction and then deconstruction of the rating system [6]. After construction, individual performance is monitored with the use of control charts (Shewhart and EWMA) enabling changes in form to be identified. This enables the detection of strengths/weakness in the individual's underlying KPI and skills, through the deconstruction of the performance rating(s).

Following the acquisition of a suitable rating and skill measures, examples are provided to demonstrate the effectiveness of these sports statistics in the proper context, monitoring the performance of individuals. The examples shown demonstrate the implementation of monitoring procedures after the completion of a specific competition, but the same approach can be adopted during the season using match data as it becomes available. Both the skills and overall performance, as outlined in Figure 1, are examined using graphical tools. Obviously, the summarized data can be explored at different levels of refinement depending on the needs of the end user. However, it is inadvisable to start with analyses of raw data first due to varied match conditions. It is preferable to assess overall performance first to see if the 'process' (player performance) is in control. In the first example, data for an entire competition is examined for each individual negating the effect of changeable match conditions by considering many 'sampling situations'.

Figure 2: Radar Plot for Comparison of Three Skills for International Halfbacks based on Super 12 2000 Performances
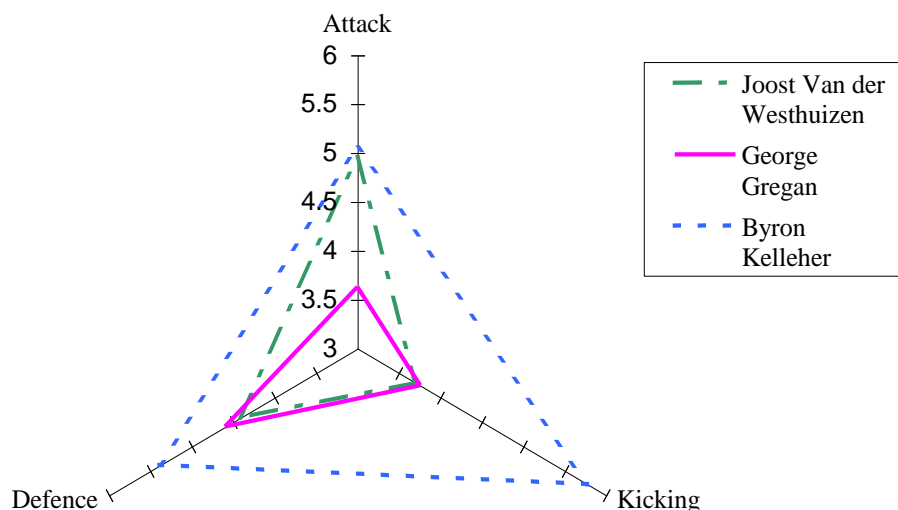
In the selection process it is useful to compare the skill sets of individuals. Figure 2 directly compares three halfbacks involved with international duties in the 2000 season. Mean scores for the three key attributes (attack, defence and kicking) are displayed for each individual (sample sizes are provided below). The data was obtained from the Super 12 competition in 2000 prior to international games. The key attributes are displayed in a radar plot. This graphical summary gives an immediate comparison of the key aspects of overall performance. An axis is dedicated to each of the three key attributes. The centre of the plot is associated with low performance. Conversely, the extremities of the axes relate to higher performance. Each of the key attributes was standardised to have a mean of 5 and standard deviation of 2. The scale of the axis is truncated to best illustrate the differences between individuals (minimum of 3 and maximum of 6). Given that at the international level slight weakness or strength will be exploited, this manipulation of the scale is justified. The larger the area of the polygon, the better the individual has performed. Further, a regular shaped polygon indicates that the player is relatively equal in each facet. This enables strengths of weaknesses to be identified quickly. The above graph shows that New Zealand's Byron Kelleher (12 matches) is the most complete halfback with all attributes being above average and equally strong in each area. George Gregan (13) from Australia and Joost Van der Westhuizen (4) of South Africa were approximately equal in their defence and kicking qualities. However, the radar plot indicates that van der Westhuizen is a better attacking player than Gregan. The data can be further mined to look for other strengths and weaknesses.

A Shewhart control chart is shown is Figure 3 for the kicking attribute of former All Black and Canterbury Crusaders first five eighth, Andrew Mehrtens. In the case of the Super 12 Tournament, there will be at most

13 matches played by any given individual, assuming that the individual's team makes it through to the play-offs. This means that a control chart has little power (opportunity) to detect "out of control" situations. Tight upper and lower control limits set two standard deviations away from the centre line are therefore appropriate [5]. In an additional attempt to increase the power, all eight run rules are applied to identify any changes in performance and form. In this case an alarm is signalled in the ninth match, the 75:27 win over the Northern Bulls. The raw statistics show that in this match Mehrtens only kicked a total of 139 metres from 7 kicks (19.9 metres per kick) before being replaced by Aaron Flynn. The high score suggests that the Bulls were completely dominated by the Crusaders. Therefore it is likely that Mehrtens was kicking to retain possession and create attacking opportunities rather than kick for territory and maintain pressure, as was the case in the Super 12 final. This difference is pronounced when the kicking statistics from the ninth match are compared to those from the Super 12 final, where Mehrtens kicked 499 metres from 14 kicks (35.6 metres per kick). It is important to note the average attribute score of 7.09 is well above the population average of 5 for this attribute (standard deviation = 2). As mentioned previously, skills are vulnerable to match volatility, hence the alarm from match nine; so it is preferable to examine overall performance first.

Figure 3: Shewhart Control Chart of Kicking Attribute for Mehrtens, Super 12 2000
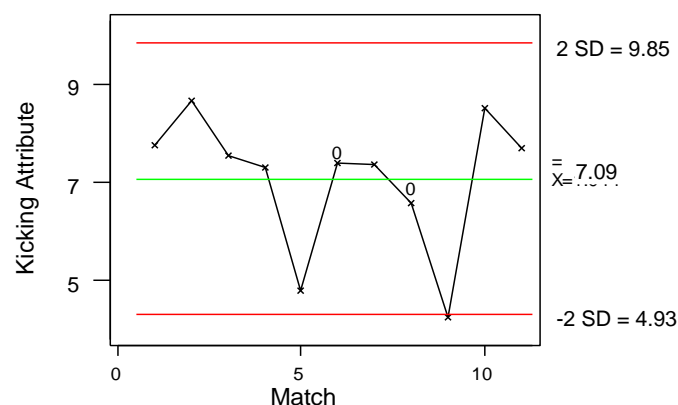
Ability cannot be established from one match alone, so in this instance the data is smoothed using an exponentially weighted moving average (EWMA) so that trends or significant changes in performance can be identified. Given the limited amount of data available, it is suggested that the smoothing constant is set equal to 0.25 and the sigma limits are set at two [5]. The Shewhart chart shown previously plots the actual match ratings. The EWMA (exponentially weighted moving average) includes past performances when giving a match rating. This has the advantage of being able to pick up small changes in performance more quickly than the Shewhart Chart.

Figure 4: EWMA Control Chart of Performance Ratings for Mehrtens, Super 12
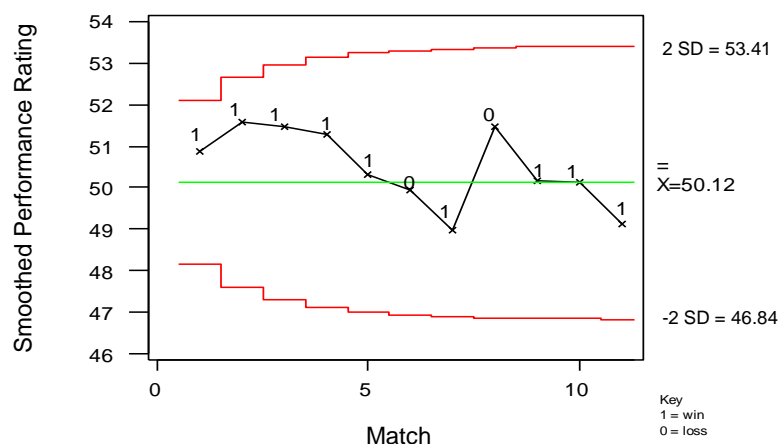
Figure 4 suggests that Mehrtens overall match performance remained consistent throughout the Super 12 as no alarms are signalled. This suggests that his ability has not changed. Whilst his kicking attribute (shown in Figure 3) showed that he was well above average in that respect, the overall performance rating of 50 is average. Further examination of Mehrtens' skill set reveals his strengths and weaknesses. The three core attributes of first five eighths are attack, defence and kicking. Mehrtens rates 6.2, 1.9 and 7.1 for these

attributes respectively. This means that Mehrtens talent with the boot is hampered by his limited defensive qualities. Mehrtens weakness on defence is well known and attacking opposition teams often target him. Consequently, the Crusaders have at times resorted to moving Mehrtens to fullback or wing when defending.

The control chart for Larkham shows that his general performance remained consistent throughout the season. It is also worthwhile noting that his average is 63, compared to the population average of 50. This supports the general perception that Larkham was one of the better first five eighths in the world.

Figure 5: Shewhart Control Chart of Performance Ratings for Larkham, Super 12 2000
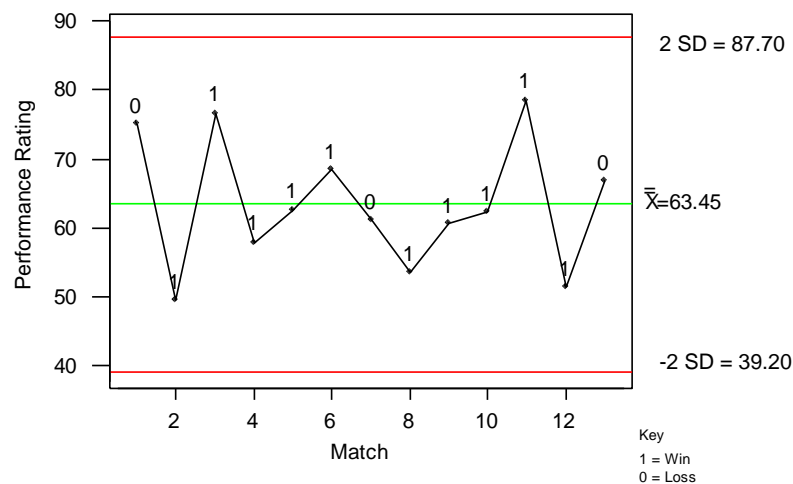
Underlying the overall match rating is the key performance attributes. These can be assessed using tools for statistical inference to identify potential match strategies. An investigation of the statistics from Larkham's Super 12 season produces a number of significant results. There is a significant correlation between his match rating and his attacking ability ($r = 0.83$). That is, the more often and further he runs, the higher the match rating. Additionally, the existence of a significant negative correlation between kicking and attacking implies that he either adopts one style of play or the other ($r = -0.90$). This claim is made because the Shewhart control chart suggests that overall performance did not change. A simple ANOVA showed that the higher Larkham's kicking attribute the more likely his side is to lose ($p=0.000$). This suggests that teams that are able to stop Larkham running freely and force him to kick are more likely to win.

6. Limitations

The limits of a statistical rating system are governed by the data available. In this study, the data available was restricted due to commercial circumstances. Furthermore, there was no information relating to the spatial, technical or team aspects of performance. Data relating to spatial and technical aspects of performance are costly and difficult to obtain. Additionally, information relating to team play is very difficult to define or collect and then relate back to a specific individual. Because of these limitations, the rating system discussed in this paper can only be viewed as a general measure of individual performance. Additionally, sport generates relatively small data sets for data mining. Thus the dimension reduction networks utilised in this paper may be prone to over-fitting when retraining occurs. This could affect the contextuality of the models, although this could be remedied by using regularisation in the training process.

When a rating system is viewed as supplemental information, then the negative aspects relating to the application of the rating system are dispelled. Incorrectly and overused statistics create the possibility of coaching in hindsight, where action is taken well after the event, as described by former New Zealand Test cricketer, Brendon Bracewell [4]. As a result coaching becomes reactive rather than proactive, a situation that is undesirable at the first class level. Additionally, as sport statistics are outcome focused, the processes (such as technical and mental) may be overlooked. Consequently, provided sports ratings (and other measures of sporting performance in complex team sports) are viewed as a general measure of performance that is a useful tool for detecting trends in performance and communicating information to the public via the media, no problems will arise.

7. Conclusion

The methodologies described in this paper are exciting given the potential use by both coaches and the media. Contextual data (KPI) with a lower dimensionality can easily be displayed, enabling visual comparisons (within-individuals and between-individuals) to be made rapidly. This is useful to the media for communicating vast amounts of information in the simplest possible way aiding the entertainment package of professional sport.

Properly implemented statistical procedures improve the quality of the secondary information available to coaches and selectors. The ratings and KPI that comprise the secondary information are important because they 'smooth' the inherent match volatility to which univariate statistics are vulnerable. The problems associated with univariate statistics due to match volatility highlight the need for 'stable' data, which is provided in the form of the overall performance measure obtained via multivariate techniques. Furthermore the time spent exploring a database is minimised by pinpointing the possible areas of anomalous behaviour.

However, to implement data effectively in any environment, caution must be exhibited when choosing a potential method. Domain knowledge is crucial to ensure that the database is capable of providing solutions to complex problems. Whilst data mining techniques are excellent at extracting answers from large and complex databases, it is up to the practitioner to ensure the validity of the results. Furthermore, the needs of the end user must be evaluated and built into the data mining process.

The philosophy described in this paper is not limited to rugby or sport data. The methodology is applicable to any circumstance where an understanding of multivariate data is required. Careful examination of the data and methods can ensure the validity of analyses, promoting greater comfort in the information age.

8. References

[1] Bennett, J. M. Did Shoeless Joe Jackson Throw the 1919 World Series? Proceedings of the Section on Statistics in Sports of the American Statistical Association, (1992), 33-38.

[2] Berkmann, J. *The Complete Guide To Test Cricket In The Eighties*. Partridge Press, London, (1990).

[3] Berry, M. J. A., & Linoff, G.S. *Data Mining Techniques,* John Wiley & Sons, New York, (1997).

[4] Bracewell, B. P. (2001). *Coaching - The 'Contract Syndrome' Disease*. Available: www.cricketnz.co.nz/page19.html (accessed 18/9/01).

[5] Bracewell, P.J. Monitoring Meaningful Rugby Ratings. *Journal of Sports Sciences*, **8**, (2003), 611-620.

[6] Bracewell, P.J. *Quantifying Individual Rugby Player Performance through Multivariate Analysis and Data Mining*. Unpublished Ph.D. Thesis. Massey University, Albany, New Zealand, (2003).

[7] Bracewell, P.J., Meyer, D.H., & Ganesh, S. Extracting Measures of Performance for Individual Rugby Players: Data Mining in Sport. *Proceedings of Artificial Neural Networks and Expert Systems Conference*. (Kasabov & Woodford eds.). Otago University, Dunedin, (2001), 38-42.

[8] de Lacy, H., & Fox, G. *Think & Play Winning Rugby*. Harper Sports, Auckland, (2000).

[9] Franks, B.D., & Deutsch, H. *Evaluating Performance in Physical Education*. Academic Press, New York, (1973).

[10] Fausett, L. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice Hall, New Jersey, (1994).

[11] Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. *Multivariate Data Analysis with Readings.* (4th ed.). Prentice Hall: Englewood Cliffs, New Jersey, (1995).

[12] Haykin, S. *Neural Networks: A comprehensive Foundation.* Prentice Hall, New Jersey, (1999).

[13] Hawkins, J.M. (Ed) *Oxford Senior Dictionary*. (7th ed.). Oxford University Press, London, (1990).

[14] Hotelling, H. Multivariate Quality Control. *Techniques of Statistical Analysis.* (Eisenhart, Hastay & Wallis eds.). McGraw-Hill, New York, (1947).

[15] Hughes, M. Notation Analysis In Football. Communications To The 2nd World Congress Of Science And Football. In Reilly, T., Clarys, J., & Stibbe, A. (Eds.) *Science And Football II*. E & FN Spon, Ipswich, (1991), 151-159.

[16] Kramer, M.A. *Nonlinear Principal Component Analysis Using Autoassociative Neural Networks.* AIChE. **37**(2), (1991), 233-243.

[17] Mackinnon, M.J., & Glick, N. Data Mining And Knowledge Discovery In Databases – An Overview. *Australian & New Zealand Journal Of Statistics*. **41**(3), (1999), 255-276.

[18] Malthouse, E.C. *Limitations of Nonlinear PCA as Performed with Generic Neural Networks.* IEEE Transactions on Neural Networks. **9**(1), (1998), 165-173.

[19] Manly, B.F.J. *Multivariate Statistical Methods: A Primer*. 2nd ed. London, Chapman & Hall, (1994).

[20] Mosteller, F. Lessons from Sport Statistics. *American Statistical Association.* **51** (4), (1997), 305-310.

[21] NZRFU. *Coaching Accreditation Manual: Level 3.* Wellington, NZRFU, (1991).

[22] Ripley, B. D. Neural Networks and Related Methods for Classification. *Journal of the Royal Statistical Society*, Series B, **56**(3), (1994). 409-456.

[23] Stewart, J.J. *Rugby: A Tactical Appreciation*. Auckland: Rugby Press, (1987).

[24] Thomas, V. *Science and Sport: The Measurement and Improvement of Performance*. Faber & Faber, London, (1970).

[25] Weiten, W. *Psychology: Themes and Variations*. 3rd ed. Brooks/Cole, Pacific Grove, (1995).

USING PERFORMANCE OR RESULTS TO MEASURE THE QUALITY OF SPORTS COMPETITORS

Michael Cameron[2]
Department of Economics
University of Waikato, Private Bag 3015
Hamilton, New Zealand

Abstract

This paper examines the theoretical framework that underlies performance-based sports ranking systems. A basic model of competitor quality is developed with special attention given to assumptions necessary to create the reduced-form performance model which is used by most sports ratings systems. The key assumptions that underlie the reduced form model include optimality of team selection, and optimality of strategy selection. The paper then tests a simple linear dynamic model applied to Super 12 rugby. The model performs adequately, accurately predicting nearly 64% of results from the 2003 season, and over 70% of late-season results. The incorrect predictions of late-season games illustrate that not only are the predictions not sensitive to deviations from optimality in strategy or team selection in the current trial, but the ratings themselves are likely to be biased (in unknown directions and of unknown magnitudes) due to deviations from optimality in previous trials. Future developments in this theory should focus on measurement of the factors that impact on player skills and elements of successful team selection and strategy, as well as further development of the theory surrounding risk preference, utility of performance and discounting.

1. Introduction

The ranking of sports competitors has long been a topic of animated discussion. Every person who holds more than a passing interest in a competitive sport has their own perception of who is the 'best', and some preference ordering or ranking of competitors. These perceptions may be subject to significant subjective bias. Sports ranking systems are mathematical models that attempt to objectively determine a ranking for competitors within a sport.

The problem with current rankings systems is that they often do not provide a solid theoretical framework to support their methodology, and in the case of numerical ratings they are not always clear what it is that the rating actually measures. That a system is performance-based suggests only that the ranking is reflective of each competitor's past results i.e. these systems are to some extent 'backwardly-predictive'. However, this interpretation differs from the generalised use of rankings systems, which is to infer who is most likely to win a future contest between any two competitors.

This paper begins by providing a theoretical framework for the ranking of sports competitors on the basis of past results, and develops a model of competitor quality, with special attention given to the assumptions that are necessary to reduce the full model of competitor quality to the 'reduced-form' model in use by most current ratings systems. The results obtained are robust to the specification of the mathematical model – they apply equally regardless of the type of rankings system. Finally a simple dynamic ratings system is used to illustrate the limitations of the reduced-form model. Predictions based on this model are not only not sensitive to deviations from optimality in strategy or team selection in the current trial, but the ratings themselves are likely to be biased (in unknown directions and of unknown magnitudes) due to deviations from optimality in previous trials.

The implications of this paper are that the ranking or ratings provided by sports rankings systems need to be interpreted carefully, and interpretations are highly sensitive to how the 'best' competitor is defined. This paper suggests that defining the 'best' competitor as the highest quality competitor, where quality is a measure of a competitor's potential performance, provides a consistent interpretation. Using this interpretation should allow the sports ranking to be used, along with additional data about team selection and strategy, to more accurately predict future results.

---

[2] The author is indebted to Steven Lim and John Tressler for their helpful comments on earlier drafts of this paper.

2. An overview of sports rankings systems and performance

There are almost as many different sports rankings systems as there are possible ways of ranking competitors. All these systems can be separated into two broad themes – those that simply create a preference ordering for competitors ('ordinal ranking' systems) and those that provide a numerical value or rating which is the basis for ranking ('cardinal rating' systems). Both systems may use simple or increasingly complex mathematical modelling to develop their ranking.

Ordinal rankings systems include 'ladders', whereby competitors move ahead of any team above them on the ladder which they beat (or possibly below any team that they lose to), or more complex systems that determine a preference ordering based on time-weighted past results[3]. Cardinal ratings systems are numerical models which compute, using variables such as results, number of points scored, home advantage, numerical 'ratings' of the competitors in that sport. The competitors can then be ranked on the basis of their rating. Both types of systems have merit, but require the same set of assumptions to be satisfied, and are subject to the same limiting principles outlined in this paper. However, the analysis provided here is probably more easily applied to a cardinal (numerical) ratings system.

3. Modelling competitor quality and performance

Most performance-based sports rankings systems claim to measure the 'performance' of competitors[4]. Performance may in fact be an extremely simple variable to measure – a competitor either wins or loses. However, the factors that determine competitors' performance in any given sporting contest, or trial[5], are complex and both the interactions between them and the mechanisms through which these factors translate into performance are not well understood (e.g. see Yilmaz and Chatterjee [11]).

First, a careful definition of quality is required. In this paper, quality will be defined as a measure of a competitor's potential performance – a result of their inherent characteristics, and in particular their inherent ability to translate potential performance into actual performance, and their propensity to deviate from optimal strategy. The quality of competitor x can be expressed as some function of the characteristics of that competitor:

$$T_x = T(\text{characteristics of competitor } x) \qquad (1)$$

In the simplest sense, a competitor of a higher 'quality' could be expected on average to beat, or perform better than, a competitor with lower 'quality' in any trial. Since our measure of quality takes into account how potential performance is translated into actual performance, then any measure of the performance of a competitor is an indirect measure of the quality of the competitor.

The only direct evidence of the 'quality' of a competitor is their performance, as measured in the results of n trials. The expected, or average, result of each trial is determined by the quality of competitors competing in the trial. In the case of a trial involving two competitors, x and y, the result R_{xy} can be simply expressed as a function of the 'quality' of each competitor, and random variation ε:

$$R_{xy} = R(T_x, T_y, \varepsilon) \qquad (2)$$

Assumption One:
> ε is normally distributed with a mean of zero, and uncorrelated with any of R_{xy}, T_x, or T_y.

Assumption one is the standard assumption in any econometric application. Violations of assumption one introduce bias.

The result of a trial can be defined in many ways. It could be a simple binary result (1 = win, 0 = loss), a margin of victory (or loss), or a combination of attacking or defensive or other statistics. However, the key

---

[3] For example, see Ross Finlayson's World Rugby Rankings at http://www.live.com/wrr/
[4] For example, see http://www.atptennis.com/en/players/entrysystem/ and http://www.fifa.com/en/rank/procedures.html for ratings systems for men's tennis and international soccer, respectively.
[5] Hereafter a single sporting contest, involving one or more competitors, will be referred to as a trial.

point is that a competitor of higher quality has a greater chance of achieving a 'better' result than a lower quality competitor. A better result in this case is a greater chance of victory, a greater margin of victory, or a 'better' combination of statistics, depending on which metric is used to measure the result.

Motivation is one important aspect of the translation of quality into performance – competitors should be motivated to succeed. This suggests a further assumption is necessary:

Assumption Two:
> *Every competitor has strictly positive marginal utility of results.*

This assumption identifies the motivation or drive to success of all competitors. It suggests that, ceteris paribus, a competitor would always strictly prefer a better result. This may seem obvious, but consider the way that results are calculated. In the simplest case, where the result is only defined as a win or loss, then we could reasonably expect a competitor to strictly prefer the 'better' result. However, if the result is defined across the margin of victory, or some combination of offensive and defensive statistics, this might not necessarily be true. If the competitor is only concerned with victory and not the margin of victory, as seems reasonable for any trial in a knockout competition, then additional margin of victory beyond some 'safety margin' might have zero marginal utility (or even negative marginal utility if there is a chance of injury for the competitor which would impact on future results), and assumption two would be violated.

If this assumption is violated, then we might not be able to infer that higher quality competitors would, on average, achieve 'better' results. Competitor quality (for both competitors in the two-competitor case) is no longer the sole determinant of the result, and this violates assumption one.

By rearranging equation (2), we can derive a probability distribution function for the 'quality' of the competitor, based on performance when the quality of competition is allowed for. This probability distribution function provides a 'rating' of competitor x, τ_x, which is an imperfect estimate of their 'quality', and this distribution could be obtained from the distribution of ε in equation (2) if ε were known. This 'rating' can be expressed as a function of their result against competitor y and the quality of competitor y:

$$\tau_x = \tau(R_{xy}, T_y) \tag{3}$$

Equations (2) and (3) are the simple basis under which most sports ranking systems, both ordinal and cardinal, operate. Of course in calculating τ_x there are a number of different methods. Most involve the use of a matrix or a dynamic progression of results in the calculation of the 'rating' (for specific examples, see Stefani [9]). All ratings systems make the assumption that the central limit theorem holds, even though T_x might not be a static parameter. Even if the quality of the team changes over time, the number of observations (trial results) increases, the rating will approach the 'quality' for each team.

Assumption Three:
$$\lim_{n \to \infty} \tau_{xn} = T_x$$

If there are a sufficient number of trials and assumption three holds, then equation (3) becomes:

$$\tau_x = \tau(R_{xy}, \tau_y) \tag{4}$$

If assumption three is violated, this suggests that what is actually being measured (τ_x) is not a statistically good estimate of T_x, and some other estimator should be used. Since quality T_x is what is intended to be measured, assumption three must hold.

The essential problem with ratings systems is that equation (2) grossly simplifies the number of factors that impact of the result of a trial, and therefore assumption one is violated, as is assumption three, regardless of the number of observations. If we were to use the simple model presented above, then factors which we know to impact on the result of trials, such as home advantage, would not affect any predictions we made. Excluding certain variables that are known to affect results (i.e. those that are not orthogonal to the results variable) introduces omitted variable bias (Verbeek [10]). This bias could be significant, and thus any predictions made using the simplified model would also be significantly biased.

A better model for the result of a trial would include all factors known or suspected to impact on the result of the trial, as in equation (5) below, which may represent the 'ideal' performance model.

$$R_{xy} = R \left(H_{xy}, S_x, S_y, \psi_x, \psi_y, \lambda_x, \lambda_y, \delta_x, \delta_y, \gamma_x, \gamma_y, \varphi_x, \varphi_y, \tau_x, \tau_y, \varepsilon \right) \qquad (5)$$

In equation (5), R_{xy} represents the result between competitors x and y; H_{xy} represents a combination of the 'home advantage' of competitor x and the 'away disadvantage' of competitor y; S_x and S_y represent the strategy employed by competitors x and y respectively; ψ_x and ψ_y represent the 'natural skill level' of the competitors; λ_x and λ_y represent the experience of the competitors; δ_x and δ_y represent the physical decline of the competitors; γ_x and γ_y represent the effect of coaching for each competitor; φ_x and φ_y represent the synergies of skills within the team, in the case of a team sport; τ_x and τ_y represent the rating of competitors x and y respectively, (which is likely to be equivalent to their quality if the number of trials is large); and ε represents random variation in results, which is assumed normally distributed with a mean of zero (see assumption one, above).

In the case where a competitor is an individual, the terms φ_x and φ_y are zero. In the case where a competitor is a team, then $\psi_x = \psi(\psi_{x1}, \psi_{x2}, \ldots, \psi_{xn})$; and $\psi_y = \psi(\psi_{y1}, \psi_{y2}, \ldots, \psi_{yn})$; where n is the number of team members. The same is also true of the variables λ, δ, and γ. Where more than two competitors are involved in the trial (for example a golf competition or an athletics race), the formula should be expanded to include values of S, ψ, λ, δ, γ, φ (where applicable), and τ for each competitor.

Each of these variables, or combinations of the variables, is discussed in further detail below.

3.1 Home advantage and conditions advantage

Home advantage is a widely acknowledged aspect of all competitive sport, and many studies have been undertaken in this area (e.g. see Courneya and Carron [3] or Nevill and Holder [8] for literature reviews, or Bray and Widmeyer [2] for more recent work). Results often suggest that a number of factors create home advantage, such as crowd, familiarity, and travel factors (Nevill and Holder [8]), or psychological factors such as perceived increases in collective efficacy (Bray and Widmeyer [2]). 'Home' advantage can probably be extended further to include the possibility that proximity to home confers some advantage to one competitor (or lack of proximity confers some disadvantage to another).

It is likely that the idea of 'home' advantage extends beyond even proximity effects, and there may be in fact a 'condition advantage'. Condition advantage could be conferred where conditions in which the trial is conducted are favourable to one competitor even where proximity or 'home' advantage is negligible – these conditions might include climate effects, wherein one competitor performs better in certain climate, location effects separate from home advantage (e.g. certain tennis players perform better on different court conditions, like clay, or grass courts), and other effects such as rules of the trial (e.g. it is likely that a rugby league team would have some advantage against a rugby union team if the trial were conducted under rugby league rules).

Home advantage (and condition advantage) is the one variable that is so robust to scrutiny that it cannot be easily dismissed or eliminated from consideration. Where home advantage is not explicitly included in the formulation of a sports ranking system, there is likely to be significant bias, particularly where the distribution of trials (between 'home' trials and 'away' or 'neutral venue' trials) for competitors is not randomly determined. For instance, where home advantage is not included in the ratings formulation, a competitor could successfully increase their ranking by competing solely in 'home' conditions. This is an obvious violation of assumption three.

3.2 Team selection, skill, experience, physical decline or injury, coaching and team synergies

Team selection is one of the key components of a team's performance (as measured by its results), and is a process undertaken by 'team management'. Team management is the entity that is ultimately responsible for both the teams financial and competitive performance – it is an integral part of the

'competitor'. Team management has a utility function that is defined over its preferences for income and results:

$$U_x = U(R,Y) \tag{6a}$$

While no further assumptions are necessary about this function, for the purposes of this discussion let us assume that the utility of results and the utility of income are additively separable[6]. This allows equation (6a) to be rewritten as:

$$U_x = U(R) + U(Y) \tag{6b}$$

It seems reasonable then that team management will select a team that maximises its utility, rather than simply maximising the expected result. In amateur or representative sports, it is possible that the marginal utility that team management derives from income is close to zero, suggesting that team management is most concerned with results. This is somewhat less true of professional sports. In fact, it could be quite conceivable for a professional sports team to prefer a higher net income (due to lower player salaries) than 'on-field' performance. The Arizona Cardinals (N.F.L. football team) have consistently failed to produce good results over the course of many seasons, despite continually receiving a preferential position on the N.F.L. draft (which is based on season performance). Note that these results are possible even if assumption two (strictly positive marginal utility of results) is satisfied (e.g. see Fort and Quirk [7] and DeGennaro [5]).

In isolation, equations (6a) and (6b) assume that team management has no time preference, and makes its decisions based on utility for each result. In fact, team management is likely to derive utility not only from the result of the next immediate trial, but from the expected result of future trials, discounted at some appropriate rate to represent additional uncertainty. The same is true of current and future income. Team management may hold any risk preference, and there is little evidence in the literature about the risk preference of team management[7]. However, given that team management often have some responsibility to outside stakeholders over their management of the team, it seems possible that team management would be risk averse. In that case, the utility of team management is described by equation (6c) below.

$$U = \sum_{m=1}^{\infty} \frac{1}{(1+i)^{m+1}} U[E(R_m)] + \sum_{n=1}^{\infty} \frac{1}{(1+j)^{n+1}} U([E(Y_n)] \tag{6c}$$

If team management were risk seeking, then their expected utility would instead be described by equation (6d):

$$E(U) = \sum_{m=1}^{\infty} \frac{1}{(1+i)^{m+1}} E(U(R_m)) + \sum_{n=1}^{\infty} \frac{1}{(1+j)^{n+1}} E(U(Y_n)) \tag{6d}$$

Note here that the discount rates applied in equations (6c) and (6d) might be different for results (*i*) and income (*j*). It is possible even that the discount rate for results is not in itself constant. It seems plausible that when the next immediate trial for a team is a championship final or some other important trial, then the discount rate for future results might be much higher than that for an 'early season' trial[8].

Team management make decisions over the selection of players that form the team (note that of course in the case of individual sports then there is no selection decision to be made). The basis for this decision is the skill level of each player, and how well those skills will translate to results, both present

---

[6] If this were true then the marginal utility of results would not be affected by income, and the marginal utility of income would not be affected by results. While the latter seems reasonable, it is somewhat less likely that the marginal utility of results is independent of the income of the team.

[7] There is certainly an opportunity for research into the risk preferences of team management.

[8] Future research in the area of 'discounting of the utility of future performance' may also be fruitful.

and future. Following equation (5), the cumulative[9] skill level of all players in the team affects performance, and cumulatively higher skill levels improve expected performance, or increase the expected result.

Obviously a players 'skill level' is an extremely difficult thing to measure. A player's skill might be defined across any number of dimensions, and includes both physical skills and mental skills. Take for example the role of halfback in rugby union – this player will have some skill at kicking, passing, running with the ball, tackling, and so on. Each of these dimensions can be further dissected into more specific skills.

A players' skill set is then defined along a number of dimensions – if the degree of each skill that a player possesses can be represented by some real number, then a player's skill level can be represented by a column vector of order m, where m is the number of (physical and mental) skills that may impact on performance. A player's skill level (vector) is affected by many other factors, including their 'natural skill level', their experience, coaching, injuries and the natural decline in physical (or possibly mental) ability that comes with age, and the synergies that exist between that player's skills and the skills of other players in the team. Each of these factors might affect the whole skill vector in the same way, or affect different skills independently.

Initially, let a player's 'natural skill level' be the skill vector that they possess before any of the other factors have been taken into account – this is the level of skill they possessed before they became involved in competitive sport. Where a player has been involved in several different competitive sports, it will be the skill vector that they possessed before entering their first competitive sport, where the two sports involve similar skill sets (for example, players switching from rugby league to rugby union, or vice versa).

It seems reasonable that a player's skill vector would be affected by their competitive experience – a more experienced player should be better prepared, both mentally and physically, for the trial (e.g. see Duda [6]). Experience appreciates the skill vector of the player. Coaching is likely to also have a positive effect on the player's skill vector, although it is possible that poor coaching could have a negative effect. Coaching takes the form of both physical and mental training.

Injuries certainly reduce a player's ability to translate their skill vector into 'on-field' performance. Injuries are not confined to physical skills though – it is entirely possible for a player to be affected by 'mental injuries'. Injuries depreciate the skill vector of the player. As a player ages, it seems reasonable that their physical skills decline naturally (note that this process could be extremely gradual) – this could also be true of mental skills. The process probably accelerates as the player ages, which is why most competitive sports are contested by young competitors (e.g. see Berry and Larkey [1]). Each of the three effects above (experience, coaching, and injuries or physical decline) can be represented as a diagonal matrix of order m.

Finally, a player's skill vector, as expressed in how it translates into performance (or results), is likely to be affected by the skill vectors of all other players in the team – positively (synergies) or negatively (dysfunctions). These effects might work individually between player's skills, rather than the same between one player and the rest of the team, so the 'synergy' effect is best represented by a matrix of order mn x mn, where m is the number of skills that are 'measured', and n is the number of players in the team.

Note that only some of the factors that may affect a player's skill vector have been represented here. Other factors, such as equipment, or the use of performance-enhancing drugs, will also have an effect. Equipment might have two different effects, depending on whether the equipment is 'skill-enhancing', or 'skill-providing'.

Skill-enhancing equipment has an appreciating effect which can be represented in a similar was to the effect of experience or coaching, above. An example of skill-enhancing equipment is gloves for rugby

---

[9] Note that the method of cumulating may be additive or multiplicative (or some other method) and no inference is made here about the specific functional form that a model of 'team skills' should take.

players, which are designed to aid the handling ability of the player. Skill-providing equipment is equipment which provides the player with additional skills that they would otherwise not have possessed (or at least not possessed at a level sufficient to compete in that sport). Note that, unlike the inherent skills that the player possesses, the additional skills provided by equipment might not be affected by any of the other factors. An example of skill-providing equipment is a formula one racing car, which provides the driver with acceleration, braking, and speed skills that they might otherwise not have possessed. These 'skills' are unlikely to be affected by the experience of the driver (but may be affected by the experience of the team), coaching, or physical decline.

Recall the 'natural skill level' of each player can be represented by a vector of order 1 x *m*. We can represent the 'natural skill level' of a team by expanding the vectors of all *n* players into a team vector of order 1 x *mn*. Each of the other factors (experience, coaching, injury, and physical decline) can similarly be represented by diagonal matrices of order *mn* x *mn*. By multiplying through all of these factors and finally by the synergy matrix mentioned above, we can derive a measure for the 'adjusted skill level' of the team:

$$\Psi_x = \psi_x \lambda_x \delta_x \gamma_x \varphi_x \qquad\qquad (7)$$

In equation (7), a value of one for any element of the vectors λ (experience), δ (physical decline or injury), γ (coaching effect), or φ (synergies) implies that there is no effect on that particular skill for that particular player. Values of less than one depreciate the player's skill, and values of greater than one appreciate the player's skill.

The adjusted skill level is the skill level that the team will make use of in the trial. In practice it is difficult to measure most of the dimensions of a player's skill, and the factors that affect it, providing a pragmatic reason why these factors are ignored in most formulations measuring the performance of teams. However, it may be possible to exclude these factors where it can be assumed that, over time, the total 'adjusted skill level' of the team does not vary.

Assumption Four:
Team management is rational in team selection, and seeks to maximise its utility defined over income and results.

Assumption four suggests that team management will always select the 'best' team available, in that the best team is the one that maximises the utility of team management. Again, no assumption is necessary here about the risk preference of team management. If this assumption always holds, then rational team selection is consistent with the definition of quality in Section 2, above. Adjusted skill level of the team, Ψ, will be optimised at Ψ^*, and results will be separately affected by team quality (as previously defined) and deviations from rationality, $(\Psi - \Psi^*)$. Since Ψ^* forms part of 'quality', then only $(\Psi - \Psi^*)$ will remain in the performance model (see equation (8), below).

Obviously assumption four is not suitable for individual competitive sports, where there is no team selection and competitors must compete using whatever skills they possess. This is still consistent with our definition of quality from section 2, since the individual competitor's adjusted skill levels is one of their inherent characteristics.

3.3 Competitor strategy

One of the key determinants of how skills translate into performance is the selection of strategy for a competitor. Competitors' strategy selection will be based on their expectations of performance, and their expectations of the strategy selection of other competitors.

Further, for teams it is probably difficult to separate the selection of strategy from team selection. There is no empirical evidence to suggest any advantage to either selecting the best strategy given the skill sets of previously selected 'best' players, or selecting the best players to fit a given strategy[10]. Team selection in itself is an important component of an overall team strategy. However, it is likely

---

[10] Indeed, this may be another area for future research.

that strategy for any particular trial evolves as the likely makeup of opposition team and strategy selection becomes apparent. Final strategy selection probably occurs *after* team selection for all competitors. This suggests a complementary assumption to assumption four is necessary:

Assumption Five:
Competitors are rational in strategy selection and seek to maximise their utility defined over the result of the current trial.

If all competitors are rational[11], then each will select the strategy that maximises their utility (defined over present and future income and results as described above), given the strategies of all other competitors. Again risk preference will be important to the selection decision, but no assumption is necessary.

In the simplest sense, if assumption five holds, then strategy will reduce to a pure strategy Nash equilibrium. In the most likely case, where a pure strategy Nash equilibrium does not exist, then the outcome may be a mixed strategy Nash equilibrium.

If assumption five does not hold, then competitors would be selecting strategies that do not maximise their expected utility (given the decisions of other competitors), which may or may not violate assumption two. For example, if a competitor attempts to hide their strategy for future trials by not playing their 'best' strategy or team (thereby not revealing their preferred strategy, and possibly lulling future opponents into a state of overconfidence) in the current trial, this violates assumption five. In this case, the competitor probably has a low discount rate for future results (possibly even a negative discount rate, since future results are more important to them) and so this does not violate assumption two. The Australian rugby team were easily beaten by the New Zealand team in Sydney in June 2003, by a score of 50-23. The Australian team played an unusual strategy of kicking away a lot of possession to the highly-skilled back three of the New Zealand team, resulting in many opportunities for New Zealand to score. Four months later, the Australian team again played New Zealand, in the semi-final of the World Cup. Australia had radically changed strategy and held on to possession of the ball – restricting the opportunities for New Zealand to score. In both games, team selection for both teams was broadly similar. Australia had obviously played a sub-optimal strategy in the early game. It may be arguable whether this was an intention long-term strategy with the intention of fooling future opponents (particularly New Zealand), but it is a clear violation of assumption five. A ratings system that did not take this into account would probably underestimate the quality of the Australian team following the earlier loss to New Zealand, and might then have predicted a similar result in the World Cup semi-final.

If assumption five always holds, then rational strategy selection is consistent with the definition of quality in Section 2, above. Competitor strategy, S, will be optimised at S^*, and results will be separately affected by team quality (as previously defined) and deviations from rationality, $(S^* - S)$. Since S^* forms part of 'quality', then only $(S^* - S)$ will remain in the performance model (see equation (8), below).

3.4 The 'ideal' performance model revisited

Given the theories presented above, even without assumptions four and five, results should only be affected by the home advantage of a team, its quality (as measured by its rating), deviations from rationality, and random variation. The ideal performance model (from equation (5) above), then reduces to:

$$R_{xy} = R(H_{xy}, \ \bar{x}, \ \bar{y}, [S_x - S_x^*], [S_y - S_y^*], [\ \ _x - \ \ _x^*], [\ \ _y - \ \ _y^*], \ ) \qquad (8)$$

---

[11] Note that here the concept of rationality really means 'bounded rationality', since no competitor could be expected to have, or be able to make use of, full information.

3.5 A note on the independence of variables

While the model presented above might suggest that these variables are independent of each other, this is most likely not the case. As suggested above, team selection and strategy are often difficult to separate. However, it is likely that strategy and team selection are not determined independently of team quality, either. For example, it is generally accepted that low quality soccer teams playing in the knockout stages of the World Cup will usually play a highly-defensive long-ball strategy – effectively attempting to hold out during regular time because they expect to have a greater chance of winning the game if it is decided in a penalty shootout.

4. The 'reduced-form' performance model

Provided that assumptions four and five hold, the 'ideal' performance model in equation (8) can be further reduced into the form that is in most common use in sports ratings systems. If there is no deviation from rationality in team selection or strategy, then the terms (S-S*) and (Ψ – Ψ*) are zero. The 'reduced-form' performance model is presented in equation (9).

$$P_{xy} = P(H_{xy}, \ \bar{}_x, \ \bar{}_y, \ ) \tag{9}$$

This 'reduced form' model offers the advantage (when compared with equation (5)) of all variables being measurable and the data relatively easily obtained. In fact, this reduced form model is the basis of many sports ratings systems.

This reduced-form model does have limitations, however. If used in isolation, it forms predictions for future results based solely on home advantage and the quality of the two competitors. While this might seem reasonable, if assumptions four or five are violated then these predictions may become significantly biased. These limitations are illustrated with a numerical example in the next section.

5. Limitations of the reduced-form performance model:
An application of a simple linear dynamic model to 2003 Super 12 rugby

One example of an application of the reduced-form performance model is the AQB ratings system[12]. This system utilises a linear dynamic process across all known results, adjusted for a static home advantage parameter, to produce ratings for all competitors. The model is zero-sum, in that a winning team's rating rises by the same amount that the losing team's rating falls, and so provides a rating for each team relative to all other teams.

The AQB ratings system is described by the following formula (10a) below. Notice that this is a reduced-form performance model, since it includes as independent variables only the ratings of the two competitors and the home advantage variable.

$$^{t}_{x} = ^{t-}_{x} + R(B - H - DK) \tag{10a}$$

R is the result variable (equal to 1 for a win, or -1 for a loss). B is the base ratings change parameter – this is the amount by which a teams rating would increase if it beat an identically rated team at a neutral venue. H is the home advantage parameter and modifies the base ratings change so that when a home team wins its rating does not increase by as much (and conversely for away teams). D is the difference in ratings between the two teams (^{t-}_{x} of losing team – ^{t-}_{x} of winning team), and K is a parameter that weights this difference.

---

[12] The AQB ratings system was developed by the author to rate international sports teams. It is outside the scope of this paper to discuss the system in any great detail, or discuss its relative merits or shortcomings when compared to other possible systems. For further information, refer to http://www.image.co.nz/aqb/about.html

For this analysis, the parameter values are somewhat arbitrary[13]. R_x^0 for all teams was set at 1000, and the base ratings change at 20, and D at 0.08. The home advantage variable was calculated on the basis that if assumption three holds, then:

$$E(\,R_x^t\,) = R_x^{t-1} \tag{10b}$$

If this is true then it is simple to form a probability for each team winning any given contest, simply using the ratings of the two teams. However, the probability of victory would need to be modified to acknowledge that home teams have a greater probability of victory. To simplify this, home advantage was assumed to be constant across all teams, and over time. Up to the end of the 2003 season, 64.58% of games were won by the home team (counting drawn games as half a win). To modify the probability of a home team winning when playing an away team with an identical rating to 64.58% the home advantage variable would have to equal 5.8333. With the numerical values of the parameters included, equation (10a) then becomes:

$$R_x^t = R_x^{t-1} + R(20 - 5.8333 - 0.08D) \tag{10c}$$

Note that the home advantage parameter of 5.8333 is equivalent to 72.9 ratings points (i.e. for the system to predict an away team victory, the away team's rating would have to be more than 72.9 points greater than the home team's rating). Using this system and data of Super 12 rugby results from seasons 1996-2002, a set of ratings for the first week of Super 12 2003 was calculated. The ratings prior to the first week of Super 12 play in 2003 are presented in Table 1 below.

Table 1: Ratings of Super 12 teams prior to the 2003 season.

Crusaders	1170.91
Brumbies	1089.09
Highlanders	1082.90
Reds	1064.76
Waratahs	1030.64
Stormers	991.14
Blues	985.91
Hurricanes	984.82
Sharks	974.47
Chiefs	955.96
Cats	872.92
Bulls	803.20

The 2002 Super 12 champion Crusaders are naturally rated the highest, with the runners-up Brumbies second. Since these ratings were based only on information prior to the beginning of the 2003 season, we might reasonably expect the accuracy of predictions to improve over the course of the season as additional (more recent) data points are included.

The ratings update after each week of results to include new information (the next round of results). A simple head-to-head prediction for each game of the Super 12 2003 could be determined by a comparison of the ratings of each pair of teams (suitably adjusted for home advantage). By the end of the season, the ratings had changed considerably, and are presented in Table 2 below.

---

[13] The values of these parameters are based on the parameters actually used in the AQB ratings system, and developed over time to produce a set of ratings that seems to conform with general expectations about the quality of the teams.

Table 2: Ratings of Super 12 teams at the end of the 2003 season.

Blues	1136.27
Crusaders	1132.93
Brumbies	1044.13
Reds	1030.65
Hurricanes	1029.58
Highlanders	1029.54
Waratahs	1026.13
Stormers	969.89
Bulls	953.34
Sharks	926.19
Chiefs	888.33
Cats	839.75

As expected, the Super 12 champion Blues are on top of the ratings, with the runners-up, the Crusaders, second. The bottom-of-the-table Cats are also at the bottom of the ratings.

In terms of accuracy of prediction, this system correctly predicted 44 of the 69 results (63.8%). This is only slightly better than a naïve system that predicts the home team to win every game, which would have correctly predicted 43 (62.3%) of the results.

As we might expect from a dynamic system, its predictions later in the season were significantly better – of the last thirty games, it correctly predicted 22 results (73.3%), including both semi-finals and the final. The naïve system would have predicted 21 results (70.0%). The games which were incorrectly predicted during the last five weeks are listed in Table 3 below (home team is listed first).

Table 3: Incorrectly predicted results from the final five weeks of the 2003 Super 12 season.

Week 8	Chiefs	25-31	Sharks
	Stormers	20-41	Reds
Week 9	Bulls	39-19	Reds
Week 10	Bulls	32-31	Crusaders
Week 11	Highlanders	23-27	Waratahs
	Hurricanes	27-35	Brumbies
Week 12	Brumbies	21-28	Crusaders
	Sharks	16-24	Bulls

Some of the incorrect predictions might be due to natural variation in results, ε, and others may be due to limitations in the dynamic model (this is probably particularly true of early-season games where $\frac{t-+}{x}$ has not correctly adjusted to changes in the team over the 'off season'). Even in later season games, if some of these incorrect predictions can be explained by variations from optimality in team selection or strategy, then this suggests that either assumption four or assumption five has been violated, and that the reduced-form model is biased. The direction and magnitude of any bias resulting from a violation of either assumption is not clear, and would depend on how far the team had deviated from optimal team and strategy selection, as well as how often they had in the past, and how often and the magnitudes of deviations from optimality of the other teams in the competition. Further, these biases cannot always be clearly identified separately from standard variations in performance.

As one example, in the Highlanders-Waratahs game from week 11, the system could be very confident that the Highlanders (rating 1079.33) would beat the Waratahs (rating 989.35), especially with home advantage. After the game, many experts were pointing to the positional change of influential back Matt Burke to fullback from outside centre where he had disappointed (e.g. see Daily Telegraph [4]). This suggests that earlier strategy choices by the Waratahs might not have been optimal, and their losses during earlier games through the season may have been due at least in part to their deviation from optimality of strategy selection. This would have resulted in an underestimation of their quality (as measured by their rating prior to this game), and an incorrect prediction for this game (where their strategy selection had returned to optimality), based on the reduced-form model.

This illustrates the problem with using the reduced-form model to formulate ratings (or as a measure of team quality). Not only are the predictions that it makes not sensitive to deviations from optimality in strategy or team selection in the current trial, but the ratings themselves are likely to be biased (in unknown directions and of unknown magnitudes) due to deviations from optimality in previous trials. However, unless some statistical measure of the closeness to optimality of team selection and strategy can be developed, the only pragmatic solution is to simply acknowledge that these problems are apparent.

6. Conclusion

This paper began by providing a theoretical framework which could be used to develop a model for the ranking of sports competitors on the basis of past results. Most current rankings systems use a reduced form of this model, but this relies on several assumptions. Two of these assumptions (rationality or optimality of team selection and strategy) are likely to be seldom satisfied. One such example from the 2003 Super 12 rugby season was provided whereby the 2003 Waratahs could have been under-rated due to sub-optimal strategy selection. Not including this sub-optimality (or the return to a more optimal strategy choice) in predicting Super 12 results not only causes errors in prediction, but is likely to cause biases of unknown direction and magnitude in the ratings of all teams.

Rankings or ratings provided by sports rankings systems need to be interpreted carefully, especially where the rankings are taken as inferences of which competitor would win a head-to-head contest. Rankings may attempt to show which competitor is the 'best', but are highly sensitive to how the 'best' competitor is defined. This paper suggests that defining the 'best' competitor as the highest quality competitor, where quality is a measure of a competitor's potential performance, provides a consistent interpretation. Using this interpretation should allow the sports ranking to be used, along with additional data about team selection and strategy, to more accurately predict future results.

Further research in this area would certainly aid in the development of rankings systems that are more accurate predictors of performance, or more quickly adjust to new situations without introducing volatility. In particular, research with a focus on measurement of the factors that impact on player skills and elements of successful team selection and strategy should be undertaken, as well as further development of the theory surrounding risk preference, utility of performance and discounting. These would aid our understanding of the optimality of team selection and strategy. Only then could we have confidence that rankings do indeed reflect the quality of sports competitors.

7. References

[1]   S.M. Berry and P.D. Larkey, The effects of age on performance of professional golfers, in *Science and golf III: proceedings of the World Scientific Congress of Golf*, M.R. Farrally and A.J. Cochran (Eds), Human Kinetics, Champaign, Il., (1999): 127-37

[2]   S.R. Bray and W.N. Widmeyer, "Athletes' perceptions of the home advantage: an investigation of perceived causal factors", *Journal of Sport Behavior*, **23** (2000), 1-10.

[3]   K. S. Courneya and A. V. Carron, "The home advantage in sport competitions: a literature review", *Journal of Sport and Exercise Psychology*, **14** (1992), 13-27.

[4]   Daily Telegraph, "Waratahs throw out welcome mat – Dwyer plans to put Rogers on fast track", Sydney, (2003, May 5), 034.

[5]   R.P. DeGennaro, "The utility of sport and returns to ownership", *Journal of Sports Economics,* **4** (2003), 145-153.

[6]   J.L. Duda, "The relationship between goal perspectives and persistence and intensity among recreational sport participants", *Leisure Sciences,* **10** (1988), 95-106.

[7]   R. Fort and J. Quirk, "Owner objectives and competitive balance", *Journal of Sports Economics,* **5** (2004), 20-32.

[8]   A. M. Nevill and R. L. Holder, "Home advantage in sport: an overview of studies on the advantage of playing at home", *Sports Medicine,* **28** (1999), 221-236.

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

[10] M. Verbeek, "A guide to modern econometrics", Wiley, Chichester, U.K., (2001), 51.

[11] M.R. Yilmaz and S. Chatterjee, "Patterns of NBA team performance from 1950 to 1998", *Journal of Applied Statistics,* **27** (2000), 555-566.

THE MATHEMATICS OF BICYCLING PART 3: THE SOMERSAULT

W. H. Cogill*
School of Civil and Environmental Engineering
University New South Wales, Sydney
PO Box 1, Kensington
NSW 2052, Australia

April 2003

Abstract

The rider of a mountain bicycle is in danger of somersaulting on account of the location of the centre of gravity of the rider. The height of the centre of gravity of the rider, and its location in the fore and aft direction, determine the stability of the bicycle and the risk of a somersault.

The forward somersault is the most common cause of serious injury. It is usually caused by the front wheel encountering an object which causes the bicycle to overturn in the forward direction. The bicycle and rider execute a forward somersault about an axis perpendicular to the plane of symmetry of the bicycle and the rider.

The forward somersault is a hazard particularly when using a mountain bicycle. In the case of a road bicycle, the load of the rider is lower and more precisely located between the front and rear wheels. This provides greater stability.

The problem is to investigate the conditions which will cause the rider to somersault forwards. Braking can initiate a somersault. Typical immediate causes include stones, and longitudinal slits and grooves in the surface of the road.

Load diagrams are given for conditions of descending and ascending riding. They are intended to be used both to influence the design of a bicycle and to configure the rider's position for optimum stability. The equations of the system, based on the conservation of momentum and of energy, are given in order to indicate the trajectory of a rider during a somersault.

The rider is safer if the front tyre is compressible and if front suspension is used.

1 Introduction

The present note is Part 3 of a series. The aim of Part 1 was to write the Hamiltonian for a bicycle[1]. The aim of Part 2 was to determine the overturning moment caused by turning the handlebar of a bicycle[2].

This note treats a form of accident to riders of bicycles. The forward somersault is the most common cause of serious injury, and occurs particularly to riders of mountain bicycles. It is caused by the front wheel encountering an object which causes the bicycle to overturn in the forward direction. This occurs about an axis perpendicular to the plane of symmetry of the bicycle and the rider.

The forward somersault is less hazardous to the rider of a road bicycle. In the case of a road bicycle, the load of the rider is lower and more concentrated on the rear wheel. This configuration provides greater stability. A

---

mountain bicycle is stable when moving on level ground without restraining objects on the surface. If the front wheel of the bicycle is restrained from moving forwards, the resulting force can cause the bicycle and rider to somersault.

Typical restraining objects include stones and grooves in the surface of the road.

The problem is to investigate the conditions which will cause the bicycle to somersault.

2 Load applied by a bicycle

A bicycle can be approximated by three masses. The largest is the mass of the rider. The bicycle frame and its moving parts can be approximated by two masses, located at the centres of the wheels. The mass of the bicycle is ignored, as it is small compared with the mass of the rider. See Figure 1. In all the diagrams, an arrow points to the front of the bicycle which is on the right of the diagram. Figure 1a shows a mountain bicycle and rider. The wheelbase b is equal to five units throughout. The mass of the rider is located towards the rear wheel. Figure 1b shows the locations of the equivalent masses of a recumbent bicycle. The centre of gravity of the rider is close to the base of the wheels and towards the rear of the two wheels. A recumbent bicycle is more stable than a mountain bicycle.

(a) Mountain Bicycle   (b) Recumbent bicycle
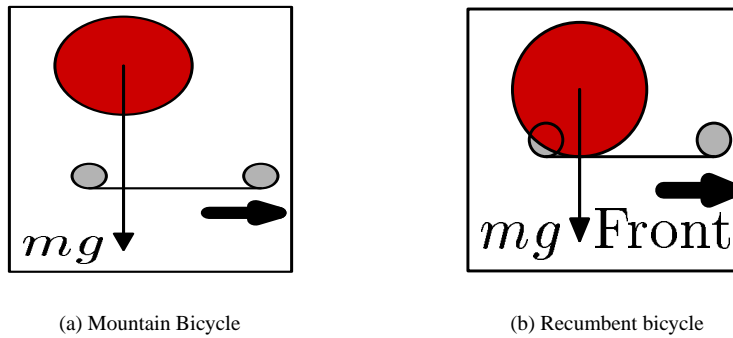
Figure 1: Load diagram of a bicycle

Figure 2 shows the notation employed. The slope of the riding surface is denoted by α, and is positive when uphill.
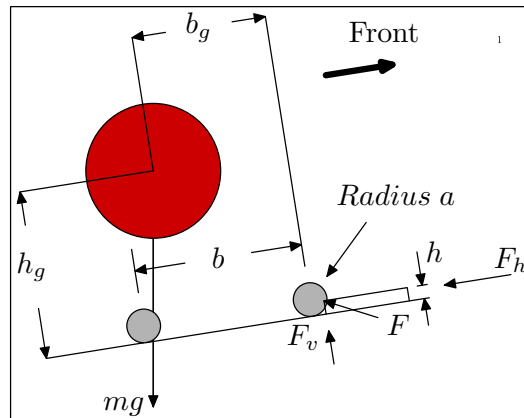
Figure 2: Diagram showing the notation employed

The diagram shown in Figure 3a represents a bicycle on an ascending gradient at an angle of 0.5 radians to the horizontal. The resultant of the load due to the weight of the rider passes outside the wheelbase. The bicycle is unstable and somersaults to the rear. This somersault is less liable to cause injury than a forward somersault. The rider is able to step off the bicycle onto the road surface.

The diagram shown in Figure 3b represents a bicycle on a descending gradient at an angle of 0.3 radians. The resultant of the load due to the weight of the rider passes inside the wheelbase. The bicycle is therefore stable.

(a) Bicycle inclined at 0.5 radians ascending

(b) Bicycle inclined at 0.3 radians descending
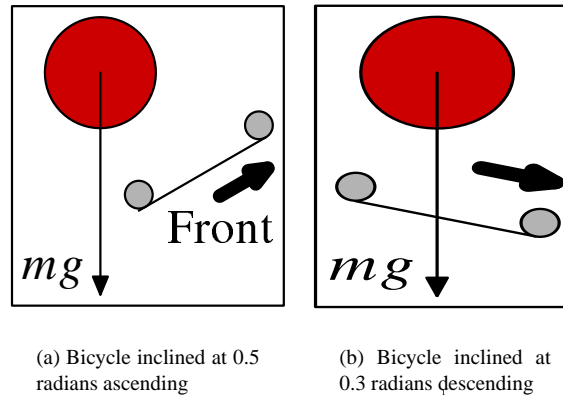
Figure 3: Load diagram of a bicycle: the effect of gradient

3 The effect of braking

(a) Bicycle inclined at 0.51 radians descending

(b) Bicycle inclined at 0.25 radians descending
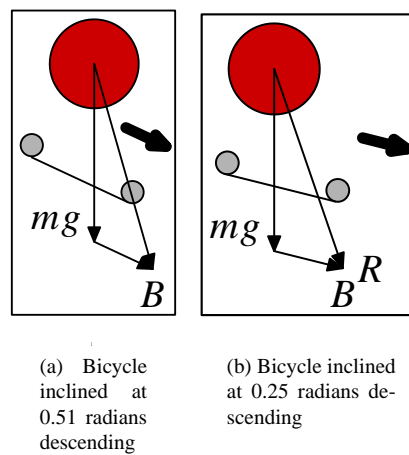
Figure 4: Load diagram of a bicycle on a descending gradient with braking

Figure 4 shows the effect of braking on the stability of a bicycle. The diagram shown in Figure 4a shows the braking force B exerted by the bicycle on a road surface. The diagram is drawn assuming a descending gradient of 0.51 radians. A braking force is shown equal to 0.4 times the normal reaction between the road surface and the wheels. The resultant of the load including the weight of the rider passes outside the wheelbase, and the bicycle is unstable.

The bicycle is more stable on a less steep gradient, as shown in Figure 4b. In this figure, the descending gradient is 0.25 radians. A braking force is applied equal to 0.4 times the normal reaction between the road surface and the wheels. The resultant of the load due to the weight of the rider passes inside the wheelbase and the bicycle is stable.

The ratio of the braking force B to the reaction normal to the riding surface will be referred to as the braking coefficient, μ_b. The bicycle is stable if the resultant reaction with the riding surface passes inside the wheelbase. The condition to ensure this can be expressed as

$$\mu_b \, h_g \, cos \, \alpha - h_g \, sin \, \alpha < b_g \tag{1}$$

where the gradient α is negative for a descending slope. If the rider sits far back, b_g is large and the bicycle is stable. To achieve greater stability, the rider must leave the saddle and balance over the rear wheel.[1]
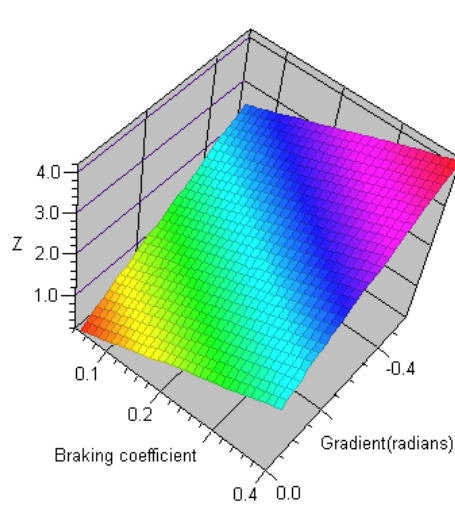
Figure 5: Diagram showing the maximum safe braking coefficient to maintain stability in relation to the offset $Z = b - b_g$.

Figure 5 shows the conditions for stability when the bicycle is descending. The maximum safe braking coefficient depends upon the gradient. For example, consider a braking effort which develops a coefficient of friction of 0.4 on the level, and renders the bicycle marginally stable(the offset $Z = b - b_g = 2.0$; $h_g = 5.0$). If the descending slope is 0.2 radians, a braking coefficient of 0.2 renders the bicycle marginally stable with the same configuration.

The Macsyma code used to produce Figure 5 is as follows. Note that placing c=0.0 enables the offset $b - b_g$ to be displayed as Z.

```
b = 5.0; (* wheelbase of bicycle *)
c = 0.0; (* forward offset of rider from rear wheel *)
h = 5.0; (* distance of c.g. of rider from riding surface *)
(c:0.0,h:5.0)$
plot3d([c+y*h*cos(x)-h*sin(x)],x,-0.5,0,y,0.05,0.4,"Gradient(radians)","Braking coefficient","Offset");
(* y is the coefficient of braking friction; x is the gradient, negative if descending *)
(*y h cos[x] is the frictional driving force *)
(*-h sin[x] is the component of the load in the direction of travel*)
```

---

[1] An increase in b_g for descending, and a decrease in b_g for ascending, yield greater stability and an increase in Z, the criterion chosen for stability in Figure 5 and Figure 7.

4 The effect of a driving force when ascending

The ratio of the driving force to the load normal to the surface will be called the coefficient of driving friction, or the driving coefficient, and denoted by μ_d.

Figure 6 shows the effect of the driving coefficient and the rider position on the stability of a bicycle when ascending. In Figure 6a the gradient is 0.2 radians. The position of the rider is at a distance of one fifth of the wheelbase forward of the rear wheel, and is given by $b - b_g = 1$. A driving coefficient shown of 0.2 times the component of the load normal to the surface causes the bicycle to become unstable.

Figure 6b the position of the rider is forward of that shown in Figure 6a. The position of the rider is given by $b - b_g = 2$, or two-fifths of the distance of the wheelbase forward of the rear wheel. As the resultant passes through the point of contact of the rear wheel, the bicycle is marginally stable. The driving coefficient shown is 0.2. A driving coefficient of greater than 0.2 leads to a backward somersault.
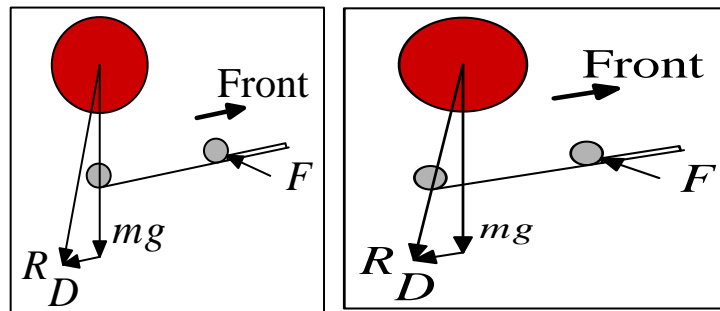
(a) Bicycle inclined at 0.2 radians uphill, driving coefficient 0.2, $b - b_g = 1$: The bicycle is unstable

(b) Bicycle inclined at 0.2 radians uphill, driving coefficient 0.2: The mass of the rider is moved forwards $b - b_g = 2$. The bicycle is marginally stable

Figure 6: Load diagram of a bicycle inclined uphill, showing the effect of the driving coefficient and the position of the rider on stability.

1

1

At the limiting driving coefficient, a small obstacle on the riding surface is sufficient to cause a backward somersault.

Referring to Figure 2, the condition to ensure that the reaction from the riding surface passes inside the wheelbase is

$$\mu_d \, h_g \, cos \, \alpha + h_g \, sin \, \alpha < b - b_g \qquad (2)$$

where μ_d is the coefficient of driving friction, and α is the gradient of the riding surface which is positive for an ascending gradient.

Figure 7 shows the driving coefficient as a function of the gradient. The figure is calculated with the aid of equation (2) for an ascending gradient. For a value of $Z = -2$, Figure 7 shows that the maximum driving coefficient when ascending a gradient of 0.2 radians is approximately 0.2. This confirms the result shown in Figure 6b.

The tractive effort can be increased if the rider moves the centre of gravity forward. This is indicated in Figure 7. If the rider moves 0.5 units further forward, the value of $-(b - b_g)$ becomes -2.5. The maximum driving coefficient on an ascending gradient of 0.2 radians is then 0.33: the maximum driving coefficient is increased from 0.2 to 0.33 on account of the change in the position of the rider's centre of gravity.
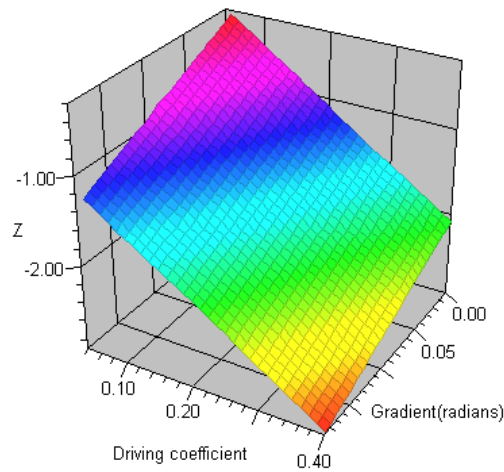
Figure 7: Diagram showing the maximum safe driving coefficient when ascending, in relation to the offset $Z = -(b - b_g)$.

The Macsyma code used to produce Figure 7 is as follows.

```
b = 5.0; (* wheelbase of bicycle *)
c = 0.0; (* forward offset of the rider from the rear wheel *)
h = 5.0; (* distance of the centre of gravity of the rider from the riding surface *)
(c:0.0,h:5.0)$
plot3d([c-y*h*cos(x)-h*sin(x)],x,0,3/16,y,0.05,0.4,"Gradient(radians)","Driving coefficient","Offset")
(* y is the coefficient of driving friction; x is the gradient in radians *)
(*y*h cos[x] is the frictional driving force *)
(*h sin[x] is the component of the load in the direction of travel*)
```

5   The effect of an obstacle

Figure 8 shows the effect of an obstacle encountered by the front wheel of the bicycle. The encounter leads to a force F applied to the tyre of the front wheel. In Figure 8a, the height of the obstacle is equal to or greater than the radius of the wheel. The force F is applied purely radially to the front wheel, and parallel to the riding surface. Unless the bicycle and rider are stopped by braking and the compression of the tyre, this situation leads to a forward somersault.

5.1   Surmounting an obstacle

We consider the maximum height h of an obstacle which allows the bicycle to surmount the obstacle, and not to commence a forward somersault. The situation is shown in Figure 8b. The dimensions are shown in Figure 2.

We take moments about the centre of gravity of the rider. The centre of gravity is located at height h_g above the riding surface, and at a distance b_g aft from the centre of the front wheel. From Figure 2, $F_v b_g - F_h(h_g - h)$ must be positive for the bicycle to be stable, and not to commence a forward somersault.

In order to surmount an obstacle, the force of the impact must have a component normal to the riding surface. The height of the obstacle must be less than the radius of the front wheel. See Figure 8b.

(a) Bicycle inclined at 0.44 radians descending

(b) Bicycle inclined at 0.25 radians descending
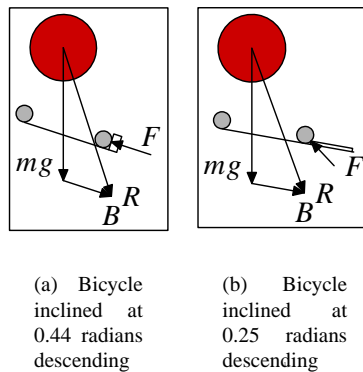
Figure 8: Load diagram of a bicycle on a descending gradient showing an obstacle

The bicycle commences to rotate about an axis perpendicular to the paper. The direction of rotation is anticlockwise in Figure 2.

While surmounting an obstacle, the bicycle is rotating about the point of contact between the rear wheel and the riding surface. The point of contact moves during the operation of surmounting the obstacle. This movement is neglected in what follows. The resisting force F shown in Figure 8b is assumed constant, as it is governed by the tyre pressure. The moment of inertia of the bicycle and rider about the axis of rotation is I, the applied torque is N, and the angle of rotation is θ. We have

$$I\ddot{\theta} = N \tag{3}$$
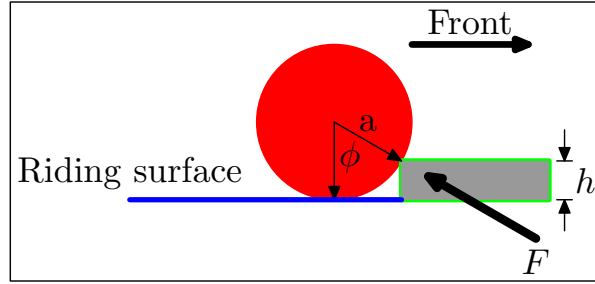
Figure 9: Diagram showing the angle ϕ subtended at the centre of the front wheel by an obstacle of height h

Referring to Figure 2 and Figure 9, the force F exerted by the obstacle is assumed to act radially on the front wheel. The torque N about the axis of rotation is given by $F\ [b + a\ sin(\phi)]cos(\phi) + F\ h\ sin(\phi) = F\ [b + a\ sin(\phi)]cos(\phi) + F\ a\ [1 - cos(\phi)]sin(\phi)$; h is the height of a prismatic obstacle in contact with the front wheel, of radius a; ϕ is the angle subtended by the obstacle at the centre of the front wheel. Equation (3) becomes

$$\ddot{\theta} = \frac{F}{I}[b + a\ sin(\phi)]cos(\phi) + a\ [1 - cos(\phi)]sin(\phi) \tag{4}$$

Integrating, we have

$$\theta = \frac{F}{I}\int\limits_{\phi\to 0}\int\limits^{\phi\to\phi}[b + a\ sin(\phi)]cos(\phi) + a\ [1 - cos(\phi)]sin(\phi)dt.dt; \tag{5}$$

We have $dt = \dfrac{dt}{d\phi}d\phi$. If the height h of the obstacle is small compared with the radius a of the wheel, then $a\ cos(\phi)\dfrac{d\phi}{dt} \approx v_0$ and $dt = \dfrac{a}{v_0}cos(\phi)d\phi$.[2]

$$\theta = \frac{F}{I}\left(\frac{a}{v_0}\right)^2\left(\frac{3\,b\ cos(\phi)}{8} - \frac{b\ cos(3\,\phi)}{24} + \frac{(-a + 4\,b\,\phi)\ sin(\phi)}{8} - \frac{a\ sin(3\,\phi)}{24}\right) \tag{6}$$

An obstacle of height h subtends an angle θ_h at the point of contact of the rear wheel with the riding surface. If the bicycle is to surmount the obstacle, we must have $\theta \geq tan^{-1}\left[\dfrac{a(1 - cos(\phi))}{b + a\ sin(\phi)}\right]$ given by equation (6). During the time $t(\theta_h)$ required for the bicycle to rotate anticlockwise by an angle θ_h, the rear wheel of the bicycle will have travelled forward a distance $a\ \phi$. See Figure 9. [3] If $t(\theta_h) < \dfrac{a\ \phi}{v_0}$, the bicycle will surmount the obstacle. If $t(\theta_h) > \dfrac{a\ \phi}{v_0}$, the bicycle and rider will commence a forward somersault.

---

[2] The integration and simplification were performed with the aid of Mathematica version 4.1

[3] This is an approximation. It is true to better than one part in a hundred, for an obstacle of height 0.17 and a wheelbase of 1.8 units.

(a) Approach speed v_0=4.2m/s   (b) Approach speed v_0=13m/s
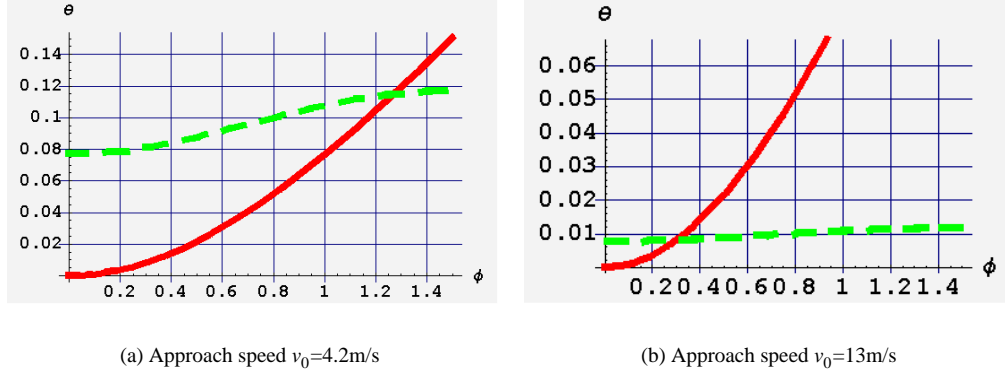
Figure 10: Diagram showing the rotation θ of a bicycle having a wheelbase of b=1.8m after encountering an obstacle of height h. The full line is the elevation of the object given by $\theta_h = tan^{-1}\left[\dfrac{a(1 - cos(\phi))}{b + a\,sin(\phi)}\right]$; the dashed line is the response of the bicycle given by equation (6).

Figure 10 shows the rotation θ of a bicycle after encountering an obstacle of height h. The obstacle subtends an angle ϕ (Figure 9) at the front wheel. For example, consider an obstacle which subtends an elevation $\theta_h = 0.04$ at the rear wheel($\phi = 0.7$). The bicycle rotates anticlockwise about an axis perpendicular to the plane of symmetry of the bicycle and rider. Figure 10a shows the response of a bicycle having an approach speed of v_0=4.2m/s ($F_\phi \approx 0.1$)[4]. The bicycle surmounts the obstacle. Figure 10b shows the response of a bicycle having an approach speed of v_0=13m/s($F_\phi \approx 0.01$). The bicycle does not surmount the obstacle.

The Mathematica commands used to produce Figure 10 are as follows

Speed into obstacle

This file shows two graphical outputs. The first is the height of an obstacle, expressed as the angle subtended by the obstacle at the rear wheel. The second is the height to which the front wheel of the bicycle can ascend, also expressed as the angle subtended at the rear wheel. The second graphical output depends upon the approach speed. This speed is expressed by $\mathcal{F}_\phi = \frac{F}{i}\left(\frac{a}{v_0}\right)^2$, where v_0 is the speed of approach. If it is required to change the speed of the approach, change F within the program.

```
In[1]:=
        \[Theta] := ArcTan[0.35(1 - Cos[\[Phi]])/(1.8 + 0.35 Sin[\[Phi]])];
        (*Wheelbase 1.8m; wheel radius 0.35m *)
        b = 1.8; a = 0.35;
        F = 0.1;
        y = Sqrt[(3/8) * b * Cos[\[Phi]] - (1/24) * b * Cos[3 * \[Phi]] + (1/8) * (-a + 4 * b * \[Phi]) * Sin[\[Phi]] -
         (1/24) * a * Sin[3 * \[Phi]] ];

        bp := Plot[{\[Theta], Evaluate[F * y]},
          {\[Phi], 0, 1.5}, TextStyle\[Rule]{FontSize\[Rule]20, FontWeight- > "Bold"},
            AxesLabel - > {"\[Phi]", " \[Theta] "},
              PlotStyle - > {{RGBColor[1, 0, 0], Thickness[0.02]}, {RGBColor[0, 1, 0],
            Dashing[{0.05, 0.05}], Thickness[0.02]}},
              Background - > GrayLevel[0.95], GridLines- > Automatic]$
```

---

[4]If for example the force $F = 600kPa \times 0.05^2 m^2 = 1500N; r = 1m; I = 100kg \times 1m^2 = 100kgm^2; a = 0.35m; v_0 = 4.2ms^{-1}$ then
$F_\phi = \dfrac{F\,r}{I}\left(\dfrac{a}{v_0}\right)^2 \approx 0.1$

```
Show[bp,FontSize- > 14,FontWeight- > "Bold",
 GridLines- > Automatic]
```

6   Trajectory of rider

We consider the trajectory of the somersaulting rider after the front wheel of the bicycle encounters an obstacle which it fails to surmount. The front wheel remains in contact with the object. The bicycle and rider rotate forwards until they make contact with the riding surface.

The centre of gravity of the rider is located with reference to the riding surface(see Figure 2). h_g is the distance of the centre of gravity normal to the riding surface, r is the radial distance and β is the initial angle of elevation of the centre of gravity from the axle of the front wheel, around which bicycle and rider are assumed to pivot. v_0 is the approach velocity of the rider; α is the gradient of the riding surface, which is positive for an ascending surface; θ is the angle of rotation of the rider, as defined in equation (3).

We have, from the laws of conservation of energy and of angular momentum about the point of impact

$$1/2 \left[\left(\frac{dr}{dt} \right)^2 + r^2 \left(\frac{d\theta}{dt} \right)^2 \right] + g\, r \left(\sin(\phi) - \sin(\phi_0) \right) = 1/2\, v_0^{\,2} \tag{7}$$

$$r^2 \left(\frac{d\theta}{dt} \right) = v_0\, h_g \tag{8}$$

where $\phi = \theta - \alpha$ and $\phi_0 = \beta - \alpha$.

By eliminating $\dfrac{d\theta}{dt}$ between equations (7) and (8) it follows that

$$\frac{dr}{dt} = v_0 \left(1 - 2qr - \frac{h_g^2}{r^2} \right)^{1/2} \tag{9}$$

where $q = \dfrac{g}{v_0^2} \left(\sin(\phi) - \sin(\phi_0) \right)$

Dividing both sides of equation (9) by both sides of equation (8) and noting that

$$\frac{dr/dt}{d\theta/dt} = \frac{dr}{d\theta} \tag{10}$$

we have

$$\frac{dr}{d\theta} = \frac{r^2}{h_g} \left(1 - 2qr - \frac{h_g^2}{r^2} \right)^{1/2} \tag{11}$$

The solution of equation (11) yields the radial distance r of the centre of gravity of the somersaulting rider from the center of rotation in terms of the angle of rotation θ.

6.1   Calculation of the trajectory of the rider

The following Macsyma code was used to obtain an indication of the movement of the centre of gravity of the rider($h_g = 0.8m$, $v_0 = 7.1m/s$, $\alpha = 0$, $\beta = \arcsin 0.9$). It represents a numerical solution of equation (11) with specific initial conditions.

In the following code, the variable 't' is the 'θ' variable in equation (11); the variable 'r' is the 'r' variable in equation (11).

```
Define the differential equation to be solved.
eq:'diff(r,t)=r^2/0.8*sqrt(1-0.2*r*(sin(t)-sin(0.9))-0.63/r^2);%Pi%e
Specify the initial conditions and solve the equation.
 ic:'at(r,t=0.9)=0.8;
 sol:runge_kutta(eq,'r,'t,ic,0.9,3.1,0.1)$
 A plot of the displacement of the rider follows.
 graph(-[assoc('r,sol)]*cos([assoc('t,sol)]),[assoc('r,sol)]*sin([assoc('t,sol)]),
 "Horizontal displacement","Vertical displacement",[15])$
```
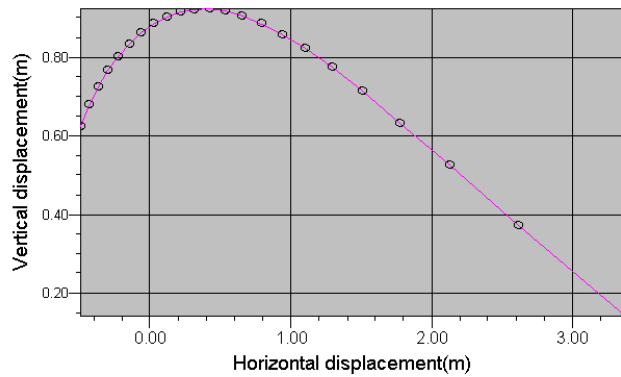

Figure 11: The calculated trajectory of the rider. The figure shows the displacement of the rider($\alpha = 0$).

Figure 11 shows the displacement of the rider after contact with an obstruction, which the bicycle and rider fail to surmount. The approach velocity of the rider v_0 is 7.1 m/s.

In Figure 11, the displacements are in metres. The plotted points are at intervals of 0.1 second. The somersault commences with the bicycle and rider rotating as a single body. The plot indicates that the rider hits the road surface at a velocity of approximately 10 metres per second.

7 Case study

Figure 12 shows the partial reconstruction of an accident. It illustrates the injurious circumstances which can occur. This case was selected from an actual event.

A longitudinal slit was present in an open road. A bicycle tyre became wedged in the slit with an effect similar to that of an upraised object of height h equal to or greater than the radius of the wheel(see Figure 2 and Figure 8a). The cyclist was obliged either to decelerate the bicycle as the tyre became wedged, an impossible task in the time available, or to execute a forward somersault.

Figure 12: Reconstructed accident situation. Forward somersault

The bicycle was a road bicycle. The cyclist was hospitalized and became paraplegic.

8   Conclusions and recommendations

The rider of a mountain bicycle is endangered on account of the location of the centre of gravity of the rider. The height of the centre of gravity of the rider, and its location in the fore and aft direction, determine the risk of a somersault.

Small obstacles on the riding surface can be surmounted if the approach speed is low.

The rider is able to prevent a forward somersault by leaving the saddle and balancing directly over the rear wheel of the bicycle.

A backward somersault is less likely to cause injury as the rider is able to step off the bicycle before the somersault is complete.

Load diagrams may be used both to design a bicycle and to configure the rider's position for optimum stability.

The equations of the system, based on the conservation of momentum and of energy, are given in order to indicate the expected behaviour during a somersault.

8.1   Recommendations

The work raises several questions.

8.1.1   Braking and climbing

The diagrams which are shown in Figure 3 to Figure 7 are special cases, involving the configuration of the bicycle, the gradient and the braking or driving coefficients. A general expression could be derived including all of these factors and avoiding the need to study each bicycle and rider combination independently.

8.1.2 Surmounting a small obstacle

The expressions leading to equation (6) are based on the assumption that the axis of rotation does not move while surmounting the obstacle. This is an approximation.

8.1.3 Trajectory of somersaulting bicycle and rider

The equation (11) governing the trajectory was solved numerically. The solution was reached after a restriction was followed: the height h_g of the rider is not consistent with a physical situation, but is adapted in order to yield a solution with the aid of Macsyma. The trajectory plotted in Figure 11 is an indication only of the path followed by the centre of gravity of the rider.

References

[1] Cogill W.H., The mathematics of bicycling, The Hamiltonian for a bicycle *Proceedings of* 2nd *Conference on Mathematics and Computers in Sport*, Bond University, Australia, 37-42, (1994)

[2] Cogill W.H., The mathematics of bicycling:Part II, Overturning moment, *Proceedings of* 4th *Conference on Mathematics and Computers in Sport*, Bond University, Australia, 245-248, (1998)

9 Appendix

The result of equation (6) indicates that an object may be surmounted. Two conditions must be satisfied.

1. The height of the object must be small

2. The approach speed must be low.

Either of these conditions permits the bicycle to rotate backwards sufficiently to mount the object, and not to perform a forward somersault.

The maximum safe approach speed depends on the geometry of the bicycle. It depends on

1. The radius of the front wheel (0.35m in the current example)

2. The length of the wheel base (1.8m in the current example)

3. The moment of inertia of the bicycle and rider about the axis of rotation(100kg-m^2 in the current example). The axis of rotation is assumed to be the point of contact between the rear wheel and the riding surface

4. The force exerted by the obstacle on the bicycle and rider (1500N in the current example,see section 5.1).
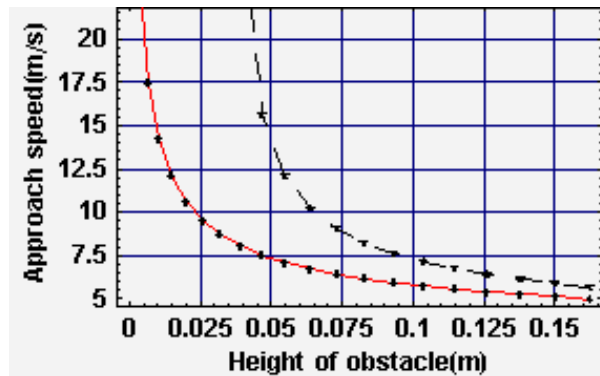
Figure 13: The maximum safe approach speed plotted against the height of the obstacle.

Figure 13 shows the maximum safe approach speed in relation to the height of the obstacle, based on Equation 6. The values of F and of I are the same as those used in Figure 10[5].Figure 13 shows two curves.The solid line shows the response for a rigid wheel. The dashed line shows the response for a bicycle having a front wheel which is able to retract by an amount which reflects a change(a decrease of 0.02 radians in the current example) in the elevation θ of the obstacle.

Figure 13 was prepared with the aid of Mathematica. The code used is as follows. The angle ϕ is the angle subtended by the obstacle and the riding surface at the axle of the front wheel, as shown in Figure 9.

Speed into obstacle

```
In[3]:= << "Graphics`MultipleListPlot`"

In[4]:= b = 1.8; a = 0.35;

In[5]:= \[Theta] := Tan^-1 ( a (1 - Cos(\[Phi])) / b + a Sin(\[Phi]) );

In[6]:= y = √( 3/8 b Cos(\[Phi]) - 1/24 b Cos(3 \[Phi]) + 1/8 (4 b \[Phi] - a) Sin(\[Phi]) - 1/24 a Sin(3 \[Phi]) );

In[7]:= tabl = Table[ {b \[Theta], Evaluate[ √( 1500 a^2 / 100 \[Theta] ) y]}, {\[Phi], 0.1, 1.1, 0.05} ];

In[8]:= tabl1 = Table[ {b θ, Evaluate[ √( 1500 a^2 / 100 (θ - 0.02) ) y]}, {φ, 0.5, 1.1, 0.05} ];

In[9]:= MultipleListPlot({tabl, tabl1}, PlotJoined → True, Frame → True, GridLines → Automatic,
        FrameLabel → {"Height of obstacle(m)", " Approach speed(m/s)"},
        TextStyle → {FontSize → 12, FontWeight → Bold, FontFamily → Arial},
        PlotStyle → {RGBColor[1, 0, 0], Dashing[{0.05, 0.05}]}, GridLines → Automatic,
        Background → GrayLevel[0.95]);
```

At speeds above those indicated in Figure 13, the bicycle and rider will commence a forward somersault as shown by equation (11). The compressibility of the tyre renders the bicycle and rider safer than is shown by equation (6) in which it is assumed that the wheel and tyre are rigid. Additional safety is provided if front suspension is fitted and is activated.

Figure 13 and the associated Mathematica code will require editing in order to represent a particular bicycle and rider combination.

---

[5]Front suspension on bicycles enables larger obstacles to be surmounted. The effect of suspension is similar to that of an increased depth of tyre.

A NEW STATISTIC IN CRICKET —
THE SLOG FACTOR

G. L. Cohen
Department of Mathematical Sciences
University of Technology, Sydney
PO Box 123, Broadway, NSW 2007, Australia
graeme.cohen@uts.edu.au

Abstract

It seems relevant that commentators would want to describe how hard a cricketer hits a ball. This may be measured by the ratio of the number of runs achieved to the number of scoring strokes made, within any particular context. The higher this *slog factor*, the harder the ball is hit. An investigation of Australian representative cricketers over a three-year period has been carried out and shows the average slog factor for them to be around 2.2 in tests and 1.8 in one-day matches. Perhaps more importantly, the slog factor has a practical use in describing the distribution of runs scored by a batsman and hence in determining the probability that the batsman scores a century, say. It turns out, however, that this probability varies little with the slog factor, but depends more on the batsman's true batting average. The result will be compared with that from a more empirical argument in a recent paper.

1. Introduction

According to *The Macquarie Dictionary*, the colloquial verb *to slog* means "to hit hard, as in boxing, cricket, etc." In cricket, some batsmen consistently hit the ball harder than others so it seems reasonable to invent a way of measuring this. Generally speaking, the harder a ball is hit, the more likely it is to beat the fieldsmen and to result in more runs than would be obtained from a similar but softer stroke. This justifies the introduction of the *slog factor*, which is the ratio, for a given batsman in a given context, of the number of runs scored to the number of scoring strokes made. So the slog factor measures how hard a given batsman hits a ball. It is quite different in concept from the batting strike rate (runs per ball faced, essentially) and the batting average (runs per wicket, essentially). These might be described as measuring respectively how effectively the ball is hit (in being able to avoid the fieldsmen) and how technically correct the ball is hit (in being able to score runs and avoid getting out).

Values of the slog factor for all Australian representative cricketers over the period 9 January 2000 to 25 January 2003, given separately for tests and one-day internationals (ODIs), will be presented and discussed in the next section. It is perhaps surprising that for every batsman the slog factor in tests far exceeds that for ODIs. Averaged for all batsmen in this period, the slog factor was about 2.2 for tests and about 1.8 for ODIs.

In a previous paper [2], the slog factor was less dramatically called the 'strike constant' so to that extent it is not new but has been given a new interpretation. It was used on the earlier occasion in justifying the expression $(A/(A + 2))^{c/2}$ as the probability of a batsman scoring at least $c > 1$ runs, where A is the batsman's true average score over all previous innings of interest. The two 2s here arise in essence by taking the slog factor as approximately equal to 2 for all batsmen. An expression was also obtained in [2] for the probability of a batsman scoring a duck (being dismissed, having made no scoring strokes).

Around the same time, a paper appeared by Tan and Zhang [4] seeking to fit the distribution of batting scores of the two great English batsmen Jack Hobbs and Herbert Sutcliffe to separate negative exponential curves. This led them to assign the quantity $e^{-c/A}$, in effect, as the probability that a batsman scores at least c runs. It is not difficult to see that the earlier expression approaches this asymptotically as $A \to \infty$, giving some validity to both approaches and largely answering questions on the distribution of cricket scores that go back at least 60 years. References to earlier papers on the topic are given in [2].

2. Australian slog factors

Over the period 9 January 2000 to 25 January 2003, Australia played 30 test matches and 68 ODIs against other test-playing nations. For tests and ODIs separately, the number of scoring strokes s and number of runs scored r are shown in Table 1 for each batsman. The ratio r/s is the slog factor. Not all batsmen played in both tests and ODIs.

Lower order batsmen have been included, since there is interest in how hard they hit the ball, but there is little meaning to be attached to those batsmen with very few scoring strokes. They have been included as contributing to the overall picture indicated by the totals. These give average slog factors of 2.19 and 1.77 for tests and ODIs, respectively, which is of interest in that they do not differ greatly from the corresponding figures near the other end of the performance spectrum, namely Sydney grade cricketers, given by Cochran [1] and quoted in [2]. Cochran's results were 2.16 and 1.82 for traditional cricket and limited-overs cricket, respectively.

The difference in the values for the two forms of cricket reflects the greater number of boundaries in tests, as the batsman waits for the loose ball, and the scrambling for singles in ODIs. This is mirrored for individual batsmen since in Table 1, in all cases where a batsman played both tests and ODIs, the slog factor for tests was considerably higher. The figures confirm common perceptions, such as that Adam Gilchrist hit the ball hardest of the recognised Australian batsmen at that time.

3. An application of the slog factor

In the earlier paper [2], the slog factor for a particular batsman, as calculated in some particular past context, is denoted by κ and is assumed to remain constant for similar future contexts. (The context may be a complete career, or against a particular team, or in some particular batting position, for example. The model is more applicable to traditional cricket than to limited-overs matches.) Letting the random variable X denote the number of scoring strokes made by the batsman in a subsequent innings, the model then allows the probability of at least k scoring strokes ($k \geq 0$) to be determined as

$$\Pr(X = k) = pq^k,$$

where $p = \kappa / (A + \kappa)$ and $q = 1 - p$. In this, A is the true average of the batsman over all innings in the original context so that this generally includes also innings in which the batsman was not out. The details are given in [2] and are reproduced in [3].

For this geometric distribution, the expected value of X is $E(X) = q / p = A / \kappa$. Then the expected value of the score R, or number of runs obtained from the X scoring strokes, is obtained as follows:

$$E(R) = E(R \mid R > 0) = E((R / X) X \mid X > 0) = E(\kappa X \mid X > 0) = \kappa E(X) = A,$$

128

since $X = 0$ if and only if $R = 0$. This provides the justification of the model and in particular of the use of the true average.

Table 1: Slog factors for Australian cricketers, 9 January 2000 to 25 January 2003

	ODIs			Tests		
	Scoring strokes s	Runs r	r/s	Scoring strokes s	Runs r	r/s
Bevan	1061	1720	1.62			
Bichel	75	131	1.75	56	128	2.29
Campbell	29	54	1.86			
Clark	24	39	1.63			
Dale	6	14	2.33			
Fleming	44	75	1.70	3	6	2.00
Gilchrist	903	1959	2.17	852	1986	2.33
Gillespie	46	63	1.37	106	233	2.20
Haddin	6	13	2.17			
Harvey	141	235	1.67			
Hauritz	12	13	1.08			
Hayden	726	1358	1.87	1246	2796	2.24
Hogg	48	89	1.85			
Kasprowicz				8	20	2.50
Katich				6	15	2.50
Langer				834	1864	2.24
B. Lee	100	190	1.90	187	451	2.41
S. Lee	83	167	2.01			
Lehmann	428	723	1.69	45	97	2.16
Love				48	95	1.98
MacGill	1	1	1.00	41	66	1.61
Maher	175	322	1.84			
Martyn	945	1490	1.58	723	1445	2.00
McGrath	13	20	1.54	53	115	2.17
Miller				38	80	2.11
Ponting	1249	2278	1.82	909	2031	2.23
Slater				326	709	2.17
Symonds	238	486	2.04			
Warne	159	259	1.63	284	625	2.20
Watson	171	262	1.53			
M. Waugh	760	1307	1.72	671	1436	2.14
S. Waugh	623	1026	1.65	816	1666	2.04
Williams	7	13	1.86			
Totals	8073	14306	1.77	7252	15864	2.19

It is then straightforward to estimate the probability that the batsman scores c or more runs ($c > 0$): putting $d = [(c - 1)/\kappa]$ (where $[\cdot]$ denotes the greatest-integer function), we have

$$\Pr(R \geq c) = 1 - \Pr(R \leq c - 1) = 1 - \Pr(\kappa X \leq c - 1) = 1 - \Pr(X \leq d)$$
$$= \sum \Pr(X = k) = q^{d+1} = (A/(A + \kappa))^{d+1}, \tag{1}$$

with the summation being for k from $d + 1$ to ∞.

It was remarked in [2] that taking $\kappa = 2$ and assuming initially that c is even allows the simpler expression

$$\Pr(R \geq c) = (A / (A + 2))^{c/2}, \qquad (2)$$

which may then be used for all $c > 0$ with, it turns out, very little variation from the more exact result in (1).

This is illustrated in Table 2 which gives three estimates for the number of career centuries and scores of 50 or more, as well as the actual results, for Adam Gilchrist. Note first that, up to April 2004, Gilchrist had played 76 innings and was not out 14 times, with a traditional average of 54.35. From this we calculate

$$A = ((76 - 14) / 76) \times 54.35 = 44.34$$

(see [2] for the details). The column headed (1) uses the 'exact' formula from equation (1) taking $\kappa = 2.33$ (from Table 1); the column headed (2) uses the expression in equation (2); the column headed (3) uses the estimate $e^{-c/A}$ for the same probability and will be referred to below. In each case, the estimated probability is multiplied by 76 to obtain the estimated frequency.

Table 2: Actual and estimated numbers of times Gilchrist has scored 50 or 100 in 76 innings

	Actual	(1)	(2)	(3)
≥ 50	26	24.6	25.2	24.6
≥ 100	10	8.4	8.4	8.0

4. Comparison with the method of Tan and Zhang

In their paper [4], Tan and Zhang fitted a curve of the form $y = ae^{-bx}$ to a graph of all of Jack Hobbs' test scores, and then all of Herbert Sutcliffe's. The method was summarised in [3], and is repeated here.

For Hobbs, for example, they plotted (x_j, j) where x_j is the jth greatest score, for $j = 1, \ldots, 102$, since 102 is the number of Hobbs' test innings. The least-squares approach gives the following estimates of the parameters: $a = 104$ and $b = 0.0185$. The aggregate score t is then given by

$$t = \int_0^\infty y \, dx = a / b$$

and their estimate of the true mean score follows as $t / a = 1 / b$, since a is an estimate of the number of innings. This value and that for Sutcliffe are included in Table 3, below, which gives a full numerical comparison of the results of the two approaches. Using A for the true mean, as above, then Tan and Zhang would have been justified in taking $b = 1/A$.

The number of innings in which a score of c or more is obtained is given (approximately) by ae^{-bc} (the ordinate of the point on the curve with abscissa c) and the total number of innings is approximately a, so by this approach

$$\Pr(R \geq c) = ae^{-bc}/a = e^{-bc} = e^{-c/A}. \qquad (3)$$

A different derivation of this result was given in [4], and reproduced in [3], but this method is more direct and simpler. The result has already been demonstrated in Table 2 and is again demonstrated in Table 3.

Table 3: Batting statistics of Hobbs and Sutcliffe

	Hobbs			Sutcliffe		
	Actual	C	TZ	Actual	C	TZ
True average A	53.04		54.05	54.23		53.48
Wicket average A_w	53.33			54.64		
Median score	40	37.1	37.5	38	38.0	37.1
Number of ducks	4	3.4	1.9	2	2.7	1.6
Number of 50s	43	40.4	41.2	39	34.0	34.2
Number of 100s	15	16.0	16.4	16	13.7	13.4
Number of 200s	1	2.5	2.6	0	2.3	2.1
Number of 300s	0	0.4	0.4	0	0.4	0.3

This table is adapted from that in [3]. The columns headed C are based on equation (2) and those headed TZ on equation (3), except with regard to the median score and the number of ducks. The estimate for the median score obtained in [2] is $A \ln 2$, and is independent of the choice of κ; Tan and Zhang's estimate is the same but the table reflects their estimated values for A. Note: Hobbs played 102 test innings, not out seven times; and Sutcliffe played 84 test innings, not out nine times.

The estimate for the number of ducks is arrived at as follows. An earlier result in this paper gives the probability of a batsman making no scoring stroke as $\Pr(X = 0) = p = \kappa / (A + \kappa)$. But this includes the probability of scoring 0, not out. It is necessary here to use a batsman's *wicket-average A_w* in place of A. This is the average of only those innings in which the batsman was out and, for international players, may be calculated directly from information on the CricInfo web site. In most cases, A_w turns out to be very close to the true average A, as would be expected if a batsman averaged much the same in completed innings as in not-out innings. Taking $\kappa = 2$, this gives $\Pr(\text{duck}) = 2 / (A_w + 2)$, and this, multiplied by the number of completed innings, has been used in Table 3. Tan and Zhang used a different approach.

Finally, we can relate the estimates contained in equations (2) and (3):
$$(A / (A + 2))^{c/2} = ((1 + (2/A))^{A/2})^{-c/A} \sim e^{-c/A},$$
as $A \to \infty$. That is, the two results should be expected to be in close agreement for large values of A, and this has been observed in Tables 2 and 3. Furthermore, a slightly more careful analysis using the calculation in equation (1) shows that this asymptotic result does not depend on the value of κ. By a totally different method, we have thus arrived at much the same result as Tan and Zhang.

6. Postscript — Brian Lara

On 12 April 2004, Brian Lara of the West Indies scored 400 not out in a test against England. That was his 187th test innings (not out six times) and raised his traditional batting average from 51.14 to 53.35. His average was greatest at 62.61 in April 1994, after scoring 375 against England in his 26th test innings (with no not-outs at that time). Using our equation

(2), at these three stages in Lara's career, namely just following the 375 and just before and just after the 400 not out, the probability that he would score 400 or more runs was 0.0019, 0.00038 and 0.00050, respectively. That is, at best about one chance in 500 and about one in 2500 when he actually accomplished it.

Our estimated frequencies, based on equation (2), and the actual frequencies of such scores are included in Table 4.

Table 4: Actual and estimated numbers of times Lara has scored 50, 100, 200, 300 or 400 in 187 innings

	Actual	Estimated
≥ 50	68	72.3
≥ 100	25	28.0
≥ 200	7	4.2
≥ 300	2	0.6
≥ 400	1	0.1

5. Acknowledgments

6. References

[1]   K. Cochran, *Comparison of the Score to Scoring Strokes Ratio in Various Forms of Cricket*, undergraduate essay, Department of Mathematical Sciences, University of Technology, Sydney (2000).

[2]   G. L. Cohen, "Cricketing chances", in *Proceedings of the Sixth Australian Conference on Mathematics and Computers in Sport*, Bond University, Queensland, 1–3 July 2002 (G. Cohen and T. Langtry, editors), UTS Printing Services (2002), 1–13.

[3]   G. L. Cohen, "The probability of scoring a century in cricket", *Math. Spectrum*, to appear.

[4]   A. Tan and D. Zhang, "Distribution of batting scores and opening wicket partnerships in cricket", *Math. Spectrum*, **34** (2001/2002), 13–16.

A MATHEMATICAL VIEW OF SPORTS BETTING

John S. Croucher
Macquarie Graduate School of Management
Macquarie University NSW 2109
Australia
jcrouche@gsm.mq.edu.au

Abstract

This paper examines the intricacies of sports betting from the point of view of both the gambler and bookmaker. From coverage of types of wagers and the organisations that provide them, a detailed outline of the role of the bookmaker is discussed and the way in which a fixed odds market operates. Consideration is also given to the operation of starting price operators such as those run by various Totalisator Agency Boards. A number of examples are provided along with tips on how to use your sporting knowledge to either beat the bookmaker or other punters who are trying to do the same thing.

1. Introduction

Betting on sporting events has become a booming business around the world and Australia is no exception. This type of wager can be made on all manner of local, national or international sporting activities, whether on or off-course, in person, by telephone or via the Internet. The first such institution in the country was *Centrebet* which was founded in Alice Springs in the Northern Territory in 1992. *Centrebet*, subsequently purchased by *Jupiters Casino* and now owned by the *SportsOdds*, has 190,000 clients and offers a whole range of betting opportunities on sporting events. The Totalisator Agency Board of NSW (known as *TAB Limited*) has *SportsTAB* as their department that offers sports betting. TABCORP also has a controlling interest in Star City Casino in Sydney. In Victoria the corresponding organization for sports betting is *TAB Sportsbet*. Many agencies accept bets live online via the Internet (see Haywood [2]) on around 40 different sports.

The basis of betting on sporting events is that odds and prices are offered to the gambler on a whole range of possibilities within each sport. A few of these include:
- Winner of the *Brownlow Medal* (for the best and fairest player) in the AFL
- First try scorer in rugby league matches
- Team leading at half time
- Winners of basketball matches
- Winners of soccer matches
- Winner of the Super 12's rugby competition
- Winner of the Wimbledon tennis finals
- Winner of one-day cricket matches and Tests
- Next batsman out in cricket

There has been much written on the subject of just how to invest money wisely on sports with Beaudoin *et al.* [1] providing strategies for managing a bankroll while Thorp [4] reviews optimal betting systems for favourable games. A somewhat different view is shared by O'Hara [3] who outlines the history of betting in Australia from its very beginnings. The task of deciding just what initial odds to offer on these contests falls to the *bookmaker* who has a

vital role to play in their determination. After that time the ebb and flow of money on the various outcomes will determine the current prices on offer.

2. Bookmakers

In order to frame a market for these events, the bookmaker must use the subjective judgement of both themselves and colleagues. Their task is to predict the likely outcome of a sporting event that can involve many variables. Once the market has opened and bets are placed, the initial prices that they offer can quickly change and it is the task of the bookmaker to (try to) make sure that they make a profit no matter what the outcome of the event. This is not always easy. The business of bookmaking takes into account a number of relevant factors that include:

- Accurate prices (the latest information, statistics, opinions)
- Price adjustments during betting
- Keeping a balanced book

Bookmakers should be able to predict the likely betting activity and set prices that will attract bets on all competitors. If bets are taken in proportion to each team's percentage, the bookmaker will show a profit regardless of the outcome of the event. A bookmaker must be prepared to create liability or risk, and in return for accepting risk, bookmaking (in theory at least) is supposed to be a profitable activity. In essence, bookmakers promote wagers made at unfavourable odds. Betting percentages will always be against the gambler, but to take advantage of those percentages the bookmaker must also be prepared to accept risk. The decision to accept risk may be subjective while the management of risk must be objective.

A given set of events will not always lead to the same response by the bookmaker at all times. Moreover, once a risk has been accepted, there is no assurance that the outcome will be favourable to the bookmaker. They therefore take on risk for what it can bring to the bottom line, since some gamblers will win and some will lose. However, a bookmaker is not immune to losses, as a number of those who have gone out of business will testify. Although it is not necessary for a bookmaker to be 100% correct all the time, a poorly managed risk invariably turns to a *trading loss* and this can spell disaster if huge losses are incurred.

A bookmaker must always be prepared to match liability or risk with the perceived importance of a particular contest. For example, suppose that the bookmaker forecasts that they will be holding a total of about $200,000 on a sporting event. An early bet of $5000 should not present risk problems since the subsequent betting should ensure that a bet of this size does not dominate the pool.

Definitions

- The *true price* is based on the team's 'true chance'. This is a theoretical value.
- A team is referred to as *over the odds* or *overs* if its price is *greater* than its true price.
- A team is referred to as *under the odds* or *unders* if its price is *less* than its true price.

A bookmaker usually quotes a *fixed price* or return to the potential gambler. When a bet is made at a particular price, this price cannot be changed. Even though prices can fluctuate during betting, once a bet has been accepted the price quoted cannot change for that particular bet.

Many sports betting agencies do not quote *odds* but rather a *price* for the return (including the original stake) for each $1 wagered. It is instructive to give a few simple examples of how

134

sports betting operates and to demonstrate the equivalence between prices and odds. This makes it easy to compare two bookmakers who may choose to display what they have on offer in different ways. Note that if the price offered is k, then the amount *won* would be $(k–1)$. This would make equivalent odds of $k–1$ to 1.

Example 1

A bookmaker offers a price of $2.50 on a particular team. If the team is successful and a punter has wagered $10, they will receive a total of 10 x $2.50 = $25. This consists of the original $10 stake plus winnings of $15.

Example 2

Conversion of prices to odds.

(a) $3.50. In this case $k = 3.50$ and so the equivalent odds are $(3.50 - 1)$ to 1 or 2.5 to 1 or *5 to 2*.

(b) $2.00. In this case $k = 2.00$ and so the equivalent odds are $(2.00 - 1)$ to 1 or 1 to 1 or *even money*.

(c) $1.25. In this case $k = 0.25$ and so the equivalent odds are $(1.25 - 1)$ to 1 or 1 to 4 or *4 to 1 on*.

The situation can also be viewed from the opposite point of view. That is, if a bookmaker is offering odds of *n to 1*, the equivalent price is $(n + 1)$.

Example 3

Conversion of odds to prices.

(a) 25 to 1. Since $n = 25$, the price would be $(25 + 1) = $26.

(b) 13 to 4. Odds of 13 to 4 are the same as 3.25 to 1, so that $n = 3.25$. The price would therefore be $(3.25 + 1) = $4.25.

(c) 9 to 4 o*n*. Odds of 9 to 4 *on* are the same as 4 to 9 or 0.44 to 1, so that $n = 0.44$. The price would therefore be $(0.44 + 1) = $1.44.

There are a number of ways in which a bookmaker can manage the risk they are taking. These include:

- Adjusting the odds of a well-backed outcome to detract any further betting action on that outcome. This is called *winding the price in*. For example, suppose a large bet is placed on a particular outcome at a price of $10. To deter further bets on that outcome, the book-maker might drop their price to only, say, $4, to discourage bets on that outcome, at least for the time being.
- Adjust the price of one or more other outcomes to attract betting action on them.
- Lay off some of the risk with a competitor organisation (e.g. another bookmaker). In this way the liability is spread around should the bet be successful.
- Suspend betting action on a particular outcome or altogether.

3. Market percentage

In any market in which a bookmaker quotes, the *percentages* of the outcomes must sum to over 100%. The sum of these percentages is known as the *market percentage* or *THP* (Theoretical Hold Percentage). This is an extremely important aspect since if the market percentage is less than 100%, all outcomes could then be backed in proportion to their apportioned chance and the gambler could be guaranteed a profit.

Definition 1

The *percentage* of an outcome $= \dfrac{100}{\text{Price}}$

It is easy to show that the market percentage must be over 100% or else it would be possible for punters to guarantee making a profit.

Suppose that an event has n possible outcomes and the bookmaker offers prices of x_1, x_2, x_3, ..., x_n respectively for them to be successful. Then:

$$\text{Market percentage} = \frac{100}{x_1} + \frac{100}{x_2} + \frac{100}{x_3} + \dots\dots + \frac{100}{x_n} \qquad (1)$$

Suppose that a gambler decides to invest an amount (in $) of $100/x_i$ on outcome i for all outcomes 1, 2, ..., n. Suppose outcome j is successful. Then:

Return to the gambler $=$ (Price for outcome j) x (Amount invested on outcome j)

$$= x_j \ \text{ x } \ \frac{100}{x_j}$$

$$= \$100$$

Note that in the above the x_j terms cancel out. That is, the gambler doesn't care which outcome is successful since their return will always be $100. The total outlay of the gambler will be the sum of the individual bets. That is:

$$\text{Total outlay} = \frac{100}{x_1} + \frac{100}{x_2} + \frac{100}{x_3} + \dots\dots + \frac{100}{x_n}$$

which is precisely the same as the market percentage in (1). It follows that, if the market percentage is *less* than 100, the gambler is guaranteed a profit each time. This profit will be (100 − market percentage). However, if they were to multiply each of their bets by a factor, their profit would also be multiplied by that factor with no risk to them. In other words, there would essentially *no limit* to the amount of guaranteed profit that the gambler could make. For this reason it would be a very foolish bookmaker who would have a market percentage less than 100.

Example 4

A bookmaker is framing a market for the four teams in the AFL (Australian Rules Football) semi-finals and displays the following prices (for a $1 wager).

Brisbane $ 2.00
Sydney $20.00

Collingwood $ 4.00
Port Adelaide $12.50

The market percentage can be found using (1) to determine whether a gambler could guarantee a profit. If so, what bets would they have to place to make a guaranteed profit of, say, $500?

Using (2), the market percentage is:

$$\text{Market percentage} = \frac{100}{x_1} + \frac{100}{x_2} + \frac{100}{x_3} + \frac{100}{x_4}$$

$$= \frac{100}{2} + \frac{100}{20} + \frac{100}{4} + \frac{100}{12.5}$$

$$= 50 + 5 + 25 + 8$$

$$= 88\%$$

Since the market percentage is less than 100%, a gambler *can* guarantee a profit of $(100 - 88) = \$12$ by investing the amounts in Table 1. From this table it can be seen that the total amount invested is $88, but whatever team wins the gambler will collect a total of $100.

Table 1: Amount to invest on each team for Example 4 to make a profit of $12.

Team	Amount invested ($)
Brisbane	100/2 = 50
Sydney	100/20 = 5
Collingwood	100/4 = 25
Port Adelaide	100/12.5 = 8
Total	88

This shows that the punter can make a guaranteed profit of $12 for every $88 that they invest (or a profit of 13.6%). If the target profit is $500, then calculate how many multiples of 12 are required to reach 500. The answer is 42 (since 42 x 12 = 504). This means that the gambler should multiply all of their bets in Table 1 by a factor of 42. This yields the quantities in Table 2.

Table 2: Amount to invest on each team for Example 4 to make a profit of $500.

Team	Amount invested ($)
Brisbane	50 x 42 = 2100
Sydney	5 x 42 = 210
Collingwood	25 x 42 = 1050
Port Adelaide	8 x 42 = 336
Total	3696

The total outlay for the gambler would be $3696. However, whatever team won, the return would be $4200, thus yielding a guaranteed profit of $504. This still represents a profit to the gambler of 13.6% on the outlay.

Even if an individual bookmaker does *not* have a market percentage of less than 100, it still might be possible for a gambler to guarantee a profit. Since it is not necessary to place all bets

with the same bookmaker, the gambler can 'shop around' to find a selection of bookmakers who, when particular prices are selected, yield a market percentage of less than 100%. In this case a betting scheme such as that outlined in Example 4 could be undertaken. For example, a gambler might bet on Brisbane with one bookmaker, Sydney and Port Adelaide with another and Collingwood with a third bookmaker.

There is another possibility, although somewhat more risky. Since it is most likely that the market percentage for an event will exceed 100%, if a gambler is prepared to rule out one of more outcomes (that is, they are prepared to bet against them), they can reduce the market percentage to below 100% and bet along a similar lines to Example 4.

Example 5

The ING Cup is a one-day domestic cricket competition in Australia played between the various states. By February 2004 there were only five teams that had a possibility of winning the final. Table 4 shows the prices offered by *Centrebet* on 12th February. Also shown are the percentages for each team.

Table 3: Prices offered by *Centrebet* on five teams to win the ING Cup

Team	Price ($)	100/Price
Western Australia	2.10	47.62
Victoria	2.75	36.36
Queensland	3.50	28.57
NSW	51.00	1.96
South Australia	51.00	1.96
	Total	116.47

As expected, the *percentages* (100/price) shown in Table 4 add up to over 100 and so there is no strategy guaranteed to win. However, if we are prepared to eliminate one or more teams such that percentage drops down to *less than* 100, then wagers may be placed along the lines of Example 4.

For instance, if we decided that Queensland (the third favourite) would not win and eliminated them from the discussion, the remaining percentages add up to only 87.90. We could then bet *100/price* on each of the remaining teams. A Queensland victory would then be the only outcome by which we would not win. There is then a risk involved to gain a profit. In this case, for every $87.90 invested, there would be a profit of $12.10 or 13.76%. However, according to their price, *Centrebet* has estimated that true probability that Queensland will be the winner at $1/(3.50 \times 1.165) = 0.25$. If they are correct in their view, there would be a 25% chance that you would lose your bet, but a gain of 13.76% if you won it.

As it happened, the 2004 ING Cup final was played between Queensland and Western Australia on 29 February in Brisbane. It was won by Western Australia by four wickets with just two balls to spare.

4. Framing an odds-based market

When a bookmaker frames an initial market, there are a number of tasks to be undertaken. Some of these include:
- Undertake research for performance indicators of each competitor.

- Weigh up the chance of success for each competitor.
- Set the market percentage (must be over 100%).
- Assign the market percentage to each competitor to reflect his or her weightings and then, if necessary, adjust for public perception.
- Convert the percentages to prices.
- Open for business.

The framing of the initial market prices is of great importance, since punters eager to snap up a bargain can pounce upon any errors. Example 6 shows one way in which such a market can be framed.

Example 6

In February 2004 there commenced a six match series of one-day international cricket between New Zealand at home and South Africa. Since South Africa had recently displayed sparkling form at home in defeating the West Indies 3–1, they were installed as firm favourites with bookmakers, despite this time playing away.

Suppose a bookmaker feels that New Zealand has a 20% chance of winning the series while South Africa has a 62% chance and there is an 18% chance of it being drawn. The bookmaker decides to set the market percentage at 110%.

To set the prices, the bookmaker multiplies each of these percentages by 1.10 as follows:

New Zealand to win   20 x 1.10 = 22.0%
South Africa to win   62 x 1.10 = 68.2%
Series is drawn   18 x 1.10 = 19.8%
         Total   110.0%

The *price* (for a $1 bet) for each outcome can now be calculated by dividing each of these percentages into 100%. This yields:

New Zealand to win   100/22.0 = $4.55
South Africa to win   100/68.2 = $1.47
Series is drawn   100/19.8 = $5.05

This would be the bookmaker's opening market that would then be adjusted as subsequent bets are placed. In fact, *Centrebet's* final prices on the series were: New Zealand ($4.20), South Africa ($1.50) and Draw ($5.00) which were all in fairly close agreement with the bookmaker's figures.

Despite being the clear underdogs in the betting, after losing the first game, New Zealand emerged clear victors 5-1 in the series when they won the next five games in a row.

5. Fluctuating prices

We have already noted that bookmakers are *not* guaranteed of making a profit, while the Totalisator (TAB) betting is organised so that they cannot lose since their cut (called a *take-out rate*) is already removed before the final prices are declared. Bookmakers must pay the punter the price provided at the time the bet is placed while the TAB only pays the price at the time the contest actually starts (at which time betting closes). For this reason, some punters do

not place their wager with the TAB until only a few minutes before betting finishes since it is only then that they can be reasonably sure that there will be little change in the price offered.

The *current price* of an outcome is based on the relative amount of money that has already been invested on it. In the case of horse racing, for example, the Totalisator display board at the racecourse, along with the screen monitors at the TAB, actually show the *current price* for a $1 bet. This current price is constantly changing according to how much money is placed on each outcome. In a similar way, a *bookmaker* displays the price they will give the punter. At the racetrack, most bookmakers now use a computer to quickly perform these calculations.

Consider a sporting event that has *n* possible outcomes. Let:

a_i = the amount bet on outcome *i* $(i = i, …, n)$

Then the *total gross pool* invested is:

$$\$P \;=\; a_1 + a_2 + \;….. \; + a_n$$

Suppose that the organisation decides to take out a fraction *x* of the gross pool for their profit $(0 \le x \le 1)$.

This leaves the *dividend pool* – the amount (in $) to distribute to the punters:

$$(1–x)P$$

Suppose that outcome *j* is successful $(1 \le j \le n)$.

Then for each $1 invested on the winning horse, the punter receives (in $):

$$\frac{(1–x)P}{a_j} \tag{2}$$

This is then the *current price* of the outcome. As the amount invested overall, *P*, changes, along with the values of the a_i, then the price also fluctuates.

Example 7

A betting agency was accepting wagers on the NBA basketball championships in the USA and toward the end of the 2004 season there were six teams left in contention. Table 4 shows the total amount of the bets that have been placed on each team up to this point in time.

Given the agency's take-out rate of 13.75%, what price should it currently be offering on each team? (Note: The *actual* price the gambler receives is that at the *close* of betting.)

Table 4: Amounts bet to date on the NBA teams in Example 7

Team	Amount bet ($)
Los Angeles Lakers	250,000
Sacramento Kings	100,000
San Antonio Spurs	20,000
Minnesota T-Wolves	80,000
Dallas Mavericks	150,000
Indiana Pacers	200,000

Total	*800,000*

In this case:

$P = 800,000$ and $x = 0.1375$

Then the *price* is given by (2):

$$\text{Price} = \frac{(1-x)P}{a_j}$$

$$= \frac{0.8625 \times 800,000}{a_j}$$

$$= \frac{690,000}{a_j}$$

(Recall that $a_j = \$$ amount bet on outcome j.)

Note that the *dividend pool* is $690,000. That is, the agency will have to pay out a total amount of $690,000 no matter which team is successful. The current prices may be easily calculated using (2). Note that, in accordance with many agency practices, the dividends have been rounded *down* to the next 10 cents. The results are shown in Table 5.

Table 5: Current prices for the NBA teams in Example 7

Team	*Current price ($)*
Los Angeles Lakers	2.70
Sacramento Kings	6.90
San Antonio Spurs	35.50
Minnesota T-Wolves	8.60
Dallas Mavericks	4.60
Indiana Pacers	3.40

It can be seen in Table 5 that the *favourite* team (at least with punters) is the Los Angeles Lakers since it has the lowest price. The *outsider* is San Antonio Spurs with the largest price. Since the amounts that are invested on each team are continually changing, fresh calculations by the agency will mean continual changes in the prices. But they will still not care which team wins since they will always make the same profit.

To illustrate that this method achieves the desired result, suppose that the Minnesota T-Wolves are successful. The agency will have taken 80,000 in wagers (in $) and for each of these $ amounts will pay out $8.60. This yields a total payout of $80,000 x 8.60 = $688,000, a result close to the $690,000 they are seeking. A similar outcome is obtained for *any* of the six teams who win. Their income, however, is still $800,000.

6. Remarks

Sports betting gives the fan a chance to show that they are not only a keen supporter of the game but have a knowledge that is superior to others who may wish to make a wager on the outcome. They may avail themselves of a wide variety of information such as recommenda-

tion of professional tipsters, likely weather conditions, injuries to key players, home ground advantage and previous performance of the competitors to name but a few. If a gambler always takes *overs*, then they are likely to show a profit in the long term. This really means that they are placing bets on outcomes that are a higher price than is really justified. In this case the bookmaker has initially (and most likely other punters as betting progresses) *underestimated* the ability of the competitors and given them a price that is higher than it should have been. If a gambler has the ability to detect this, they are well on the way to being successful. A bookmaker must try to avoid initially giving overs, but this is difficult since they can only *approximate* the true price with a subjective judgement. It may only be after the event that they realise they were wrong.

The modern sports betting agency will use a computer system that allows their betting managers to make adjustments in real time, since many bets may be placed *during* a particular event. This might include, for example, the next goal or try scorer in football or next batsman out in cricket.

Such a system would enable the manager to:
- Set prices as desired.
- Monitor all bets that are made
- Make price changes in real time
- Set parameters that trigger off alerts. This includes any particularly large bets that might be viewed with suspicion.

Betting on sports has become very popular and has created a great deal of interest among many sports followers. The television channels are also very pleased with this aspect since they feel that, even if a sporting contest is becoming one sided and the winner is well known before the completion, the interest of viewers will be maintained until the end just so they can see if they were able to come up with correct final score. Even if they did not do so themselves, they are still often interested in what was the final price for being correct. It seems that (almost) everyone is happy!

References

[1] D. Beaudoin, R. Insley and T. Swartz, "Studying the bankroll in sports gambling", in *Sixth Conference on Mathematics and Computers in Sport*, G. Cohen and T. Langtry (editors),University of Technology, Sydney (2002), 69-77.

[2] H. Haywood, *"BeatWebCasinos.Com: The Shrewd Player's Guide to Internet Gambling"*, RGE Publishing, Oakland, California, USA (2000).

[3] J. O'Hara, *A mugs game: a history of gaming and betting in Australia*, New South Wales University Press, (1988), Sydney.

[4] E.O. Thorp, "Optimal gambling systems for favourable games", *Review of the International Statistical Institute*, (1969), 37, 273-293.

BOUNCING A WATER POLO BALL

Neville de Mestre
Faculty of Information Technology
Bond University
Gold Coast, Queensland 4229

Abstract

During a game of water polo, players sometimes endeavour to deceive the opposing goalkeeper by bouncing the water polo ball off the water surface and into the goalmouth. The dynamics of this attacking strategy is investigated in the paper.

1. Introduction

When a ball bounces on a hard surface the ball is usually deformed during its contact time with the surface. Most elastic balls (eg. tennis, squash, basketball) recover their original shape after the bounce, but their velocities are always changed. The addition of top spin or back spin creates additional changes to the velocity of the ball after impact [2].

Bouncing an object off a water surface is a more difficult task to achieve and also to analyse. Recently Bocquet [1] has analysed the skipping of a fast stone off the surface of a lake or river. The water surface deforms and the stone does not. Stone skipping is not always achievable, and the stone may plummet to the bottom without skipping at all. The way to make a stone skip is set out in Jerdone Coleman - McGhee's book [3] and the web page of the North American Stone Skipping Association (NASSA).

Horatio Nelson is supposed to have discovered by accident that cannon balls, which do not float, ricochet off the surface of the sea in a series of hops and skips. He began to use this to great effect in striking enemy vessels near the water line. The idea of bouncing solid spheres or cylinders off the surface of water was later developed by Barnes-Wallis during World War II for the famous Dambuster Raid of 1943 [4].

In this paper the bouncing of a water polo ball off the surface of the pool will be considered, the principal difference between this object and stones, or bouncing bombs, is that the ball can float.

During the game of water polo, some players can project the ball down onto the water surface so that it bounces once and travels upwards into the high parts of the goalmouth. The opposing goalkeeper is frequently deceived by the downward action of the throw and stays low in the water to prevent the shot from going under his outstretched arms. The change of direction of the ball from down to up after the bounce can trick an unwary goalkeeper. In this sporting situation very little spin is imparted to the ball.

2. Floating Ball

When a water polo ball floats, its weight is balanced by the upwards buoyancy force from Archimedes' Principle. Thus

$$mg = \frac{\pi}{3} z^2 \left(a - z \right) \rho_w g$$

where m denotes the mass of the ball, mg its weight, ρ_w the density of water, a the radius of the ball, and

$\pi z^2 \left(a - z \right) / 3$ represents the volume of water displaced by the ball when its lowest point is distance z below the water surface. Rearranging this yields

$$z^3 - 3az^2 + 3m/(\pi\rho_w) = 0$$

For a water polo ball $a = 0.11(m)$, $m = 0.425(kg)$, $\rho_w = 1000(kg\ m^{-3})$ and hence the above cubic equation becomes

$$z^3 - 0.33z^2 + 0.0004 = 0$$

with solution $z = 0.037$. That is, the ball floats with 3.7cm under the water, and the remaining 18.3cm above the water, and this has been checked experimentally.

3. Vertical Impact

The vertical impact of a ball on water will be considered first of all. Frequently, the start of each quarter in a water polo game involves the referee dropping the ball vertically at the halfway line near one side of the pool, while the opposing teams race for the ball from their respective goal lines. The whole motion of this vertically-dropping ball can be checked easily by experiment to see if the mathematical model of its motion is correct.

The first part of the motion involves the ball falling vertically through the air. The governing equation is

$$m\ddot{z} = mg - mk\dot{z}^2 \quad .... (1)$$

where the last term denotes the drag due to the air, and a dot denotes differentiation with respect to time. The Reynolds number of the air flow past the ball is in the region where the drag force can be represented by a quadratic model with a resistance coefficient

$$mk = \frac{1}{2}\rho_A A C_D$$

where ρ_A denotes the density of air, A denotes the area of cross-section of the ball, and C_D denotes the drag coefficient. The OZ axis is vertically down with its origin at the centre of mass of the ball just as it reaches the water surface (see Figure 1).

Figure 1: Forces and co-ordinate system for a falling water polo ball:

Dividing equation (1) by m and writing

$$\ddot{z} = \frac{d}{dz}\left(\frac{1}{2}\dot{z}^2\right)$$

yields

$$\frac{d}{dz}(\dot{z}^2) = 2g - 2k\dot{z}^2$$

The solution

$$\dot{z}^2 = \frac{g}{k}\left(-\exp\left(k\left(z_0 - z\right)\right)\right) \quad\text{... (2)}$$

is obtained by separating the variables and integrating using the initial condition $\dot{z}=0,\ z=z_0\ (<0)$.

With $C_D = 0.47$ and the density of air $\rho_A = 1\ (kg\ m^{-3})$, and if the ball is dropped from a height of 1.5m $(z_0 = -1.5)$, the velocity of the ball as it hits the surface is calculated from result (2) to be 5.34 ms^{-1}.

The ball then starts to penetrate into the water surface, but new forces come into play. These are the drag force of the water medium and the buoyancy force from Archimedes' Principle. The governing equation of motion is now

$$m\ddot{z} = mg - \frac{1}{2}\rho_w\left(2\pi az\right)\varsigma_D\dot{z}^2 - \frac{\pi}{3}\rho_w z^2\left(a - z\right)g \quad\text{.............................. (3)}$$

where $2\pi az$ is the wetted area of the ball when it has penetrated a depth z into the water. Note that when the ball is fully immersed its wetted area is $4\pi a^2$, but its cross-sectional area is πa^2. Hence, when the wetted area is used, the drag coefficient C_D has to be divided by 4.

Equation (3) can be recast in the form
$$\ddot{z} + Kz\dot{z}^2 = g - Bz^2\left(a - z\right)$$
where $K = \pi\rho_w a C_D / m$ and $B = \pi\rho_w g / (m)$. The measured values for a water polo ball produce
$$\ddot{z} + 95.54z\dot{z}^2 = 9.81 - 24172z^2\left(0.33 - z\right)$$
with initial conditions $z=0,\ \dot{z}=5.34$ when t = 0. The software package MATHEMATICA yields
$\dot{z}=0,\ z=0.14$ when t = 0.052,
that is, the ball penetrates to a depth of 14cm before it starts to rise again. This has been checked experimentally by dropping a dry ball and retrieving it after one bounce to observe the wetted region.

As the ball rises, the governing equation now has to have the drag reversed in direction, but all other forces stay in their original directions. Thus the new equation for the motion of the ball is
$$\ddot{z} - 95.54z\dot{z}^2 = 9.81 - 24172z^2\left(0.33 - z\right)$$
with initial conditions those at the final penetration depth 0.14m. MATHEMATICA now yields
$z=0,\ \dot{z}=-1.40$ when t = 0.093,
indicating the speed of the ball as it leaves the water surface. The total time immersed is 0.145 seconds which is much longer than the contact time if the ball had been dropped onto a hard surface.

The ratio of outwards speed to inwards speed is (1.4/5.34) yielding a value for the coefficient of restitution 0.26.

4. Oblique Impact

Suppose that the ball strikes the water obliquely with velocity components $\dot{x}=u_o$, $\dot{z}=v_o$ at t = 0. The origin of the OX, OZ co-ordinate system is again the position of the centre of mass of the ball at the first impact (Figure 2).

Figure 2: Initial conditions at impact:

145

As before, when the ball stops penetrating, it is seen that $\dot{z}=0$ at $t=\tau_1$, and when the ball leaves the water surface $z=0$ at $t=\tau_2$. The condition for the ball not bouncing, but ploughing into the water and stopping, is clearly $\dot{x}=0$ for some $t \le \tau_2$.

As the ball penetrates the surface there will be viscous forces and pressure forces due to the water acting at all points of the wetted area. These can be combined to give a resultant drag acting in the direction opposite to the ball's velocity, and a resultant lift acting perpendicular to this direction (Figure 3).

Figure 3: Forces on the ball during oblique impact:

The governing equations of motion for inward penetration are

$$\ddot{z} = g - Bz^2 (a - z) - Kz\dot{z}\sqrt{\dot{x}^2 + \dot{z}^2} - (C_L / C_D) z\dot{x}\sqrt{\dot{x}^2 + \dot{z}^2} \ \ldots\ldots\ldots (4)$$

$$\ddot{x} = -Kz\dot{x}\sqrt{\dot{x}^2 + \dot{z}^2} + (C_L / C_D) z\dot{z}\sqrt{\dot{x}^2 + \dot{z}^2} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$
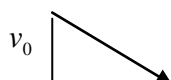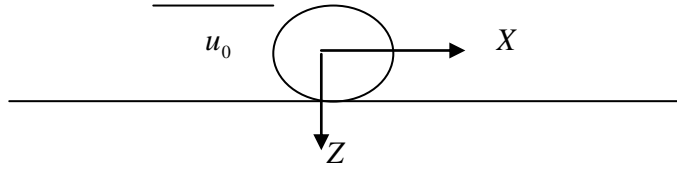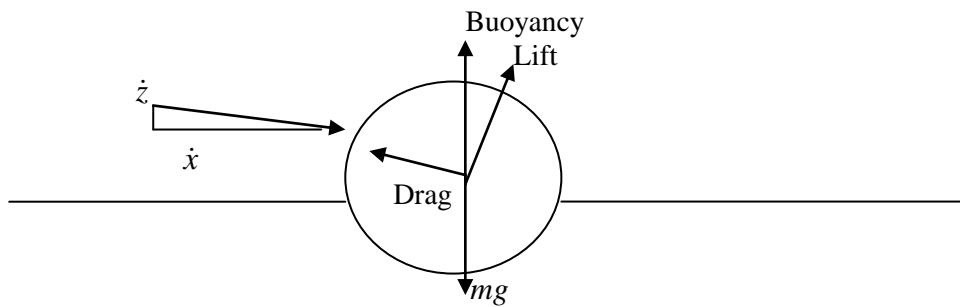
where B and K are defined previously, and C_L is the lift coefficient, which is not well documented for a bouncing water polo ball off a water surface. The values K and B are as previously used, namely K=95.54 and B= 24172, while $C_D = 0.47/4$ with C_L not yet specified.

The equations for the outwards section of the penetration of the ball beneath the water surface are the same as for the inwards section, because the change of direction of the drag is now taken care of by the change in the sign for $\dot{z}$ and no change to $\sqrt{\dot{x}^2 + \dot{z}^2}$.

5. Experiments

The motion of a water polo ball bouncing off a water surface was captured on video, and the relevant information measured. The values obtained were

$u_o = 10.7 ms^{-1}$, $v_o = 6.2 ms^{-1}$

for the ball just before contacting the water, and

$u_2 = 6.6 ms^{-1}$, $v_2 = -1.8 ms^{-1}$

as the ball was just leaving the surface.

Various values were tried for the ratio C_L/C_D and the output from the numerical solution of equations (4) and (5) using MATHEMATICA compared with the experimental output (u_2, v_2). The results are shown in Table 1.

Table 1: Numerical calculation for Exit Speed:

C_L/C_D	$\tau_1 (\ell = O)$	z_1	$\dot{x}_1$	$\tau_2 (\ell = O)$	$\dot{x}_2 = u_2$	$\dot{z}_2 = v_2$
1	0.01	0.053	7.2	0.048	3.60	-2.05
1.5	0.012	0.043	8.8	0.031	5.60	-2.50
2	0.010	0.037	9.4	0.025	6.80	-3.5
2.4	0.009	0.034	10.0	0.021	7.0	-4.0

Each calculation shows that the ball is bouncing, since $\dot{x}_2$ is positive as the ball leaves the water. No calculation achieves agreement for both $\dot{x}_2$ and $\dot{z}_2$ with the experimental measurements, although $C_L/C_D = 2$ gives good agreement for $\dot{x}_2$. Note however that the depth of penetration (ℓ_1) predicted for this case is exactly the same as the depth when the ball floats.

There are a number of errors that could have arisen with the measurements. The video ran at 25 frames/second, so that successive frames jump every 0.04 seconds. This means that the position of the ball in the frame before contact is not at the surface, and the position of the ball in the next frame shows that it has left the surface. Extrapolations forwards and backwards were needed to determine approximate measured values for u_o, v_o, u_2, v_2. Further experiments are planned with a high speed camera scheduled to arrive before the conference is held.

A verification of the model for bouncing a ball off a water surface may have an application to the putting of a golf ball. The golf ball has been noticed to bounce in little hops across the green when observed closely, riding on the tops of the grass and not rolling purely as many suspected. Investigations are still underway to obtain a useful model for the motion of a golf ball putted across the surface of a green.

6. References

[1]   L. Bocquet, "The Physics of Stone Skipping", *Am.J.Phys.*, **71** (2003), 150-155.

[2]   H. Brody, "That's how the ball bounces", in *The Physics of Sport,* A. Armenti Jr.(Ed), Am.Inst.Phys, New York, (1992),155-158.

[3]   J. Coleman-McGhee, *"The Secrets of Stone Skipping"*, NASSA, (1995).

[4]   I. Hutchings, "Bouncing bombs of the Second World War", *New Scientist*, (2 March 1978), 563-566.

A SEVEN-STATE MARKOV PROCESS FOR MODELLING AUSTRALIAN RULES FOOTBALL

Donald Forbes & Stephen R. Clarke
School of Mathematical Sciences
Swinburne University
Melbourne, Australia

Abstract

The official Australian Football League data as collected by Champion Data for the 185 matches played in season 2003 has been used to develop a seven-state Markov process for modelling Australian Rules football matches. Transition probabilities for each match have been derived and the resulting matrix of probabilities powered up to ascertain the steady state distribution for each match. Expected totals for each state have been deduced using these distributions and compared back to the observed totals to ascertain the appropriateness of such a model.

1. Introduction

Australian rules football is the predominant winter sport played in the southern states of Australia. A national competition run by the Australian Football League (AFL) contains 16 clubs and includes at least one team from each of the 5 mainland states. The clubs compete against each other during the home and away season over 22 weeks before the top eight sides play in the finals series over four weeks to determine the premiership winner.

The game is played on grounds that are oval in shape of varying dimensions. At each end of the ground is a semi-circle that signifies 50 metres to the goal posts. Inside the arc of this circle is known as the forward zone for the attacking team and defensive zone for the opposition. Between the two 50m arcs is known as the midfield zone. AFL matches consist of four quarters that are each 20 minutes in length of actual playing time. Once time off is included, a quarter will run for at least 28 minutes in real time.

At each end of the ground is a set of four upright posts, which the attacking team uses to score. Scoring can be done by the addition of either six points or one point. A six point score is known as a goal and occurs when a player kicks the ball between the two centre posts. The ball is then returned to the centre of the ground for a bounce. If the ball is touched or passes between the two outside posts, a behind worth one point is. If the opposition kick the ball or punch it between the two outside posts, the attacking side registers a rushed behind. After the scoring of a behind, the opposition kick the ball back into play from the goal square via what is known as a kick-in. A diagram of an AFL ground is included in Figure 1 with relevant features described.

Since 1996 the data collection company Champion Data (CD) has been responsible for recording and storing the occurrences of each match. Their systems and processes have resulted in more detailed information being available than ever before. Previously one had to rely on summary statistics at the end of each match that gave no indication of the quality or position on the ground of match occurrences. Presently, CD records the position of each statistic according to its zone on the field and time codes each event within the game. The level of detail inherent in their collection has seen them become the official information provider to the AFL as well as supplying data to the AFL clubs and media outlets alike.

Figure 1: Important features of an AFL playing field

Attacking goal posts

Forward 50m
Zone

Length
135 – 185m

Midfield
Zone

Centre Square
50m x 50m

Defensive 50m
Zone

Goal Square

Width
110 – 155m
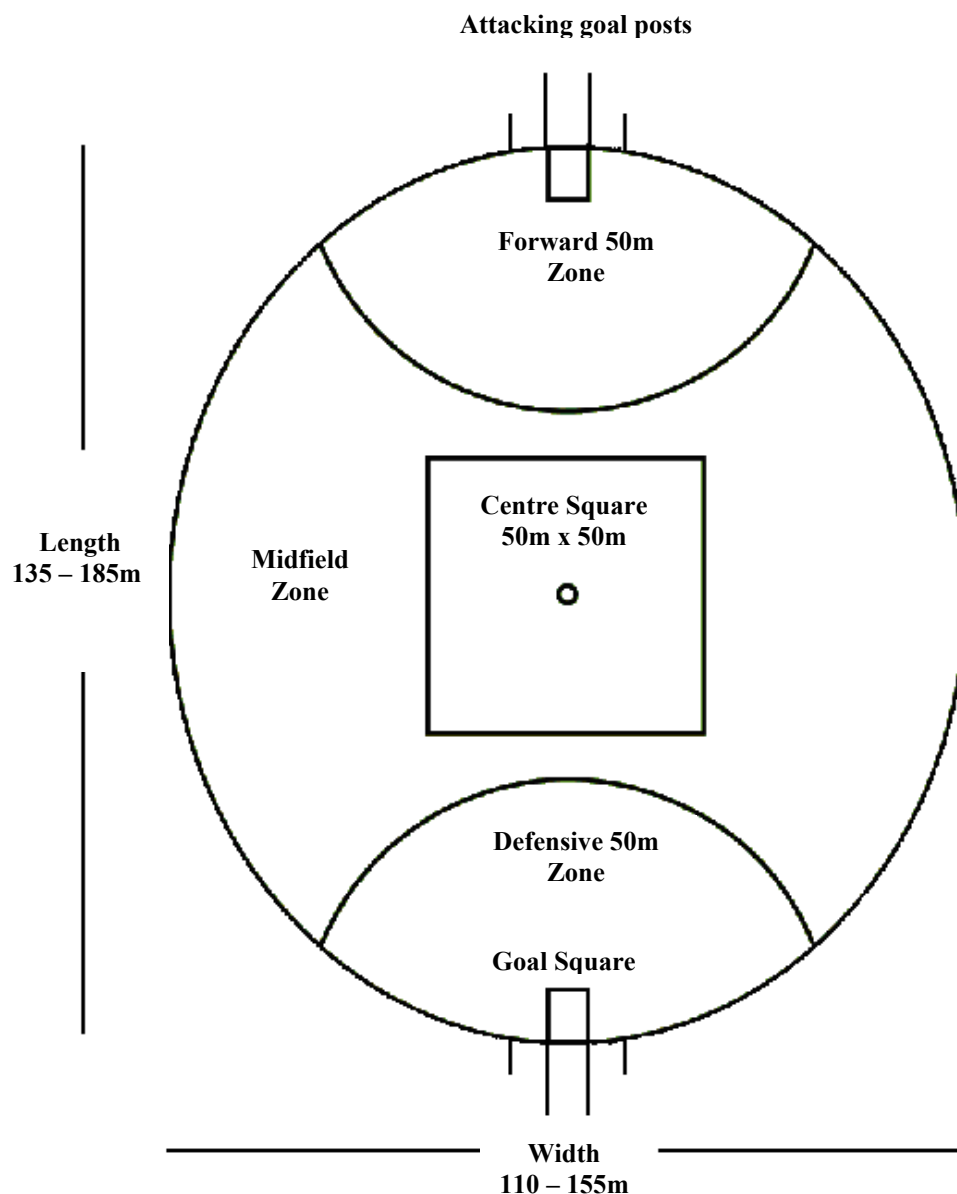
This model stems from a research collaboration with CD, who have provided access to their data bases stemming back to the 1998 season. This has resulted in uniform data and definitions of statistics for every match since the beginning of 1998. The advent of CD's level of detail and information has enabled a Markov model to be investigated and ultimately implemented as opposed to different techniques that may not rely on the level of information that this model requires.

Attempts at sports' modelling have involved many different mathematical techniques over the years. Least squares techniques have been popular with researchers since Stefani [1] modelled American football and basketball results. Bailey [2, 3] used the same technique in different applications to AFL football with his modelling of Brownlow medal votes and match predictions. A summary of papers was compiled by Norman [4] in which he looked at 17 papers concerned with ways to utilise stochastic processes for modelling sport. The use of Markov chains for sport modelling begun in the 1970s with the work of Bellman [5] however, it has predominantly revolved around the sport of baseball. Clarke and Norman[6]

utilised stochastic techniques to investigate different decision processes in the game of cricket. More recently, the Markov chain and its application to baseball have been revived through the work of Bukiet [7]and Hirotsu [8].

Bukiet's research proposed a Markov chain model for baseball that found optimal batting orders, run distributions per half inning and per game and the expected number of games a team should win. This involved a 25 state model and therefore a 25 x 25 transition matrix for each player consisting of that player's probabilities of shifting the state of the game to any other during an appearance at the plate. These probabilities are dynamic in the sense that they can be adjusted as the season progresses and form strengthens or wanes. Hirotsu extended the twenty-five state model described above in a number of ways including a 1,945 state model for expected runs using non-identical players and a 1,434,673 state model to obtain the probability of victory from any state in the game. He also addressed strategy issues such as optimal pinch-hitting and substitution for pitchers based on the handedness of the pitcher and player at bat.

It must be noted that the game of baseball is a discrete event sport similar to cricket and differs from continuous sports such as Australian rules or soccer. In a continuous time Markov chain the process makes a transition from one state to another, after an interval of time has been spent in the preceding state. This interval is defined as the state holding time. For a discrete time Markov chain the holding time is 1, while in continuous time Markov chains it is exponentially distributed. Continuous sports have also been modelled using a Markov chain approach and the methods involved are not greatly different. Hirotsu [9] utilised a 4-state Markov chain for approximating the game of soccer and used this model to evaluate the expected number of goals in a match as well as the expected number of league points a team was likely to obtain. It was also useful for investigating strategy issues such as when to substitute or commit a deliberate foul in order to increase the chances of victory. It was this work that provided the inspiration for this model.

The purpose of this paper is to demonstrate the validity of a seven-state Markov model to AFL matches. Clarke and Norman [10] investigated the decision process in an AFL game of when to rush a behind in order to maximise a team's chance of victory. Their model did not utilise actual data, instead the authors chose to assume transition probabilities based on their knowledge of the game. With CD's entry into the information collection market, this work can be revisited and actual data used.

An AFL match is a more complicated modelling proposition than soccer due mainly to the scoring system and the quantity of time the ball is not in either team's possession. Whilst Hirotsu's model was relatively simple, with only four states, this model has a minimum of seven states with the potential for many more based on the location of the ground that the event of interest takes place. CD records this.

2. The Model

A seven-state model has been investigated. It is made up of the following states:

> State 1 – Team A in possession
> State 2 – Team B in possession
> State 3 – Ball in dispute
> State 4 – Team A goal   (6 points)
> State 5 – Team B goal
> State 6 – Team A behind (1 point)
> State 7 – Team B behind

Up to 84 different match occurrences are called by CD for each game. The model uses 26 of the 84 to assign transition probabilities between each state. The statistics had to be coded according to what transition they constituted. This was done using the Champion Data and AFL accepted event definitions. For example, a short kick is defined as a kick of less than 40m that finds a team-mate and as such guarantees the ball stays with the team in possession. The watching of matches off tape also assisted in best approximating the events of play with the proposed model. Table 1 lists the occurrences included in the model, as well as a description of each and the state transition associated with it.

There are several reasons why the remaining statistics were not used in the model. Firstly, only count data was used. This ensured categorical variables were omitted e.g. inside 50 and rebound 50, interchange on or off. These variables added no numerical value to the model. Secondly, some variables offer no evidence of what has taken place, as far as possession and scoring goes, within the game. Examples of these are free kicks, bounces and tackles. Finally, not all of the statistics are mutually exclusive and may be recorded twice. The statistical package, SAS 8.01, was used to transform the raw data into a form that constituted individual transition matrices for each game of the 2003 season. In order to do this redundant statistics had to be removed from the analysis. In certain instances some statistics will be included in other codes as well as their own. This happens with derived statistics such as a long kick to advantage, which will also be included as a long kick. For instance, in a match if Team A had 30 long kicks and 10 long kicks to advantage, the system would record them as having had 40 long kicks. The same issue arose with goals and the kick that resulted in the goal. The data will record each goal scoring kick within the kick code as well as recording the goal. These doubled up occurrences had to be eliminated so that what happened in the game was reflected as accurately as possible by the numbers used for transition probabilities. Furthermore, a transition that is defined by the characteristic of the event taking place, e.g. a KKSH guarantees possession, did not need to have the associated possession gather included as well. Watching games off tape, accompanied by the transaction files allowed for decisions to be made on what to include in the analysis and what to leave out.

Of the 49 transition probabilities, 23 have zero probability, as the associated transition cannot occur. An example would be team A having the ball and team B kicking a goal. The remaining 26 transition probabilities need to be calculated. An example of some of these transitions and the relevant match statistics that comprise them are given below.

- *Possession → Possession:* Team A (or B) has possession and the disposal ensures they retain possession (KKSH, HBEF, KKLA)

  The definition of these disposal types guarantees that the team that has the ball retains it, either via the foot or hand, for the next play.

- *Goal → Possession:* Team A (or B) kicks a goal and Team A (or B) gains the next possession (CBFP)

  In an initial model the scoring of a goal meant a reversion to state 3 with probability 1. However, projections are more accurate when the resulting possession after the goal is attributed to the relevant team, as evidenced by a CBFP, or to dispute if there was no clear possession.

- *Behind → Possession:* Team A (or B) scores a point and Team B (or A) gains the next possession (KILA, KISH, KISE)

  Similar to when a goal is scored, the model was more accurate when probabilities were assigned to transitions based on the kick-in statistics for the match rather than

assuming the non-scoring side gained possession with probability 1. The kick-in codes above result in the team that kicks the ball back into play retaining the ball. Other codes may lead to the ball going 'in dispute' or ensuring the opponents gain possession.

Table 1: Statistic codes, descriptions and transition for match occurrences contained in model

Stat. Code	Description	Transition
BEHI	Behind (1 point)	POSS → BEHI
CBFP	Centre Bounce First Possession	GOAL → POSS
GATH	Gather of Loose Ball	DISP → POSS
GEHA	Hard ball get	DISP → POSS
GELO	Loose ball get	DISP → POSS
GERU	Gather from a ruck	DISP → POSS
GOAL	Goal (6 points)	POSS → GOAL
HBCL	Clanger Handball	POSS → OVER
HBEF	Effective Handball	POSS → POSS
HBIN	Ineffective Handball	POSS → DISP
KIBU	Kick in resulting in a ball-up	BEHI → DISP
KICL	Clanger Kick-in	BEHI → OVER
KIIN	Ineffective Kick-in	BEHI → DISP
KILA	Long Kick-in to advantage	BEHI → POSS
KILO	Long Kick-in	BEHI → DISP
KISE	Kick-in to self	BEHI → POSS
KISH	Short Kick-in	BEHI → POSS
KKCL	Clanger Kick	POSS → OVER
KKGK	Ground Kick	DISP → DISP
KKIN	Ineffective Kick	POSS → DISP
KKLA	Long Kick to advantage	POSS → POSS
KKLO	Long Kick	POSS → DISP
KKSH	Short Kick	POSS → POSS
MACO	Contested Mark	DISP → POSS
MAUN	Uncontested Mark	DISP → POSS
RUSH	Rushed Behind (1 point)	DISP → BEHI

The transition probabilities are then calculated by using the total number of statistics that start in the same state and dividing the count for each transition by this total. These probabilities comprise the transition matrix for each game by inserting them in the relevant row and column. Once the transition matrices for each game from the 2003 season (185 games) have been calculated, the long-term behaviour of each matrix is of interest. The Markov process from each match has an associated limiting probability distribution:

$\pi = (\pi_0, \pi_1, ...,\pi_7)$, where $\pi_j > 0$ for $j = 0, 1, ...,7$ and $\sum_j \pi_j = 1$

This convergence means that in the long run, the probability of finding the Markov chain in state j is approximately π_j no matter in which state the chain began at time 0. Relying on this knowledge, the limiting distributions for each match were derived by raising the transition

matrix to the power 10,000. We can also think of π_j as the proportion of time in each match that the chain spends in state j. Alternatively, they are the probabilities of team A having possession, team B having possession, the ball being in dispute, team A kicking a goal or behind and team B kicking a goal or behind at any time in the match.

3. An Example

To illustrate, the match data from the round 18 game between Hawthorn and Port Adelaide has been used. Table 2 includes the raw data from the match summary file and the adjusted totals after the data has been cleaned. The difference in totals between *raw* and *adjusted* indicates the redundant occurrences that have been removed from the analysis.

Table 2: Match stats and adjusted totals for round 18 game: Hawthorn v. Port Adelaide

Statistic	State	Hawthorn		Port Adelaide		Dispute
		Raw	Adjusted	Raw	Adjusted	
BEHI	POSS	11	11	9	9	0
CBFP	GOAL	10	9	13	11	3
GATH	DISP	24	16	31	19	0
GEHA	DISP	15	13	24	22	0
GELO	DISP	57	40	53	47	0
GERU	DISP	0	0	4	2	0
GOAL	POSS	11	11	12	12	0
HBCL	POSS	5	5	2	2	0
HBEF	POSS	73	73	81	81	0
HBIN	POSS	6	6	11	11	0
KIBU	BEHI	0	0	1	1	0
KICL	BEHI	0	0	0	0	0
KIIN	BEHI	0	0	1	1	0
KILA	BEHI	3	3	2	2	0
KILO	BEHI	3	0	9	7	0
KISE	BEHI	2	2	2	2	0
KISH	BEHI	7	7	1	1	0
KKCL	POSS	13	12	10	9	0
KKGK	POSS	2	1	3	2	0
KKIN	POSS	21	20	29	28	0
KKLA	POSS	17	16	12	11	0
KKLO	POSS	86	55	77	57	0
KKSH	POSS	58	51	48	36	0
MACO	DISP	18	18	14	14	0
MAUN	DISP	57	3	41	2	0
RUSH	DISP	3	3	3	3	0
Total		502	375	493	392	3

The adjusted occurrences are then used to calculate the number of times the ball has been in each state. For the state, 'DISP', the totals for both sides are combined, as it does not make sense to refer to Hawthorn or Port Adelaide 'in dispute', but instead that the ball is 'in dispute'. Therefore the 'DISP' total for the match is 202. Using the transitions from Table1, the totals for each transition are calculated and both of these totals are used to calculate the relevant transition probability. Table 3 contains the transition probabilities for this match as well as the relevant totals used to calculate these probabilities.

Table 3: Transition probabilities and totals for round 18 game: Hawthorn v. Port Adelaide

Transition	Hawthorn			Port Adelaide		
	State Total	Transitions	Probability	State Total	Transitions	Probability
BEHI DISP	12	0	0.0000	14	9	0.6429
BEHI POSS	12	12	1.0000	14	5	0.3571
DISP BEHI	202	3	0.0149	202	3	0.0149
DISP POSS	202	90	0.4455	202	106	0.5248
GOAL DISP	23	3	0.1304	23	3	0.1304
GOAL POSS	23	9	0.3913	23	11	0.4783
POSS BEHI	261	11	0.0421	258	9	0.0349
POSS DISP	261	82	0.3142	258	98	0.3798
POSS GOAL	261	11	0.0421	258	12	0.0465
POSS OVER	261	17	0.0651	258	11	0.0426
POSS POSS	261	140	0.5364	258	128	0.4961

Using the values from the probability columns gives the transition matrix shown in Table 4.

Table 4: Transition matrix for round 18 game: Hawthorn v. Port Adelaide

	State Description	1	2	3	4	5	6	7
		Ha Poss	PA Poss	Dispute	Ha Goal	PA Goal	Ha Behi	PA Behi
1	Ha Poss	0.536	0.065	0.314	0.042	0	0.042	0
2	PA Poss	0.043	0.496	0.38	0	0.047	0	0.035
3	Dispute	0.446	0.525	0	0	0	0.015	0.015
4	Ha Goal	0.391	0.478	0.13	0	0	0	0
5	PA Goal	0.391	0.478	0.13	0	0	0	0
6	Ha Behi	0	0.357	0.643	0	0	0	0
7	PA Behi	1.000	0	0	0	0	0	0

This matrix of probabilities can then be powered up to ascertain the limiting distribution for this match. Once the limiting distribution is derived, the probabilities for each state can be applied to the number of transitions, in this case 770, for the match to calculate the expected number of events from each of the seven states. Table 5 contains the limiting probabilities for each state and the expected number of events for each state. The actual numbers for this match have also been included for comparison.

Table 5: Limiting distribution for round 18 match between Hawthorn and Port Adelaide

State	Hawthorn			Port Adelaide		
	Limiting Prob.	**Expected**	**Actual**	**Limiting Prob.**	**Expected**	**Actual**
Behind	0.0179	13.8	14	0.0159	12.2	12
Dispute	0.2529	194.7	202	0.2529	194.7	202
Goal	0.0141	10.9	11	0.0162	12.5	12
Possession	0.3349	257.9	261	0.3481	268.0	258

These totals were relatively close to the actual totals indicating that there is merit in the model and the assumptions that have been made are legitimate. In fact, the projected score for this game is 79 points for Hawthorn and 87 points for Port Adelaide. The actual score in the match was Hawthorn 80 points and Port Adelaide 84 points. As this is only one game, the signed errors were investigated for the whole season.

4. Results

The expected totals for each match were compared to the actual totals to investigate model accuracy. The resulting error was calculated by subtracting the expected total from the observed total. These errors were then averaged for the season for each state to gain an understanding of the accuracy of the model. It was hoped that confidence intervals for the error from each state would contain zero, indicating that the model has the potential to exactly reflect what had taken place in the game. The summary statistics for the error associated with each state are presented in the Table 6.

Table 6: Summary statistics for the model error from each state

State	n	Mean	Std. Dev.	95% C. I.
Behind	370	-0.003	0.3	[-0.04, 0.03]
Dispute	185	-0.9	8.9	[-2.22, 0.35]
Goal	370	0.002	0.5	[-0.05, 0.05]
Possession	370	0.5	10.7	[-0.63, 1.56]

It is evident from Table 6 that the mechanics of the model perform very well when compared to actual occurrences. In each case the absolute value of the mean error is less than one and for scoring events the errors are virtually zero. An average negative signed error indicates that the expected number of observations is less than the actual number whilst an average positive error reflects observed occurrences being greater than expected occurrences. The mean error of –0.9 for dispute suggests that on average, the model is underestimating the number of disputed occurrences. On the other hand, the model overestimates the number of possessions each team gets in a game.

This scenario is most likely to result due to an inability to accurately distinguish each individual occurrence. It is hoped that in time the data manipulation processes will allow the model to eliminate every superfluous occurrence from analysis. Furthermore, any lack of fit may also be due to the fact that the underlying physical process of an AFL game is not a simple first order Markov process. Transition probabilities may depend on the previous two or three possession types. 95% Confidence intervals were derived for the error from each state to ascertain whether an error of zero could be reasonably expected. These confidence intervals all contain zero giving some indication that the theory behind the model has a sound

basis in terms of reflecting the events of the match. It is hoped that further refinements in cleansing the data will result in the errors being reduced further.

5. Applications

All applications included in this paper are concerned solely with post match analysis. Several applications of the model will be discussed, however, these are in no way an exhaustive list. It is believed that the model would be useful to coaches and technical staff alike for pinpointing which match occurrences have the greatest impact on chance of victory. For example, one would be able to adjust the match statistics (e.g. halve clanger kicks and adjust long or short kicks accordingly) to identify by how much poor disposal and ball retention has contributed to a loss. The results may then enable a team to better focus their resources to ensure their match tactics and set-up maximise their chance of victory.

Similarly media outlets may be interested in identifying so called 'match turning' events. By taking the game at its present state and scenario, the offending event could be altered to see what effect this has on the ultimate result. Such an approach would be best suited to end game scenarios with relatively few transitions remaining. A direct extension of this would be to reproduce the work of Clarke and Norman [10] relating to when it is advantageous to add to your opponent's score by rushing a behind. As the transition probabilities for this model are based on actual data it would be more realistic than their work.

Simulation would be required to carry out the above applications and it is also believed that simulating matches would be beneficial in identifying just how random the results of a game of football can be. It is believed that the general public and elements of the media are relatively uninformed when it comes to the variability implicit in a game of AFL football. Statements such as 'they were lucky', 'we were robbed' or 'everything is going right' would be able to be justified or disputed by simulating games and observing how variable the final margin can be with the same transition probabilities. Final season competition ladders could be re-calculated and a team's finishing performance quantified by how lucky or unlucky they may have been during the season.

Finally, the transition probabilities for each club from this season can be investigated with a view to identifying where powerful teams derive their success. Conversely for the perceived weaker teams their shortcomings can be identified in a similar manner. This analysis could give an informed and quantitative view of which occurrences in a match are beneficial and detrimental to the final result and where resources can be best directed.

6. Extensions

At the moment such a model is only useful for post game analysis since all input data is only gleaned after the event. The model could be used for forecasting by predicting the input data prior to each match. This information could then be used to assign winning probabilities and projected scores for each match. Furthermore disposals and possession types could be calculated for each team from each match with the information able to be used for betting purposes.

In attempting to best predict the input data one has to be mindful of whether it is best to predict the raw input statistics such as long kicks, or the teams' transition probabilities. It may be the case that teams maintain consistent transition probabilities regardless of how many disposals and possessions they have had in a match or the volume of transitions in a match

affects the transition probabilities. Either way, it is hoped that a prediction technique can be implemented, allowing the model to be used for forecasting.

A further extension of using the model for prediction purposes would be to ascertain whether a dynamic model would outperform a pre-match static model. It may be the case that at half time, the events of the first half could be used to better approximate what takes place in the rest of the match as opposed to keeping the initial predictions gleaned before the start of the match. Such an approach could change AFL prediction by updating live as events take place on the field.

It is intended to use the recording of statistics by CD according to location to increase the number of states in the model. These extra states would be dependent upon the location of the ground and would give a more informed view of which teams are good in attack/defence and which teams struggle. Furthermore, it could be utilised to identify the important characteristics for various locations on the ground that either contribute to or reduce a team's chances of victory. It may also aid teams at the selection table by allowing them to place players in the right position consistent with where they perform best.

A second-order Markov process could provide an improved fit to the data by including more transition probabilities. After further investigation of better data coding techniques, the trade off between adding more transition probabilities to the model and how much the fit could be improved will be assessed.

7. Conclusion

A Markov process model using AFL match statistics as input data provides an excellent approximation to actual events of a match. The model can be used for numerous post-match analysis purposes in its present state, which would be beneficial to teams, the media and the public. It is anticipated that the model can be refined further by adding extra states according to position on the ground before developing it into a dynamic prediction model that uses best estimates of input data.

References

1. Stefani, R.T., Football and basketball predictions using least squares. *IEEE Transactions on systems, man, and cybernetics*, 1977. **7**: p. 117-121.

2. Bailey, M., Identifying arbitrage opportunities in AFL betting markets through mathematical modelling, in *Proceedings of the Fifth Australian conference on Mathematics and Computers in Sport*, G. Cohen and T. Langtry, Editors. 2000, University of Technology Sydney: Sydney. p. 37-42.

3. Bailey , M. and S.R. Clarke, Predicting the Brownlow medal winner, in *Proceedings of the sixth Australian conference on mathematics and computers in sport*, G. Cohen and T. Langtry, Editors. 2002, University of Technology, Sydney: Gold Coast. p. 56-62.

4. Norman, J.M. Markov Process Applications in sport. in *IFORS conference*. 1999. Beijing China.

5. Bellman, R., Dynamic Programming and Markovian decision processes, with application to baseball, in *Optimal strategies in sports*, S.P. Ladany and R.E. Machol, Editors. 1977, North Holland: Amsterdam. p. 77-85.

6.   Clarke, S.R. and J.M. Norman, Dynamic Programming in cricket: Protecting the weaker batsman. *Asia Pacific Journal of Operational Research*, 1998. **15**: p. 93-108.

7.   Bukiet, B., E. Harold, and J. Palacious, A Markov chain approach to baseball. *Operations Research*, 1997. **45**(1): p. 14 - 23.

8.   Hirotsu, N. and M. Wright, A Markov chain approach to optimal pinch-hitting strategies in a designated hitter rule baseball game. *Journal of the Operations Research Society of Japan*, 2003. **46**(3): p. 353 - 371.

9.   Hirotsu, N. and M. Wright, An evaluation of characteristics of teams in association football by using a Markov process model. *Journal of the Royal Statistical Society: Series D*, 2003. **52**(4): p. 591 - 602.

10.   Clarke, S.R. and J.M. Norman, When to rush a behind in Australian Rules football: a Dynamic Programming approach. *Journal of the Operational Research Society*, 1998. **49**(5): p. 530-536.

Augmented Reality User Interfaces
for
Biomechanical Data Overlays

Richard Green

Computer Science, University of Canterbury,
Christchurch 8020, New Zealand
richard.green@canterbury.ac.nz

Abstract

Biomechanical analysis and sport coaching aids need to automatically overlay (augment) components of human body movement onto video data without a need for joint markers or set-up procedures. This paper presents a Continuous Human Movement Recognition (CHMR) system for recognising a large range of specific sport skills from continuous 3D full-body motion. A new methodology defines an alphabet of dynemes, units of full-body movement, to enable temporal segmentation and recognition of diverse skills. Using multiple Hidden Markov Models, the CHMR system attempts to infer the sport skill that could have produced the observed sequence of dynemes. This approach enables the CHMR system to track, quantify, recognise and augment hundreds of full-body sport skills from gait to twisting somersaults. This extends the sport coach-computer user interface to automatically augment recognised full-body sport skills with overlaid biomechanical data.

1 Introduction

Research and commercial interest in the design and development of Augmented Reality User Interfaces (ARUIs) for sport is growing rapidly. This paper provides a technical analysis of the problems associated with accurately recognising and tracking body and limb position, and describes a Continuous Human Movement Recognition (CHMR) system that is based on a generalisable "dyneme" model of human movement. Although the focus is on the technical challenges of recognising movement, doing so is the first component in many of the new breed of ARUIs for coaching sport. Beyond merely overlaying biomechanical data by tracking joint angles, this research extends ARUIs to more effectively support coaches by recognising skills from continuous full body motion. This skill aware ARUI now knows when each skill begins and ends with such temporal segmentation enabling recognition of hundreds of skills to progress ARUIs toward automated augmented coaching and skill analysis.

The CHMR framework (Figure 1) forms a basis for recognising and understanding full-body movement in 3D for a large diverse range of sport skills.
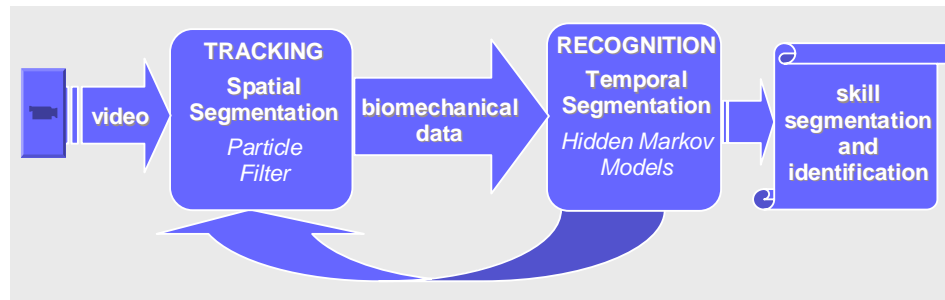
Figure 1. Overview of the continuous human movement recognition framework for a biomechanics ARUI.

Human movement is commercially[1] tracked by requiring subjects to wear joint markers/identifiers, an approach with has the disadvantage of significant set up time. Such an invasive approach to tracking has barely changed since it was developed in the 1970s. Using a less invasive approach free of markers, computer vision research into tracking and recognising full-body human motion has so far been mainly limited to gait or frontal posing [21]. Various approaches for tracking the whole body have been proposed in the image processing literature using a variety of 2D and 3D shape models and image models as listed in Table 1.

---

[1] Commercially available trackers are listed at www.hitl.washington.edu/scivw/tracker-faq.html

Authors	Shape model	Image model
Hogg	Cylinders	Edge
Rohr	Cylinders	Edge
Gavrila & Davis	Superquadrics	Edge
Drummond & Cipolla	Conics	Edge
Goncalves et al.	Cones	Edge
Kakadiaris & Metaxas	Deformable	Edge
Wren & Pentland	2D colour blobs	Skin colour blobs
Ju et al.	Patches (2D)	Flow
Bregler & Malik	Cylinders	Flow
Wang et al.	Cylinders	Flow
Cham & Rehg	Patches (2D)	Template
Wachter & Nagel	Cones	Flow + Edge
Plänkers & Fua	Deformable	Silhouette + Disparity
Deutscher et al.	Cones	Edge + Silhouette
Brand	Outline	Silhouette moments
Rosales & Sclaroff	Outline	Silhouette moments
Liebowitz & Carlsson	Outline	Hand-marked joints
Taylor	Outline	Hand-marked joints
Leventon & Freeman	Outline	Hand-marked joints

Table 1   Comparison of different human body models.

Computer-human ARUIs for sport will become increasingly effective as computers more accurately recognise and understand full-body movement in terms of specific skills. Stokoe began recognising human movement in the 1970s by constructing sign language gestures (signs) from hand location, shape and movement and assumed that these three components occur concurrently with no sequential contrast (independent variation of these components within a single sign). Ten years later Liddel and Johnson used sequential contrast and introduced the movement-hold model. In the early 1990s Yamato et al began using HMMs to recognise tennis strokes. Recognition accuracy rose as high as 99.2% in Starner and Pentland's work in 1996. Constituent components of movement have been named cheremes [33], phonemes [37] and movemes [3].

Author   - Recognition approach
Stokoe 78 – transcription: sign = location (tab) + hand shape (dez) + movement (sig)
Tamura and Kawasaki 88 - cheremes to recognise 20 Japanese signs
Liddell and Johnson 89 - use sequences of tab,dez,sig => Movement-Hold model
Yamato, Ohya and Ishii 92 - HMM recognises 6 different tennis strokes for 3 people
Schlenzig, Hunter and Jain 94 - recognises 3 gestures, *hello, goodbye* and *rotate*
Waldron and Kim 95 - ANN recognises small set of signs
Kadous 96 - recognises 95 Auslan signs with data gloves – 80% accuracy
Grobel and Assam 97 - ANN recognise finger spelling - 242 signs, coloured gloves
Starner and Pentland 96 - HMM recognises 40 signs in 2D, constrained grammar
Nam and Wohn 96 - HMM very small set of gestures in 3D – movement primes
Liang and Ouhyoung 98 - 250 Taiwanese signs, continuous, temporal discontinuities
Vogler & Metaxis 97 - HMM continuous 53 signs, models transitions between signs
Vogler & Metaxis 98 - HMM continuous 53 signs, word context with geometrics
Vogler & Metaxas 99 - 22 signs, define tab,dez,sig as phonemes - one hand

Table 2   Human movement recognition research.

As can be seen from Table 2, most movement recognition research has been limited to frontal posing of a constrained range of partial-body motion usually only tracking the 2D location of one hand. By contrast, this paper describes a computer vision based ARUI framework that recognises continuous full-body motion of hundreds of different sport skills. The full-body sport skills in this study are constructed from an alphabet of 35 *dynemes* – the smallest contrastive dynamic units of human movement. Using a novel framework of multiple HMMs the recognition process attempts to infer the human movement skill that could have produced the observed sequence of dynemes.

Because the temporal variation of related joints and other parameters also contains information that helps the recognition process infer dynemes, the system computes and appends the temporal derivatives and

second derivatives of these features to form the final *motion vector* for each video frame. Hence the motion vector includes joint angles (32 DOF), body location and orientation (6 DOF), centre of mass (3 DOF), principle axis (2 DOF) all with first and second derivatives.

2. Recognition

To simplify the design, it is assumed that the CHMR system contains a limited set of possible human sport skills. This approach restricts the search for possible skill sequences to those skills listed in the skill model, which lists the candidate skills and provides dynemes – an alphabet of granules of human motion – for the composition of each skill. The current skill model contains hundreds of skills where the length of the skill sequence being performed is unknown. If *M* represents the number of human movement skills in the skill model, the CHMR system could hypothesise M^N possible skill sequences for a skill sequence of length *N*. However these skill sequences are not equally likely to occur due to the biomechanical constraints of human motion. For example, the skill sequence *stand jump lie* is much more likely than *stand lie jump* (as it is difficult to jump while lying down). Given an observed sequence of motion vectors y_1^T, the recognition process attempts to find the skill sequence $\hat{s}_1^N$ that maximises this skill sequence's probability:

$$\hat{s}_1^N = \arg\max_{s_1^N} p(s_1^N / y_1^T) \equiv \arg\max_{s_1^N} p(y_1^T \mid s_1^N) \, p(s_1^N) \tag{1}$$

This approach applies Bayes' law and ignores the denominator term to maximise the product of two terms: the probability of the motion vectors given the skill sequence and the probability of the skill sequence itself. The CHMR framework described by this equation is illustrated below in Figure 2 where, using motion vectors from the tracking process, the recognition process uses the dyneme, skill, context and activity models to construct a hypothesis for interpreting a video sequence.
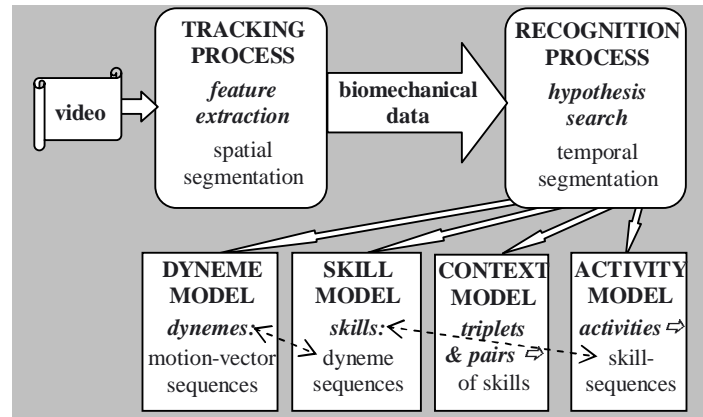
Figure 2   Continuous Human Movement Recognition system for a biomechanics ARUI. The dyneme, skill, context and activity models construct a hypothesis for interpreting a video sequence.

In the tracking process, motion vectors are extracted from the video stream. In the recognition process, the search hypothesises a probable movement skill sequence using four models:

- the *dyneme model* models the relationship between the motion vectors and the dynemes.
- the *skill model* defines the possible movement skills that the search can hypothesise, representing each movement skill as a linear sequence of dynemes;
- the *context model* models the semantic structure of movement by modelling the probability of sequences of skills simplified to only triplets or pairs of skills as discussed in Section 2.3 below.
- The *activity model* defines the possible human movement activities that the search can hypothesise, representing each activity as a linear sequence of skills (not limited to only triplets or pairs as in the context model).

Three principle components comprise the basic hypothesis search: a dyneme model, a skill model and a context model.

As the phoneme is a phonetic unit of human speech, so the dyneme is a dynamic unit of human motion. The word *dyneme* is derived from the Greek *dynamikos* "powerful", from *dynamis* "power", from *dynasthai* "to be able" and in this context refers to motion. This is similar to the phoneme being derived from *phono* meaning sound and with *eme* inferring the smallest contrastive unit. Thus *dyn-eme* is the smallest

contrastive unit of movement. The sport skills in this study are constructed from an alphabet of 35 dynemes which HMMs use to recognise the skills. This approach has been inspired by the paradigm of the phoneme as used by the continuous speech recognition research community where pronunciation of the English language is constructed from approximately 50 phonemes

<div align="center">a   b</div>
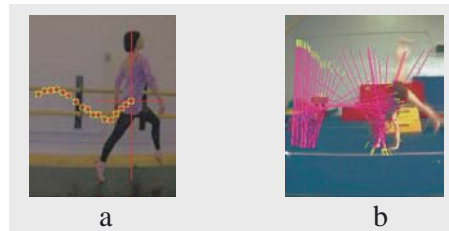
Figure 4. Biomechanics ARUI overlaying COM parameters during running and principle-axis parameters through a cartwheel.

A centre-of-mass (COM) category of dyneme is illustrated in the biomechanics ARUI in Figure 4a where each running step is delimited by COM minima. A full 360° rotation of the principle axis during a cartwheel in Figure 4b illustrates a rotation dyneme category.

3. Performance

Hundreds of skills were tracked and classified using a 1.8GHz, 640MB RAM Pentium IV platform processing 24 bit colour within the Microsoft DirectX 9 environment under Windows XP. The video sequences were captured with a JVC DVL-9800 digital video camera at 30 fps, 720 by 480 pixel resolution. Each person moved in front of a stationary camera with a static background and static lighting conditions. Only one person was in frame at any one time. Tracking began when the whole body was visible which enabled automatic initialisation of the clone-body-model.

The skill error rate quantifies CHMR system performance by expressing, as a percentage, the ratio of the number of skill errors to the number of skills in the reference training set. Depending on the task, CHMR system skill error rates can vary by an order of magnitude. The CHMR system results are based on a set of a total of 840 sport movement patterns, from walking to twisting saltos. From this, an independent test set of 200 skills were selected leaving 640 in the training set. Training and testing skills were performed by the same subjects. These were successfully tracked, recognised and evaluated with their respective biomechanical components quantified where a skill error rate of 4.5% was achieved.

| A: stretch & step | B: quarter rotation (unknown skill) | C: half rotation (cartwheel identified) |
| D: finished cartwheel | E: right step | F: left step |
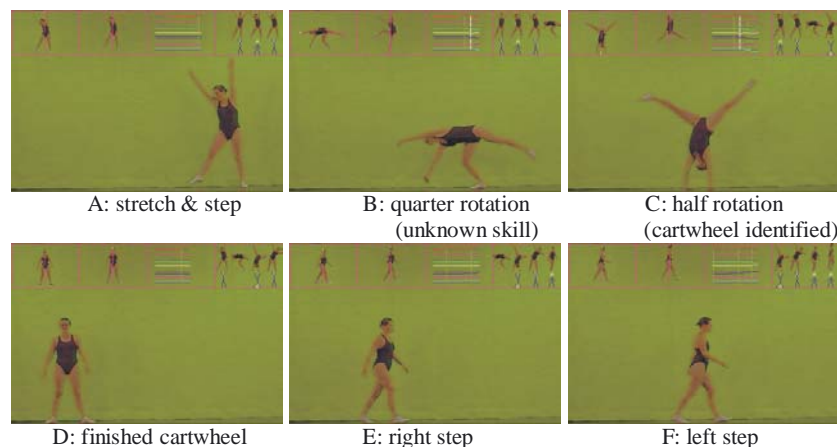
Figure 5 A ARUI coaching system recognising stretching into a cartwheel followed by steps.

In Figure 5, an ARUI augments each picture with four overlaid tiles displaying CHMR processing steps:
- Tile 1: Principle axis through the body.
- Tile 2: Body frame of reference (normalised to the vertical).
- Tile 3: Motion vector trace (subset displayed).
- Tile 4: Recognising *step*, *stretch* and *cartwheel* indicated by stick figures with respective snapshots of the skills.

This coach ARUI system is recognising the sequence of skills *stretch & step, cartwheel, step* and *step* from continuous movement. As each skill is recognised, the scene is augmented with a snapshot of the

corresponding pose displayed in the fourth tile. Below each snapshot is a stick figure representing an internal identification of the recognised skill. Notice that the cartwheel is not recognised after the first quarter rotation. Only after the second quarter rotation is the skill identified as probably a cartwheel. This ARUI system enables, for example, a list of all cartwheels across training sessions to be compared for tracking progress. Figure 4 illustrates another ARUI coaching system augmented by biomechanical data.

Recognition was processed using the (Microsoft owned) Cambridge University Engineering Department HMM Tool Kit (HTK) with 96.8% recognition accuracy on the training set alone and a more meaningful 95.5% recognition accuracy for the independent test set where H=194, D=7, S=9, I=3, N=200 (H=correct, D=Deletion, S=Substitution, I=Insertion, N=test set, Accuracy=(H-I)/N). 3.5% of the skills were ignored (deletion errors) and 4.5% were incorrectly recognised as other skills (substitution errors). There was only about 1.5% insertion errors – that is incorrectly inserting/recognising a skill between other skills.

The HTK performed Viterbi alignment on the training data followed by Baum-Welch[2] re-estimation with a context model for the movement skills. Although the recognition itself was faster than real-time at about 120 fps, the tracking of 32 DOF with particle filtering was computationally expensive using up to 16 seconds per frame.

4. Conclusions

This paper has provided a technical analysis of the problems of accurately recognising and tracking body and limb position, and demonstrated a Continuous Human Movement Recognition based ARUI system for sport skill analysis and coaching that is based on a generalisable "dyneme" model of human movement. A ARUI system successfully recognised and augmented skills and activities with a 95.5% recognition accuracy to validate this CHMR framework and the dyneme paradigm.

However, the 4.5% error rate attained in this research is not yet evaluating a natural world environment nor is this a real-time system with up to 10 seconds to process each frame. Although this 95.5% recognition rate was not as high as the 99.2% accuracy Starner and Pentland [31] achieved recognising 40 signs, a larger test sample of 200 full-body skills were evaluated in this paper.

The following improvements seem most important for generalising to an ARUI system encompassing many partial body movement domains:

- Expand the dyneme model to improve discrimination of more subtle movements in partial-body domains. This could be achieved by either expanding the dyneme alphabet or having domain dependent dyneme alphabets layered hierarchically below the full-body movement dynemes.
- Enhance tracking granularity using higher resolution cameras with a higher frame rate.

Although the focus has been on the technical challenges of automating the recognition and augmenting of sport skills, doing so is the first component in many of the new breed of ARUIs for sport.

5. References

[1] N I Badler, C B Phillips, B L Webber, *Simulating humans*, Oxford University Press, New York, NY, 1993.
[2] M Brand, V Kettnaker, Discovery and segmentation of activities in video, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), August 2000.
[3] C Bregler, Learning and recognizing human dynamics in video sequences, *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 1997.
[4] C Bregler, J Malik, Tracking people with twists and exponential maps, *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 8–15, 1998.
[5] T J Cham, J M Rehg, A multiple hypothesis approach to figure tracking, *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 1: 239–245, 1999.
[6] T Drummond, R Cipolla, Real-time tracking of highly articulated structures in the presence of noisy measurements, *IEEE International Conference on Computer Vision*, ICCV, 2: 315–320, 2001.
[7] D M Gavrila, L A Davis, 3-D model-based tracking of humans in action: a multi-view approach, *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 73–80, 1996.
[8] L Goncalves, E Di Bernardo, E Ursella, P Perona, Monocular tracking of the human arm in 3D, *IEEE International Conference on Computer Vision*, ICCV, 764–770, 1995.
[9] K Grobel, M Assam, Isolated sign language recognition using hidden Markov models, *IEEE International Conference on Systems, Man and Cybernetics*, 162–167, Orlando, 1997.
[10] D Herbison-Evans, R D Green, A Butt, Computer Animation with NUDES in Dance and Physical Education, *Australian Computer Science Communications*, 4(1): 324-331, 1982.
[11] D C Hogg, Model-based vision: A program to see a walking person, *Image and Vision Computing,* 1(1): 5–20, 1983.

---

[2] An excellent discussion of HMMs and application of Viterbi alignment and Baum-Welch re-estimation can be found in the extensive HTK documentation of HTKBook: http://htk.eng.cam.ac.uk/docs/docs.shtml

[12] A Hutchinson-Guest, *Choreo-Graphics; A Comparison of Dance Notation Systems from the Fifteenth Century to the Present*, Gordon and Breach, New York, 1989.

[13] F Jelinek, *Statistical Methods for Speech Recognition*, MIT Press, Cambridge, Mass., 1999.

[14] S X Ju, M J Black, Y Yacoob, Cardboard people: A parameterized model of articulated motion, *IEEE International Conference on Automatic Face and Gesture Recognition*, 38–44, 1996.

[15] M W Kadous, Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language, *Workshop on the Integration of Gesture in Language and Speech*, WIGLS, 165–74, Applied Science and Engineering Laboratories, Newark, 1996.

[16] I Kakadiaris, D Metaxas, Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection, *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 81–87, 1996.

[17] M E Leventon, W T Freeman, Bayesian estimation of 3-d human motion from an image sequence, *TR–98–06*, Mitsubishi Electric Research Lab, Cambridge, 1998.

[18] R H Liang, M Ouhyoung, A real-time continuous gesture recognition system for sign language, *Third International Conference on Automatic Face and Gesture Recognition*, 558–565, Nara, 1998.

[19] S K Liddell, R E Johnson, American Sign Language: the phonological base, *Sign Language Studies*, 64: 195–277, 1989.

[20] D Liebowitz, S Carlsson, Uncalibrated motion capture exploiting articulated structure constraints, *IEEE International Conference on Computer Vision*, ICCV, 2001.

[21] T B Moeslund, E Granum, A survey of computer vision-based human motion capture, *Computer Vision and Image Understanding* 18: 231–268, 2001.

[22] Y Nam and K Y Wohn, Recognition of space-time hand-gestures using hidden Markov model, *ACM Symposium on Virtual Reality Software and Technology*, 1996.

[23] A Pentland, B Horowitz, Recovery of nonrigid motion and structure, *IEEE Transactions on PAMI*, 13: 730-742, 1991.

[24] S Pheasant, *Bodyspace. Anthropometry, Ergonomics and the Design of Work*, Taylor & Francis, 1996.

[25] R Plänkers, P Fua, Articulated soft objects for video-based body modeling, *IEEE International Conference on Computer Vision*, ICCV, 1: 394–401, 2001.

[26] J M Rehg, T Kanade, Model-based tracking of self-occluding articulated objects, *Fifth International Conference on Computer Vision*, 612-617, 1995.

[27] J Rittscher, A Blake, S J Roberts, Towards the automatic analysis of complex human body motions, *Image and Vision Computing*, 20(12): 905-916, 2002.

[28] K Rohr, Towards model-based recognition of human movements in image sequences, *CVGIP - Image Understanding* 59(1): 94–115, 1994.

[29] R Rosales, S Sclaroff, Inferring body pose without tracking body parts, *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 2000.

[30] J Schlenzig, E Hunter, R Jain, Recursive identification of gesture inputers using hidden Markov models, *Second Annual Conference on Applications of Computer Vision*, 187-194, 1994.

[31] T Starner, A Pentland, Real-time American Sign Language recognition from video using Hidden Markov Models, *Technical Report 375*, MIT Media Laboratory, 1996.

[32] W C Stokoe, *Sign Language Structure: An Outline of the Visual Communication System of the American Deaf*, Studies in Linguistics: Occasional Papers 8. Linstok Press, Silver Spring, MD, 1960. Revised 1978.

[33] S Tamura, S Kawasaki, Recognition of sign language motion images, *Pattern Recognition*, 31: 343-353, 1988.

[34] C J Taylor, Reconstruction of articulated objects from point correspondences in a single articulated image, *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 586–591, 2000.

[35] C Vogler, D Metaxas, Adapting hidden Markov models for ASL recognition by using three-dimensional computer vision methods, *IEEE International Conference on Systems, Man and Cybernetics*, 156–161, Orlando, 1997.

[36] C Vogler, D Metaxas, ASL recognition based on a coupling between HMMs and 3D motion analysis, *IEEE International Conference on Computer Vision*, 363–369, Mumbai, 1998.

[37] C Vogler, D Metaxas, Toward scalability in ASL recognition: breaking down signs into phonemes, *Gesture Workshop 99*, Gif-sur-Yvette, 1999.

[38] S Wachter, H Nagel, Tracking of persons in monocular image sequences, *Computer Vision and Image Understanding* 74(3): 174–192, 1999.

[39] M B Waldron, S Kim, Isolated ASL sign recognition system for deaf persons, *IEEE Transactions on Rehabilitation Engineering*, 3(3): 261–71, 1995.

[40] J Wang, G Lorette, P Bouthemy, Analysis of human motion: A modelbased approach, *Scandinavian Conference on Image Analysis*, SCIA, 2: 1142–1149, 1991.

[41] C Wren, A Azarbayejani, T Darrell, A Pentland, "Pfinder: Real-time tracking of the human body", *IEEE Transactions on PAMI*, 19(7): 780-785, 1997.

[42] J Yamato, J Ohya, K Ishii, Recognizing human action in time-sequential images using hidden Markov models, *IEEE International Conference on Computer Vision*, ICCV, 379-385, 1992.

COMPARISON OF THE METHODS TO RESET TARGETS FOR INTERRUPTED ONE-DAY CRICKET MATCHES

Manohita Gurram and S. Narayanan
Department of Biomedical, Industrial, and Human Factors Engineering
Wright State University, Dayton, Ohio, USA

Abstract

Target resetting in interrupted one day international cricket matches is an important issue in establishing fairness in the competition. The Duckworth Lewis method has been adopted by the International Cricket Council for this purpose and has been in vogue for over five years. A metric to measure the fairness of any method used to reset targets is defined. Using a historical data set, this paper compares and tabulates the value of the fairness metric for different methods proposed for resetting targets in one-day cricket matches. Results indicate that the existing methods have limited capability in generating a fair target. Various factors that potentially lead to this limitation are described in this article.

1. Introduction

Cricket is a very popular game in Africa (South Africa and Zimbabwe), Australasia (Australia and New Zealand), the Indian subcontinent (Bangladesh, India, Pakistan, Sri Lanka), and the West Indies. Mathematical modeling and operations research (OR) are being extensively used in sport in general, and cricket in particular. Mathematical modeling was used by Clarke and Allsopp [4] to statistically analyze the binary data from the matches and to quantify the magnitude of victory. Logistic regression was applied by de Silva, Pond, and Swartz [6] for estimating the effect of home team advantage in one day international cricket matches. Mathematical formulations and distribution functions are built by Carter and Guthrie [2] for the probability of wide ball, no ball, wickets and scoring of runs. Regression equations from the statistical analysis of the data set were used by Jayadevan [11] in resetting the target for rain interrupted matches. The application of OR in cricket dates back to 1990 when Mike Wright [14] used OR techniques for scheduling English umpires. Wright [16] has also developed methods to schedule the fixtures for the games using OR techniques such as tabu search, intensification and diversification. English Cricket Board (ECB) has been using Wright's methods for scheduling since 1990 [15, 17]. Apart from scheduling, OR is also used in developing methods for resetting the targets for rain interrupted matches by Allsopp and Clarke [1] and Duckworth & Lewis [7, 8, 9]. Dynamic programming was used by Clarke [3] and Preston & Thomas [13] to assist in determining optimal scoring rate strategies and by Johnson, Clarke and Noble [12] in assessing comparative player performances between batting and bowling.

One-day matches are intolerant of the interruptions due to weather. If the match is interrupted by rain and there is not enough time for the two teams to play their initial assigned set of overs, suspending the match or continuing the match another day is not appropriate. So the number of overs and the target should be reduced appropriately. Since the basic reason why one–day matches have become popular is their ability to end in a decisive result, some rule must be applied to decide on the winner when there is an interruption. The advances in science and technology are used to make many of the important decisions in cricket with great precision. Decisions on run-out, stumping, leg before wicket, boundaries and catches are taken only after thoroughly analyzing the situations. When compared to these events, the target scores are much more important. There is therefore a need for a fair scientific method which tackles the target resetting problem when there is an interruption.

2. Existing Methods

2.1 A brief description

Over the years many rules for resetting the targets in interrupted matches have been proposed, and some have been put to use. A brief description of some of the methods is presented in the next section along with their merits and demerits.

2.1.1 Average Run Rate Rule

According to this rule, the winner is decided based on the comparison of the average run rates of the teams. Average run rate is the ratio of the total number of runs scored by a team to the number of overs in which the team scored the runs.

$$\text{Average Run rate} = \frac{\text{Total Score}}{\text{Total Overs}}$$

Merits:
1)   Very easy to apply.

Demerits:
1)   The influence of the number of wickets lost is not considered.
2)   Pattern of scoring of the runs by the teams is not taken into account.
3)   Favors the team batting second.

2.1.2 Most Productive Overs

According to this method, the target set for the team playing second is determined by totaling the runs scored by the team playing first in the same number of highest-scoring overs as the number of overs that the second team has been assigned.

Merits:
1)   The pattern of scoring of the runs by the team playing first is taken into account.

Demerits:
1)   Requires substantial amount of bookwork.
2)   Favors the team batting first.
3)   The influence of the wickets lost is not considered.

2.1.3 Discounted Most Productive Overs

According to this method, the total from the most productive overs is discounted by 0.5% for each over lost. Though this method slightly reduces the advantage given to the team batting first by the Most Productive Overs rule, it still has the same demerits as that of the Most Productive Overs rule.

2.1.4 Duckworth and Lewis method

This method considers the two-factor relationship between the proportion of total runs and the two resources, overs remaining and the wickets at hand. The average total score is fit to an exponential equation and the target is reset on that basis.

Merits:
1) This is the first scientific method that was applied in the case of target resetting.
2) This method is very easy to understand and to apply.
3) The method can be easily applied to all the different types of interruptions and for the matches with multiple interruptions too.

Demerits:
1) The method does not take into account the batting strategy of the teams and does not typically account for the general scoring pattern in the one day matches.
2) As the character of one-day games is changing, both in terms of players' skills at the game and the tactics that are adopted, the constants in the Duckworth Lewis formula need to be changed periodically.

2.1.5 Jayadevan Method

In this method mathematical models of the development of the innings are formed based on the regression analysis of a number of matches that have been played. A minimum percentage of runs to be achieved corresponding to the fall of wickets was obtained so that the effect of the fall of the wickets was also considered.

Merits:
1) This method can handle any number of interruptions at any point of the game.
2) This method considers the scoring pattern of the teams.

Demerits:
1) The calculations can get tedious and complicated if there are many interruptions.
2) This method is not sufficiently responsive to the fall of wickets.

2.2 Metric for comparison

There has been a lot of outcry about the fairness of the existing methods, including the Duckworth Lewis method which has been in use for over five years in international cricket. In many crucial matches where there was an interruption due to rain, these methods gave unexpected and unsatisfying results. Additionally, post-hoc analysis has shown that the target resetting methods that were in use had, or would have, affected the result in a drastic way. A couple of such instances are listed below:

In the 1992 World Cup semifinal match between South Africa and England, South Africa had four wickets in hand and required 22 runs of the final 13 balls to win the game. South Africa had a good chance of winning the game with their two batsmen at the crease in a comfortable position. There was a rain interruption at that stage and according to the Most Productive Overs rule that was in use, the target was set to 22 runs required off 1 ball. According to the rule, there was no change in the target as the two maiden overs in the England innings were reduced. The revised target was unachievable under any circumstances.

In the final of the 2003 World Cup between India and Australia, Australia set a target of 360 runs in 50 overs by scoring very quickly in the early stages. When rain threatened the match when India was 145 for 3 after 23 overs, India was only 5 runs behind the Duckworth Lewis target. Losing no more wickets and scoring 12 runs in the next two overs followed by a terminal downpour would have handed the title of World Cup champions to India. But rain was not in favor of India, and Australia went on to clinch the title by dismissing the Indians for 234.

Since the results in the matches like the ones illustrated above are very important, there is a need for a method which sets a fair revised target. Before attempting to propose new methods, it is important to systematically assess how fair the current methods have been in resetting targets. A fair method in our point of view is one which gives the same result regardless of an interruption. As we cannot prevent interruptions like rain or crowd disturbance, we need a method which will reduce the influence of these effects on the result of the match. The metric we have used to compare different methods is to determine if the winner that a method would declare, assuming there is an interruption, is the same as the winner at the end of an uninterrupted match.

2.3 Comparison of methods

For the purpose of comparison of the different methods, a data set which consisted of 377 one-day international tournament matches from 1998 - 2003 which had "no interruption" of any sort was considered. By "no interruption" we mean that each innings of the match was concluded only when the team batting had either finished their quota of overs or had lost all the wickets at their disposal. The details of these games were obtained from the archives of www.cricket.org [18], in the form of the ball-by-ball commentary. These files were stored in text format. A program was written in Java ^{TM} to read a set of text files and generate a scoreboard file for each. The scoreboard file contained the score and the wickets lost after each over in both the innings. These were taken as the input data for the different methods. A hypothetical interruption was assumed to occur at the end of an over and a method for target resetting was applied to find the winner at that time. Comparison was then made with the actual winner of the game. This was repeated with each of the five mentioned methods for all the matches in the data set.

Since a minimum of 25 overs have to be bowled to the side batting second before rain interruption rules can be applied, for each method a program was written to apply the method at the end of each over in the second innings from the 25th over to the end of the match. A comparison was made between the winner declared by the method to the original winner of the match and the number of mismatches was noted. The ratio of the number of mismatches to the number of overs subtracted from one was taken as a measure of the fairness of the method.

Fairness Metric =   1 -   Number of overs for which there is a mis-match
                          Number of overs used for comparison

If, for a few overs, the winner specified by the target resetting method is not the same as the actual match winner, then it may not be a serious concern. Therefore, a limit must be put on the number of overs for which there is a mismatch. For more than 20% of the overs considered i.e., five out of the 25 overs, if there is a mismatch then it can be taken as the failure of the method.

The type of interruption considered for evaluating the fairness of the methods is that the match stops within second innings after a certain over and no further play is possible. When the matches that have already been finished are considered, this is the only type of interruption which can be applied.

3. Results and discussion

The data set consisted of a set of 377 matches. In 325 of these matches, the second innings went beyond 25 overs. The percentages of the failure of the before mentioned methods when applied to this data set are tabulated below.

Table 1: Performance of the existing methods

	Method	Fairness Metric	Fairness metric when 20% of the mismatches are forgiven
1	Duckworth Lewis	0.705	0.762
2	Jayadevan	0.699	0.791
3	Most Productive Overs	0.603	0.606
4	Discounted Most Productive Overs	0.605	0.638
5	Average Run Rate Method	0.708	0.736

From the above table we can see that all the methods which have been proposed till now have low values for fairness metric. Thus, there is a need for a new method which has a higher fairness metric.

In order to devise a new method which will be fairer than the existing methods, an analysis was completed on the matches for which the existing methods would have failed. An attempt was made to ascertain the similarities in those matches which could have caused the failure of the methods. A set of 60 matches which would have failed for at least three of the methods were taken to observe patterns. A few patterns were proposed and they were tested for their presence in the failed matches. These patterns included:

1) The teams playing: Both the teams were evenly matched in their capability to win the match against each other.
2) Loss of wickets: Both the teams lost more than 5 wickets in their respective innings.
3) Batting line up: When separating the team into three groups, with the first three batsmen in the first group and the 4, 5, 6th batsmen in the second group and the last 5 batsmen in the third group. The average score of the second group or the third group was greater than or equal to the average score of the first group.
4) Fall of Wickets: For one of the teams, at least three wickets fell within the space of 20-25 runs before 45 overs.
5) Irregular bowler taking wickets: An irregular bowler took more than two wickets.
6) Short of bowlers: A irregular bowler got more than three overs.
7) Target: The target was higher than 300.
8) Match Conditions: The matches were played under floodlights.
9) Chasing: The team batting second was the winner.

Table 2: Percentage of matches which support the hypotheses

Patterns	Percentage of matches in which the patterns were present
Teams playing	78
Loss of wickets	70
Batting line up	56
Fall of Wickets	74
Irregular bowler	23
Short of bowlers	49
Target	2
Match conditions	43
Chasing	61

Table 2 shows that most of the patterns which are proposed could affect the failure of the methods. Some of the patterns like the teams playing, the loss of wickets, the batting line up, the match conditions, and chasing are method-specific. The proposed method should thus take those aspects into account. The fall of wickets dictates for a change in the play strategy of the teams so that rain does not affect the outcome of the result of the match. The short of bowlers factor recommends a change in the selection strategy of the team. The target factor was taken into account in a new method called the Duckworth Lewis Professional Edition. This method has a correction factor for matches which have targets higher than 300. Table 2 shows that, in our data set, only 2 of the matches had a target greater than 300 so this new method was not considered in our analysis of the existing methods. Table 2 also shows that there are factors other than the target which have greater influence on the fairness of the methods.

Our future work will focus on developing a new method by considering the salient aspects of the nine patterns. The new method will be tested with the fairness metric and an attempt will be made to develop a method which is fairer than any current method.

4. References

[1] P. Allsopp and S. R. Clarke, "Measuring the performance of teams in one-day cricket", Proceedings of the 5th Conference, Mathematics and Computer in Sport, (2000), 28-36.

[2] M. Carter and G. Guthrie, "Cricket interruptions: fairness and incentive in interrupted cricket matches", Proceedings of the 6th Conference, Mathematics and Computers in Sport, (2002).

[3] S. R. Clarke, "Dynamic programming in one-day cricket-optimal scoring rates", Journal of Operational Research Society, (1998), **39**, 331-337.

[4] S. R. Clarke and P. Allsopp, "Fair Measures of performance: the World Cup of cricket", Journal of the Operational Research Society, (2001), **52,** 471-479

[5] S. R. Clarke and P. Allsopp, "Response to Duckworth and Lewis, comment on Clarke and Allsopp (2001). Fair measures of performance: The World Cup of Cricket", Journal of the Operational Research Society, (2002), **53,** 1160-1161

[6] B. M. de Silva, G. R. Pond and T. B. Swartz, "Estimation of the magnitude of the victory in one-day cricket", Proceedings of the 5th Conference, Mathematics and Computer in Sport, (2000).

[7] F. C. Duckworth and A. J. Lewis, "A fair method for resetting the target in interrupted one –day cricket matches", Proceedings of the 3rd Conference, Mathematics and Computers in Sport, (1996), 51-68.

[8] F. C. Duckworth and A. J. Lewis, "A fair method of resetting the target in interrupted one-day cricket matches", Journal of the Operational Research Society, (1998), **49**, 220-227.

[9] F. C. Duckworth and T. Lewis, "Your comprehensive guide to the Duckworth/Lewis method of target resetting in one-day cricket", University of the West of England, Bristol, (1999).

[10] F. C. Duckworth and A. J. Lewis, "Comment on Clarke and Allsopp (2001). Fair measure of performance: The World Cup of Cricket", Journal of the Operational Research Society, (2002), **53,** 1159-1160

[11] V. Jayadevan, "A new method for the computation of target scores in interrupted, limited – over cricket matches", Current Science, (2002), **83**, 577-586.

[12] M. I. Johnston, S. R Clarke and D. H. Noble, "Assessing player performance in One Day Cricket using Dynamic Programming", Asia-Pacific Journal of Operational Research, (1993), **10**, 45-55.

[13]   I. Preston and J. Thomas, "Batting Strategy in Limited Overs Cricket", Journal of the Royal Statistical Society, (2000), **49**, 95-106.

[14]   M. B. Wright, "Scheduling English cricket umpires", Journal of the Operational Research Society, 1991, **42**, 447-452.

[15] M. B. Wright, "A fair allocation of county cricket opponents", Journal of the Operational Research Society, 1992, **43**, 195-201.

[16] M. B. Wright, "Timetabling county cricket fixtures using a form of tabu search", Journal of the Operational Research Society, 1994, **45**, 758-770.

[17] M. B. Wright, "School timetabling using heuristic search", Journal of the Operational Research Society, 1996, **47**, 347-357.

[18] www.cricket.org/link_to_database/ARCHIVE/

USING PERFORMANCE ANALYSIS TECHNOLOGY TO EVALUATE THE INSTRUCTIONAL PROCESS IN SPORT

John Hammond,

School of Education,
Southern Cross University, PO Box 157,
Lismore, New South Wales,
Australia.

Abstract

This paper examines two case studies in sport. Firstly, in the teaching of swimming, comparisons are made between instructors' perceived importance for facets of swimming teaching and time allocated during a lesson. Secondly, two soccer coaching accreditation courses are analysed, to provide a comparison of the aims of the instructor coach and events taking place during the course itself. Each study employed computer and video analysis in combination with other approaches, as part of an holistic aproach to evaluation. The studies demonstrated that performance analysis techniques, normally employed for the enhancement of athlete performance, can be a useful tool in assessing instructional processes in sport. Additionally, in the conference presentation itself there will be a further description of the performance analysis tools.

1. Introduction

In recent years technology has become more important and valuable in the analysis and development in sport, for the purpose of increasing performance levels [9]. In 1983 Franks et al. [6] argued that technology such as this would ultimately refine the analysis of sport through improved manipulation and interpretation of data. Their prediction has been largely correct, as computerised systems of the notation of events occurring in sport (notational analysis), to aid in the analysis of performance, have now been developed throughout the world [1].

Notational analysis has been described as a computer-video interactive system that enables an observer to systematically and objectively record and view the performance and behaviour of athletes within team sports. This system allows for viewing and analysis of selected video instances at a time convenient to coaches or others who may have a purpose in viewing such a recording [8]. Original notation systems in sport were prepared by hand with little technical assistance [12]. This type of notational analysis varied considerably, depending on the sport and the specific aims of any single observation. As Hughes [16] indicated, forms of hand notation are generally accurate, however, they tend to have more disadvantages than the more modern forms of analysis. The primary disadvantage of hand notation is time, not only to learn how to use set systems, but also the time taken in processing the information into understandable and relevant material for the coach, athlete or scientist [16]. Additionally, this type of analysis is relatively subjective [5].

Hughes [14] stated that there were four main purposes of modern notational analysis. These being to: analyse the movements of athletes; evaluate tactics; evaluate technique; and, compile statistics related to a sport. Franks et al. [6] claimed that these purposes could be extended to include: the provision of immediate feedback; developing databases; indicating areas which require improvement; evaluation; and, acting as a mechanism which would enable selective searching through video recordings games. These are all functions that have potential to aid the coaching process. Furthermore, Hughes [15] noted that research of McDonald [18] and Franks and Miller [7], devoted to establishing a definite need in sport for more objective forms of analysis, stressed the importance this may have on coaching practices.

Coaches can seldom assess a player's performance without additional technical assistance. It is impossible for one person to either assess or remember the involvement of individuals in the constantly changing environment of a playing field during complex team games [23]. Notational analysis of sporting performances has been recognised as having the capability to fill an important role in supporting the coaching process [4]. Notational analysis even though it has been the vehicle for considerable research, was founded in practical applications [17]. However, this type of analysis is still yet to be widely adopted in any systematic fashion for the evaluation of instructional processes in sport [10]. Preliminary research and development by Dickson et al. [3] and Hammond et al. [11], proposed computer assisted learning in both soccer and golf respectively, using a computerised feedback system. Begbie and Hammond [2] reported on a customised event-recording system in conjunction with video footage, to evaluate the effectiveness of swimming instruction. Whilst Sproule et al. [21] combined computer and video technology to aid in the evaluation of trainee physical education teachers during sport lessons. Further, Perry and Hammond [20] have adapted a commercially produced performance analysis system to assist in the evaluation of coach education practices.

This conference paper will describe two case studies which have investigated instructional processes in sport using performance analysis techniques. The purpose of the paper is to illustrate how these techniques can be used in different settings suggesting that further development of the application of computer and video technology would be beneficial to sport education in the future. Although each study used a combination of data collection methods and analyses, for the context of this paper reporting of their result will focus more on the findings from the video/computer data. Additionally, in the conference presentation itself there will be further and more graphic descriptions of the performance analysis tools being employed in the field. This will include a demonstration, using brief coaching video footage, of how performance analysis technology, can be used to provide statistical evidence of the effectiveness of instruction.

2. Case Study 1 - Swimming Teaching

This study was designed to evaluate the skills and understandings that AUSTSWIM qualified instructors believe are most important and the relevance of these to practices of an exemplar swimming and water safety school. Two primary groups of participants were involved: instructors based in country areas of New South Wales (NSW); and, instructors and children associated with an exemplar swim-teach program. The study progressed through four phases that resulted in both qualitative and quantitative data being collected. To gain the required data, three research instruments were employed. These included postal questionnaires, a computerised event-analysis program and interview schedules.

Questionnaires were administered to a selection of AUSTSWIM qualified instructors. The questionnaires were aimed at obtaining instructors' perceptions and attitudes regarding the skills and understandings taught within swimming and water safety programs. For the purpose of gaining deeper insight into the attitudes of swimming instructors (to be observed at an exemplar program), it was decided that interview schedules would be designed. In addition, children who participated in lessons and one of the owners of the swimming school were interviewed. The interviews were formally structured and recorded on a cassette tape so they could be analysed at a later time.

In order to collect data related to the content of swimming lessons within the exemplar program, an analysis program was designed to record lesson events as they occurred. It was decided that computerised event analysis would be the most efficient method of data collection for the study, as not only would it allow for collection of data during 'real time' but it would also be more accurate and faster than hand forms of recording. In addition, the use of the computer program meant that the same process was strictly followed for the collection of data for each swimming lesson, overcoming the subjective nature of hand recording. Therefore, a customised event analysis program was designed to collect comparative data to that obtained from the questionnaires. This method of data collection maintained several advantages for the study as it meant that swimming lessons could be observed and individual instructors actions recorded as the lesson took place.

The study required the use of a video camera (Sony Video 8 Handycam), monitor (National Panasonic TC 800 EU) and laptop computer (Macintosh Power Book 500, Power PC) – see Figure 1. To run the customised video/computer program, two people were required. This included one person to input data into the computer and an additional person to manipulate the video camera. The computer program used during the study was developed to identify how much time is being spent on specific tasks during swimming and water safety lessons. Initially, the hypercard grid did not contain an event square to represent kicking, this was amended with squares being placed on the grid to represent front and back kicking. In addition, squares were developed for skills including general movement through water, submerging, instruction and waiting. All event squares originally placed on the grid remained. It was decided that, as during a pilot study there was ample time to record events, additional event squares would not cause a great problem.

QuickTime™ and a
Photo - JPEG decompressor
are needed to see this picture.

Figure 1: Collecting the Data at the Poolside

The final hypercard grid, as set out on the computer keyboard, can be seen in Figure 2. The computer program was used to record events during nine lessons conducted by three different instructors (three lessons each). Each of the lessons involved the teaching of two or three children.

1 A Surv. Begin	7 G Dive Begin	13 M Breast Begin	19 S F.Kick Begin	25 5 Movement
2 B Survival End	8 H Dive End	14 N Breast End	20 T F.Kick End	26 6 Eyes
3 C	9 I	15 O	21 1	27 7

Float Begin	Free Begin	Fly Begin	Enter	Waiting
4 D Float End	10 J Free End	16 P Fly End	22 2 Exit	28 8 Instruction
5 E Safety Begin	11 K Back Begin	17 Q B.Kick Begin	23 3 Face	29 9 Misc.
6 F Safety End	12 L Back End	18 R B.Kick End	24 4 Submerge	30 Z Error 1

Figure 2: Grid Designed for Notational Analysis

The questionnaire results illustrated which skills and understandings, in the opinion of AUSTSWIM trained swimming instructors, are most important in the instruction of beginner swimmers. To identify the most important skills and understandings in a manner that allowed additional comparison to data generated from the event analysis responses were placed into five categories, based on previous studies conducted by Harrod and Langendorfer [13]. Following this, it was concluded that the instructors who participated in the study felt that Buoyancy skills and Water Adjustment skills are the most important skills for beginner swimmers to develop during swimming and water safety lessons. This was followed by Safety and Survival skills and Entry and Exit skills. Those skills and understandings categorised as Coordinated Strokes maintained the lowest median rankings. Figure 3 illustrates the categories of skills and understandings regarded as most important by the instructors surveyed.
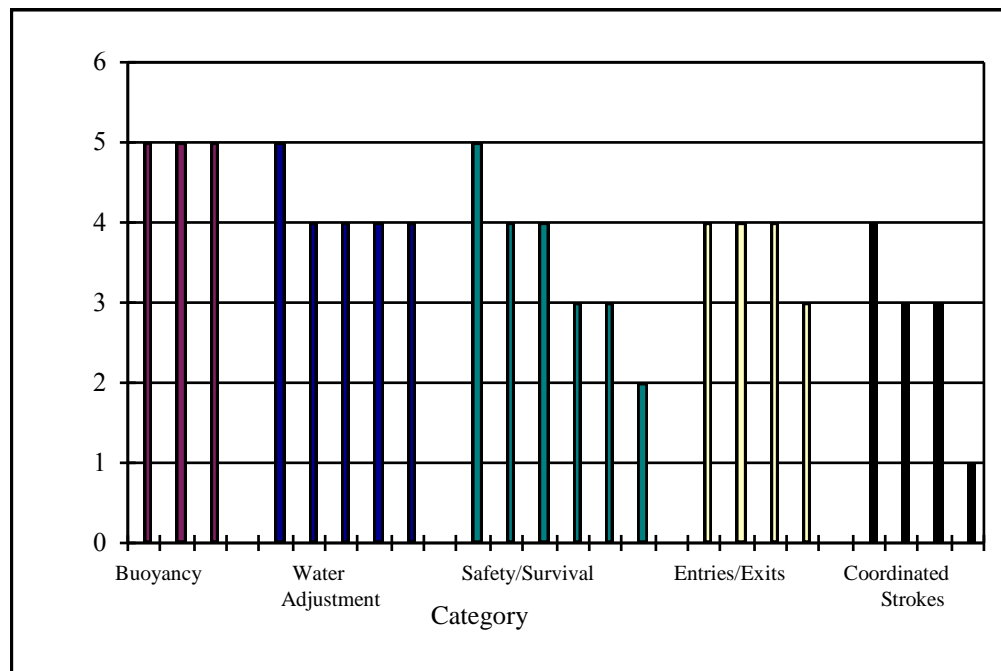
Figure 3: Responses from the 'Perceived Importance' Perspective

Figure 3 suggests that skills and understandings recorded in the Buoyancy category maintained consistently high rankings, indicating that the instructors taking part in the study felt that this category encompasses the most important skills and understandings for beginner swimmers to develop. This was followed, in order of

importance, by Water Adjustment and Safety/Survival skills. Skills and understandings perceived by these instructors to be least important for beginner swimmers to develop are those associated with Entries and Exits and Coordinated Stroke Development. It should be noted that no categories were considered as unimportant, however, some individual skills and understandings received low rankings.

Although the questionnaire responses provided valuable data to assist in addressing the study's focus, it was necessary to observe whether these beliefs were reflected in current practice. For this reason, swimming lessons conducted in an exemplar swimming school were observed and analysed. Data collected from the event analysis were examined and reported through the use of descriptive statistics. The results of these observations were used to determine if the skills and understandings perceived by AUSTSWIM instructors to be most important are reflected in the current practice of an exemplar program. Initially, lesson time was compiled to see how much time each instructor spent on developing various skills and understandings. Lesson time was then combined for all three instructors over all three lessons, as analysis of individual lessons would be somewhat limited. Table 1 describes the skills and understandings taught and the distribution of the total time during observation.

Table 1: Distribution of Instructor's Time During Observation

Skill/Understanding	Time	% Time	Skill/Understanding	Time	% Time
Freestyle	40m.45sec.	16%	Back Kick	29m.54sec	12%
Backstroke	16m.24sec.	6%	Float	6m.49sec	3%
Breastroke	31m.45sec.	12%	Movement	17m.44sec	7%
Butterfly	31m.55sec.	12%	Submerge	1m.19sec	1%
Survival	9m.47sec	4%	Face	1m.44sec	1%
Safety	-	0%	Entries	33sec	0%
Misc.	16m.13 sec	6%	Exit	4m.31sec	2%
Dive	15m. 4sec	6%	Instruction	13m10 sec	5%
Front Kick	14m.20sec.	6%	Waiting	4m.53 sec	2%

One trend that is evident in Table 1 is that no instructor included any discussion of water safety during their lesson time. Table 1 demonstrates that most time in the swimming lessons was spent developing specific stroke skills. Although this provides a guide to the various skills and understandings outlined on the event analysis program it does not provide data that can be directly compared to that within the questionnaires. Therefore, data collected during the event analysis were divided into the same categories as those used for analysis of the questionnaire data. Once the various skills and understandings had been divided into these categories, the distribution of each instructor's time was calculated.

Teacher A spent more than half of lesson time developing skills related to Coordinated Strokes. In addition, a significant amount (21 percent) of total time was spent on Water Adjustment skills. Ten percent of time over the three lessons involved Entry and Exit skills whilst Safety and Survival Skills and understandings occupied six percent of time. Three percent of time during the three lessons included the development of Buoyancy skills. During the three lessons observed, instruction and waiting time in between activities amounted to five percent of lesson time. Although Teacher B spent less time (than Teacher A) developing Coordinated Strokes (45 percent of time), it still occupied the major component of his lesson time. This was followed by 23 percent of time engaged in Safety and Survival skills. Ten percent of Teacher B's total lesson time was spent on Water Adjustment skills whilst time developing Entry and Exit skills amounted nine percent. The least amount of time during B's lessons involved Buoyancy (five percent of time). For eight percent of the total lesson time, the children were either receiving instruction or waiting.

In Teacher C's lessons, the major component of lesson time (56 percent) involved developing skills associated with formal strokes. This was followed by Safety and Survival skills, which occupied 18 percent of time and Water Adjustment, 12 percent of total time. Six percent of observed lesson time involved the practice of Entry

and Exit skills. It is interesting to note that no time was devoted to the development of formal surface Buoyancy skills during C's lessons.

To examine this further, lesson time of the three instructors was combined. Figure 4 highlights the fact that when the lesson time of the three instructors was combined, the greatest amount of time was spent developing Coordinated Strokes. This was followed by Safety and Survival, Water Adjustment and Entries and Exits. Safety and Survival and Water Adjustment skills received very similar degrees of attention. It is interesting to note that the element occupying the least amount of lesson time (from each instructor separately and the lessons combined) was Buoyancy. Activities aimed at instructing children in formal floating skills occupied only three percent of the total time observed. In the nine lessons combined, children only spent seven percent of time waiting or receiving additional instruction.
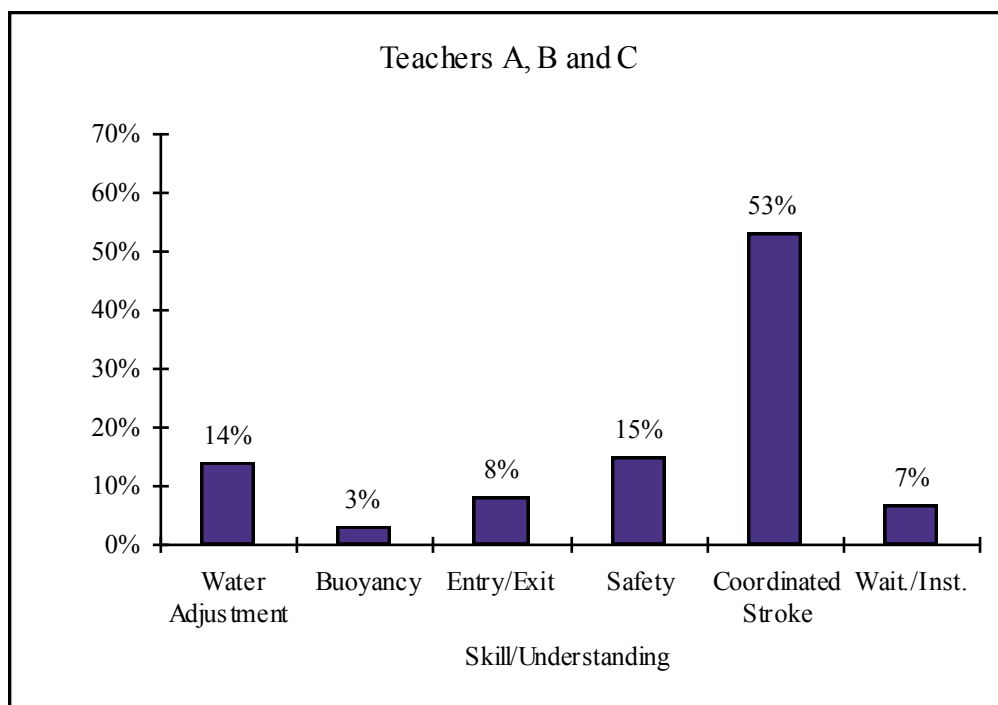
Figure 4: Distribution of Time During Nine Swimming Lessons

The event analysis indicated that, in the exemplar program, time distribution does not reflect the perceived importance of different aspects of swimming teaching. It was found that most instructor time was spent on Coordinated Stroke development. This suggests that instructor perception is not reflected in current practice. However, it is important to examine this carefully. It was evident when reviewing data collected at the exemplar program that, although lessons concentrated on specific stroke development, Water Adjustment, Safety and Survival and Buoyancy Skills were claimed to be integrated into many aspects of lesson structure. It could be said that the lessons were conducted with a more holistic view of development in mind than that reflected in the questionnaire responses. For this reason the results cannot be as clearly interpreted as they might first appear.

3. Case Study 2 – Coach Education

The purpose of this study was to determine the relationship between the aims of course providers and events during the delivery of two soccer coaching accreditation courses. A second purpose was to evaluate performance-analysis methods for assessing the course instructor's performance. A multi-dimensional case analysis approach was developed to evaluate course delivery and the research process. This approach was chosen to amalgamate sources of evidence, giving a holistic view of course quality and delivery. Data collection

methods included: hand-notation; computer logging of events; together with video. The hand-notation and video analysis was employed for five sessions of the first course and the hand-notation was replaced with computer logging of events for four sessions of the second course. This enabled video footage to link directly to the computer in real time (see Figure 5).

QuickTime™ and a
Photo - JPEG decompressor
are needed to see this picture.

Figure 5: Collecting Data on the Second Coaching Course

Questionnaires were given to the course participants, about perceptions of course effectiveness. Interviews were conducted with the instructors, about course delivery. Analysis of course documents provided further pointers to the course aims and objectives.

Two Soccer Coaching courses in Northern NSW were the focus of the study. Two course instructor's documents were systematically analysed for the study. At the conclusion of both of the courses, questionnaires were administered to all participants to gain some insight into participants' perceptions and impressions of the effectiveness of the course delivery. Specifically, the questionnaires focused on the effectiveness of the presenter, as well as the course itself. They were also asked to rank their overall impression of the course as a whole. A follow-up interview was conducted with the principal course instructor after the first course, in order to explore further issues arising from the course. The instructor was asked about his perceptions of the aims and focus of the course, what he believed were important aspects of the conduct of the course and the processes for reflection on the effectiveness of course delivery.

On the first course, a hand-notation recording system was designed to log six events identified as significant to the study. This ran concurrently with the recording of video footage to log the events as they occurred. Two Sony Digital Handycams were mounted on a tripod. One from a stationary aerial position on a balcony situated near to the course participants and the other adjacent to the area that followed the instructor and/or significant events. The second camera was located immediately adjacent to the person completing the hand notation. Three people were required to carry out this procedure, one to operate the stationary video camera, one to operate the roving camera and one to record the specific

events on the hand notation sheets. For the second course, the hand-notation recording system was replaced with a lap-top computer to log events. The GameBreaker 4.0.4 Sports Analysis software package was customised for the study's purposes. This system enabled the operator to log 12 events, providing a more detailed breakdown of events during the four sessions. Again this ran concurrently with the recording of video footage, which was synchronised with the computer logging in real time (as indicated in Figure 5). In addition to collecting the hand-notation, video footage and computer-logged data, the operators were asked to log their views of the process after each session.

Qualitative and descriptive statistical techniques were used in analysing the various forms of data. With this amalgamation of the sources of evidence some pointers to trends have emerged. A focus of the interview analysis was to uncover the instructors' main priority for course direction and content. He identified that the main focus was for the participants to gain understanding of how they could go about the development of young people in the game. Further, how they would structure and plan a training session to enable children to be placed in an environment where they can learn how to understand and play the game. This was consistent with syllabus documents prescribing course aims and content. The instructor also indicated that there was great importance placed on the technical aspects of how to teach specific soccer skills.

The NNSW Soccer Federation 'Instructor's Manual' [19] specifies aspects to be included in the course at the local level. It divides the course content into specific sessions. There is a clear emphasis on the specific skills of soccer, such as tackling or dribbling, reflecting a strong emphasis on the technical aspects of soccer. There is no direct reference to the inclusion of the principles of coaching methodology. Close inspection of the 'Soccer Australia Instructor's Manual' for the 'Junior Licence' [22] supports this emphasis with the majority of the practical sessions based on coaching juniors in the fundamentals of soccer skill and technique. A noteworthy aspect of the analysis is the majority of the methodology of coaching is dealt with theoretically. It can be argued that such an applied facet would be better dealt with in a practical context. The practical modules of the course are mainly devoted to techniques of the game. The aims of the course suggest that it offers coaches the knowledge and competency to structure/prepare/organise developmental practices for juniors. Results of the video analysis suggest a difference between these aims and the events of the courses.

Analysis of questionnaires indicated favourable perceptions of course content and delivery. This evidence is discussed in relation to intent and practice in coach education and the efficiency of employing performance-analysis techniques in logging instructional events. Overall, the participants in the two courses ranked the quality of delivery very favourably with the majority of rankings for all questions falling between 4 and 5 (5 being the highest). The highest ranked responses were related to the organisation of the course and the appropriateness of the workload demands. The lower ranked questions dealt with the appropriateness of the course for the prior knowledge/skills of the participants. This positive view was supported with positive responses to the open-ended questions. In these participants identified the fun and practical nature of the course as the most enjoyable aspect. An overall perception of these results is that the participants valued the practical elements of the course. The participants believed that the presenter did a good job for most of the time. Rankings suggested that the presenter was well prepared. Other aspects that were ranked highly included: provision of constructive feedback; professional approach; ability to pitch teaching at appropriately; and, clarity of instructions. The lowest ranked response, still considered to be true 'most of the time', was the presenter being a role model for future practice. These positive views of the presenter were supported

for the responses to the open-ended question, 'what were the best aspects of the instructor'. Participants identified the communication skills and professionalism of the presenter as the most notable aspects of his delivery.

Data collected from the hand notation/computer logging of events/video evidence were recorded and interpreted using descriptive/summary statistics. The results of observations were used to determine if the syllabus aims were reflected in the delivery of the course. Table 2 shows the percentage of time spent on categories of activity regarded as important for this study on the first course. Categories recorded as taking less than 10% of time were gathered under the 'other' category. A trend that emerged from data in Table 2 is that there is little difference in the time spent on instruction about technique or skill and that spent on coaching organisation and methodology. The syllabus mentions that the major focus of the course should be about delivery of a practical focus of coaching practices and methodology. The presenter on this course gave technique and skill equal priority. A point to note relates to the perceptions of the candidates, indicated in the results from the questionnaires. They ranked participation in fun activities as the best aspect of the course. Yet, the results from the hand notation show that there is only a small amount of time spent on active participation. The majority of time spent by the course presenter is delivering information while students listen passively.

Table 2: Results of the hand notation from the first course

Event: category of activity	Percentage time spent on event
Instruction about coaching organisation and methodology	33
Instruction about technique and skill	28
Peer coaching	12
Small group of students actively participating	11
Other (various)	16

Table 3 shows the percentage of time spent on categories of activity during the second course. A similar trend to that in Table 2 can be seen here, in that there is little difference in the percentage of time spent on technical aspects of soccer to that spent on coaching organisation methods. In addition, the majority of time spent on the course is relatively passive, whilst syllabus documents make clear that this approach is not recommended.

Table 3: Results of the computer logging of events from the second course

Event: category of activity	Percentage time spent on event
Instruction about technique and skill	32
Instruction about coaching organisation and methodology	29
Asking group for questions and feedback	17
Other (various)	22

Camera, computer and hand notation operator logs were brought to a 'roundtable' discussion about the process of logging and recording events after each course. The discussion took the form of shared reflective practice. As result of the first discussion, the suggestion to seek funding for performance-analysis software, initiated the acquisition of the 'Gamebreaker' system. In addition to the most

obvious advantage of being able to log more events by one researcher, the ability to log concurrent events was seen as the key finding in comparing a system of hand-notation with a computerised recording system. That is, in order to provide more sophisticated analysis of coaching practices, multiple activities for the same group or individual can be logged which differentiates more subtly when the instructor uses combinations of learning experiences to address more than one objective/content area.

The purpose of this study was to determine whether there was a matching relationship between the aims of soccer coaching course providers, as indicated in syllabus documents and interviews with the instructor, and experiences during the delivery of two accreditation courses. This relationship was discussed in light of the evidence from a variety of data collection methods, which provided a holistic view of course delivery. The main indicators suggest that the relationship between intent and practice in soccer coach education, in these particular courses, may not always be evenly matched. Clearly further investigation across a wider range of courses could provide coach educators with greater insight into instructional practices. These suggest that areas of high priority for the course instructors were not given proportional priority during the delivery of the course. A second purpose was to evaluate performance-analysis methods for assessing the course instructor's performance. As a consequence of this study, it was found that techniques employed in performance-analysis, normally used in analysis of players, could also be a useful method of assessing coach performance.

4. Conclusion

This paper has demonstrated that performance analysis techniques can be usefully employed to assess the instructional process in different sporting contexts. Further, adopting this approach may enhance systems of quality assurance, in the more formal sport education setting. In helping to record and analyse instructional events, computer and video technology can assist sports teachers, coaches and coach educators to monitor the content of their lessons or coaching methods. This approach, used in conjunction with other forms of evidence, can be a valuable means of providing objectivity to the more subjective and intuitive processes of reflection in evaluating the effectiveness of instruction in sport. In addition, data collected in the field can be recorded directly into a format that allows sophisticated analysis, without further data input, is an important advantage over the traditionally painstaking video analysis of pedagogical practices that has typically been post-event.

5. References

[1]   B. Appleby and B. Dawson, "Video analysis of selected game activities in Australian rules football", *J. Sci. Med. Sport,* **5** (2002), 129-142.

[2]   C. Begbie and J. Hammond, "Perception and Practice in Rural AUSTSWIM Schools", *ICHPER 8th European Congress*, London (1998), 72-77.

[3]   S. Dickson, J. Hammond, A. Brown and P. Webb, "Computer Assisted Learning in Coach Education", in *Viewing the Year 2000,* K. Hawkins and R. Nastasi (Eds), Mosby-Williams, St.Louis (1995), 22-29.

[4]   I. Franks, 'Qualitative and Quantitative Analysis', *Coaching Review*, **8** (1983), 48-50.

[5]   I. Franks, "The structure of sport and collection of relevant data", in *Computer Science in Sport,* A. Baca (Ed), VelagsgmbH & Co., Vienna (2000), 226-240.

[6]   I. Franks, D. Goodman and G. Miller, "Analysis of Performance: Qualitative or Quantitative", *Sci.Per. Res. and Tech. in Sport*, **March** (1983), GY-1.

[7]   I. Franks and G. Miller, "Eyewitness Testimony on Sport", *J. Sport Behav.*, **9** (1986), 39-45.

[8]   I. Franks and P. Nagelkerke, "The Use of Computer Interactive Video in Sport Analysis", *Ergonomics*, **31** (1988), 1593-1603.

[9]   J. Hammond, "Using Computer and Video Analysis in Sport", *Together on Sport*, **4** (1996), 1.

[10]   J. Hammond and C. Begbie, "Using Notational analysis to Reflect on Instructional Strategies in Sport"*, 2ⁿᵈ International Symposium on Computer Science in Sport,* Vienna (1999).

[11]   J. Hammond, S. Dickson, D. Parsons, "The Application of Multi-Media to the Coaching of Golf", *AISEP International Conference*, Lisbon (1996).

[12]   J. Hammond, D. Hosking and C. Hole, "An Exploratory Study into the Effectiveness of Rule Changes in Netball", *Fourth International Conference on Sport, Leisure and Ergonomics*, Liverpool (1999).

[13]   D. Harrod and S. Langendorfer, "A Scalogram Analysis of Item Order in the American Red Cross Beginner Swimmer Program", *The National Aquatics J.,* **Summer** (1990), 10-16.

[14]   M. Hughes, "A Review of Patterns of Play in Squash" in *Sports Science*, G. Atkinson, T. Reilly and L. Burwitz (Eds), E and Fn Spon, London (1986), 363-368.

[15]   M. Hughes, "Notational Analysis in Football" in *Science and Football II*, G. Atkinson and T. Reilly (Eds), E and Fn Spon, London (1993), 151-159.

[16]   M. Hughes, "Computers and performance analysis in sport", in *Computer Science in Sport,* A. Baca (Ed), VelagsgmbH & Co., Vienna (2000), 83-102.

[17]   M. Hughes and I. Franks, *Notational Analysis in Sport,* E & FN Spon, London (1997).

[18]   N. McDonald, "Avoiding the Pitfalls in Player Selection", *Coaching Sci. Update*, (1984) 41-45.

[19]   Northern New South Wales Soccer Federation *Soccer Australia Junior Licence Course Instructor's Manual,* Newcastle, Australia (n.d.).

[20]   J. Perry and J. Hammond, "Soccer Coaching Course Effectiveness: a case study in progress", *Researching in Sport, Physical and Health Education*, Wollongong (2003).

[21]   J. Sproule, G. Kinchin, M. Yelling, T. McMorris and M. McNeill, "Using computer technology to compare student teaching behaviours", *Eur. J. Phys. Ed.,* **7** (2002) 123-136.

[22]   Soccer Australia, *Instructor's Manual for the Junior Licence Course of Instruction*, Sydney (n.d.).

[23]   W. Winkler, "Match Analysis and Improvement of Performance in Soccer with the Aid of Computer Dual-Video-Systems (CCDVS)", *Sci. and Football*, **4** (1991), 6-10.

THE MECHANICS AND PERCEPTIONS OF JUDGING AN OUTFIELD CATCH

Chris Harman
Department of Mathematics & Computing
University of Southern Queensland
Toowoomba 4350, Australia

Neville de Mestre
Faculty of Information Technology
Bond University
Gold Coast 4229, Australia

Abstract

In baseball, as in other ball sports, an outfielder's most difficult fly ball to judge is the one hit in the same vertical plane defined by the fielder and the batter. A quick decision needs to be made on whether to run in, run back, or indeed to wait for the ball to arrive. The initial one or two seconds of the ball's flight is evidently crucial in providing information for the fielder to assimilate and make a decision. Experienced players are good at doing this and they become better with practice. So what are the important cues which they use? When this is better understood, training methods might be used to accommodate them. This paper critically overviews current models related to this issue and it is shown that they are somewhat limited in practical outcomes. A model is developed here which incorporates a more detailed air drag sub-model and which also includes the confounding effects of ball spin. Simulations of a fielder's perspective of the ball's flight indicate the complexity of the decision process.

1. From the bat through the air to the outfielder's glove

As any baseballer, softballer, or cricketer will tell you, the most difficult fly ball for an outfielder to judge is the one hit by the batter 'straight' towards him or her at an angle to the horizontal and in the vertical plane defined by the batter and the fielder. In that case the outfielder initially just sees the ball rising, but the distance is so great that any change in the perceived size of the ball cannot quickly be discerned to enable judgement to be made on how fast it is approaching. In contrast, if the ball is hit in a direction which is not towards the outfielder, the problem of judging the flight of the ball is much easier. In that case the trajectory of the ball is seen in at least two dimensions, rather than just a rising and falling ball as seen if it is hit in a plane directly at the outfielder. In this paper, consideration will only be given to the harder problem of judging the ball hit straight towards the outfielder, and baseball will be used as the example sport for discussion and modelling purposes. The model could also be used for cricket and softball. Since official length measures in baseball are generally discussed in feet, the imperial system is adopted in this paper.

With experience, a good outfielder will be able to tell quickly whether the ball hit directly towards him, will go over his head, drop in front of him, or come directly to his position. The best outfielders can usually carry out a rapid estimation of where the ball will land, accelerate forwards or backwards and run quickly to the predicted landing position, keeping their eyes on the ball and making any corrections to their estimation as they go. Running back while watching the ball over the shoulder is somewhat slower than turning completely and then sprinting away without looking back. But the latter strategy is fraught with danger. It does not allow much time for refinement of initial estimated landing point when the outfielder finally turns around expecting the ball to be arriving on that spot. However, this tactic can be used when the game situation is desperate; the alternative is not to reach the landing point in time, and miss any chance of catching the ball.

As detailed in the next section, others have studied the problem of judging a fly ball hit straight at an outfielder, but the trajectory mechanics and strategies have been rather simplistic. For example, constant air drag coefficients have been used and the effects of spin on the ball seem not to have been considered,

even though most batted balls will have some backspin or topspin imparted which can significantly affect the trajectory and confound the judgement of the outfielder. In this paper, spin is included in the model and shown to be an important factor in making fly ball judgements. Variation of drag coefficient with speed will also be incorporated, allowing effects of the drag crisis to be simulated. Examples are to be given here which illustrate how drag and spin affect the trajectory spatial profile, the trajectory speed profile, the time of flight, the range, and importantly, the perception of the outfielder.

The key motivation in arriving at an understanding of processes in sport should be to formulate coaching hints to help all players from beginners through to professional sports men and women. This paper raises more questions than it answers, but in the last section some rudimentary tips are given for outfielders and coaches of outfielders.

2. Earlier insights into judging a fly ball

When a baseball is hit straight towards an outfielder from over 300 ft away, a number of cues have the potential to help the fielder to decide which way to run. These include the speed of the pitch, the speed of the batter's swing, the sound of the ball striking the bat, and the early dynamics of how the ball moves, as seen by the fielder. The first two of these cues can be misleading, because the ball's departing speed from the bat depends on how nearly the ball hits the bat along a line through the centre of percussion of the bat. Also, the sound of the bat hitting the ball can often not be heard against the background crowd noise and yet an experienced outfielder will usually still judge correctly. So it appears that the dynamics of the ball for the first second or so might be the key ingredient in judging the catch.

Chapman [6] seems to be the first to carry out seriously a geometric analysis of the way in which an outfielder should run to complete a catch of a ball hit straight towards him. If ϕ is the outfielder's line-of-sight angle of elevation of the ball, he found that a ball hit with no air drag straight towards an outfielder, such that it will land precisely where the outfielder is standing, has the property that $d(\tan \phi)/dt$ is a constant. Moreover if the ball is hit such that it will fall short, or go over the fielder's head, the fielder should run at an appropriate constant speed, in which case $d(\tan \phi)/dt$ can be made to be the same constant size and the ball and the outfielder will arrive at the same time at the same spot. In his comprehensive paper on the catching problem, Brancazio [4] pointed out that, although interesting, there is a serious flaw in Chapman's analysis in that the effects of drag on a baseball are very significant and should be included. He went on to give an example of how a baseball hit in a vacuum at an angle of $60°$, and an initial speed of 147 ft/s, would travel 80% further than in air. He then showed that when air drag is taken into consideration, Chapman's constancy of $d(\tan \phi)/dt$ is no longer valid. Brancazio made an analysis of other potential cues, including the increase in the perceived change in size of the ball, but found that this contributes little or no useful information in the early part of the flight of the ball. He did find that in the latter part of the flight, a sensible strategy is to run such that the angle ϕ approaches a constant. This ensures that the outfielder and the ball will arrive at the same time. However, there is a problem with this. Making use of such a strategy is only possible when the fielder has time to adjust his speed to smoothly reach the landing point. Experienced outfielders certainly do this in order to be in smooth and balanced motion when catching the ball. But the game winning catches are taken by those who read the ball off the bat and immediately accelerate and sprint at top speed towards their predicted landing point in order to maximise their chance of arriving in time to make the catch. In this case they do not have the luxury of adjusting their speed to keep angle ϕ a constant during any part of the flight. For example, see [11] for a video of the legendary catch by the legendary outfielder Willie Mays[1] in the eighth innings of game one of baseball's 1954 World Series between the New York Giants and the Cleveland Indians. The crucial factor in that situation, apart from his remarkable athleticism in making the catch, was Mays' rapid analysis of the first part of the flight of the ball which had been hit straight towards his fielding position and which

---

[1]Godfather of the present home run king Barry Bonds

was going well over his head. Experienced and expert outfielders are good at making this quick analysis.

Brancazio [4] introduced air drag to the trajectory model, using a constant drag coefficient, and then looked at three key perceptual parameters which have the potential to provide cues when observed by the fielder as the ball is hit at $t = 0$. They are (i) the angular speed $\omega_0 = \phi'(0)$, (ii) the angular acceleration $\alpha_0 = \phi''(0)$, and (iii) the vertical acceleration $v'_{P0} = v'_p(0)$, where v_p is the fielder's observed velocity of the ball perpendicular to the line of sight to the ball. For various initial speeds and angles of the batted ball, Brancazio calculated the values of these parameters for two cases: (i) a ball hit straight towards an outfielder standing at $300\,\text{ft}$ from the batter, and (ii) for an outfielder standing at the landing point determined by the initial conditions. Comparing the perceived values from the two positions, he claimed incorrectly that "The differences are considerably larger for α_0 than for ω_0 or v'_{P0}". He apparently meant that the *relative* differences were larger for α_0. In fact ω_0 and v'_{P0} showed the largest differences. Brancazio then compared the ratios of the two values of each parameter and correctly found that the ratios of the values of α_0 when fielding at $300\,\text{ft}$, divided by the values when fielding at the landing point, were by far the largest. His conclusion was that "... the angular acceleration provides the most valuable initial cue as to the location of the landing point and the direction in which the fielder must move". However, the values of α_0 are very small in comparison to ω_0 and v'_{P0} and of course the ratios of very small numbers naturally tend to be large. Also, small values of parameters should be more difficult for the fielder to perceive, even if the relative differences are large. Hence Brancazio's conclusion warrants further analysis. But before doing this, we need to consider a more comprehensive model for the drag force and introduce to the analysis the considerable confounding influence of ball spin.

3. Does the drag crisis make the ball lurch?

The drag force F_D on a ball is reliably modelled by

$$F_D = 0.5\rho A C_D v^2,$$

where ρ is air density, A and v are respectively the cross-sectional area and speed of the ball, and where C_D is the drag coefficient for the particular ball. Most researchers take the value of C_D for a baseball to be a constant, independent of speed, and with a value somewhere between 0.3 and 0.5. Briggs [5] reported that the terminal speed of a baseball suspended in a vertical wind tunnel, is about 95 miles/hour. This corresponds to $C_D \approx 0.3$ for a baseball without much spin. Consequently, Brancazio [4] used $C_D = 0.3$. But this wind tunnel measurement does not shed much light on possible changes of C_D with speed. More recently, Mehta and Pallis [7] reported results of wind tunnel measurements of C_D for a non-spinning baseball at various speeds. Their results, illustrated by the 18 measurements in Figure 1, show the variation with speed and the obvious presence of the so called *drag crisis* where the flow past the ball suddenly changes from turbulent (lower drag) to laminar (higher drag) as the speed of the ball slows. Figure 1 also shows our least squares cubic spline fit which is used here to model variations of C_D with speed for trajectories when the baseball is hit without much spin. The characteristics of the curve suggest that when a ball is hit with little spin at high speed, the value of C_D will start with a value of about 0.35, dip suddenly to about 0.3, and sharply rise to about 0.53 as the ball slows down to below 60 miles per hour. This should cause lurching in the ball's motion as the ball's deceleration suddenly decreases and then increases. It is not clear if the fielder perceives any lurching, but using the trajectory model that follows, Figure 2 illustrates the obvious change in the drag crisis induced curvature characteristic in a typical profile of speed against time for a baseball with little spin, with initial speed 160 ft/s (109 mph), and hit at an angle of $30°$ to the horizontal. Also the effect of the drag crisis can clearly be seen in the plots of ϕ'' (the acceleration of ϕ) in Figures 3, 4, and 5 following.

The effect of spin on the drag coefficient C_D has been studied by Alaways [1] who used high speed videos of pitched balls with initial speeds between about 40 and 80 mph and spin rates between 15 and 70 Hz. For these conditions, Alaways estimated the value of C_D to be between about 0.3 and 0.45. Note
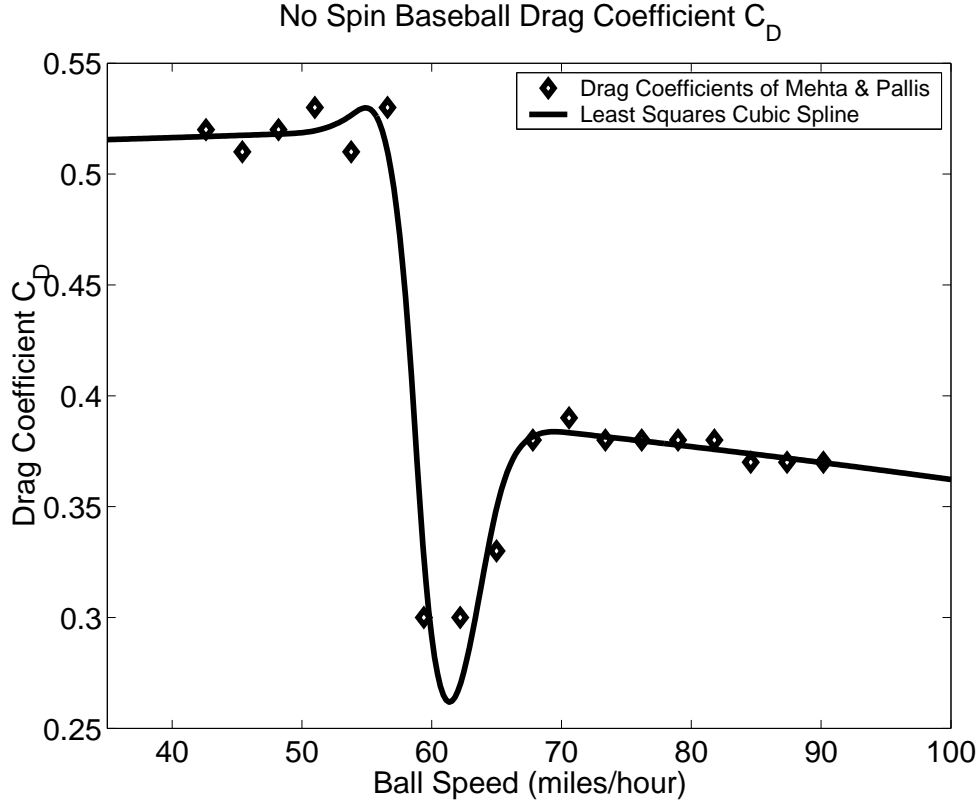
Figure 1: Baseball drag coefficient data of Mehta & Pallis and our least squares fit

that it is estimated that an expert pitcher can throw the ball with a spin rate of up to 35 revs/sec. Briggs [5] cites an experiment where tape was attached to a baseball and the baseball then pitched with spin. By counting the turns of the tape, it was estimated that the spin rate was about 27 revs/sec.

For the early part of the flight of a batted ball, the ball is travelling at high speed and the wake will be turbulent whether or not the ball is spinning. In this high speed region, the values obtained by Alaways for C_D are not very different from those illustrated in Figure 1 and so for the purposes of deciding possible cues perceived by the fielder during the first one or two seconds of flight, the least squares cubic splines fit to the data will be used in our following trajectory model as a basis for the variation of C_D with speed.

In another study, Alaways et al. [3] used cameras to estimate values of C_D for baseballs pitched in the normal way with spin. This was done during competition at the 1996 Olympic Games. In that study the speed of the ball could be accurately measured, but it was difficult to estimate the spin of the ball. Their results also suggested that the drag crisis might sometimes be experienced during the flight of a pitched ball.

4. Batted ball spin induces lift or dive

Earlier work by Briggs [5] and Rex [8] assumed that the lift force is proportional to the velocity squared times the spin rate. This seems not to be the case, see Watts & Ferrer [10], Watts & Baroni [9], Alaways & Hubbard [2], and Alaways et al. [3], and it is now recognised that the lift force is proportional to the velocity times the spin rate.
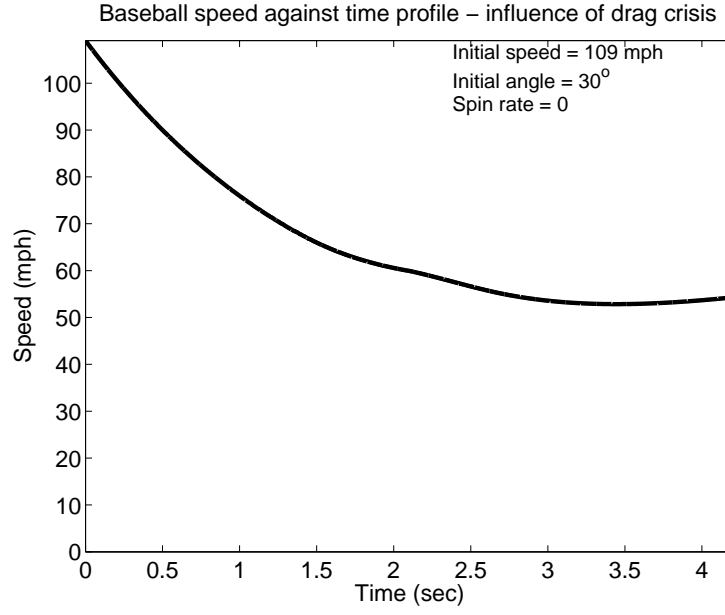
Figure 2: Non-spinning baseball speed against time profile showing effects of the drag crisis

The lift force on a baseball, at speeds typical of a batted ball, is given by

$$F_L = 0.5\rho RA\omega v,$$

where ρ is density of air, R is radius of the ball, A is the ball's cross-sectional area, ω is the rotation rate in radians per second, and v is the speed of the ball, see Watts & Baroni [9].

The rotation rates of batted balls have apparently not been experimentally measured, but Watts & Baroni estimated that a batted ball could have at least as much spin as a pitched ball, that is it could achieve at least 35 revs/sec. Now the batter generally swings in a roughly horizontal plane, so we assume here that any spin imparted by the bat to the ball will either be backspin, or topspin, with a horizontal axis of rotation. The effects of spin on the flight of a baseball are quite pronounced. For example, using the trajectory model of §5 following, a baseball hit with little or no spin, at 30° with initial speed 160 ft/s, has a range of 375 ft. Hit with the same speed and at the same angle, but with backspin of 35 revs/sec, yields a range of 440 ft, an increase of 65 ft. Hit with topspin of 35 revs/sec, the ball goes 300 ft, a reduction of 75 ft.

5. A spin on the ball model in drag

Consider a baseball travelling at speed v at an angle θ to the horizontal. The equations of motion, including air drag and spin-induced lift, are then determined by the force equations

$$m\frac{d^2x}{dt^2} = -F_D\cos\theta - F_L\sin\theta$$
$$m\frac{d^2y}{dt^2} = -F_D\sin\theta + F_L\cos\theta - mg,$$

where m is the mass of the ball and g is the acceleration due to gravity and where the lift force is perpendicular to the direction vector of the ball. If $v_x = x'(t)$ and $v_y = y'(t)$ are the components of v in the x and y directions respectively, then $v = \sqrt{x'^2 + y'^2}$ and $\cos\theta = v_x/v = x'(t)/v$ and $\sin\theta = v_y/v = y'/v$. Hence the equations of motion become

$$x'' = -Kx'\sqrt{x'^2 + y'^2} - Sy'$$
$$y'' = -Ky'\sqrt{x'^2 + y'^2} - Sx' - g,$$

where $K = 0.5\rho A C_D/m$ and $S = 0.5\rho R A \omega/m$. Note that if ω is negative, then the lift force parameter S is also negative and the ball will dive rather than lift (topspin case). For a given initial angle of projection θ, these equations can then solved for the position of the ball, x and y and the speeds v_x, v_y at any time t using a fourth order Runge-Kutta program with four initial conditions determined by the starting position and speed $(x(0), y(0), x'(0), y'(0))$.

6. Simulating the outfielder's perception

At time t after the ball is hit, ϕ is the line-of-sight angle of elevation to the ball of a fielder who will be assumed to be standing 350 ft from the batter. Then it is easy to show that

$$\phi = \tan^{-1}\left(\frac{y}{350 - x}\right),$$

and that

$$\omega = \phi' = \frac{y'(350 - x) + yx'}{(350 - x)^2 + y^2},$$

and an even more complicated expression for $\alpha = \phi''$, all of which can be numerically calculated from computed values of x, y, x', y' at each time step of the Runge-Kutta solution of the above trajectory equations.

In his analysis of the outfielder's perceived changes in ϕ, Brancazio [4] placed the fielder in two positions,(i) at 300 ft, and (ii) at the landing point of the particular ball that was hit. So the fielder's view is simulated for two different fielding positions for a single trajectory in each simulation. This is not what happens in a game, where the fielder is in a set position, and must quickly make a judgement on whether the trajectory will carry the ball over his head or drop short. In our following simulations, the fielder will be placed at 350 ft and the analysis of changes in ϕ will be made for two trajectories, one representing a ball falling short at 300 ft and the other a long ball over the fielder's head and landing at 400 ft. This better represents the game situation.

In the first example simulation, the confounding influence of ball spin becomes very clear. Figure 3 shows the trajectories in each case (with the position of the fielder at 350 ft marked), and the three corresponding plots of the values of ϕ, ϕ', and ϕ'' for the first two seconds. As indicated in the figure caption, the ball falling short has backspin and is hit at a lower speed than the long ball which has no spin. Because the trajectories look so different, one might conjecture that it would be easy to quickly judge the ball. But this is not the case. If we use Brancazio's suggestion that the value of ϕ'' at $t = 0$ is a possible cue, we get the values of approximately $0.5°/\sec^2$ for the short ball and $2°/\sec^2$ for the long ball. The ratio of these is similar in magnitude to changes in ϕ'' which Brancazio postulates as a potentially useful cue. However, these are very small rates of change in the angular speed and it is most unlikely that a fielder could perceive them. The plot of the fielder's observed angle ϕ (see second plot in Figure 3) clinches this. For the first second or so, the short ball and the long ball rise against the background with almost exactly the same angle ϕ at any time t. During this time the fielder would have no way of judging between the possibilities that the ball could fall short or go long. Hence any cues determined by angle or rates of change of angle would be quite useless for at least the first second. After this, cues would probably be found by observing the differences in ϕ'. For example, after 2 seconds have elapsed, the value of ϕ' for the long ball has risen to about $20°/\sec$ and the short ball has dropped to $10°/\sec$. An experienced fielder

6
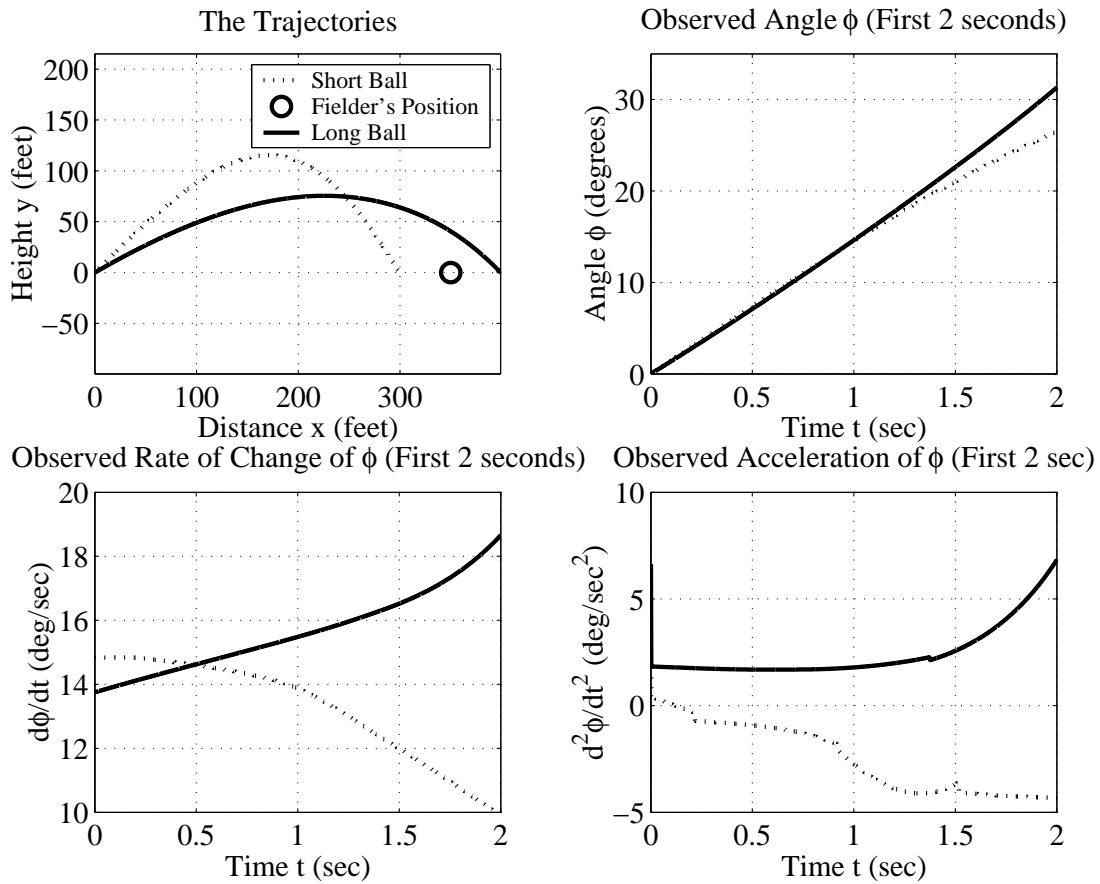
Figure 3: Short ball: backspin 30 revs/sec, launched at $45°$ with speed 128 ft/sec, lands at 300 ft, flight time 5.94 sec. Long ball: no spin, launched at $30°$ with speed 168 ft/sec, lands at 400 ft, flight time 4.36 sec.

would undoubtedly start to perceive these changes after the first second or so, and move accordingly. Note that after the first 1.5 seconds, the fielder still has 4.5 seconds to reach the short ball and about 3 seconds to reach the long ball. In this time there would be no trouble running forwards 50 ft to reach the short ball, and a good sprinter might be able to run back 50 ft to catch the long ball [2].

A 'sinking line drive' is the result of hitting a baseball with some topspin, at high speed, and at a low trajectory. If hit straight towards a fielder, this is one of the most difficult balls to judge and catch. A second example simulation, illustrated by Figure 4, shows the trajectories of two balls hit at $30°$. The short ball was hit with topspin of 20 revs/sec at 147 ft/sec and 'sinks'. The long ball was hit with backspin of 20 revs/sec at a slightly higher speed of 156 ft/sec and it 'flies' a further 100 ft. The differential effects of spin on the range of a baseball are thus clearly seen.

In contrast, if there is no spin, the situation is much simpler. In the final example simulation, see Figure 5, the two balls are hit with little or no spin at the same angle and with initial speeds 135 and 169 ft/sec respectively. They fall at 300 ft and at 400 ft. It is clear from the plots of ϕ that after about half a second, the observed angles and their derivatives start to differ noticeably. The absolute differences between the two values of ϕ' are appreciably higher than those between the values of ϕ''.

---

[2]Maurice Greene was timed at 2.76 sec for the first 20 m in the 100 m final of the 1999 World Athletics Championships.
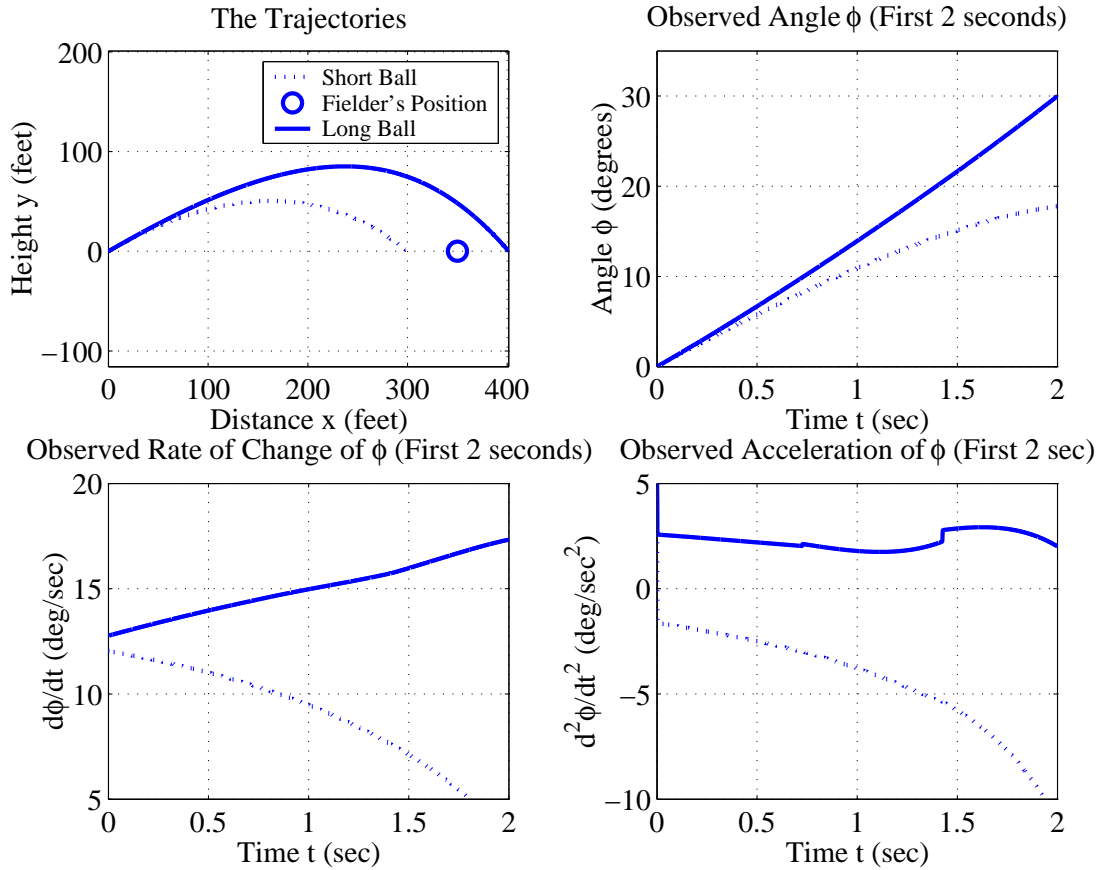
Figure 4: Short ball: topspin 20 revs/sec, launched at $30°$ with speed 147 ft/sec, lands at 300 ft, flight time 3.16 sec. Long ball: backspin 20 revs/sec, launched at $30°$ with speed 156 ft/sec, lands at 400 ft, flight time 5.12 sec.

7. How does the outfielder judge?

So what does the fielder do in these situations? The ball hit straight towards him is often tough. He should use any potential cues available at the time, such as the speed of the pitch and the batter's swing, the sound of the ball hitting the bat, and most importantly, the perceived dynamics of the changes in the line-of-sight angle ϕ.

In contrast to the conclusion of Brancazio on the importance of ϕ'', it can be argued that the value of ϕ', which measures the rate at which the angle is changing, gives a better cue. The absolute differences of ϕ' are bigger. Of course, as the example simulations show, in tough situations the fielder cannot make a decision immediately at $t = 0$, but must accumulate more information as the ball flies.

8. Coaching tips and discussion

What can coaches do? Most of the perceived cues are probably internalised by the fielder subconsciously, so that an academic discussion with the fielder on the values of ϕ', would probably not help much. Coaches generally give repetitive fly ball practice by hitting the ball with a 'fungo' bat (a lighter than normal bat), but it might be better to use a pitching machine to eject balls in the direction of the fielder. This would enable the range (and spin if possible) to be prescribed so that the fielder can learn to perceive the short or the long ball, and thus the above example simulations could be put into practice. This would be particularly useful if the important cues are to be found in the early dynamics of ϕ, and not so much

8
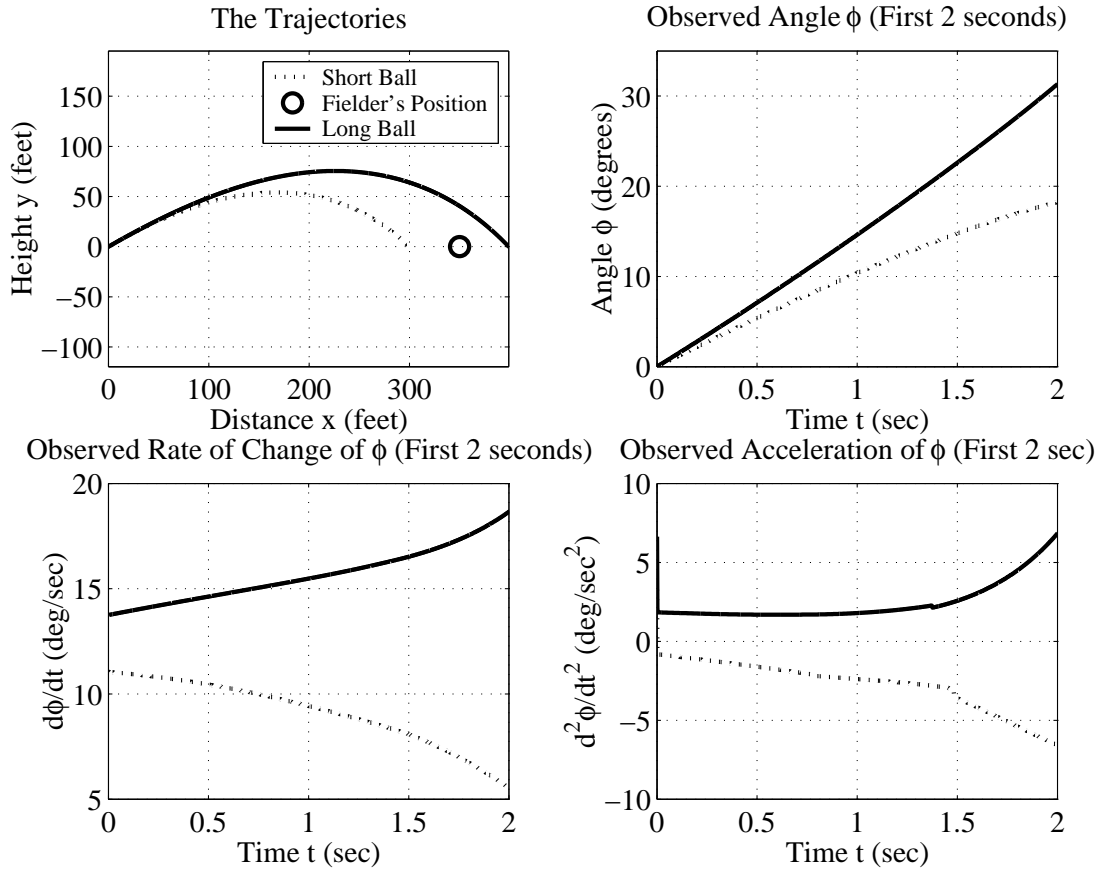
Figure 5: First plot: Fielder at 350 ft. Short ball: no spin, launched at $30°$ with speed 135 ft/sec, lands at 300 ft, flight time 3.70 sec. Long ball: no spin, launched at $30°$ with speed 168 ft/sec, lands at 400 ft, flight time 4.36 sec. Three other plots show changes in ϕ, ϕ', ϕ'' for $t \leq 2$ sec.

in the cues perceived by the swing of the bat and the sound of it striking the ball.

It may well be true that batters are more likely to hit with some backspin on the ball. This can be judged from the seemingly larger number of 'foul' balls (balls hit, but not into play) which come off the top of the bat rather than the bottom. Long ball hitters certainly know the advantage of hitting the ball with some backspin. In the above first example simulation (see Figure 3) where the short ball is hit with backspin and the long ball without spin, it is impossible for the fielder to use the changes in ϕ to judge the ball for the first one or two seconds. But the short ball's flight time is much longer. Hence, in situations like this it would be better for the fielder to start moving back rather than stay still until more information on the changes in ϕ are assimilated. There are two reasons for this: (i) in this case the ball's flight time is much longer for the short ball, giving more time for fielder's corrections, and (ii) the fielder can run faster forwards than backwards, enabling even more time for corrections if the ball is short.

In terms of future research directions for modelling and simulating fielder's perceptions of outfield fly balls, experimental work on batted ball spin needs to be done in game situations. This will help to set realistic parameters for the computed simulations. Also there needs to be more experimental work on the difficult task of measuring the drag on spinning balls moving at high speed.

It might be useful to compute real-time computer graphical simulations of "what the fielder sees". This could then be tested with the fielder seated in front of the computer screen. By showing, in real time, the first one or two seconds of a ball's flight, it might be possible to train the fielder to judge the ball. It is

9

intended to experiment with this and to report the outcomes in due course.

References

[1] L. W. Alaways, "Aerodynamics of the curve-ball: an investigation of the effects of angular velocity on baseball trajectories", PhD dissertation, University of California, Davis, (1998).

[2] L. W. Alaways and M. Hubbard, " Experimental determination of baseball spin and lift", *J. Sports Sci.*, **19** (2001), 349-358.

[3] L. W. Alaways, S. P. Mish and M. Hubbard, " Identification of release conditions and aerodynamic forces in pitched-baseball trajectories", *J. Appl. Biomechanics*, **17** (2001), 63-76.

[4] P. J. Brancazio, " Looking into Chapman's homer: the physics of judging a fly ball", *Am. J. Phys*, **53** (1985), 849-855.

[5] L. J. Briggs, " Effect of spin and speed on the lateral deflection (curve) of a baseball: and the Magnus effect for smooth spheres", *Am. J. Phys*, **27** (1959), 589-596.

[6] S. Chapman, " Catching a baseball", *Am. J. Phys*, **36** (1968), 868-870.

[7] R. D. Mehta and J. M. Pallis, " Sports ball aerodynamics: effects of velocity, spin and surface roughness", in *Materials and Science in Sports Conference*, Keynote Paper, Coronado, California, USA (2001).

[8] A.F.Rex, " The effect of spin on the flight of batted baseballs", *Am. J. Phys*, **53** (1985), 1073-1075.

[9] R. G. Watts and S. Baroni, " Baseball-bat collisions and the resulting trajectories of spinning balls", *Am. J. Phys*, **57** (1989), 40-45.

[10] R. G. Watts and R. Ferrer, " The lateral force on a spinning sphere: aerodynamics of a curveball", *Am. J. Phys*, **55** (1987), 40-44.

[11] "The catch", see `http://www.psslbaseball.com/multimedia.html`

An investigation of rugby test results

Ray Hoare, Ph.D.
HRS Ltd, P O Box 4153
Hamilton East, New Zealand

Abstract

Data from 18 rugby tests is analysed by various methods to illustrate the usefulness of statistical software to derive patterns even in such a small data set. Two tree based techniques (C&RT and CHAID) are demonstrated in order to encourage people to use these extremely powerful but infrequently used methods.

1. The data

Data on 18 rugby tests, mostly involving the All Blacks, was supplied by Dr. Paul Bracewell of Offlode. This data was collected commercially and supplied to national print media, notably the New Zealand Herald, Christchurch Press and Dominion Post. For each game, data was available for both sides, so there were 36 cases in the data set. The score for each side was used to compute a variable with two levels – 0 for a lose or draw, and 1 for a win. Labels were attached to the computed value, to make it easier to interpret.

There were 14 variables that described each game – measurements such as Territory, Possession, Rucks & Mauls, Lineouts Won, Lineouts Lost, Errors, Turnovers Won and so on.

2. The objective of this paper

This analysis will try to see if there are aspects of the game data that are strongly related to the final result, treated as a Win/Lose situation. This is not the objective of the paper, because there is not sufficient data to reliably build a predictive model. Rather, the purpose of the paper is to see if the tools available in a commercial statistical analysis program can tease out enough indications of success to make it worth while someone else tackling the task of assembling sufficient data and applying these techniques in the future.

3. Using one variable at a time to predict success

It would be unlikely that any one measure of a game would measure the probability of success well. However, by looking at variables individually some idea of the relevant data can be gained.

3.1 t-tests

The t-test is the classic statistical measure of whether a measurement has a different average

Table 1: A t-test of the Win / Lose state, sorted on p-value

Variable	Mean Lose or draw	Mean Win	p
Terriority	0.42	0.57	0.00004
Possession	0.44	0.55	0.00011
Total Metres Run	314	532	0.00021
Rucks & Mauls	64	79	0.02645
Lineouts Won	12	15	0.07
Average Metres Kicked	26	27	0.08
Average Metres Run	6	6	0.09
Lineouts Lost	2	3	0.13
Scrums Won	10	8	0.15
Errors	9	11	0.20
Penalties Conceded	9	8	0.23
Turnovers Won	12	11	0.31
Scrums Lost	1	1	0.41
Total Metres Kicked	1027	960	0.41

192

value in two groups. In this case, we would be interested if, for example, the Territory was different between games in which teams won or lost.

Table 1 is copied from *STATISTICA* [1], after removing unnecessary data and sorting by p-value. The first four variables are automatically coloured red by *STATISTICA*, signifying that the difference between the averages for the two situations are significant at the 5% confidence level or better. For better readability when printed in black and white the backgrounds of these cells have been manually shaded.

3.2   Box and Whisker plots

One should never do a t-test without looking at the distribution of the data to make sure it is approximately normally distributed, especially with the small number of observations that we are using here.

As an approximate guide, and a way of looking for outliers, the Tukey-type box and whisker plots have been produced for the four measurements that show significant t-tests and are shown in figure 1. They don't show excessive departures from normality, and graphically illustrate the ranges of the data in win and loss situations.

Figure 1: Tukey-style Box and Whisker plots
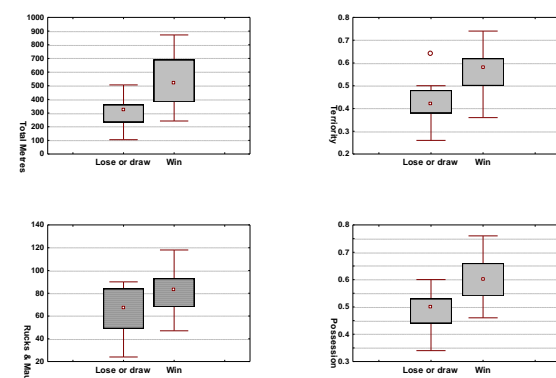
4. Multivariate analysis

4.1   Correlations

It may be interesting to see what correlations exist between the measurements.

I have prepared the table of all correlations, and then sorted it on the correlation coefficients for Territory and discarded some of the columns in the correlation table. The results are shown in Table 2 - correlations in shaded cells are statistically significant at the 5% level.

One surprise is that Errors (presumable of the team being evaluated) are positively correlated with the Territory! I suppose that if you don't have the ball you can't make mistakes.

Figure 2 shows graphically the effect of possession and territory on the win or lose situation. Only one game was lost if the Territory was >0.51, and only 2 were won if the territory was <0.5. Possession is closely related to Territory, but is not so good at discriminating the win/lose situation.

Table 2: Correlations between some of the variables

Variable	Terriority	Possession	Rucks & Mauls	Lineouts Won
Correlations (test summary data.sta) Marked correlations are significant at p < .05000 N=36 (Casewise deletion of missing data)				
Terriority	1.00	**0.85**	**0.45**	**0.61**
Possession	**0.85**	1.00	**0.66**	**0.57**
Total Metres Run	**0.73**	**0.70**	**0.41**	**0.35**
Lineouts Won	**0.61**	**0.57**	0.17	1.00
Errors	**0.46**	**0.38**	-0.00	0.24
Rucks & Mauls	**0.45**	**0.66**	1.00	0.17
Average Metres Kicked	**0.35**	0.21	-0.02	0.11
Average Metres Run	0.31	0.22	-0.05	0.09
Lineouts Lost	-0.01	0.08	0.11	-0.06
Scrums Lost	-0.08	-0.01	-0.09	0.02
Scrums Won	-0.13	0.05	0.14	-0.22
Total Metres Kicked	-0.13	-0.11	-0.28	-0.04
Turnovers Won	-0.23	-0.23	-0.14	**-0.38**
Penalties Conceded	**-0.39**	**-0.40**	-0.11	**-0.39**

4.2 Regression analysis

When it is clear that many variables are related to each other, and to the variable of interest, then multiple regression should be considered. In the special case where the dependent variable is binary, then Probit and Logit analyis are possible.

I have applied Logit analysis to this data, and experimented with a number of combinations of variables - because of multicollinearity you cannot just put all variables into the analysis.

Figure 2: The relationship between Possession and Territory

Figure 2: Possession by Territory, for won and lost games
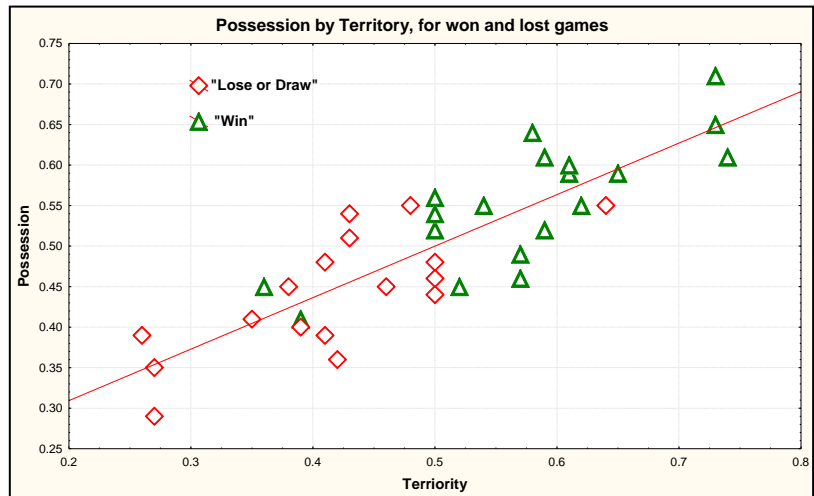
Table 3a: Predicting wins with Logit - p values

	DF	Wald Stat.	p
Intercept	1	5.76	0.016
Possession	1	6.99	0.008
Scrums Won	1	4.38	0.036

Table 3b: Predicting wins with Logit - results

Classification of cases - Log odds ratio: 2.5			
	Predicted Win	Predicted Lose	Percent correct
Win	15	4	79
Lose	4	13	76

Some statistics of the analysis are shown in tables 3a and 3b - we can predict about 80% of the games on the basis of possession and the number of scrums that are won.

4.3 Tree analysis

There are a number of ways to investigate problems of the type we have here, based on algorithms that can present their results as "trees". The general result of all the algorithms is to derive predictions from few simple if-then conditions.

In my opinion, these methods are not as well known as they should be, so I will present some results as an indication for further work. The number of cases is really too small to come to in-depth conclusions, but with a few hundred cases real results should become evident.

4.3.1 Categories and regression trees

This method, abbreviated as C&RT, [2] and known in one commercial implementation as CART, was developed in 1984. Despite this 20-year history, it is not well known, except by those specialising in data mining.

As with any kind of regression, you specify a "dependent" variable, in this case the binary one, "Win or Lose". You also specify a set of variables that are used to predict the dependent one. This method, like all tree methods, first picks the variable that best predicts the dependent variable. It reports a relationship that makes the first split of the data. The algorithm then seeks other variables that can improve the classification.

Figure 3 is an example of a C&RT plot. It gives a set of rules for correctly classifying every game in the set. The first split is based on Territory being less than 0.51, and shows that all 14

games where the territory is larger than this and the penalties are <=11.5 were won, and the one game where the territory was >0.51 and the penalties were >11.5 was lost.

Where the Territory was <= 0.51, there were only 5 games won, and it takes a combination of 4 variables to find ways to separate them from the others. This is a good example of over-classification – although the algorithm allocates only "pure" nodes, it is unlikely that the details below the first branch of the tree would replicate when more data is obtained.

There are other ways to run this analysis that give results that are different in detail, but this can give you the general idea.

Figure 3: A tree created with the Classification and Regression trees algorithm.
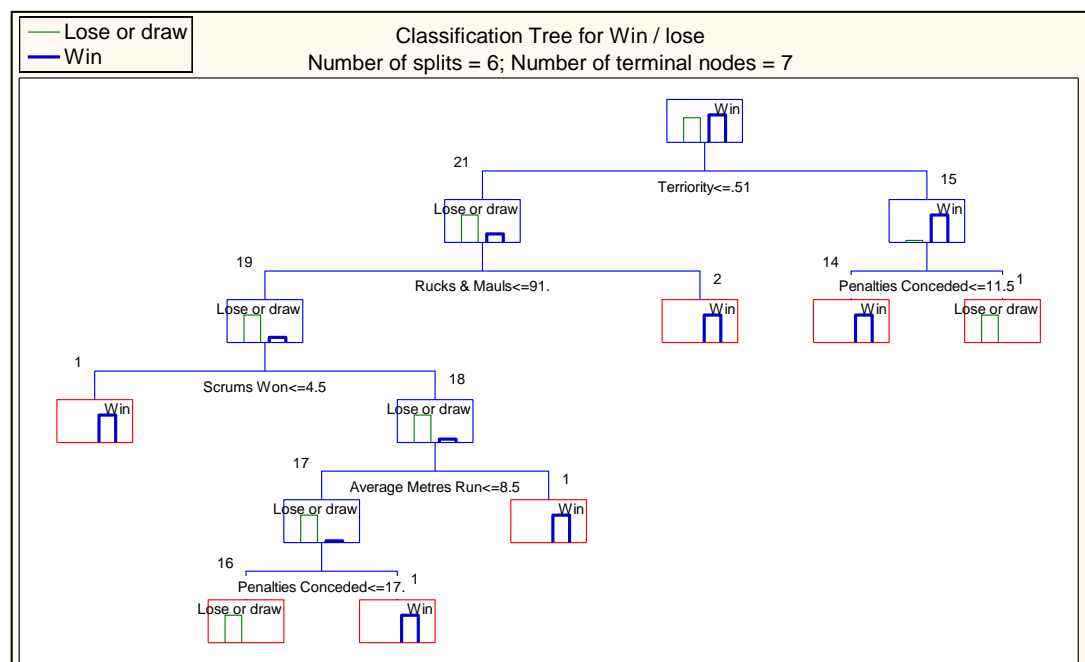
4.3.2 CHAID

This technique [3] gives results that can be represented as trees, similar to those of C&RT, but whereas C&RT always gives binary splits (two branches), CHAID can give multiple splits.

An example of a split on territory is shown here – there are 2 pure splits and one mixture. (A variation known as the exhaustive CHAID algorithm was used.)

A classification matrix is shown in Table 4 - this model is rather better at predicting losses than wins.

One option in *STATISTICA* is to make an interactive tree – you can see the statistical test used by the algorithm to make the split, and if something else makes more sense to you, you can override the algorithm and force a

Figure 4: A tree made with the CHAID algorithm
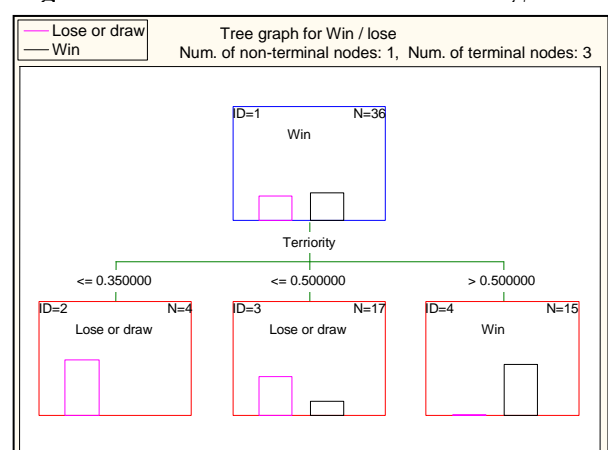
Table 4: Classification of cases with CHAID

	Pred. Lose or Draw	Pred. Win	Percentage Correct
Lose or draw	16	1	94
Win	5	14	67

different split.

In the case above, the initial split was on Territory, but Table 5 shows that the next most significant variable for that split is Possession. We can force a split on this variable, and see in figure 5 that a tree results, although the separation is not as good as the first one.

Table 5: P-values for the first spl

	Predictor Information for Node 1 (test summary data.sta) The order of predictors according to Statistic/df				
	Number of node	Split type	Chisquare Statistic	degree of freedom	Adjusted p
Terriority	3	Automatic	18.09375	2.000000	0.015073
Possession	3	Automatic	14.11427	2.000000	0.220478
Total Metres Run	3	Automatic	12.09288	2.000000	0.605766

Figure 5: A CHAID tree forced to split on Possession.
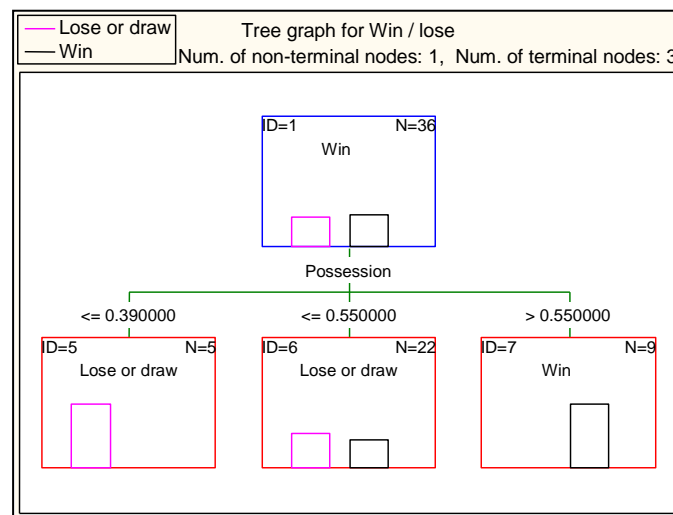
4.4   Lift charts

The lift chart provides a visual summary of the usefulness of the information provided by one or more statistical models for predicting a binomial dependent variable. The chart summarizes the utility that one may expect by using the respective predictive models, as compared to using baseline information only. With the results obtained so far, the Logit model seems to perform better than the tree-based models, if you want to forecast the results for a large proportion of the games, but C&RT is best if you are prepared only forecast up to 50% of the games.

Figure 6: A lift chart comparing 3 different models
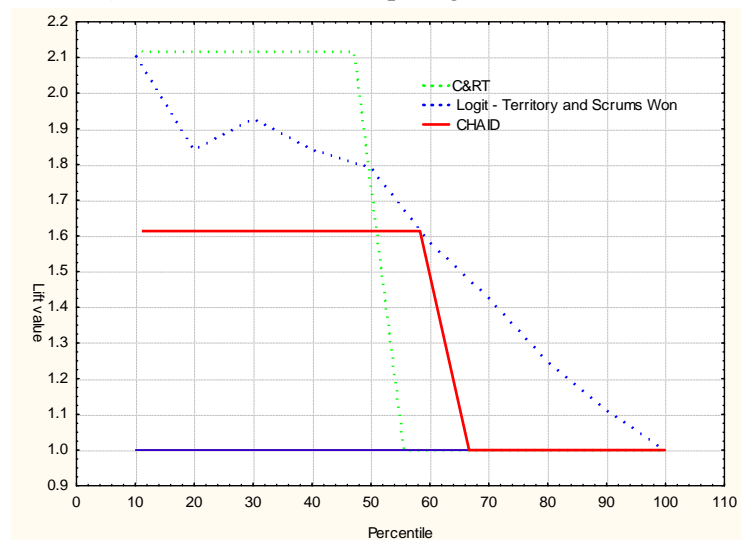
5. Summary

With a small data set such as this the conclusions reached by exploring the data interactively with simple tools are as good as those reached by using CHAID and C&RT. However, if the data set were larger then the tree based algorithms would be expected to show more subtle relationships between the variables than those that that can be seen with traditional methods. I believe that it would be worth while to assemble a larger data set and to use the CHAID and C&RT to find more interesting relationships.

6. References

[1]   StatSoft, Inc. *STATISTICA* (data analysis software system), version 6. (2004).

[2]   L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, "Classification and regression trees" Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software (1984)

[3]   G. V. Kass, "An exploratory technique for investigating large quantities of categorical data" *Applied Statistics,* **29** (1980), 119-127.

STEPS TOWARDS FAIRER ONE-DAY CRICKETING MEASURES OF PERFORMANCE

Tony Lewis
The Business School, Oxford Brookes University
Wheatley, Oxford OX33 1HX
United Kingdom

Abstract

The traditional batting measures of performance in one-day cricket are the average number of runs for each dismissal and the batsman's rate of scoring runs per 100 balls received, known as his strike rate. For a bowler the traditional measures are the average number of runs per wicket taken, the average number of balls between wickets, known as the strike rate, and thirdly, the economy rate representing the average number of runs conceded per six-ball over. These measures are known to be inadequate and rarely tell the full story of player performances as the context in which runs are scored or conceded and wickets lost or taken are not taken into account. The methodology of the Duckworth/Lewis (D/L) method that is used to reset targets in interrupted one-day matches can be utilised to evaluate better the performances of players, having regard for the stages of an innings that runs are earned and conceded and wickets lost or taken. The D/L method uses the concept of a team's combined resources of wickets and overs available in order to produce fair revised targets. This paper applies these proposed measures both to a single match and to a series of matches. It shows that the traditional measures can be seriously misleading in evaluating comparative player performances over a series, and totally inadequate in comparing batsmen and bowlers, and in combining an individual's performance with both the bat and the ball.

1. Introduction

It is assumed that the reader is familiar with the game of one-day cricket and its traditional measures of performance of both batsmen and bowlers. For readers not so acquainted, the rules and terminology of one-day cricket can be read in Engels [1], which, traditionally, is referred to as 'Wisden'. Matches produce traditional measures of performance, usually referred to as 'averages', which include runs scored per completed innings and batting strike rate for batsmen. Bowlers' 'averages' include runs conceded per wicket taken, strike rate and economy rate, namely the runs conceded per six-ball over. It is recognised within the game of cricket that these averages are insufficient to assess player performance and are sometimes misleading. The context in which runs are scored or conceded and wickets lost or taken is not taken into account. In addition, the comparison of batsmen with bowlers is impossible with these measures.

As an example of the relevance of context, for a batsmen to lose his wicket off the last ball of a one-day innings is no worse a result than scoring zero runs, yet his batting average suffers by the addition of one to the denominator. Similarly, a bowler taking the wicket off the last ball of the innings deserves no more credit than that of restricting the batsman to no runs. Yet the bowler adds one to the denominator of his average to reduce the number of runs conceded per wicket and improves his strike rate.

Such is the concern for these personal averages by the players that they rarely adopt logical team tactics at the end of their innings. Usually, wickets in hand at the end of the innings are not relevant to the score required by the team batting second to win the match. Consequently, for the good of the team, batsmen should take all risks to score as many runs as possible from the last few balls, to the extent of continuing to run until one of the batsmen is run out on the last ball (unless the ball has been hit for a boundary).

2. Illustrative Examples

The VB (Victoria Bitter) Series of 2003 involving Australia, England and Sri Lanka will be used to illustrate some of the issues in evaluating player performances and also, later in the paper, features of the proposed measures. Table 1, provides the standard batting and bowling averages for the top players for this competition. It will be seen from this table that Lehmann and Hogg, headed the batting averages. Both of these players scored a reasonable quantity of runs in a few innings but were not dismissed in two of these innings. Consequently their respective batting averages are somewhat distorted and 101.00 as an 'average' number of runs scored each innings by Lehman somewhat exaggerates his performance. Similar comments are appropriate to the contributions of Hogg and Samaraweera, and even to Bevan's contribution whose 221 runs actually came from seven innings that included three 'not outs'.

The contributions of Hayden, Jayasuriya, and Knight are based on many more dismissals and so can be considered as more reliable indicators of these players' batting performances over the whole of the 14-match VB 2003 series.

A deficiency, however, of the standard batting averages is that a player does not get a ranking until he has been dismissed at least once; an 'infinite' batting average is not acceptable. Clarke and Nissanka are placed at the bottom of the players in the list of batting averages in Table 1 for this reason.

The bowling averages of Table 1 place Bracken at the head of the list with an average of 13.50 runs conceded per wicket taken. Behind him are several more of the regular bowlers in Muralitheran, Lee and Caddick in addition to 'all-rounder' Lehmann. Bracken's low average comes partly from his playing in only two games. In one of these [1] he took 3 wickets for 21 runs in a low-scoring innings. His top-place ranking has come about from one very good performance in his small total input to the tournament. This illustrates, further, how averages are prone to such distortions when based on small volumes of data. The regular bowlers Lee and Caddick played in nine matches each, consequently the effects of single performances are smoothed over the larger number of games. This means that bowling averages provide a meaningful measures of performance only for bowlers whose input has been substantial. Bowlers with low volume input cannot be fairly compared and ranked when using the standard bowling average.

Further aspects of Table 1 include the issue that Lee was the most prolific wicket taker in the series but was ranked fourth in the averages. Also Jayawardene and Mubarak were placed at the bottom of the list of players in Table 1. Neither of these bowlers took any wickets and so is not ranked in terms of bowling average regardless of the number of runs conceded, even though their 'averages' could be regarded as being infinite. Yet, a player who takes no wicket but bowls very economically is almost as valuable in the context of one-day cricket as a bowler who takes many wickets. Bowling averages do not reflect this aspect of a bowler's performance.

When viewing the averages of Table 1 one asks to what extent are these rankings truly representative of player's contributions to the teams' efforts? Although Lee took many wickets, examination of the scorecards for these matches [1] shows that many of these wickets were those of the lower-order batsmen, which are regarded as not as valuable as the wickets of the higher-order batsmen.

Table 1: Batting and bowling averages of the VB Series 2003

Batting Averages	Bowling Averages

Rank	Player	Country	Runs	Innings	Times out	Avge	Player	Country	Runs	Wkts	Avge
1	Lehmann	Aus	202	4	2	101.00	Bracken	Aus	54	4	13.50
2	Hogg	Aus	89	3	1	89.00	Muralith'n	SriL	107	7	15.29
3	Hayden	Aus	458	9	7	65.43	Lehmann	Aus	83	5	16.60
4	Bevan	Aus	221	7	4	55.25	Lee	Aus	338	18	18.78
5	Sam'aweera	SriL	55	2	1	55.00	Clarke	Aus	24	1	24.00
6	Jayasuriya	SriL	359	7	7	51.29	Caddick	Eng	324	13	24.92
7	Knight	Eng	461	10	9	51.22	Williams	Aus	178	7	25.43
8	Collingwood	Eng	289	8	6	48.17	Watson	Aus	262	10	26.20
9	Martyn	Aus	276	8	6	46.00	Fernando	SriL	355	13	27.31
10	Gunaw'ene	SriL	45	1	1	45.00	Blackwell	Eng	303	11	27.55
11	Gilchrist	Aus	310	8	7	44.29	Warne	Aus	167	6	27.83
12	Ponting	Aus	306	8	7	43.71	Hogg	Aus	341	12	28.42
13	Watson	Aus	86	5	2	43.00	Colling'd	Eng	57	2	28.50
14	White	Eng	121	4	3	40.33	Anderson	Eng	373	13	28.69
15	Bichel	Aus	39	2	1	39.00	Lakshitha	SriL	65	2	32.50
16	Jayaw'ene	SriL	261	8	7	37.29	Bichel	Aus	230	7	32.86
17	Sangakkara	SriL	234	8	7	33.43	Vaughan	Eng	67	2	33.50
18	Atapattu	SriL	265	8	8	33.13	Nissanka	SriL	174	5	34.80
19	Vaughan	Eng	130	4	4	32.50	Jayas'ya	SriL	300	8	37.50
20	Trescothick	Eng	316	10	10	31.60	de Silva	SriL	197	5	39.40
21	Maher	Aus	94	4	3	31.33	Vaas	SriL	352	8	44.00
22	Hussain	Eng	278	10	10	27.80	Harmison	Eng	228	5	45.60
23	de Silva	SriL	123	5	5	24.60	Gunar'ne	SriL	184	4	46.00
	Clarke	Aus	39	1	0	-	Jayaw'ne	SriL	8	0	-
	Nissanka	SriL	12	1	0	-	Mubarak	SriL	29	0	-

A further aspect of the standard averages is that they do not provide a way of evaluating objectively, whether the top batsman has performed better than the top bowler has. Neither do they have the capability of combining a player's performances with bat and ball, such as those of Lehman who features in the top three of both the batting and bowling averages.

Johnston *et al* [2] proposed a player evaluation mechanism using Clarke's [3] dynamic programming model in one-day cricket but this mechanism has not had any impact on the standard performance measures outlined above. This paper looks at two alternative approaches to the evaluation of player performances that take into account the context of runs scored or conceded and wickets taken and lost. One of these approaches has the facility to be able to compare directly the performances of batsmen and bowlers.

3. The Duckworth/Lewis method

The mechanisms to be proposed in this paper for evaluating player performances are based on the Duckworth/Lewis (D/L) method of target resetting in one-day cricket [4,5]. The standard D/L model is represented by

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

In this model, $Z(u,w)$ represents the average further runs obtained in the u ($0 \le u \le 50$ in typical one-day matches) remaining overs when w ($0 \le w \le 9$) wickets have been lost where Z_o and b are positive constants. $F(w)$ is a positive decreasing step function with $F(0) = 1$. This function is interpreted as the proportion of runs that are scored with w wickets lost compared with that with no wickets lost and, hypothetically, infinitely many overs available. That is $F(w) = \lim_{u \to \infty} Z(u,w) / Z(u,0)$. As subsequent discussion will focus on the ball-

by-ball contribution of players then writing $i = 6u$, $(0 \leq i \leq 300)$ will represent the ith remaining ball of the innings.

3.1 Players' nett contributions

At any stage of an innings, the worth of a player's contribution per ball can be evaluated using (1). If there are i balls remaining and w wickets have been lost then the expected runs, r_i, from the next ball will be either

$$r_i = Z(i, w) - Z(i - 1, w) \qquad\qquad \text{(2a), or}$$

$$r_i = Z(i, w) - Z(i - 1, w + 1) \qquad\qquad \text{(2b)}$$

depending respectively on whether the batsman survived the ball or lost his wicket.

When i is not much less than 300 and $w=0$ then (2a) will be small and (2b) will be large reflecting the high value of dismissing a top-order batsman early in an innings. Conversely as i approaches zero (2a) will be relatively large and (2b) will be relatively small provided w is not approaching 9 which is the situation when the last men are then at the wicket.

If the batsman scores s_i runs from that ball then his nett contribution, c_i, for that ball is either

$$c_i = s_i - r_i = s_i - \{Z(i, w) - Z(i - 1, w)\} \qquad\qquad \text{(3a), or}$$

$$c_i = s_i - r_i = s_i - \{Z(i, w) - Z(i - 1, w + 1)\} \qquad\qquad \text{(3b)}$$

depending on whether or not the batsman survives that ball.

Early in a team's innings, because expectation of run scoring is less than a run a ball, a single run will make a positive contribution. On the other hand, towards the end of an innings, with wickets in hand, a single run will result in a negative contribution as more than a run a ball is then expected from the model in (1).

The bowler's contribution to the team effort can be evaluated as the negative of the batsman's contribution since cricket is a zero-sum game. However because 'extras' are treated unequally between batsman and bowler there needs to be a slight adjustment to the contributions of both sets of players in order that the total nett contributions of completed innings is indeed zero.

3.2   Allowing for 'extras'

There are several forms of 'extras' (alternatively called 'sundries'). The bye and leg bye are runs that count neither for the batsman nor against the bowler yet are aggregated with other runs that form the total for the team. The wide ball and no ball result in runs that are debited against the bowler, are added to the batting side's score but are not credited to the batsman facing at the time. It is possible that a batsman, in addition, scores runs at a no ball but not at a wide ball. Such balls do not count in the total of number of balls bowled in the innings to date. Consequently, the ball is bowled again until it is legitimate during which time several runs may be added to the team total and to the players' individual totals.

Let e_i be the number of extras conceded by the ith legitimate ball remaining $(1 \leq i \leq 300)$. Of these let g_i be the combined number of byes and leg byes and h_i be the combined number of wides and no balls. Thus $e_i = g_i + h_i$ and the team's total $S = \Sigma s_i + \Sigma e_i = \Sigma s_i + \Sigma g_i + \Sigma h_i$, the summations being over all 300 balls of the typical one-day innings. For an average one-day international (ODI), the total score of Team 1 is $Z(50,0) = 235$ [http://www.cricket.org/link_to_database/ ARCHIVE/ABOUT_CRICKET/RAIN_RULES/ Duckworth_Lewis_2002.html]. If $S > Z(50,0)$ then the batting team have exceeded expectations from the D/L model and conversely if $S < Z(50,0)$. If there were Σe_i extras in a particular innings then the batting side have performed above average if the sum of the batsmen's contributions exceeds $\{Z(50,0) - \Sigma e_i\}$. Thus, the adjusted contributions of batsmen

for the ith ball remaining are achieved from $c_i = s_i - r_i\{Z(50,0) - \sum e_i\}/Z(50,0)$. Since the average of the total runs, $\sum r_i$, is Z(50,0) it is easily seen that $\sum c_i = S - Z(50,0)$.

In the main, the nett contribution for bowlers is the negative of that for the batsmen facing the same balls, except that extras weigh more heavily against bowlers. The extras not counting against the bowler, in a particular innings, are represented by Σg_i. Thus, the nett contribution by a bowler on any ball is represented by $\bar{c}_i = \dfrac{r_i\{Z(50,0) - \sum g_i\}}{Z(50,0)} - (s_i + h_i)$. It is again easily seen that the total of the nett contributions of the bowlers over the 300 balls of the innings is $\sum \bar{c}_i = Z(50,0) - S$. Consequently $\sum c_i + \sum \bar{c}_i = 0$ and, after allowance for the different treatment of extras between batsmen and bowlers, the zero-sum nature of the game is captured.

If an innings is not completed, because the team batting second has won the match within their 300 balls, then $\sum c_i + \sum \bar{c}_i$ may differ slightly from zero. This discrepancy is ignored in subsequent discussion but the effect on conclusions is likely to be negligible.

4. Average batting resource consumption and bowling resource contribution

The D/L methodology [4,5] converts the *Z(u,w)* in (1) into proportions of the average total in 50 overs, *Z(50,0)* to produce

$$P(u, w) = Z(u, w)/Z(50,0).$$ (4)

This function is tabulated in percentages and used for target resetting purposes in the Duckworth/Lewis method [http://www.cricket.org/db/NATIONAL/ICC/RULES/ODI_PLAYING _CONDITIONS_D-L_METHOD.pdf]. The function in (4) is interpreted as the percentage of resources (relative to those for a 50-over innings) that a team has remaining for the *u* overs it has left when *w* wickets have been lost.

Again writing *i=6u,* the consumption of resource, p_i, for the ith ball remaining is represented by either

$$p_i = P(i, w) - P(i - 1, w)$$ 5(a), or
$$p_i = P(i, w) - P(i - 1, w + 1)$$ 5(b)

depending, respectively, on whether or not the batsman survives that ball.
Comments that were made previously in relation to the effects of extras are relevant, in a similar way, in the discussion of resource percentages. Resources are not consumed by the batsman nor contributed by the bowler for a no ball or a wide but the extra runs are debited against the bowler. Runs made by the batsman off the no ball are credited to the batsman and debited against the bowler as is normal.

For any ball *i* remaining, the batsman scores s_i runs and consumes p_i resources. His average run contribution to the team's total per unit of resource consumed can be assessed by $\sum s_i / \sum p_i$ where the sums are over all the number of legitimate balls that the batsman has faced.

202

The bowlers' performances are evaluated using the total of the runs that are conceded, s_i, plus extras debited against the bowler, h_i, so that a bowler's average of runs conceded per unit of resource contributed is measured by $\sum (s_i + h_i) / \sum p_i$ where the sums are over the number of legitimate balls the bowler has bowled.

Beaudoin and Swartz [6] have produced similar measures to these batting and bowling resource averages. Their measures involve multiplying both of these terms by 100 and calling them 'Runs per Match' (RM). These numbers can be interpreted separately as the number of runs earned or conceded had players done all of the batting or bowling themselves.

5. Example

As an example of the two proposed measures, they are applied first in a single match from the VB Series 2003 on 25 Jan 2003 between Australia and England in Melbourne. This match was the second of the best-of-three finals in which Australia scored 7/229 in their 50 overs. England scored 10/224 in 49.3 overs and lost by 5 runs. Table 2 summarises the Australian innings in terms of the players' batting performances according to the standard measures and the two proposed measures.

Table 2: Australia's innings - batting measures, 25 Jan 2003

Batsman	Runs	Out?	Balls faced	Strike rate	Nett batting contribution	Batting resources	Batting resource average
Gilchrist	26	1	32	81.25	+0.99	10.8	2.40
Hayden	69	1	91	75.82	+21.39	20.6	3.34
Ponting	1	1	8	12.50	-16.78	7.7	0.12
Martyn	11	1	14	78.57	-10.45	9.3	1.18
Bevan	10	0	21	47.61	+2.78	3.1	3.22
Symonds	8	1	29	27.58	-16.95	10.8	0.74
Hogg	71	0	77	92.20	+16.90	23.4	3.03
Warne	0	1	1	0.00	-7.38	3.2	0.00
Lee	18	1	17	105.88	+2.70	6.6	2.72
Bichel	11	0	10	110.00	+0.81	4.4	2.50
Williams					0	0	
Byes & Leg Byes	3						
Wides & No balls	1						
Totals	229	7			-6.0	99.9	

In this match, since only one innings is considered, the usual batting average (total runs divided by total number of dismissals) cannot be captured; only the total runs. Hayden scored 69 runs in 91 balls, a strike rate of 75.82 per hundred balls. Hogg scored 71 runs in 77 balls and was not out. His strike rate was 92.20. (Note that cricketing measures are traditionally rounded down to two decimal places.)

Of the two batsmen who has performed the better? Based on runs for their wicket and strike rate, Hogg comes out on top, but these measures ignore the stage of the innings that the runs were scored. Hayden batted during more than 40 of the 50 overs but Hogg batted during the last 22 overs and received the very last ball of the innings. As one of the opening batsmen Hayden needed to ensure that wickets were not given away cheaply at the beginning of the innings thereby avoiding pressure on the later batsmen. On the other hand Hogg had the 'luxury' of knowing that, particularly in the last few overs, he could throw caution to the wind

and take inordinate risks to boost the Australian total, which became 229. Such issues are seen as shortcomings in evaluating the worth of a player's contribution to his team's results and consequently evaluating his performance.

On the standard measures, Hogg is ranked as having performed better than Hayden. When the stage of the innings is taken into account, however, via the D/L methodology, Hayden is ranked ahead of Hogg. Hayden's nett contribution is +21.39 runs compared with Hogg's +16.90. Using the batting resource average Hayden has contributed 3.34 runs per unit of resource consumed but Hogg contributed only 3.03. Using the two measures proposed in this paper, Hayden's performance would be ranked as superior to that of Hogg.

It will be noted that Bevan made a modest contribution of 10 runs, yet his batting resource average is almost as high as Hayden's. Bevan was not dismissed, having retired hurt in this match, thereby not consuming resources for the loss of his wicket. However, his nett batting contribution of +2.78 places his small contribution into the right perspective and places him lower down the rankings of player batting performance.

It will also be noted, in Table 2, that Australia's total of 229 was six runs below the average of $Z(50,0) = 235$. The total nett contributions of the batsmen correspond to this slight underachievement to the average, by six runs.

Table 3: VB Series 25 Jan 2003 England's innings - bowling measures

Bowler	Runs	Wkts	Average	Balls bowled	Strike rate	Nett bowling contribution	Bowling resources	Bowling resource average
Hogg	41	1	41.00	60	60	-3.35	16.5	2.48
Warne	58	2	29.00	60	30	-8.87	21.5	2.69
Lee	30	5	6.00	57	11.4	+35.28	28.6	1.04
Bichel	42	0	-	60	-	-16.16	11.3	3.71
Williams	46	1	46.00	60	60	+0.64	20.5	2.24
run out		1				+3.50	1.5	
Byes & Leg Byes	7							
Totals	224	10				+11.00	99.9	

Table 3 illustrates some of the issues concerning bowling performance measures. If a batsman is run out then he is debited with the loss of the wicket and the corresponding resource consumption. However, the bowler is not credited with the wicket nor its resource contribution. Table 3 includes these items for the purposes of accounting for the number of wickets lost and associated resources.

Note that in Table 3, England's total of 224 was 11 runs below the 235 ODI average, consequently the nett contributions of Australia's bowlers was a positive 11.00.

The various bowling performance measures, both traditional and proposed, show consistency in this match and there is unlikely to be any dispute in their message; Lee's performance is assessed as being the best by all the measures. Warne had a better bowling average and strike rate than Hogg, although according to the D/L measures, Hogg's performance is rated better than Warne's, reflecting the notion that bowling economy, and not just wickets taken, is a very important factor in bowling performance.

Williams had an inferior bowling average to that of Hogg but by the D/L measures Williams'

performance has been rated as superior reflecting the stage of the innings when Williams conceded his runs and took his wicket compared with Hogg. Bichel would not normally be ranked as he took no wickets but, by his nett contribution and bowling resource average, his performance can be validly placed in context with his fellow bowlers.

6. Comparison of batsmen and bowlers

As indicated earlier, the traditional measures of performance cannot be used to compare or combine the separate performances for batting and bowling. The nett contribution, however, possesses these qualities so that batsmen and bowlers can be compared directly and overall performance in a particular match can be assessed objectively. This is an important characteristic as nearly every match has a 'man-of-the- match' award. In major tournaments, a media personality usually makes the judgement of this award. The nett contribution would be a valuable measure to assist in this decision.

Table 4 summarises the nett batting and bowling contributions (and the totals of these) of all the players from both Australia and England that played in the match on 25 January 2003. Looking first at the columns of nett batting contributions and nett bowling contributions, it can be seen that Lee's bowling contribution is the best of the bowler's column Just below this, however, it will be noted that Caddick made a strong bowling contribution for England.

Of the batsmen in this match the best nett contribution was from Vaughan with +24.35. In comparing the batsmen and the bowlers Lee's bowling contribution is not only the best of the bowler's but also is better than any nett batting contribution. As a consequence, his performance can be regarded as the best of the batsmen and the bowlers combined.

Table 4: Nett player contributions, Aus v Eng, 25 Jan 2003

Player	Country	Nett batting contribution	Nett bowling contribution	Match contribution
Lee	Aus	+2.70	+35.28	+37.98
Caddick	Eng	-2.00	+31.20	+29.20
Hayden	Aus	+21.39	0	+21.39
Stewart	Eng	+19.91	0	+19.91
Vaughan	Eng	+24.35	-8.49	+15.86
Hogg	Aus	+16.90	-3.35	+13.55
Collingwood	Eng	+4.27	0	+4.27
Irani	Eng	-12.54	+16.33	+3.79
Bevan	Aus	+2.78	0	+2.78
Gilchrist	Aus	+0.99	0	+0.99
Williams	Aus	0	+0.64	+0.64
Hussain	Eng	-1.06	0	-1.06
Blackwell	Eng	-3.50	-0.98	-4.48
Flintoff	Eng	+1.87	-8.23	-6.36
Martyn	Aus	-10.45	0	-10.45
Bichel	Aus	+0.81	-16.16	-15.35
Warne	Aus	-7.38	-8.87	-16.25
Ponting	Aus	-16.78	0	-16.78
Symonds	Aus	-16.95	0	-16.95

| Trescothick | Eng | -17.00 | 0 | -17.00 |
| Knight | Eng | -20.41 | 0 | -20.41 |

6.1 Combining batting and bowling performances

An aspect of the nett contributions is that they have a common, ratio scale which provides an additive feature to assess the combined batting and bowling performances of players. Table 4 provides a summary of nett batting, bowling and combined match contributions of the players from Australia (Aus) and England (Eng) that took part in the match on 25 January 2003.

It will be noted that in addition to his bowling contribution Lee made a small positive contribution when batting that added to his strong overall contribution to the match.

The measure of total nett contribution as an indicator of all-round performance would be a potentially valuable aid to the adjudicator in determining the man-of-the-match award, especially when there are several players who have made positive contributions with either bat or ball or both.

Overall, the quantitative evidence of this match supports the decision of the adjudicator to award the 'man of the match' to Lee. Although the match scorecard [1] shows that four of his five wickets were the 'cheaper' lower order batsmen, which slightly devalues his bowling average, nevertheless he was instrumental in turning around a match that England looked very likely to win. No doubt the adjudicator was influenced not only by Lee's superior bowling average, as in Table 3, but also, subjectively, by his significant effect on the outcome of the match.

7. The proposed measures applied over the whole VB Series 2003

The complete VB series of 2002-03 was a 'round robin' competition between Australia, England and Sri Lanka, which resulted in a best-of-three finals series between England and Australia. After 14 matches the measures proposed in this paper can be compared with traditional performance measures. And the additional combined resource measure can be used to evaluate the 'player of the tournament' award. Squads for each team contained up to 18 players so that some 54 players took part in at least one match. For brevity only the top 25 are presented in order to discuss the relative merits of the measures.

7.1 Batting contribution

Table 1 provides the traditional batting averages and Section 2 highlights their shortcomings. Table 5 summarises the batting performances of the top 25 players of the tournament as ranked by their nett batting contributions. Table 5 also includes the standard batting averages, with rankings, and the batting resource averages, with rankings, for these players. The reader will be able to make several interesting comparisons of the relative rankings of players by these several measure. Some of these aspects will first be highlighted after which some general conclusions are drawn on the viability of the proposed measures as compared with the standard batting performance measures.

(i)   Clarke and Nissanka are listed amongst the players in Table 5 and are ranked 17th and 18th respectively according to nett batting contribution. It will be recalled that these players were not able to ranked by the standard batting average, but perhaps somewhat alarmingly, Nissanka heads the batting resource average rankings and Clarke is also highly placed by this measure.

(ii) According to nett contribution, Lehman falls down the rankings compared with the standard batting average as the effect of being 'not out' is reduced. Similarly for Hogg who descends to 20th place after the reduction of the effect, through the D/L methodology, of only having been dismissed once in his three innings.

(iii) Bevan's ranking of 4th in the standard batting average suffers partly from the issue in (ii). However, since the D/L model (1) expects more runs per ball towards the end of the overs available when there are wickets in hand then nett contributions from the later overs are lower. Bevan scores many of his runs towards the end of the team's innings and is frequently not out. Under the nett contribution measure the value of a wicket, in terms of loss of run-scoring potential, becomes less and less important as the team's innings progresses and declines to zero right at the end. These further aspects of the nett contribution measure are reflected in Bevan slipping from 4th to 22nd in the rankings of the tournament.

(iv) Conversely several recognised batsmen, who were quite low in the rankings according to the standard batting average, have been 'promoted' using the nett batting contribution measure. Hayden becomes top and Jayasuriya, Collingwood and Gilchrist are promoted to the next three positions. Similarly other recognised batsmen have their performances more highly rated than through the simple batting average.

Table 5: VB Series 2003: Batting measures of performance

Player	Country	Average	Standard rank	Res avge	Res avge rank	Nett contrib.	Nett contrib rank
Hayden	Aus	65.43	3	3.33	6	152.76	1
Collingwood	Eng	48.17	8	3.75	4	121.05	2
Jayasuriya	SriL	51.29	6	3.11	7	116.65	3
Gilchrist	Aus	44.29	11	3.39	5	105.25	4
White	Eng	40.33	14	4.52	3	62.29	5
Lehmann	Aus	101.00	1	2.85	9	46.28	6
Ponting	Aus	43.71	12	2.57	12	43.14	7
Trescothick	Eng	31.60	20	2.49	13	37.58	8
Knight	Eng	51.22	7	2.39	18	36.18	9
Bichel	Aus	39.00	15	5.56	2	23.22	10
Sangakkara	SriL	33.43	17	2.36	19	20.49	11
Jayawardene	SriL	37.29	16	2.31	21	20.19	12
Martyn	Aus	46.00	9	2.33	20	18.04	13
Watson	Aus	43.00	13	2.75	10	17.22	14
Atapattu	SriL	33.13	18	2.24	24	16.35	15
Vaughan	Eng	32.50	19	2.47	14	14.24	16
Nissanka	SriL	-	-	8.10	1	8.75	17
Clarke	Aus	-	-	2.85	8	7.97	18
Samaraweera	SriL	55.00	5	2.58	11	7.72	19
Hogg	Aus	89.00	2	2.45	16	7.46	20
Gunawardene	SriL	45.00	10	2.41	17	4.30	21
Bevan	Aus	55.25	4	2.19	25	3.58	22
Flintoff	Eng	16.00	30	2.46	15	1.87	23
Caddick	Eng	15.00	31	2.26	23	1.73	24
Muralitheran	SriL	3.00	42	2.29	22	0.31	25

7.2 Comments on batting measures of performance

(i)   Cricketing sense suggests that in this tournament the two proposed measures are better at ranking the players than the standard batting average is doing.

(ii)   A weakness in the batting resource average can be identified through Nissanka's topping of these rankings and the surprise elevation of Bichel to second place by this measure. One aspect of this measure is that the resource averages of these two players are based on few innings and so, in a similar way to the standard batting average, cannot be regarded as truly representative. Further, but perhaps more importantly, these two players batted late in their respective innings and in matches where there was no hope of winning. Runs earned have thus come from the consumption of very little resource. When most of the wickets have been lost and plenty of overs are in hand the D/L curves (1) and (4) have virtually zero gradient so that little is contributed to the denominator of the resource percentage average for the balls they received. This suggests that the measure is prone to the similar problems of distortion arising from single above-average performances within very few innings.

Examination of the commentary [http://uk.cricinfo.com/db/ARCHIVE/2002-03/OD_ TOURNEYS/VBS/SCORECARDS/AUS_SL_VBS_ODI6_09JAN2003_BBB- COMMS.html] shows that in the match between Australia and Sri Lanka Bichel came to the wicket as last man in when there were 8.3 overs left and the match itself was already effectively lost. Under no pressure, he scored 28 runs and was dismissed with three balls left when his wicket consumed practically zero resource. This one innings, in only two for Bichel during the tournament, emphasises the resource average measure's susceptibility to distortion from small amounts of data input. It also highlights the issue that the D/L percentage curves are flat when most of the wickets have been lost with many overs left so that runs earned by batsmen are almost free of resource consumption

This issue is also relevant to the measure of nett contribution but its effect is potentially less pronounced than in the batting resource average measure as it is additive and not part of a divisor.

(iii)   Subjectively, the nett batting contribution appears to be the more reliable as a measure of batting performance when used in 'round-robin' tournaments in which some players may not play in many matches.

(iv)   Nett batting contribution is more fairly accounting for all of a batsman's innings and putting much less weight on whether or not the batsman loses his wicket, which is the result of only one ball. It is also making allowance for the stage of the innings that batsmen score their runs.

(v)   The consumption of resource by losing a wicket at the final ball of the innings is no greater than the receipt of the final ball, encouraging risk-taking tactics that are appropriate at the end of an innings.

8. Bowling performance measures

The bowlers' performances are ranked under the two measures suggested by the D/L methodology and compared with the rankings by the standard bowling averages.

8.1 Bowling Averages

Table 1 provides the standard bowling averages of the VB Series 2003. Table 6 summarises all the main performance measures of the top 20 bowlers, as ranked by the nett bowling contribution measure proposed in this paper. The rankings by the standard bowling average and Bowling Resource Average are also included. It will be seen that there is some agreement in the rankings of these players under the three mechanisms but also significant variations that warrant comment.

(i)   Under the standard bowling average, Mubarak and Jayawardene do not qualify for a finite average, as they did not take any wickets, yet their performances can be evaluated, and ranked, by using the two D/L-based measures.

(ii)   The England bowler Caddick was ranked sixth under the standard bowling average but ranked first using nett bowling contribution. As the opening bowler, many of his wickets included the higher order batsmen with greater run scoring potential and whose dismissal attracts greater merit than the dismissal of the less skillful lower-order batsmen. This contributes to his top ranking by nett contribution

Table 6: VB Series 2003 Bowling measures and rankings

Player	Country	Bowl avge	Std avge rank	Bowl res avge	Bowl res rank	Nett contrib	Nett cont rank
Caddick	Eng	24.92	6	1.96	4	54.27	1
Bracken	Aus	13.50	1	1.22	1	48.29	2
Muralitheran	SriL	15.29	2	1.55	2	47.56	3
Lee	Aus	18.78	4	2.03	6	41.94	4
Blackwell	Eng	27.55	10	2.03	7	37.25	5
Williams	Aus	25.43	7	1.96	5	29.77	6
Lehmann	Aus	16.60	3	1.70	3	29.34	7
Bichel	Aus	32.86	16	2.08	8	22.68	8
Watson	Aus	26.20	8	2.23	9	6.65	9
Warne	Aus	27.83	11	2.24	10	3.78	10
Jayawardene	SriL	-	-	2.48	15	-0.40	11
Vaughan	Eng	33.50	17	2.33	11	-0.89	12
Clarke	Aus	24.00	5	2.56	18	-2.17	13
Mubarak	SriL	-	-	2.62	19	-2.97	14
Martyn	Aus	96.00	30	2.46	12	-6.88	15
Flintoff	Eng	56.00	26	2.72	21	-8.23	16
Symonds	Aus	82.00	28	2.52	17	-8.32	17
Collingwood	Eng	28.50	13	2.75	22	-10.11	18
White	Eng	54.00	25	2.46	13	-11.26	19
Lakshitha	SriL	32.50	15	2.83	24	-13.11	20
McGrath	Aus	51.33	24	2.50	16	-13.18	21
Arnold	SriL	102.00	31	3.71	33	-17.49	22
Jayasuriya	SriL	37.50	19	2.79	23	-19.47	23
Hogg	Aus	28.42	12	2.46	14	-24.97	24
Batty	Eng	120.00	33	2.99	29	-27.84	25

(iii)   The Australian bowler Lee, although the most prolific wicket-taker as seen in Table 1, is not ranked in the top three by any of the three measures – yet he received the 'Player-of-the-Series' award. Analysis of the scorecards [1], shows that many of his 18 wickets were taken towards the end of the oppositions' innings and so consisted of many 'tail-end' batsmen not possessing the batting expertise of the higher order batsmen.

(iv)　Bracken is ranked first by the standard bowling average and the bowling resource average. His major input came in one of only two games that he played which is distorting this 'average' whereas the averages of most other bowlers are based on several more matches. The nett bowling contribution measure makes some allowance for this and ranks Bracken second but well short of the top-placed Caddick.

8.2 Comments on bowling measures of performance

(i)　Cricketing sense suggests that the two proposed measures are better at ranking the players than the standard bowling average. A more appropriate and lower value is being placed on lower-order wickets than the equal value placed on wickets by the standard bowling average.

(ii)　The bowling resource average is still prone to distortion due to single outstanding achievements in low-volume input to a tournament

(iii)　Wickets taken towards the end of an innings are accorded lesser weight and hence lesser merit in the D/L-based measures so that on the last ball, bowling a 'dot' ball has equal merit to the taking of a sacrificed wicket.

9. A combined performance measure

As outlined in relation to the VB Series final, nett contribution can evaluate combined performances in both batting and bowling. Table 7 provides the batting and bowling nett contributions, and their sums totals, for the top 20 players over the whole of the 14 matches of the VB Series 2003.

Table 7: VB 2003: Nett contributions for batting, bowling and their total

Player	Country	Batting contribution	Bowling contribution	Total nett contribution	Rank
Hayden	Aus	152.76	0.00	152.76	1
Collingwood	Eng	121.05	-10.11	110.94	2
Gilchrist	Aus	105.25	0.00	105.25	3
Jayasuriya	SriL	116.65	-19.47	97.18	4
Lehmann	Aus	46.28	29.34	75.62	5
Caddick	Eng	1.73	54.27	56.00	6
White	Eng	62.29	-11.26	51.03	7
Bracken	Aus	0.00	48.29	48.29	8
Muralitheran	SriL	0.31	47.56	47.87	9
Bichel	Aus	23.22	22.68	45.90	10
Ponting	Aus	43.14	0.00	43.14	11
Trescothick	Eng	37.58	0.00	37.58	12
Knight	Eng	36.18	0.00	36.18	13
Lee	Aus	-7.95	41.94	33.99	14
Williams	Aus	0.00	29.77	29.77	15
Watson	Aus	17.22	6.65	23.87	16
Sangakkara	SriL	20.49	0.00	20.49	17
Jayawardene	SriL	20.19	-0.40	19.79	18

| Atapattu | SriL | 16.35 | 0.00 | 16.35 | 19 |
| Vaughan | Eng | 14.24 | -0.89 | 13.35 | 20 |

This table provides information to assist in making judgements concerning players' performances over the complete VB Series. It must be stressed that just the batting and bowling performances are included in this analysis. Other aspects of players' games such as wicket keeping, fielding and catching are not accounted for and are more difficult to quantify. Consequently, assessment of overall player performances needs to make subjective allowance for these further aspects of the game. Nevertheless, it would seem that the Australian bowler Lee's award as the 'Player of the Series' ignored the strong contributions made by many other players in the series. In Table 7, Lee is ranked 14th overall with a total nett contribution of +33.99 well below the nett contributions of countrymen Hayden and Gilchrist, below Englishman Collingwood and below Sri Lankan Jayasuriya. The Australian player Lehman was the highest ranked player who made positive nett contributions in both batting and bowling. Consequently, it could be argued that he should be regarded as being the best all round player in the tournament.

It must be noted that the batting and bowling resource averages, being ratios, do not easily lend themselves to the aggregation of performances in the two disciplines. Consequently they cannot easily be used to evaluate all-round performance. Beaudoin and Swartz [6] hint that the differences of the RMs of a player provide some indicator of the players' all-round performance, but this first requires an arbitrary decision on whether or not a player purports to be an 'all-rounder'.

10. Summary

The standard measures of batting and bowling average are simple to calculate and have been used to assess player performances over many years. In the one-day form of cricket, they provide some indication of a player's abilities in batting and bowling, for a particular match, for a series, and even throughout their careers. These measures, however, are insufficient in that they do not take into account the context in which runs are made or conceded and wickets are taken or lost. Compared with the longer version of the game, there are more occasions in which the performance of the individual needs to take second place to the strategy for the team in its objective in trying to win the match. Further, these batting and bowling averages are incompatible for the purposes of comparing performances in the two disciplines and combining them in order to assess all-round performance.

This paper has proposed performance measures based on the Duckworth/Lewis model of one-day cricket and its associated resource percentage table. Using several matches over a series, players' performances can be more reliably evaluated taking into account the context in which runs and wickets are earned or conceded. The measure nett batting and bowling contribution has the considerable virtue of being in the same units and hence provides a common and additive scale of measurement for the comparison and aggregation of batting and bowling performances.

The second measure proposed in this paper is the resource average. This is the average of runs earned or conceded per unit of resource consumed by batsmen or contributed by bowlers. This measure, equivalent to the runs per match measure of Beaudoin and Swartz [6], may well have considerable virtue in the longer term where the ratios are not based on small divisors. Evidence from the tournament considered in this paper, however, suggests that it is not as reliable a measure of performance as the nett contribution in individual matches or in a short series. Further, it does not easily lend itself to the task of comparing, or combining the disciplines of batting and bowling.

Further work is ongoing into evaluating the two measures proposed in this paper in relation to their viability over the longer term. This includes addressing not only whether the more a cricketer plays, affects his ranking , but also an investigation into the relevance of an aspect of one-day cricket in that bowlers are limited to one-fifth of the overs available but batsmen have no such limit on their opportunities to contribute to the team performance.

The methodology within this paper requires data on the outcome of every ball of a match. Although this is available in verbal form on the Internet [http://www.cricket.org/db/ARCHIVE/], such availability of ball-by-ball data in electronic form is an essential element in undertaking the extensive analysis of this methodology. With the advent of more user-friendly software interface systems and the now ubiquitous laptop computer, serious consideration could well be given to the development of the provision of such data by comprehensive cricket scoring software. For the short to medium term, this facility may well be able to be negotiated through existing sports data-collection organisations such as Press Association (Sports), UK and Champion Data, Australia. A further necessity is the development of software that will process this ball-by-ball data accurately and quickly using the methodology of this paper. This is regarded as essential so that the summary statistics and player rankings can be produced in a timely manner following the conclusion of the many one-day matches that are now played regularly around the world.

If these developments are successful, then it is felt that the nett contribution measure, in particular, perhaps has the potential to be included into the methodology of the popular PriceWaterhouseCoopers (PwC) ratings of player performances. [http://www.pwcglobal.com/uk/ eng/ins-sol/spec-int/cricket/ info/hfrw/ t_htrw.html]

Acknowledgements

11. References

[1]   M. Engels (Ed), Wisden Cricketers' Almanack, John Wisden & Co. Ltd.: Hampshire.(2004)

[2]   M.I. Johnston, S.R. Clarke and D. Noble, " Assessing player performance in one-day cricket using dynamic programming", *Asia Pac J of Op Res,* **10** (1993), 45-55

[3]   S.R. Clarke, "Dynamic Programming in one-day cricket: optimal scoring rates", *J Opl Res Soc*, **39** (1988), 331-337

[4]   F.C. Duckworth and A.J. Lewis, "A fair method for resetting the target in interrupted one-day cricket matches", *J Opl Res Soc* (1998), **29** (1998), 220-227

[5]   F.C. Duckworth and A.J. Lewis, "A successful Operational Research intervention in one-day cricket", *J Opl Res Soc* **35** (Forthcoming 2004)

[6]   D. Beaudoin and T.B. Swartz, "The best batsmen and bowlers in one-day cricket", *J SAfr Stat Soc,* **37** (2003), 203-222

1.1. Copyright

ON OPTIMAL RACE PACE

R. Hugh Morton
Institute of Food, Nutrition and Human Health
Massey University, Palmerston North
New Zealand
H.Morton@massey.ac.nz

Abstract

Conventional wisdom has been that for an athlete to cover a given race distance in minimum time, a constant pace strategy is required. In practice, to win a real race tactical reasons usually dictate pace changes. Nevertheless most successful athletes have set personal best times in races where pace has been observed to vary somewhat. Mathematical proofs of constant pace optimality have assumed a certain very simple model of energy availability, and this seems to be their crucial determinant. However, more complex and better fitting bioenergetic models are now available. For one of these, the 3-parameter critical power model, constant pace is far from optimal, simple step changes in pace are better, and the optimal race pace appears to be an all-out strategy. While few athletes would accept this, some recent physiological evidence suggests that this notion may not be as far-fetched as it seems.

1. Introduction

It is usually assumed that an athlete competing to win a race (running will be considered here, but the results apply equally to swimming, cycling, etc.) will attempt to cover the specified distance, d, in minimum time, subject to certain constraints on the consumption and availability of expendable energy. If these aspects are well posed, the selection of the optimal race pace can be determined mathematically without great difficulty. An athlete who, in so doing, achieves a personal best yet does not win the race has obviously been beaten by a better athlete. However winning is the more usual objective, in which case maybe the athlete should be attempting to cover the specified distance in maximum time subject to the additional constraint that no other competitor completes the distance in a lesser time. This is not such an easy problem to solve, and will be postponed to another time.

Nevertheless, in the context of time minimisation, the question of optimal race pace selection naturally arises, and it is widely believed that a constant pace strategy is optimal. However athletes in real races are invariably seen to vary their pace, either for tactical or bioenergetic reasons. This paper addresses the role that bioenergetic assumptions make on the selection of an optimal pace strategy, with the intent of questioning the constant pace belief.

2. Prior energetic models and constant pace optimality

The simplest bioenergetic model of energy availability is one of a rate limited aerobic energy source, together with a capacity limited anaerobic source. Normally these limits are specified in units of power (watts) and work (joules) respectively, but for running their analogues are expressed in velocity and distance units; b metres/sec and a metres respectively. Additionally it is assumed that the limit b is achievable instantaneously, and that a is completely expendable. Exhaustion is assumed to occur when the anaerobic source has declined to zero. This model is known as the Critical Power Concept, formulated originally by Monod and Scherrer [9] and reviewed by Hill [5]. Under these conditions it can be easily shown that if the athlete runs at a steady pace v metres/sec ($v > b$) until exhaustion, then endurance time, t sec, and the other quantities are interrelated by the following equations and their rearrangements:

$$t = a/(v - b)$$

$$d = a + bt \qquad \text{and}$$

$$v = d/t$$

Fukuba and Whipp [4] demonstrate that this set of equations is optimal, in that no better time for given distance can be achieved. More interestingly perhaps, they also show that if v is allowed to vary in a stepwise fashion, up and/or down and in few or many steps, exactly the same minimum time is achievable provided v never drops below b.

However there are a number of unrealistic assumptions embedded in the Critical Power Concept. For one, in fact b is not achievable instantaneously. When expressed in watts, b takes several minutes to become available through the oxygen delivery system, and even when expressed in metres/sec it still takes several seconds to be reached by virtue of body acceleration requirements. Keller [6] has addressed both these issues, demonstrating that the appropriate optimal velocity profile consists of three segments; a short period of a few seconds of maximal acceleration, followed by a long period of constant (but not all out) pace until exhausted, finishing with a very short period in which the athlete collapses over the finish line.

Thus as a generalisation, constant pace still appears optimal. We shall see shortly, that addressing other shortcomings of the Critical Power Concept by making a minor but very realistic change in the bioenergetics, constant pace is no longer optimal.

3. The 3-parameter Critical Power Model

The original Critical Power Concept implies that if the athlete exercised at or below the limiting aerobic power (or ran at or below the limiting velocity) b, endurance time tended to infinity. It has been frequently demonstrated, for example by McLellan and Cheung [8], that few individuals can endure at b for more than about 30 to 60 minutes. Another implication of the Concept is that very high powers or velocities may be achieved for very short periods. This quite clearly is not so, for everyone's power output or running velocity is limited above, even if exercise is of very short duration. Another fallacy is that all of a is expendable. Indeed as long ago as 1971, Saltin and Karlsson [13] showed that at the point of exhaustion not all of a had been expended. Most intriguingly they also discovered that the higher had been the power output requirement at exhaustion, the more of a remained unexpended, and vice versa. This observation has been confirmed many more times since then.

Thus Morton [12] proposed a modification to the Critical Power Concept by adding a third parameter P_{max}, representing a maximal instantaneous power output, and being the upper limit to the muscular strength capability of the individual. Correspondingly there would exist some v_{max} (> b) representing the upper limit to the individual's running pace. The bioenergetics are then modified such that the maximal achievable power output (or running velocity) declines linearly from P_{max} (or v_{max}) to b as the anaerobic energy store is expended down from a to zero respectively. Exhaustion occurs at the instant when the power output (or running velocity) demanded equals the maximal that could be achieved with the anaerobic store as it is at that same instant. It will be readily appreciated that a cannot now be completely expended, except by exercise only just above b, and that the higher the power output above b at exhaustion, the more of a that remains unexpended. Morton [12] showed that if the athlete runs at constant pace v (> b) throughout,

then **d, t** *and* **v** *are interrelated by the following equations and their rearrangements:*

$$d = at/(t - k) + bt$$
(1)

$$t = a/(v - b) + k \text{ and}$$

$$d = v/t$$

where $k = a/(b - v_{max}) < 0$, is the horizontal asymptote of the general rectangular hyperbolic equation:

$$(t - k)(v - b) = a$$

which generates the system. It will be noted that as P_{max} (or v_{max}) tends to infinity, k tends to zero from below and the above equations revert to the original three.

The question of interest now is whether some non-constant pace strategy can result in a time less than that given by the solution for t from equation (1).

4. Pace changes under the 3-parameter Critical Power Model

Consider the situation where the athlete is allowed to make deliberate changes in pace during the course of a race, with the intent of reducing the time taken to cover the required distance. In a real race such variations may be many and varied, but for the ease of mathematical analysis, I shall restrict variations to the simplest kind.

4.1 Stepwise changes

Following the example of Fukuba and Whipp [4], suppose the athlete is allowed one stepwise change in pace by first running a distance d_1 $(0 < d_1 < d)$ at the constant pace v_1 $(> b)$ given by $v_1 = d_1/t_1$ where t_1 is the appropriate solution for t from equation (1).

At this point t_1 the anaerobic store has then been run down from a to $a_1 = a - (d_1 - b\,t_1)$.

Next the athlete runs the remaining distance $d_2 = d - d_1$ at another constant pace v_2 $(> b$, and necessarily different from v_1), given by $v_2 = d_2/t_2$ where t_2 is the corresponding solution for t from equation (1). At t_2 the anaerobic store has now run down further to $a_2 = a_1 - (d_2 - bt_2)$.

The question now is whether $t_1 + t_2$ given by the sum of these two specific solutions for t from equation (1) is less than, equal to, or greater than, t given by the solution for t from the general form of equation (1). These equations, while not unduly complex, are quite messy. Rather than attempt an algebraic comparison, let us look at a numerical example.

Consider an individual with attributes $a = 200$ metres, $b = 3$ metres/sec and $v_{max} = 8$ metres/sec. From these we can evaluate $k = -40$ sec, and these are all values appropriate to an average healthy adult. The table below gives the results over a range of selected values of d_1. From these t_1 and v_1 are found from equation (1), then a_1, etc.

Table 1: Effects of a single step change in pace on time to run 300 metres

d_1	t_1	v_1	a_1	d_2	t_2	v_2	a_2	$t_1 + t_2$
300	60.00	5.00	80.00	0	0.00	0.00	80.00	60.00
290	57.38	5.05	82.15	10	2.02	4.96	78.20	59.40
250	47.24	5.29	91.71	50	10.37	4.82	72.82	57.61
200	35.38	5.65	106.13	100	21.11	4.74	69.46	56.49
170	28.77	5.91	116.32	130	27.53	4.72	68.90	56.30
150	24.61	6.10	123.83	150	31.74	4.73	69.05	56.35
100	15.08	6.63	145.24	200	41.90	4.77	70.94	56.98
50	6.88	7.27	170.64	250	51.36	4.87	74.72	58.24
10	1.27	7.58	193.82	290	58.34	4.97	78.84	59.61
0	0.00	0.00	80.00	300	60.00	5.00	80.00	60.00

We note immediately that any single step change in running pace results in an improved 300 metres running time. It is clear that this is achieved through expending more of *a* than when no such change is made. The largest such improvement, a reduction in elapsed time of about 3.7 sec (6%) occurs when d_1 is about 57% of *d*, though the curve of total time against d_1 is fairly flat in the region of the optimum. We could now further subdivide each interval separately in the same way, but it is clear what this would lead to. It suggests that the athlete should try to cover a reasonable portion of each such distance interval at a more rapid pace, finishing off more slowly. Following this further subdivision tactic to its extreme, we should obviously next consider the consequences of an all-out effort.

4.2 Running at an all-out pace

Morton [12] shows that under the assumptions of the 3-parameter model described above

$$v_{max}(t) = b - a(t)/k \quad \text{at any time } t$$
(2)

where $v_{max}(t)$ represents the maximal velocity that could be achieved if required at time *t* when the anaerobic store is of amount $a(t)$. The runner may of course elect to run more slowly than this, but could not run any faster, and for the purposes of this illustration runs at exactly that pace. It will be realised that when $a(t) = a$, as is usual when $t = 0$, $v_{max}(0) = v_{max}$; and when $a(t) = 0$, $v_{max}(t) = b$.

Also, the anaerobic store $a(t)$ during this time is being supplemented at rate *b* from the aerobic source, whilst simultaneously being drawn down at rate $v_{max}(t)$ for maximal running, and so it's derivative is given by:

$$a'(t) = b - v_{max}(t)$$

The solution to these equations shows that the anaerobic store at time *t* is given by:

$$a(t) = a.\exp(t/k)$$

and hence the all-out running pace at time *t* is given by:

$$v_{max}(t) = b - a.\exp(t/k)/k$$

and that by integration the distance covered up to time *t* is given by:

$$d(t) = a(1 - \exp(t/k)) + bt$$
(3)

So for any fixed distance *d*, the time taken to cover it under an all-out effort is given by the solution to an appropriate version of equation (3).

In particular when we consider our illustrative runner, equation (3) yields a 300-metre time of 51.66 sec. This is clearly better than the constant pace time of 60 sec. In fact it can be shown as follows that $d(t)$ given by equation (3) is always greater than *d* given by equation (1) for any value of *t*.

$$\exp(-t/k) > 1 - t/k = (k - t)/k \qquad \text{for all } t > 0 \text{ since } k < 0, \text{ hence}$$
$$-\exp(t/k) > k/(t - k) \qquad \text{hence}$$
$$1 - \exp(t/k) > t/(t - k)$$

which on multiplication by $a > 0$ and adding *bt* establishes the superiority of an all-out effort.

By implication of monotonicity therefore if *d* is fixed, the solution for *t* from equation (3) is always less than the solution for *t* from equation (1).

The time of 51.66 sec is also better than any of the strategies in Table 1. Clearly the times in Table 1 can be improved on by further subdivision of each interval in the same way, and repeatedly so *ad infinitum*. However it is unclear whether the limiting time in so doing is equal to 51.66 sec (which I suspect it is) or whether there is any other different strategy that can improve on 51.66 sec. I suspect not, but cannot prove it. We note that $a(t)$ at $t = 51.66$ has declined to 54.97, lower than any value in Table 1. Once again the improvement in time to run 300 metres has been achieved by the greater expenditure of anaerobic store that the all-out strategy accomplishes.

5. *Some physiological considerations*

Despite what one might think, there appears to be very little published empirical research evidence on the optimality, or otherwise, of a constant pace strategy in running or other competitive racing events. What there is suggests pace variation may be better.

Foster *et al.* [3] studied nine well-trained cyclists who each performed five 2-km time trials. The pace was set differently for the first half of each distance, varying from 5% slower to 5% faster than the constant pace corresponding to their best 2-km performance, with the second half pace being reversed. The extremes in pace variation produced the slowest times, while the best time was produced with the first half at 51% of their best pace.

In a study of 11 long distance runners, Billat *et al.* [1] found that a self selected variable pace strategy led to a small but significant improvement in performance only at the fastest (and shortest) of the four different race paces considered (90, 95, 100 and 105% of the pace associated with maximal oxygen uptake).

In a study of eight well trained kayak paddlers, Bishop *et al.* [2] found that 2-minute kayak ergometer performance was superior following an all-out strategy when compared to an even paced strategy. This effect was attributed to faster oxygen uptake kinetics.

Kennedy [7] undertook an observational study of the performances of 19 elite male and 19 elite female rowers over a simulated 2000 metre race. Velocity was measured every 200 metres and the resulting profiles characterised. Kennedy found that after 10 weeks of training, the three fastest males adopted a constant pace profile, whereas the three fastest females adopted an all-out profile, concluding the presence of gender differences.

6. Remarks

It seems evident that the change in optimality from constant pace to all-out effort is a consequence of the maximal output feedback that is added to the bioenergetic model by means of equation (2), rather than simply the addition of a third parameter to essentially the same simple energy demand and supply system. In this respect it seems likely that more complex models, say with three components rather than two [10, 11], with a similar maximal output feedback, would also suggest an all-out strategy as optimal.

7. References

[1]   V. L. Billat, J. Slawinski, M. Danel and J. P. Koralsztein, "Effect of free versus constant pace on performance and oxygen kinetics in running", *Med. Sci. Sports Exerc.,* **33** (2001), 2082-2088.

[2]   D. Bishop, D. Bonetti and B. Dawson, "The influence of pacing strategy on VO_2 and supramaximal kayak performance", *Med. Sci. Sports Exerc.,* **34** (2002), 1041-1047.

[3]   C. A. Foster, C. Snyder, N. N. Thompson, M. A. Green, M. Foley and M. Schrager, "Effect of pacing strategy on cycle time trial performance", *Med. Sci. Sports Exerc.,* **25** (1993), 383-388.

[4]   Y. Fukuba and B. J. Whipp, "A metabolic limit on the ability to make up for lost time in endurance events", *J. Appl. Physiol.,* **87** (1999), 853-861.

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

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

[7]   M. D. Kennedy, "Development of race profiles for the performance of a simulated 2000-m rowing race", *Can. J. Appl. Physiol.,* **28** (2003), 536-546.

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

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

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

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

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

[13] B. Saltin and J. Karlsson, "Muscle glycogen utilisation during work of different intensities", in B. Pernow and B. Saltin (eds), *Muscle Metabolism During Exercise* (Plenum Press, New York, 1971), 289-300.

SERVICE NEUTRAL SCORING STRATEGIES IN TENNIS

P.K. Newton
Department of Aerospace & Mechanical Engineering
and
Department of Mathematics
University of Southern California
Los Angeles, CA 90089-1191
and
G.H. Pollard
School of Information Sciences and Engineering
University of Canberra
Australia

1.2.Abstract

We prove that the probability of winning a set in tennis is independent of which player serves first, as long as points are treated as independent and identically distributed random variables. Thus, both classical and tiebreaker scoring strategies are service neutral strategies. Other scoring strategies are proposed that retain this desirable feature but might be considered fairer from other points of view.

1. Introduction

Consider a tennis match between two players each of whom has a constant probability of winning a point on their own serve, call these probabilities p_A and p_B. Suppose the values obtained under these probabilities are independent and identically distributed (iid) random variables. This approach is used in Pollard [7] and Newton and Keller [6] and has been used to model other racquet sports as well. We give a simple proof that the probability of winning a set (and thus a match) does not depend on which player serves first, using either classical scoring or 13 point tiebreaker scoring. We call scoring systems with this feature *service neutral* strategies. We then propose alternative scoring strategies that are also service neutral, but might be considered fairer from other points of view.

2. Tennis scoring is service neutral
4.

We denote the probability that player A(B) wins a set when serving first by p_A^S (p_B^S). Our goal is to prove that $p_A^S = q_B^S$, where $q_B^S = 1 - p_B^S$, i.e. player A's probability of winning the set when serving first is the same as his probability of winning the set when player B serves first.

It is useful to consider two mutually exclusive events, one in which a conventional set ends with fewer than 11 games played (i.e. scores of 6-0, 6-1, 6-2, 6-3, or 6-4), and one in which it ends with more than 11 games played (i.e. scores of 7-5, 8-6, 9-7, … using classical scoring, or 7-5 or 7-6 using tiebreaker scoring). First, consider the case in which the set ends with fewer than 11 games. Imagine a fictitious scoring scheme in which the players play exactly 10

games, alternating serves, thus each player serves exactly 5 games.[1]

After 10 games are played, either there is an outright winner and loser (i.e. the score is 10-0, 9-1, 8-2, 7-3, or 6-4), or the score is 5-5. The outright winner and loser of this fictitious set would also have been the outright winner and loser of a conventional set,
and we prove that the fictitious set is service neutral. Since each player has served exactly 5 games and since the probability of winning each game is independent, either player could have served first and the serves could be arranged in any order. The probability of reaching a score of 5-5 is also independent of which player served first for the same reasons. Thus, if a conventional set ends in fewer than 11 games, it is service neutral.

Now suppose the score has reached 5-5. With classical scoring, the players play two additional games, with each player serving one of the games. If one of the players wins both games, the set ends 7-5, and the order in which the players served is immaterial because of the independence hypothesis. Thus, this two game extension is service neutral. If the score becomes 6-6, repeat the two game sequences using any serving order, as long as each player serves one of the games and keep repeating these two game extensions until one of the players wins both games. This process is service neutral.

If the set score reaches 6-6 and classical scoring is not used, a 13-point tiebreaker is played in which the serving sequence is AB,BA,AB,BA,AB,BA ...As in the set, consider two mutually exclusive events, one in which the tiebreaker ends with fewer than 13 points (i.e. scores of 7-0, 7-1, 7-2, 7-3, 7-4, or 7-5) and one in which it ends with more than 13 points (i.e. scores of 8-6, 9-7, ...). First, consider the case in which the tiebreaker ends with fewer than 13 points and imagine the fictitious scoring scheme in which the players play exactly 12 points, alternating serves as AB,BA,AB,BA,AB,BA. After the 12 points are played, either there is an outright winner and loser (i.e. the score is 12-0, 11-1, 10-2, 9-3, 8-4, or 7-5), or the score is 6-6. The outright winner and loser of the tiebreaker using this fictitious scoring scheme would also have been the outright winner and loser of a conventional tiebreaker. Since each player has served exactly 6 points and since the points are independent, either player could have served first and serving could be arranged in any order. The probability of reaching a score of 6-6 also does not depend on who serves first, for the same reason. Thus, if a conventional tiebreaker ends with fewer than 13 points played, it is service neutral.

Now suppose the tiebreaker score has reached 6-6. The players play two additional points, with each player serving one of the points. If one of the players wins both points, the tiebreaker ends 8-6 and the order in which the players served the two points is immaterial, as long as each serves one of them. If the score becomes 7-7, keep repeating these two point sequences in any serving order until one of the players wins both points of the sequence. These two point extensions are each service neutral.

Thus, we have proven that tennis scoring, either using classical or tiebreaker schemes, is service neutral, as long as points are treated as iid random variables. Furthermore, the players

[1] This type of argument is used by Anderson [1] to prove a result regarding

a scoring system introduced by Kingston [3].

could serve in any order, as long as neither serves more than 5 games if the set goes up to 10 games.

3. Alternative scoring strategies

There are many other scoring strategies that are also service neutral which have desirable features not present in conventional scoring schemes. For example, with conventional scoring, the player who serves first, if he holds serve, is never behind in the match, and this is sometimes thought to create a psychological advantage. Alternatively, if a `back-to-the-wall' effect is present, as described in Jackson and Mosurski [2], it may give the first receiver a psychological edge since he is always serving from behind. There is also evidence of a `first game effect' in tennis, documented in Magnus and Klaassen [5], i.e. the first game of the match seems to be the hardest to break serve, thus favoring the player who serves first. The following scoring strategies are ones that are service neutral but are designed to mollify these effects.

Option 1. Revised classical scoring:

Consider the serving sequence ABBA,ABBA,ABBA,..., instead of the conventional sequence AB,AB,AB,AB,AB,AB.... Once the set ends, the next set begins with player B serving, with the sequence BAAB,BAAB,BAAB....This scheme is service neutral but has the advantage over the conventional one that as long as serve is held, both players begin their two game serving sequence playing from behind. This eliminates the asymmetry inherent in the conventional sequence in which the first server might never be serving from behind, while the first receiver always serves from behind. In addition, alternating which player serves first from set to set partly offsets any advantage gained by the first server. The server who begins each set could alternate as ABABA or ABBAA in a 5 set match.

Option 2. Revised tiebreaker scoring:

Use the option 1 serving sequence ABBA,ABBA,ABBA through the first 12 games until the set ends. If the score reaches 6-6, play a 13 point tiebreaker using the sequence BAAB,BAAB,BAAB... Once the set ends, the next set begins with player B serving first.
This scheme is service neutral, has the desirable features of option 1 scoring through the first 12 games, and allows player B to start serving the tiebreaker, which partly mollifies the possible advantage player A gained from having served the first game of the set.

4. Other generalizations

A few service neutral generalizations are somewhat surprising. Firstly, the methods of Section 2 can be used to show that *any* service order for the first 10 games, provided each player is planned to serve on (up to) 5 occasions each is service neutral. Thus, as an `extreme' example of only academic interest, the planned game sequence
AAAAA,BBBBB,AB,AB,... is service neutral.

Correspondingly, the methods of Section 2 can also be used to show that any service order within the tiebreaker game for the first 12 points, provided each player is planned to serve (up to) 6 points each is service neutral. Thus, as an `extreme' example, the planned point sequence
AAAAAA,BBBBBB,AB,AB,... is also service neutral.

As a final comment, we note that even though points, games, and sets in tennis are not, strictly speaking, independent or identically distributed random variables (see the statistical study of Klaassen and Magnus [4]), there is not yet a large enough body of evidence supporting systematic or recurrent non-iid effects to recommend viable alternatives to the service neutral strategies currently in use.

5. Example

Here we consider an example in which player A has a probability of 0.7 of winning his service game, and player B has a probability of 0.6 of winning his service game. Assuming player A serves for the first game, and then service games alternate as is the present practice, a tree or branching diagram, or the recursion formulas in Newton and Keller [6] can be used to compute the probability player A wins 6- i (i=0,1,2,3 and 4). The results are shown below.

Prob (A wins 6-0) = 0.021952
Prob (A wins 6-1) = 0.0889056
Prob (A wins 6-2) = 0.09567936
Prob (A wins 6-3) = 0.187463808
Prob (A wins 6-4) = 0.109593388
SUM = 0.503594156

If we now assume player A serves the first 5 games, and player B serves the next 5 games if necessary, the corresponding probabilities are shown below.

Prob (A wins 6-0) = 0.067228
Prob (A wins 6-1) = 0.0979608
Prob (A wins 6-2) = 0.11310768
Prob (A wins 6-3) = 0.115704288
Prob (A wins 6-4) = 0.109593388
SUM = 0.503594156

Thus, this example demonstrates one of the generalizations mentioned in the above paragraph. Note that the probabilities A wins 6-4 are equal (as they should be) under the two service regimes, but the others are not, however the sums are equal.

6. References

[1]   C.L. Anderson ``Note on the advantage of first serve'', *J. of Combinatorial Theory, Series A*, **23** (1977) , 363.

[2]   D. Jackson and K. Mosurski , ``Heavy defeats in tennis: Psychological momentum or random effects'', *Chance* **10** (1997), 27-34.

[3]   J.G. Kingston, ``Comparison of scoring systems in two-sided competitions'', *J. of Combinatorial Theory (A)*, **20** (1976), 357-362.

[4]   F.J.G.M. Klaassen and J.R. Magnus, ``Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model'', *Journal*

of the American Stat. Ass. **96**, no. 454, June (2001), 500-509.

[5]   J.R. Magnus and F.J.G.M. Klaassen, ``On the advantage of serving first in a tennis set: four years at Wimbledon''*,The Statistician*, **48** (1999), 247-256.

[6]   P.K. Newton and J.B. Keller, ``The probability of winning at tennis'', University of Southern California, Center for Applied Mathematical Sciences preprint (2003).

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

DYNAMIC PROGRAMMING IN CRICKET: BATTING ON A STICKY WICKET

J. M. Norman	Stephen R. Clarke
The Management School	School of Mathematical Sciences
University of Sheffield	Swinburne University
Sheffield	PO Box 218, Hawthorn
S1 4DT, England	Vic. 3122 Australia

Abstract

In cricket, a rain-affected pitch (sticky wicket) can make batting more difficult than normal. On sticky wickets, lower order batsmen are often sent in to "hold the fort" until the wicket improves. In this paper, a stochastic dynamic programming model is used to examine the appropriateness of this policy. The model demonstrates the tactic is often optimal. The decision to send in a lower order batsman is optimal more often than when compared to the similar practice of using a nightwatchman.

1. Introduction

In cricket, the difficulty of batting depends to a large extent on the state of the pitch. This can change during the course of the match, particularly if it is affected by rain. A rain affected or 'sticky' wicket may be very difficult to bat on, and take several overs of play to dry out and return to normal batting condition. When his side was faced with a sticky wicket, Sir Don Bradman [1] would adopt the policy of "sending in tail-end batsmen to hold the fort until the wicket improved". While this tactic reduces the exposure of the good batsmen to the poor conditions, it increases the chance they will be left at the end of the innings without a batting partner. In this note, we examine the correctness of the policy of sending in a lower order batsman on a sticky wicket.

2. The model

We adopt the model described in Clarke and Norman [2], which is described again here, in slightly adapted form. We assume there are two types of batsman. Teams usually contain 6 or 7 recognised batsmen, selected primarily for their batting skills, and 4 or 5 lower order batsman (or duffers) generally selected for their bowling and with problematical batting ability. Recognised batsmen gradually improve their expected performance as they gain experience of the conditions (play themselves in), so their rate of scoring increases and their chance of dismissal decreases in the early part of their innings. Duffers don't improve, and their expected performance depends only on the pitch. The batsman's rating x is an indicator of expected performance and may change only at the end of an over, depending on the pitch and the number of overs batted. During the over, a batsman with a rating of x scores $r(x)$ runs but may be dismissed (on the last ball of the over) with probability $p(x)$. At the start of the next over his rating is x'. Note that $r(x)$ must be even to avoid changes of end during an over. $r(.)$ could also be thought of as the expected number of runs, for a distribution of runs that contained only even values.

The numerical data assumed are given in Table 1. A duffer always has a rating of zero while a recognised batsman has a rating of 1 initially, rising by one each over (whether he is on strike or not) up to a maximum of 5 as he plays himself in. On a sticky wicket all dismissal probabilities for a recognised batsman are 50% greater than those for a normal wicket. For a duffer we consider 3 scenarios: in the first the dismissal probability on a sticky wicket is the same as for a normal wicket; in the second it is 50% greater than on a normal wicket; and in the third it is double that for a normal wicket.

The $r(.)$ and $p(.)$ values for a normal wicket imply an expected score for a duffer of 10 and for a recognised batsman of nearly 40. This generates an average total score during an innings of about 272 runs, a reasonable figure for test and county matches. Note that scores in the second innings are usually lower than those in the first innings, and are also the more likely to be rain affected.

Table 1: State transitions, scoring rates and dismissal probabilities as a function of batsmen's ratings.

Type of Batsman	Rating at start of over x	Rating at start of next over x'	Runs scored $r(x)$	Probability of dismissal $p(x)$	
				normal wicket	sticky wicket
Duffer (scenario 1)	0	0	2	0.2	0.2
Duffer (scenario 2)	0	0	2	0.2	0.3
Duffer (scenario 3)	0	0	2	0.2	0.4
Recognised batsman, who plays himself in when not on strike as well as when he is on strike. It takes him four overs to play himself in	1	2	0	0.2	0.3
	2	3	0	0.167	0.25
	3	4	2	0.125	0.1875
	4	5	4	0.100	0.15
	5	5	4	0.083	0.125

As captain of the side batting on a sticky wicket, you are at the start of an over, with n overs left until it is expected that the wicket will return to normal. The batsman on strike has a rating of x, and the batsman at the other end has a rating of y. If there is no dismissal, at the start of the next over their respective ratings will be x' and y', and the strike will rotate to the other batsman. If a dismissal occurs, the decision to be taken regarding the incoming batsman, who will not be on strike next over, is whether he should be a recognised batsman or a duffer. This decision depends on a, the number of recognised batsmen not dismissed, and b, the number of duffers not dismissed, in both cases excluding those at the wicket. While various objective functions are possible in cricket (see, for example, Clarke [3] or Preston and Thomas [4]), here we maximise the expected number of runs scored.

Let $f_n(a,b,x,y)$ be the expected number of runs scored in the remaining part of the innings with n overs remaining until the pitch returns to normal, using an optimal policy, i.e. one that maximises the expected runs scored.

Then $f_n(a,b,x,y) = \max$ [
> A: send out a recognised batsman if a dismissal occurs (provided $a \geq 1$)
> B: send out a duffer (tail-ender) if a dismissal occurs (provided $b \geq 1$)]

Where: - $A = r(x) + [1-p(x)]f_{n-1}(a,b,y',x') + p(x) f_{n-1}(a-1,b,y',1)$
- $B = r(x) + [1-p(x)]f_{n-1}(a,b,y',x') + p(x) f_{n-1}(a,b-1,y',0)$

$f_n(0,0,x,y) = r(x) + [(1-p(x)] f_{n-1}(0,0,y',x')$

When $n = 0$ the wicket returns to normal, and we have a supplementary dynamic programming problem concerned with the optimal strategy and maximum expected score for a normal pitch. This dynamic program is similar to the above formulation, but with no limit on the number of overs left, since we are concerned with test cricket. While the generally optimal strategy to this problem is well known (always send in the recognised batsmen first), here we need the limiting values of the objective function to use as the starting values $f_0(a,b,x,y)$ for the sticky wicket problem. These can be found by using backward recursion for a large number of overs, often called value iteration (see, for example, Hastings [5] or Smith [6]). So for example, $f_0(4,5,1,1)$ is the expected score when opening with two recognised batsmen on a normal pitch with 4 recognised batsmen and 5 duffers still to come in. The above procedure resulted in a value of 267, near the 272 calculated earlier by a less exact method. We assume that when play resumes on a normal wicket, recognised batsmen have to play themselves in, as would certainly be the case if a sticky wicket persisted for the rest of the day's play but improved overnight.

3. Results

Since the optimal strategy at each stage (number of overs) depends on three variables (a, b, and the rating of the other batsman) a three-way table is necessary to show the optimal strategy. However, since in the

solution obtained the dependence on the rating of the other batsman falls into one of only 7 categories, we can display the results using the following matrix format.

		Number of duffers remaining *b* 1 2 3 4 5
Number of recognised batsmen remaining *a*	1	x x x x x
	2	x x x x x
	3	x x x x x
	4	x x x x x
	5	x x x x
	6	x x x

The entries in the matrices have the following meanings:
- *R* send in a recognised batsman
- *D* send in a duffer
- *S* send in a duffer unless the other batsman is a duffer, in which case send in a recognised batsman
- *J* send in a recognised batsman unless the other batsman has a rating of 3 or more
- *A* send in a duffer unless the other batsman has a rating of 0 or 1
- *B* send in a recognised batsman unless the other batsman has a rating of 4 or 5
- *C* send in a recognised batsman unless the other batsman has a rating of 1, 2 or 3

- *n* is the number of overs left until the wicket gets back to normal.

Table 2: Optimal strategy with *n* sticky wicket overs left for scenario 1, (*p(0)* normal, others raised 50%)

n=1	n=2	n=10	n=20	n=30	n=40
RRRRR	RRSDD	RSDDD	RRRSD	RRRRR	RRRRR
RRRRR	RRRDD	RDDDD	RRRDD	RRRRS	RRRRR
RRRRR	RRRDD	DDDDD	RRDDD	RRRAD	RRRRR
RRRRR	RRRDD	DDDDD	RRDDD	RRRDD	RRRRR
RRRR	RRRD	DDDD	RDDD	RRBD	RRRR
RRR	RRR	DDD	RDD	RRD	RRR

Table 3: Optimal strategy with *n* sticky wicket overs left for scenario 2, (every *p(.)* raised by 50%)

n=1	n=2	n=10	n=20
RRRRR	RRDDD	RRRSD	RRRRR
RRRRR	RRRDD	RRJDD	RRRRR
RRRRR	RRRDD	RRSDD	RRRRR
RRRRR	RRRDD	RRDDD	RRRRR
RRRR	RRRD	RRDD	RRRR
RRR	RRR	RRD	RRR

Table 4: Optimal strategy with *n* sticky wicket overs left for scenario 3, (*p(0)* doubled, others raised 50%)

n=1	n=2	n=10	n=20
RRRRR	RRDDD	RRRRC	RRRRR
RRRRR	RRRDD	RRRRS	RRRRR
RRRRR	RRRDD	RRRRS	RRRRR
RRRRR	RRRDD	RRRRD	RRRRR
RRRR	RRRD	RRRR	RRRR
RRR	RRR	RRR	RRR

4. Discussion

It is not surprising that in the first scenario, it is common for a duffer to be sent in – in this case, duffers bat as well on a sticky wicket as they bat on a normal one. However, in the second scenario, when the dismissal probabilities of duffers are raised by the same percentage as those of recognised batsmen, duffers are still frequently used. This is in line with the common practice as used by Bradman, and noted in the introduction. In the third scenario, when duffers handle the sticky wicket much worse than recognised batsmen, it is still optimal to sacrifice them when the wicket will improve quickly.

Tail-end batsmen are often sent in, not only to hold the fort on a sticky wicket, but to hold the fort towards the end of a day's play, to save a recognised batsman from having to play himself in twice, that is, to act as a night watchman. The two policies, for a sticky wicket and for the night watchman situation, may be compared, using the scoring rates and dismissal probabilities of scenario 2. The optimal policies are shown in Tables 5 and 6. Note that in the sticky wicket policy, n refers to the number of overs until the wicket gets back to normal, and in the night watchman policy n refers to the number of overs left until close of play at the end of the day. In either case, recognised batsmen have to play themselves in when play resumes at n = 0.

Table 5: Optimal strategy with n sticky wicket overs left for scenario 2.

n=1	n=2	n=3	n=4	n=5
RRRRR	RRDDD	SDDDD	SDDDD	RRSDD
RRRRR	RRRDD	RDDDD	DDDDD	RJDDD
RRRRR	RRRDD	RDDDD	DDDDD	RSDDD
RRRRR	RRRDD	RDDDD	DDDDD	RSDDD
RRRR	RRRD	RDDD	DDDD	RSDD
RRR	RRR	RDD	DDD	RSD

n=6	n=7	n=8	n=9	n=10
RSDDD	RRRSD	RRSDD	RRRRD	RRRSD
RDDDD	RRADD	RRDDD	RRRSD	RRJDD
RDDDD	RRSDD	RRDDD	RRRDD	RRSDD
RDDDD	RRDDD	RSDDD	RRRDD	RRDDD
RDDD	RRDD	RDDD	RRRD	RRDD
RDD	RRD	RDD	RRR	RRD

Table 6: Optimal strategy with n overs left in the day for night watchman policy for scenario 2

n=1	n=2	n=3	n=4	n=5
RRRRR	RRSDD	RSDDD	SDDDD	RRSDD
RRRRR	RRRDD	RRDDD	RDDDD	RRSDD
RRRRR	RRRDD	RRDDD	RDDDD	RRDDD
RRRRR	RRRDD	RRDDD	RDDDD	RRDDD
RRRR	RRRD	RRDD	RDDD	RRDD
RRR	RRR	RRD	RDD	RRD

n=6	n=7	n=8	n=9	n=10
RSDDD	RRRSD	RRSDD	RRRRS	RRRSD
RDDDD	RRRSD	RRDDD	RRRRS	RRRDD
RDDDD	RRRDD	RRDDD	RRRRD	RRRDD
RDDDD	RRRDD	RRDDD	RRRRD	RRRDD
RDDD	RRRD	RRDD	RRRR	RRRR
RDD	RRR	RRD	RRR	RRR

Even a cursory comparison of the policies reveals that the decision to send in a duffer is much commoner in the sticky wicket policy. This is somewhat surprising, in that nowadays it is common practice to send in a night watchman towards the end of a day, but maybe less common to send in a lower order batsman to hold the fort on a sticky wicket.

5. An alternative model

Suppose, however, that a sticky wicket does not last until the end of the day's play. It might then be reasonable to suppose that a recognised batsman would not need to play himself in from scratch, as at the start of the following day's play. A sticky wicket does not suddenly become a normal wicket between one over and another and a recognised batsman may play himself in just as well on a sticky wicket as on a normal wicket. Thus we might make the limiting values of $f_0(a,b,x,y)$ take account of the actual ratings of the two batsmen at the end of the sticky wicket period, instead of making them revert to 1 (if recognised batsmen). Results corresponding to those in tables 2, 3 and 4 are shown below in Tables 7, 8 and 9.

Table 7: Optimal strategy with n sticky wicket overs left for scenario 1, ($p(0)$ normal, others raised 50%)

n=1, 2	*n=3*	*n=10*	*n=20*	*n=30*	*n=40*
RRRRR	*RRSDD*	*RRSDD*	*RRRSD*	*RRRRS*	*RRRRR*
RRRRR	*RRRDD*	*RRDDD*	*RRADD*	*RRRRD*	*RRRRR*
RRRRR	*RRRDD*	*RRDDD*	*RRDDD*	*RRRDD*	*RRRRD*
RRRRR	*RRRDD*	*RRDDD*	*RDDDD*	*RRRDD*	*RRRSD*
RRRR	*RRRD*	*RSDD*	*RDDD*	*RADD*	*RRJD*
RRR	*RRR*	*RDD*	*RDD*	*RDD*	*RRD*

Table 8: Optimal strategy with n sticky wicket overs left for scenario 2, (every $p(.)$ raised by 50%)

n=1, 2	*n=3*	*n=10*	*n=20*
RRRRR	*RRRSD*	*RRRRA*	*RRRRR*
RRRRR	*RRRRD*	*RRRRS*	*RRRRR*
RRRRR	*RRRRD*	*RRRSD*	*RRRRR*
RRRRR	*RRRRD*	*RRRDD*	*RRRRS*
RRRR	*RRRR*	*RRRD*	*RRRA*
RRR	*RRR*	*RRR*	*RRR*

Table 9: Optimal strategy with n sticky wicket overs left for scenario 3, ($p(0)$ doubled, others raised 50%)

n=1	*n=2*	*n=10*	*n=20*
RRRRR	*RRRRR*	*RRRRR*	*RRRRR*
RRRRR	*RRRRR*	*RRRRR*	*RRRRR*
RRRRR	*RRRRR*	*RRRRR*	*RRRRR*
RRRRR	*RRRRR*	*RRRR*	*RRRRR*
RRRR	*RRRR*	*RRRR*	*RRRR*
RRR	*RRR*	*RRR*	*RRR*

In each scenario it is now much less common for a duffer to be sent in; not surprisingly, in the third scenario duffers are not sent in at all. The same result is evident when the sticky wicket policy for scenario 2, as shown in Table 10, is compared with the night watchman policy as previously given in Table 6.

Table 10: Optimal strategy with *n* sticky wicket overs left for scenario 2.

n=1	*n=2*	*n=3*	*n=4*	*n=5*
RRRRR	RRRRR	RRRSD	RRRRD	RRRSD
RRRRR	RRRRR	RRRRD	RRRRD	RRRSD
RRRRR	RRRRR	RRRRD	RRRRD	RRRDD
RRRRR	RRRRR	RRRRD	RRRRD	RRRDD
RRRR	RRRR	RRRR	RRRR	RRRD
RRR	RRR	RRR	RRR	RRR

n=6	*n=7*	*n=8*	*n=9*	*n=10*
RRRRS	RRRRS	RRRRS	RRRRS	RRRRA
RRRSD	RRRSD	RRRRD	RRRRD	RRRRS
RRRSD	RRRDD	RRRSD	RRRSD	RRRSD
RRRSD	RRRDD	RRRDD	RRRDD	RRRDD
RRRD	RRRD	RRRD	RRRD	RRRD
RRR	RRR	RRR	RRR	RRR

It does seem that much depends on whether a sticky wicket is expected to last until the end of the day's play. If so, then the use of duffers to hold the fort is very often optimal, but otherwise, not. There is often a case for using a night watchman towards the end of a day's play: if the wicket is sticky the case is stronger still.

6. Conclusion

In *Farewell to cricket*, Bradman [1] puts the issues clearly.

> "It is all very well to [be] gallant and heroic, but the captain's job embodies the welfare of the team, and if his own personal success is an integral part of victory, he should not act accordingly. On several occasions I was compelled to rearrange our batting order as a matter of tactics because of the state of the wicket. It almost invariably succeeded.
>
> Some were unkind enough to suggest that my purpose was to avoid batting on a wet wicket. Of course it was, but only because such avoidance was in the interests of the team.
>
> Cricket pitches behave in a variety of ways after rain. The man never lived whose judgement was infallible. Not the least difficulty is to decide how long a wicket will remain bad. Under Australian conditions sufficient rain on a hard wicket, followed by a hot sun, will generally produce a glue pot. Some are worse than others. But will it remain sticky for an hour or a day? One cannot tell.
>
> In 1936-7 against Allen's team, we worked like beavers to try and get quick wickets. Then I found out the pitch was drying more slowly than anticipated and I had to tell my bowlers not to get England out"

We don't expect modern captains would be brave enough to instruct their team not to try to dismiss the opposition. Bradman hints at the criticism captains can invite by following strategies which could be mis-construed by less informed observers. The benefits of a strategy that is optimal in the long run may be outweighed by the criticism bound to follow when it does not succeed in the short term. While Bradman was happy to open the innings with his number 10 and 11 batsman, most observers would probably see this as a sign of weakness by the regular openers. Clearly strong evidence would be needed to convince captains this might be a reasonable tactic in some circumstances.

While there are several improvements that could be made to our model, it is a first step in investigating the merits of alternative batting orders. Allowing dismissals off any ball, and incorporating a distribution of runs to allow for a change of ends for batsmen during overs, would make the model more realistic, but probably intractable. While these models may be difficult to solve for optimality, a more realistic

simulation model could be used to investigate fixed strategies. A simulation model might also carry more weight with cricket players and administrators, and might even be used as a learning tool.

While covered wickets now reduce the frequency of rain affected pitches, there are still many other situations in today's cricket that produce variable batting conditions. The first session of a test match often produces a lively green wicket that is expected to become less dangerous after the first hour or so. At the start of any day overcast weather might provide ideal conditions for swing bowling. At any time cloud cover might result in poor light. The opposing team's main strike fast bowler(s) may become less dangerous as tiredness sets in, and ultimately must be replaced by possibly more benign bowlers. In such cases the use of lower order batsman to protect the recognised batsmen from the more dangerous conditions may be advantageous to the team.

References

[1]   D. Bradman, *Farewell to cricket*, ETT Imprint, Sydney, (1997).

[2]   S. R. Clarke and J. M. Norman, "Dynamic programming in cricket: choosing a night watchman", *J. Oper. Res. Soc*, **54,** (2003), 838-845.

[3]   S. R. Clarke, "Dynamic Programming in One-Day Cricket – Optimal Scoring Rates", *J. Oper. Res. Soc*, **39**, (1988), 331-337.

[4]   I. Preston and J. Thomas, "Rain rules for limited overs cricket and probabilities of victory", *The Statistician*, **51**, (2002), 189-202.

[5]   N. A. J. Hastings, *Dynamic programming with management applications*, Butterworth, London, (1973).

[6]   D. K. Smith, *Dynamic programming: a practical introduction*, Ellis Horwood, Chichester, (1991).

IS THERE A DIFFERENCE IN THE PREDICTABILITY OF MEN'S AND WOMEN'S BASKETBALL MATCHES?

Pam Norton
School of Mathematical Sciences
Monash University, VICTORIA, 3800
pam.norton@sci.monash.edu.au

Abstract

It has been observed that correct predictions can be made for approximately 80% of women's netball matches in the Commonwealth Bank Trophy, but only at best 70% for AFL football matches. In this paper we apply a simple model to men's and women's basketball matches at the elite level in Australia for the 2003-4 season to investigate whether the results of women's matches are more predictable than those for men. The model uses ratings for each team and includes a common home ground factor. The ratings are adjusted during the season depending on performance. It was found that there was no significant difference in the percentages of men's and women's basketball matches that could be correctly predicted nor in the average error of the predictions.

1.   Introduction

There are many predictions published for AFL matches. Each of the main Melbourne newspapers has a group of "experts" who each week give their tips. Swinburne Sports Statistics (www.swin.edu.au/sport) also publish several sets of tips. Typically the best tipsters manage to get about 70% of tips correct for the home and away games. Swinburne Sports Statistics also publishes tips for netball, and have been doing this since 2000. Typically around 80% of matches in the home and away series can be predicted (as to outcome) correctly. Table 1 gives the percentage of correct predictions and the average error for the AFL predictions done by the Swinburne computer (S. Clarke) and for the Commonwealth Bank Trophy (CBT) (Australian netball league) predictions for the past four seasons.

Table 1: Percentage of correct predictions and average error for AFL (Swinburne computer tips) and CBT (as published on Swinburne Sports Statistics Website).

Year	AFL Percentage of correct predictions	AFL Average error	Netball (CBT) Percentage of correct predictions	Netball (CBT) Average error
2000	69	31.8	80	9.3
2001	65	31.5	82	8.9
2002	67	29.0	87	9.0
2003	67	27.5	75	9.0

This invites the question "Is there a difference in the predictability of men's and women's matches?" A sport at elite level common to both is basketball, and so a simple prediction model was applied to both the men's National Basketball League (NBL) and Women's National Basketball League (WNBL) for the home and away fixtures in 2003-4, to determine the percentage of correct predictions and the average error of the predictions.

1.1 The NBL

Australia's NBL was formed in 1979 with 9 teams. There are now 12 teams in the competition, and next season the fixture will move to a "conference" format, with two groups

of six playing a home and away fixture, followed by a final playoff between winners of the groups.

Between 1979 and 1997, the NBL season was played between April and September, but in 1998 the league moved its season to summer, with games being played between October and April. In the 2003-2004 season, twelve teams competed in a home and away series, playing each other team three times, either once at home and twice away, or twice at home and once away. The fixture was not balanced, but tried to minimise travelling costs where possible. The teams contesting the 2003-2004 season were the Adelaide 36ers, Brisbane Bullets, Cairns Taipans, Hunter Pirates, Melbourne Tigers, New Zealand Breakers, Perth Wildcats, Sydney Kings, Townsville Crocodiles, Victoria Giants, West Sydney Razorbacks and Wollongong Hawks. Of these, only Brisbane and Wollongong competed in the first season (in 1979), and both the Hunter Pirates and New Zealand Breakers were new in 2003-4.

Matches last for four quarters of 12 minutes. If the score is tied after this time, extra time is played. Players can score either 1, 2 or 3 points for a goal, depending on whether it results from a free throw, a standard field goal or a field goal scored from behind the three point line. The match winner is the team that has scored the most points at the end of the game.

1.2 The WNBL

In 1981, a two round competition, called the Women's Interstate Basketball Conference, was held between three teams from South Australia, three teams from Victoria, two from Sydney and one from the AIS. One round was played in Adelaide and one in Melbourne to reduce costs. A similar competition continued annually until 1986. In 1986 the National Women's Basketball League was formed and played its first full home and away competition. In 1998, the WNBL also moved its season to summer.

There are now eight teams, and they play each other three times during the season, either twice at home and once away, or once at home and twice away. In 2003-2004 the teams were the same as in 2002-2003 - AIS, Adelaide Lightning, Bulleen Melbourne, Canberra Capitals, Dandenong Rangers, Perth Lynx, Sydney Flames and Townsville Fire. Of these, only the AIS competed in the first season (in 1981). Again the fixture is not balanced, but tries to reduce the travelling costs where possible, and so it is common in a round for a team to play several games away on tour. The past season's draw also allowed the winners of the previous year's competition, the Canberra Capitals, to represent Australia at the inaugural FIBA Women's World Cup in Russia last year where they finished fifth.

Matches last for four quarters of 10 minutes, with extra time being played if the match is drawn at the end of the forty minutes. Scoring is the same as for the men's competition.

Data for some previous seasons was obtained, and indeed there had been efforts by the author to predict the results in seasons 2000-2001 and 2002-2003 using the simple model discussed below. Table 2 lists the average number of points per match, the average home team winning margin, the percentage of home wins, the percentage of correct predictions, and the average error of these predictions. The average home team winning margin can be taken as an estimate of a generic home advantage. The ratio, R, of the total number of points in a match to the average home team winning margin is also given. This can be used to compare the effect of home team advantage in sports. Stefani and Clarke [9] determined the values of R for soccer (3 European cups) – 3, hockey (USA) – 10, professional football (USA) – 12, Australian rules football – 21, and baseball (USA) – 34. A value of R of 53 was calculated for Australian women's netball by Norton and Clarke [7]. A value of R of 71 can be calculated from the data given in Snyder and Purdy [7] for collegiate basketball. The values given for R

in Table 2 indicate that women's basketball in Australia has a low home advantage compared to other sports, but seems to be about twice that for collegiate basketball.

Table 2: Summary of WNBL match statistics for 2000-2001, 2001-2002 and 2002-2003 seasons

Season	2000-2001	2001-2002	2002-2003
Number of matches	84	84	84
Average number of points per match	147.1	143.5	143
Average home team winning margin	4.77	3.26	3.88
Percentage of home team wins	55	56	61
Percentage of correct predictions	76	n.a.	74
Average prediction error per match	11.9	n.a.	10.6
R	31	44	37

2. The Prediction Model

There have been many studies of team ratings and home advantage in sport. Literature reviews include Courneya and Carron [3] and Stefani [8]. Specific examples include Australian rules football [9], basketball [4], netball [6] and soccer [2].

The simplest linear model for a home and away competition which incorporates home advantage involves an ability rating for each team in the competition on a neutral ground, together with a common home advantage for all teams. This model is used as the basis for our predictions. The model differs from those used to determine team ratings and home advantage for a season (or more) in that the team ratings will be adjusted after each match to reflect current performance.

Let w_{ijk} be the predicted winning margin (in points) when team i plays team j on team i's home ground in match k of the rounds. In the WNBL, k takes values between 1 (corresponding to the first match of the rounds) and 84 (corresponding to the last match of the rounds), and in the NBL, from 1 to 198. Let u_{ik} be the rating for team i in match k, which is a rating of team i on a neutral ground. This is essentially a one number summary of a team's ability, form or level of performance. Let h be the home ground advantage of a team, which includes all that is advantageous for a team playing at home and all that is disadvantageous for any other team playing at that team's home ground. Then

$$w_{ijk} = u_{ik} - u_{jk} + h$$

The ratings u_{ik} in the model are relative. It is only the differences in ratings that are important. For this application, the ratings have been chosen so that an "average" team would have a rating of 100. (If all teams were of equal ability, they would each have a rating of about 100. Any other number could have been chosen here.)

For prediction purposes, the team ratings should be adjusted as the season progresses to account for differing team performance. How are the initial ratings selected? Since there was no prior data available for the men's competition other than last year's ladder, each team was given an initial rating based on the previous season's ladder position and percentage, or, in the case of teams new to the competition, from player statistics and media reports. For the WNBL, each team started with its rating from the end of the previous season. The initial ratings are given in Tables 5 (for the NBL) and 6 (for the WNBL). The common home team advantage for both leagues was chosen to be 4, based on previous WNBL seasons. Harville and Smith [4] also reported an estimated home advantage of 4.68 for college basketball.

After each match, the ratings of the two teams participating were adjusted, according to each team's performance. If m_{ijk} was the actual home team winning margin when team i plays team j at home in match k, then the revised ratings $u_{i(k+1)}$ and $u_{j(k+1)}$ of teams i and j which participated in match k were as follows:

$$u_{i(k+1)} = u_{ik} + \alpha(m_{ijk} - w_{ijk})$$
$$u_{j(k+1)} = u_{jk} - \alpha(m_{ijk} - w_{ijk})$$

where α is a constant, $0 \leq \alpha \leq 1$. For the season investigated, the value of α was chosen to be 0.2, as this value has been found to generally yield the highest number of correct predictions for netball matches using a similar model.

There was an implementation problem with this. Predictions are made once a week for the coming week's matches. Teams often play several matches in a week, and so the predicted results for all these matches are based on performance prior to the week, and not taking into account the results of matches played earlier in the week.

One other adjustment was made to the model for the WNBL. The WNBL has an "equalisation" policy. This presents a problem when considering the Canberra Capitals team. Table 3 below shows the top 10 players in terms of points per match for the season in 2003-2004. The table for the previous season was similar. Lauren Jackson is well above her nearest rival in terms of points scored in a match, so if she did not play, (and she was out for five weeks mid-season,) the rating of the Canberra Capitals was decreased by about 15, since any replacement for her was likely to average no more than 13 points per game.

Table 3 : Top 10 Points per Game, WNBL, 2003-2004 (Source: www.wnbl.com.au)

1. Lauren Jackson   28.1
2. Belinda Snell   17.6
3. Shelley Hammonds   15.4
4. Katrina Hibbert   15.3
5. Jacinta Hamilton   15.1
6. Gina Stevens   14.8
7. Laura Summerton   14.0
8. Jodie Datson   13.8
9. Gabrielle Richards   13.5
10. Hollie Grima   13.4

Finally, at the end of each season, the results were used to determine fixed team ratings for the season and a common home advantage. This was accomplished by fitting a linear model of the form

$$w = u_i - u_j + h + e$$

where w is the predicted winning margin (in points) when the home team i plays away team j, u_i is the rating for team i on a neutral ground, h is the home ground advantage of a team and e is a random error assumed to have a mean of zero. The ratings are relative, so for the WNBL, the rating of Townsville was set to 0, and for the NBL, the rating of Hunter was set to zero. For comparison, these fixed seasonal ratings were then used retrospectively to predict the results of all matches over the season.

3. Results

A summary of the results for the season is given in Table 4 . It appears that the choice of home advantage of 4 points was a reasonable one, being close to the average home team

winning margins. In the NBL rounds, home teams won 61% of matches, compared to 63% for the WNBL. Overall, 76% of women's matches were predicted correctly (in terms of winning), but only 70% for men. A (two-sided) two-proportion z-test gave a *P*-value of 0.26, so the difference in proportions was not significant. The average error in the predictions for the WNBL was 12.2 points per game and for the NBL 11.5 points per game. A (two-sided) independent two sample t-test gave a P-value of 0.54, so this difference was not significant.

Table 4: A Comparison of WNBL and NBL predictions for Season 2003-4

	WNBL	NBL
No of matches	84	198
Average no of points per match	135	198
Average home team winning margin	4.73	4
Percentage of matches won by home team	63	61
No. of correct predictions	64	138
Percent of correct predictions (standard error)	76(5)	70(3)
Average error in predictions (standard error)	12.2(1.0)	11.5(0.6)
Home advantage used in model	4	4

Figure 1 gives side-by-side boxplots of the margins for matches in the WNBL and NBL in the 2003-2004 season. The margin is the winning team score − losing team score. The distribution of margins in both is skewed to the right, with some outliers, with the median margin for the NBL being 12 and the median margin for the WNBL being 14. (The two outliers in the boxplot of margins for the WNBL both came from matches involving the Perth team.) A two-sided Mann-Whitney test for the difference in medians showed that this difference was significant, with P-value 0.03. Hence the median margin in WNBL matches was bigger than the median margin in NBL matches.

Figure 1: Boxplots of margins for matches in NBL and WNBL for 2003-2004 season
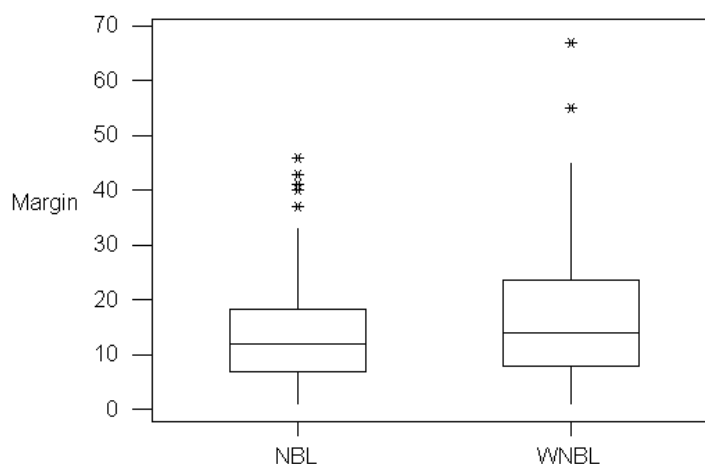
Figure 2 gives side-by-side boxplots of prediction errors for the WNBL and NBL in the 2003-2004 season, where, for each match, prediction error = |actual margin − predicted margin|. Again, the distributions are both skewed to the right, with some outliers, and with a median

prediction error of 9 for the NBL and a median prediction error of 10.5 for the WNBL. A two-sided Mann-Whitney test for the difference in medians showed that this difference was not significant, with P-value 0.46.

Figure 2: Boxplots of prediction errors for matches in NBL and WNBL for 2003/4
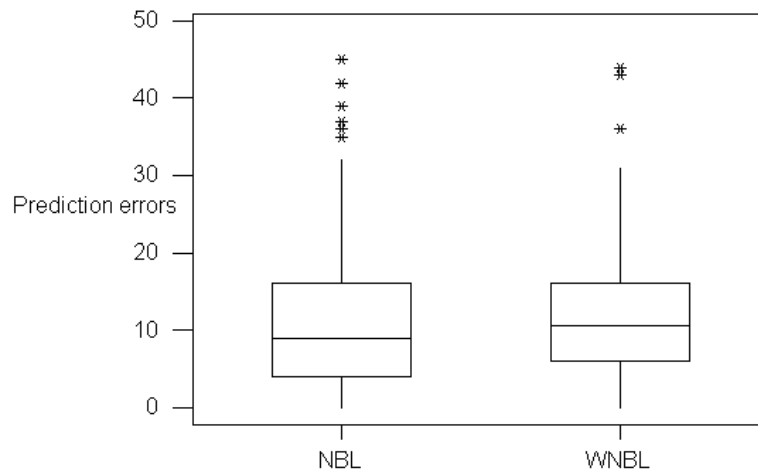
Tables 5 and 6 show the initial and final ratings assigned to each team, together with the ladder position at the end of the rounds, the number of games won in the season and the percentage of points (that is, the ratio of the total points scored by the team to the total points scored against the team expressed as a percentage). Final ratings generally reflected ladder positions, with two exceptions – the Melbourne Tigers and Townsville Crocodiles for the NBL and the Canberra Capitals for the WNBL. It is interesting to note that in the NBL, the final ratings of the top six teams were larger than the initial ratings, whereas the final ratings of the bottom six teams were less than their initial ratings. In the WNBL, the poor performance of the Perth Lynx side was probably a major contributor to the ratings of all teams except that of the Canberra Capitols rising.

Table 5: Ratings and ladder details for NBL, season 2003-2004

	Initial rating	Final rating	Position	Number of games won	Percentage
Sydney Kings	107	113.4	1	26	113.1
Wollongong Hawks	104	107.8	2	25	111.4
West Sydney Razorbacks	99	104.6	3	22	105.0
Brisbane Bullets	93	104.4	4	22	107.5
Melbourne Tigers	101	107.9	5	20	101.8
Cairns Taipans	98	102	6	16	102.1
Perth Wildcats	107	99.2	7	15	98.6
Adelaide 36ers	102	96.2	8	14	97.4
Townsville Crocs	105	101.6	9	13	97.4
New Zealand Breakers	103	95.8	10	12	94.3
Victoria Giants	95	91.4	11	11	91.9
Hunter Pirates	97	88.8	12	2	84.1

Table 6: Ratings and ladder details for WNBL, season 2003-2004

Team	Initial rating	Final rating	Position	Number of games won	Percentage
Dandenong Rangers	100	108.8	1	17	123.8
Sydney Flames	106	106.8	2	13	109.5
Adelaide Lightning	103	106.4	3	13	111.7
Canberra Capitals	116	111	4	13	110.3
Townsville Fire	104	104.2	5	12	105.7
Bulleen Melbourne	91	99.4	6	11	102.4
AIS	88	90.6	7	5	84.4
Perth Lynx	91	67.9	8	0	65.6

At the end of the season, the constant team ratings and a common home advantage was determined. These are given in Tables 7 and 8 for the WNBL and NBL respectively. Note that for the WNBL, five of the eight teams have a rating within 5 (the home advantage) of each other. So for any match involving two of these teams, the home team would be predicted to win. For both the NBL and WNBL, the overall model was significant at the .001 level, with $R^2 = 0.43$ for NBL and 0.62 for WNBL. The values of R^2 reflect the higher variability in the NBL. Harville and Norman [4] obtained a value of R^2 of 0.56 for college basketball, showing that it is less variable than the NBL but more variable than the WNBL. Clarke and Norman [2] obtained a value of R^2 of 0.19 for English soccer, reflecting its high variability, Clarke [1] obtained a value of R^2 of about 0.40 for Australian Rules Football, and Norton and Clarke [6] obtained a value of R^2 of about 0.70 for the Commonwealth Bank Trophy (netball).

Table 7: Team rating and common home advantage for WNBL, 2003-2004

Team	Rating
Dandenong Rangers	8.9
Sydney Flames	2.6
Adelaide Lightning	2.8
Canberra Capitals	2.3
Townsville Fire	0
Bulleen Melbourne	-1.5
AIS	-13.0
Perth Lynx	-27.5
Home advantage	5.0

Table 8: Team rating and common home advantage for NBL, 2003 - 2004

Team	Rating
Sydney Kings	28.8
Wollongong Hawks	25.1
West Sydney Razorbacks	20.3
Brisbane Bullets	22.7
Melbourne Tigers	17.6
Cairns Taipans	18.3
Perth Wildcats	14.5
Adelaide 36ers	13.1
Townsville Crocodiles	13.5
New Zealand Breakers	10.8
Victoria Giants	8.2
Hunter Pirates	0
Home advantage	4.2

If the ratings and common home advantage in Tables 7 and 8 had been known in advance, and kept fixed for the season, then the outcome of 82% of matches in the WNBL would have been correctly predicted with an average error of 10.95 points per match and the outcome of 74.7% of NBL matches would have been correctly predicted with an average error of 9.67 points per match.

4. Conclusion

There was no significant difference in the percentage of matches in the NBL and in the WNBL that could be predicted correctly. Overall about 70% of matches were correctly predicted in each league. If the predictions were done at random, then we would expect a 50% success rate, and the models used certainly improve on this. On the other hand, if the percentage of correct predictions is in excess of 80%, then there would be very few surprise results, and interest in the sport may decline. In this case, team equalisation policies should be put into play, not only to generate interest in the sport, but also to make teams more competitive.

The median margin in the NBL was less than that for the WNBL, which means that matches in the NBL are closer than in the WNBL. This is similar to the situation in tennis tournaments, where there are usually more women's matches with low scorelines, where one player is dominant. At the 2002 Australian Open, for example, the average number of games in a set of men's singles was 9.67, compared to 8.93 for a set of women's singles.

The basic prediction model could possibly be improved by having different home advantages for some teams. In the NBL, it was clear that Perth and Adelaide had a larger home advantage than the others, while for the WNBL, Adelaide in particular seemed to have a larger home advantage than other teams. It would also be useful to investigate the effect of changing the value of α, the constant used in adjusting the ratings after each match. Other methods of adjusting ratings could also be considered.

5. References

[1]   S. R. Clarke, "Performance modelling in sport", *Ph. D Thesis, Swinburne University* (1997).

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

[3]   K. S. Courneya and A. V. Carron, "The home advantage in sport competitions: a literature review",
J. Sport. Exer. Psychol., **14** (1992), 13-27.

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

[5]   A. M. Nevill and R. L. Holder, "Home advantage in sport: an overview of studies on the advantage of playing at home", *Sports Med.,* **28** (1999), 221-236.

[6]   P. Norton and S. R. Clarke, "Home ground advantage in the Australian netball league", *MathSport* **5**, G. Cohen and T. Langtry (Eds), University of Technology, Sydney, (2000), 168-173.

IF IT RAINS, DO YOU STILL HAVE A SPORTING CHANCE?

A critical review of two target resetting methods as applied to One-Day cricket

Matthew J. V. Ovens BSc.
School of Mathematical Sciences
Clayton Campus, Monash University, Australia

E-mail: Matthew.Ovens@sci.monash.edu.au

ABSTRACT
With the introduction of One-Day cricket and the objective of producing a result within a predetermined time frame, many methods have been suggested and used to recalculate the target score after a stoppage. Currently, the Professional Edition of the Duckworth/Lewis method is used, via computer software called CODA. Detailed analyses of the earlier methods, by other authors (e.g. [1]), have been published previously and will not be repeated in this work. This paper critically reviews the current method of target resetting and discusses a potential alternative.

1. Introduction

One-Day (or Limited Overs) cricket began in England in the 1960s and eventually spread to the member countries of cricket's governing body, the International Cricket Council (ICC). Originally the two sides were restricted to between 40 and 65 overs, however the standard for One-Day International (ODI) matches is now 50 overs. The game is played between two teams of 11 players. When batting, each side receives 50 six-ball overs unless they lose 10 wickets before such time. The side batting first (referred to as 'Team 1') attempts to score as many runs as possible whilst trying to ensure it will receive the full 50 overs. The side batting second (referred to as 'Team 2') must then attempt to score one more run, within 50 overs, than Team 1's total to win the game.

2. The Problem

The rules of One-Day International cricket do not allow for play to extend over to another day, in most tournaments, and as such, any stoppages (e.g. rain or sandstorms) affect the match time available. Even when play can be extended to a second day, such as in the upcoming ICC Trophy in England, this second day may also be affected by stoppages. In an effort to maintain fairness to both sides, the number of overs to be bowled is reduced. In the event that Team 2 has fewer overs to face than Team 1, the required target is adjusted to reflect this loss of overs. The problem then is, "How can the target be set in such a way as to not greatly advantage one team?"

3. Objectives A Solution Should Satisfy

Any suggested solution to the problem of fairly setting the target in a reduced overs match should, in this author's opinion, endeavour to meet (or attempt to come close to) as many of the following objectives, listed in order of decreasing importance, as possible.

Objective 1: Simplicity

The proposed method should be simple enough that any player, umpire, official, commentator or spectator can compute the new target, either mentally or with the aid of a "ready reckoner".

Objective 2: Adaptability
The proposed method should take into account all available variables (e.g. wickets in hand, balls remaining, etc.) with regard to the state of play at the time of the stoppage.

Objective 3: Lack of Constants
The proposed method should not require the use of arbitrary constants, except where these constants are derived from commensurate historical data and are used as a form of benchmarking. Constants derived from historical data (e.g. G50 in the Duckworth/Lewis method) should be updated after every commensurate game or at least after every series.

Objective 4: Repeatability
The proposed method should be capable of handling multiple stoppages within any one game. Additionally, each application of the method should be independent of how many stoppages have occurred previously (i.e. the method is invariant with respect to the number and manner of stoppages).

Objective 5: Fairness
The proposed method should set a fair and reasonable target with respect to the state of play at the time of the stoppage. The method should not be capable of being used as a playing strategy by either team (i.e. knowing that a stoppage is likely should not alter playing strategies, like choosing whether to bat or not after winning the toss).

Objective 6: Maintenance of Probability of Winning
The proposed method should attempt to maintain the probability of winning for each team across the stoppage. This objective assumes that the probability of winning is determinable at every stage of the match, which may not be true.

Objective 7: Maintenance of Excitement
The proposed method should maintain the competitive excitement for the spectators. The method should not be capable of producing a target that is clearly unachievable when a team is potentially capable of winning prior to a stoppage (e.g. Team 2 needing 22 runs off 13 balls prior to the stoppage revised to 21 runs required from 1 ball! {England v South Africa, World Cup 1992}. See Note 1).

4. The Current Solution: Duckworth / Lewis Method

The Duckworth/Lewis method comes in two varieties, Professional and Standard editions. The Professional Edition is to be used in all ICC recognised matches, according to ICC rules. The Standard Edition may be used in all other matches or in ICC recognised matches when the Professional Edition software is not working or unavailable. The mathematics and mechanics of the two methods are essentially the same. Since the full mathematical details behind the method are not available due to commercial reasons and the CODA software is currently only available to full ICC members, for the purpose of this paper, we refer only to the tables for the Standard Edition (hereafter known as the D/L method).

Although the full mathematical details of the current method are not available, the original version of the method was described in full mathematical detail in [1]. It is assumed that the derivation of the method remains unchanged and that only some of the parameter values have been altered to update it to the current method.

The central concept behind the method is that each team has a certain amount of resources with which to play the game. These resources consist of both the number of overs and the number of wickets available to each team. When a stoppage occurs, it causes the number of overs to be reduced; this is a reduction in resources. The amount of resources lost is

dependent upon the state of the match at the time of the stoppage (i.e. which team is batting, how many overs remain in their innings prior to the stoppage and how many wickets have been lost).

The D/L method (as described in [1]) sets the revised target by using the percentage of resources available to each team, the resources available to Team 1 are called R1 and similarly for Team 2 they are called R2. The values of R1 and R2 are obtained either from the software or from the D/L method table. When using the table, the resources are determined by the following procedure:

1. Note the resource percentage, from the table, corresponding to the number of overs left and the wickets lost. Call this Ra.
2. Note the resource percentage, from the table, corresponding to the number of overs remaining after the stoppage and the same number of wickets lost. Call this Rb.
3. The difference Ra- Rb gives the resource percentage lost. Call this Rc.
4. The resource available to Team A (RA) is then the resource percentage available to Team A at the start of their innings less Rc (e.g. R1=100- Rc, in the case where the first stoppage occurs during Team 1's innings)

In the case of multiple stoppages, the resource percentage lost due to a stoppage is subtracted from the resources available prior to that stoppage to get the new value of resources available to the current batting team.

These are then used in the following formula:

$$
T = \begin{cases} \left(S \times \dfrac{R2}{R1}\right) + 1 & if \quad R2 < R1 \\ S + 1 & if \quad R2 = R1 \\ S + \left(\dfrac{(R2 - R1) \times G50}{100}\right) + 1 & if \quad R2 > R1 \end{cases} \text{ (See Note 2)}
$$

where T is the target for Team 2 (rounded down to an integer), S is the number of runs scored by Team 1 and G50 is the agreed average total score in a 50-over innings (currently equal to 235).

With reference to the objectives defined in section 3 of this paper, the D/L method satisfies, to some extent, almost all of these objectives. Duckworth and Lewis ([2], [3]) identify that the objective of simplicity is of paramount importance to ensuring the acceptance of any method by the cricket-going public. Thus, Duckworth and Lewis appear to prefer that this objective must be met first and the Standard Edition meets this criterion. The method certainly satisfies the objective of adaptability and repeatability. The method does have one "constant" in it, namely G50, which is the agreed average score for that level of play. With regards to the remaining three objectives, namely Fairness, Maintenance of Probability of Winning and Maintenance of Excitement, this method is often criticised. Many spectators and commentators believe that the method favours the team batting at the time of the stoppage. This author tested this hypothesis using available World-Cup, ICC Trophy and other One-Day tournament data where the D/L method had been used and a match result obtained and found no significant result, but the data set was small. To the outside observer though, it does not appear that the potential use of the D/L method alters the playing strategies of opposing teams (i.e. choice of batting or bowling first, made at the toss).

Duckworth and Lewis [2] identify that the "Probability of Winning" is difficult to define and determine, but suggest that using "degree of difficulty" in place of probability is reasonable.

This allows the D/L method definition of resources to be used to maintain the probability of winning across a stoppage. The main difficulty associated with this objective still comes from stoppages during Team 1's innings, since the "probability" of winning is even harder to find.

The Maintenance of Excitement objective, whilst important in this increasingly commercial age, should only be used as a final criterion when evaluating methods. The D/L method's main aim is to maintain the margin of advantage across a stoppage. This sometimes leads to a game being already won when play could be recommenced, a hardly exciting prospect for the spectators who have waited around in case play could recommence. In some cases, the D/L method has maintained the excitement for the fans by setting an achievable, but difficult target. Thus, this method does, on occasion, satisfy the last objective.

5. An Alternative Solution: Jayadevan's Method

An alternative to the Duckworth/Lewis method is that proposed by V. Jayadevan in his paper published in *Current Science* [4]. Jayadevan, a civil engineer, took the novel approach of looking at the scoring patterns of teams playing in closely fought matches. These scoring patterns were then expressed as percentages of innings and runs to create normal (or par) scores. Regression analysis performed on the data produced a cubic equation (Equation 2, where R is cumulative percentage runs and O is the cumulative percentage overs) for the normal scores. This cubic equation resembles the match development much better than the parabola (Equation 1, Figure 1) used in the Parab method (as described in [1]). A second cubic equation (Equation 3) is found, again using regression, to provide target scores. Figure 2 shows the two curves plotted on the same axis. It should be noted that the normal curve corresponds to the expected number of wickets lost at each stage of the match, whereas the target curve corresponds to nine wickets down, since the match can still be won.

$$R = 7.46 \times O - 0.059 \times O^2$$

Equation 1: Parab Method Equation
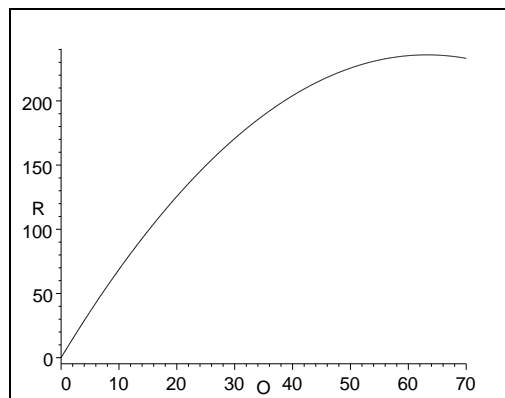
Figure 1: Graph of the Parab Method Equation

$$R = 1.305717007 \times O - 0.013782 \times O^2 + 0.0001069 \times O^3$$

Equation 2: Jayadevan's Normal Score Equation

$$R = 1.6631192 \times O - 0.009254 \times O^2 + 0.0000261 \times O^3$$

Equation 3: Jayadevan's Target Score Equation
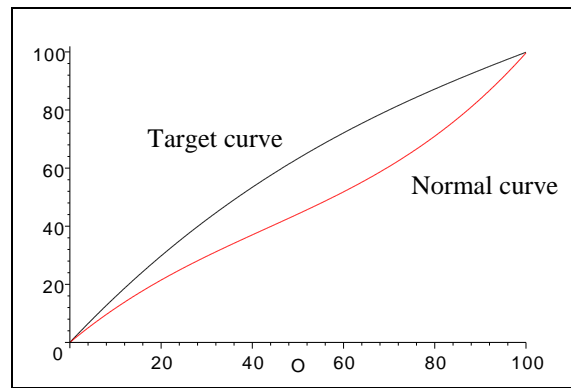
Figure 2: Jayadevan method cubic curves

As Figure 2 shows, the "normal" development of an innings is: a quick start followed by a "survival and accrual" period and ending with a "slogging" session. Jayadevan provides details of how these equations were derived and the assumptions underlying the method.

Jayadevan states that any stoppage in a match can be classified into one of three mutually exclusive possibilities. The three possibilities are: a stoppage between innings; a stoppage during Team 2's innings; or a stoppage during Team 1's innings. Stoppages that occur prior to commencement of play do not affect the application of the method. Multiple stoppages are handled by repeated application of the method.

5.1 A stoppage between innings:

In the case of a stoppage occurring after the end of Team 1's innings and prior to Team 2 commencing their innings, Jayadevan's method is applied as follows:
1. Determine the percentage of overs for Team 2.
2. Look up the corresponding target score percentage in the target table.
3. Multiply this percentage by Team 1's final score to obtain the new target score for Team 2.

5.2 A stoppage during Team 2's innings:

In the case of a stoppage occurring during Team 2's innings, Jayadevan's method is applied as follows:
1. Determine the percentage of overs already completed by Team 2.
2. Look up the corresponding normal score percentage in the normal table for the number of wickets fallen.
3. Multiply this percentage by Team 1's final score to obtain the par score for Team 2 (called PAR1).
4. Determine the percentage of remaining overs to be played at recommencement.
5. Look up the corresponding target score percentage.
6. Multiply this percentage by the result of Team 1's score minus the par score (PAR1) found in step 3.
7. Add PAR1 to the result of step 6 to find the new target score for Team 2 (net target-1). In the event of another interruption to Team 2's innings:
8. Using the table, determine N1, the normal score percentage corresponding to the first stoppage with respect to the total number of overs available to Team 2 prior to this stoppage. Then find N2, the normal score percentage corresponding to the current state of play with respect to the total number of overs available to Team 2 prior to this stoppage. PAR2 is then given by:

$$PAR2 = PAR1 + \frac{\left(N2 - N1\right)}{\left(100 - N1\right)}$$

9. Determine Ta, Tb and Tc. Ta is the target score percentage corresponding to the number of overs to be bowled after the stoppage (C) divided by the number of overs remaining when the first stoppage occurred (A). Tb is the target score percentage corresponding to the number of overs to be bowled before the stoppage (B) divided by the number of overs remaining when the first stoppage occurred (A). Tc is the ratio of Ta and Tb. Subtract PAR2 from net target-1 and multiply the result by Tc.
10. Add PAR2 to the result of step 9 to find net target-2. This is the new target for Team 2.

5.3 A stoppage during Team 1's innings:

In the case of a stoppage occurring during Team 1's innings, Jayadevan's method is applied as follows:
1. Determine the percentage of overs completed by Team 1.
2. Look up the corresponding normal score percentage in the normal table for the number of wickets fallen.
3. Determine the percentage of remaining overs after the stoppage with respect to the original number of overs remaining.
4. Look up the corresponding target score percentage in the target table.
5. Multiply the target score percentage by the difference between 100% and the normal score percentage.
6. Add this percentage to the normal score percentage obtained in step 2 to get the Effective Normal Score (ENS) in the total percentage of overs played.
7. Look up the target score percentage for the total percentage of overs played.
8. Multiply this target percentage by the ENS from step 6 to get the Multiplication Factor (MF).
9. Multiply the score made by Team 1 with MF to get the target for Team 2.
   In the event of another interruption to Team 1's innings:
10. Compute PAR2 as:

$$PAR2 = PAR1 + \frac{\left(N2 - N1\right)\left(ENS - PAR1\right)}{\left(100 - N1\right)}$$

11. Compute target percentage for the remaining overs as:

$$T_c \times \left(ENS - PAR2\right)$$

12. Add the result from step 11 to PAR2 to get the new ENS.
13. Look up the target score percentage for the total percentage of overs played.
14. Multiply this target percentage by the new ENS to get the new Multiplication Factor (MF).

Whilst, at first glance, Jayadevan's method appears complex, it is actually relatively simple to apply and is only slightly more complicated than the D/L method. Jayadevan's method requires one lookup in a table and one multiplication in the case of a stoppage between innings, up to a maximum of three lookups and four arithmetic operations. By contrast, the D/L method requires two lookups and a minimum of five arithmetic operations. Thus, in terms of table lookups and simple arithmetic operations, there is little difference in the work required to implement either method. The use of computer software makes it even easier, but this would not be required to employ the Jayadevan method (c.f. the Professional Edition of D/L).

With reference again to the objectives defined in section 3 of this paper, the Jayadevan method satisfies, to some extent, most of these objectives. The method, as seen above, requires the same skills and approximately the same number of operations as the D/L method and thus also meets the objective of simplicity. The method certainly satisfies the objective of adaptability and repeatability as any state of the game can be handled by the method. The method does not have any constants in it, however the regression equations used to derive the method do. This is not a major problem as these constants, like the G50 in the D/L method, would need to be updated from time to time to reflect the changing nature of the game. Unlike the G50 though, they are not required when applying the method to the current state of play.

Owing to the fact that this method has not yet been applied in the international cricket arena, it is difficult to determine if the method would satisfy the Fairness objective. The method presents no obvious way in which it could be used as a playing strategy and, due to its similarity to the D/L method; one would estimate that the Jayadevan method would meet the objective in the same way. The Maintenance of Probability of Winning objective would also be met in a similar way to that of the D/L method, but again, the determination of the probability of winning is difficult, if not impossible, at every stage of a match.

Similarly to the D/L method, the objective of Maintenance of Excitement, the Jayadevan method can sometimes lead to a game being already won when play could recommence. The method can, in some cases, maintain the excitement for the fans by setting a reasonable target. Thus, this method should, on occasion, satisfy the last objective.

6. Comparison of Solutions

Comparing both of the methods reviewed in this paper, we can construct Table 1, which provides a quick summary of the objectives, as defined in section 3, met by each method.

Method	Objective						
	Simplicity	Adaptability	Lack of Constants	Repeatability	Fairness	Maintenance of Probability of Winning	Maintenance of Excitement
D/L	✓*	✓	✓	✓**	✓'	⊠''	✓
Jayadevan	✓	✓	✓	✓	✓?	?	?

*	The Professional Edition requires software & a PC
**	The constant is only updated periodically
'	Playing strategy is sometimes altered because a stoppage may occur
''	The probability of winning is not maintained
✓?	The method appears not to be able to be used as a playing strategy
?	These properties are undetermined due to a lack of real world data

Table 1: Method Comparison via Objectives

The true test of any method, though, is in the ability of the method to set fair and reasonable targets in real world (or theoretical) situations. To compare the methods, the examples used are taken from various papers by Jayadevan and Duckworth and Lewis and a recent match between Australia and India. These examples are listed, in no particular order, and summarised in Table 2.

6.1 Example 1:

Team 1 scores 300 in their allotted 50 overs. Team 2 has lost 2 wickets after 25 overs when play is abandoned. What is the target for Team 2 to be declared the winners?

6.2 Example 2:

Team 1 scores 300 in their allotted 50 overs. A stoppage then occurs causing Team 2's innings to be shortened to 25 overs. What is the new target for Team 2

6.3 Example 3:

Team 1 scores 250 in their allotted 50 overs. Team 2 in reply are 75 with no wickets down after 20 overs when a stoppage occurs and their innings is shortened to 30 overs. What is the winning target for Team 2?

6.4 Example 4:

Team 1 scores 250 in their allotted 50 overs. Team 2 in reply are 75 with 2 wickets down after 20 overs when a stoppage occurs and their innings is shortened to 30 overs. What is the winning target for Team 2?

6.5 Example 5:

Team 1 scores 250 in their allotted 50 overs. Team 2 in reply are 180 with 4 wickets down after 30 overs when a stoppage occurs and play is abandoned. What is the winning target for Team 2?

6.6 Example 6:

Team 1 scores 100 in 25 overs when a stoppage occurs, terminating their innings. Team 2's innings is also shortened to 25 overs. What is the winning target for Team 2?

6.7 Example 7:

{ODI #1091} Team 1 has scored 226 for the loss of 8 wickets in 47.1 overs when a stoppage occurs, terminating their innings. Team 2 have their innings shortened to 33 overs. What is the winning target for Team 2?

6.8 Example 8:

{ODI #1812} Team 1 has scored 176 for the loss of 5 wickets in 36.5 overs when a stoppage occurs, shortening their innings to 46 overs. Team 1 scores a further 5 runs from 6 balls (i.e. 181/5 after 37.5 overs) when a stoppage shortens their innings to 40 overs. Team 1 finish their innings at 193 for the loss of 6 wickets. What is the winning target for Team 2?

6.9 Example 9:

{ODI #2086} Team 1 has scored 296 for the loss of 4 wickets in their rain interrupted full 50 overs. Team 2 are 73 with the loss of 1 wicket when a stoppage occurs, shortening their innings to 34 overs. What is the winning target for Team 2?

Ex.	D/L[*]	Jayadevan	Target Set
1	119	130	-
2	201	191	-
3	143	159	-
4	159	165	-
5	139	144	-
6	178	157	-
7	192	187	187
8	220	212	223
9	233	215	225

Table 2: Method comparison by examples

* Duckworth/Lewis targets calculated with the Standard Edition tables

From the table above we can see that in all cases there is a difference of at least 5 runs between the two target resetting methods. For the real-world cases, Examples 7, 8 and 9, the actual targets set in the match are also given. It should be noted that the rain-rule used in Example 7 was not the D/L method and from the target set alone, since this author has not discovered any other information, the method most likely used was the MPO method. In this case, Team 2 managed to reach the target with 5 overs to spare. Examples 8 and 9 are both cases where the D/L method was actually used and the targets were set using the Professional Edition of the D/L method. In the case of Example 8, Team 2 was all out 33 runs short of the target and having faced all but 2 of their overs. In Example 9, Team 2 reached the target with 1 ball to spare and 2 wickets in hand.

It is difficult to quantify what constitutes a fair and reasonable target for Team 2, rather it is generally a very subjective choice made be the individual. This author believes that the targets set by the Jayadevan method are more fair and reasonable than those set the D/L method in the cases demonstrated here. As mentioned earlier, several hundred simulations have been conducted that, in most cases, show there is little difference between the targets set by the two methods. Only in some extreme cases do the methods differ and always the D/L method sets a higher (and perhaps less reasonable) target.

7. Conclusion

Whatever the method chosen for resetting the target in a one-day international match, there will always be criticism. The fans will usually claim the method is unfair to their team, regardless of the true state of the match. Commentators will complain that the method is too complicated to understand or that it is too simple and fails to account for one factor or another. The method, whatever it is, should be transparent and open to scrutiny from any interested party. This is one of the main criticisms of the current Duckworth / Lewis method, in that the mathematical details and CODA software are not widely available due to commercial reasons, which leads to some fans believing that the method relies upon black magic rather than logic. Jayadevan method, on the other hand, has detailed the methodology and the mathematics behind the creation of his method, allowing everyone to scrutinise the details. According to Jayadevan, "a large number of players, umpires, cricket administrators, critics and cricket enthusiasts," have found the method superior to that of Duckworth and Lewis [4].

This paper set out to critically evaluate the current method and Jayadevan's proposed alternative method for resetting the target in a one-day international cricket match after a stoppage. The results of this evaluation lead to the conclusion that Jayadevan's method is perhaps superior to that of Duckworth and Lewis, even though the method does not meet all of the objectives as defined in section 3. Like the Duckworth / Lewis method before it, this

method is the best of what is currently on offer as an alternative rain-rule method and should therefore be quickly and thoroughly analysed by the ICC and its member nations and adopted as the new rain-rule.

It is interesting to note that the Jayadevan method was presented to the BCCI Technical Committee Conference on 11 July 2000 and to the Umpire's Seminar in September 2000 as well as at the BCCI meeting on 7 April 2001. The BCCI forwarded the proposal to the ICC, but, as yet, the ICC have not taken up the method.

8. References

[1]   F. Duckworth, T. Lewis, "A Fair Method For Resetting the Target in Interrupted One-day Cricket Matches", *MathSport* **3**, N. de Mestre, Ed., Bond University, Queensland, Australia, (1996), 51-68.
[2]   F. Duckworth, T. Lewis, "Can the Probability of Winning a One-day Cricket Match be Maintained Across a Stoppage?", *MathSport* **6**, G. Cohen, T. Langtry, Eds., Bond University, Queensland, Australia, (2002), 141-154.
[3]   F. Duckworth, T. Lewis, "Review of the Application of the Duckworth/Lewis Method of Target Resetting in One-Day Cricket", *MathSport* **6**, G. Cohen, T. Langtry, Eds., Bond University, Queensland, Australia, (2002), 127-140.
[4]   V. Jayadevan, *Current Science* **83**, 577-586 (2002).
[5]   F. Duckworth, T. Lewis, "Developments in the Duckworth-Lewis (D/L) Method of Target - Resetting in One-Day Cricket Matches", *MathSport* **4**, N. de Mestre, K. Kumar, Eds., Bond University, Queensland, Australia, (1998), 131-151.
[6]   F. Duckworth, T. Lewis, "Comments on Matthew Ovens 'If it rains…'", Personal Correspondence, 14 May 2004.

9. Notes

9.1 Note 1:
England vs South Africa, World Cup 1992. This application of the rain-rule has been much debated. Many have questioned why the full 13 deliveries weren't bowled given that there was time available. If the D/L method were used then the target would have been 4 runs from 1 ball [6]. By comparison, if the Jayadevan method were used then the target would have also been 4 runs from 1 ball. This target is more reasonable than the one set by the rain-rule in place at the time. This extreme case should not be used as the basis for selecting a method, but does highlight other issues that should be looked at with regards to when a rain-rule should be applied. Details of the match can be found on the web at CricInfo or via the URL:

http://aus.cricinfo.com/link_to_database/ARCHIVE/WORLD_CUPS/WC92/ENG_RSA_WC92_ODI-SEMI2_22MAR1992.html

9.2 Note 2:
The case where resources are equal was not included in [1] and the original paper on the method, as published in the Journal of Operational Research. This case was included after further development of the method ([5], [6]).

10. Additional Resources

1.   P. Allsopp, S. Clarke, Methods for Quantifying Performances in One-Day Cricket, G. Cohen, T. Langtry, Eds., 5th Australian Conference on Maths and Computers in Sport, University of Technology, Sydney, Australia (2000).
2.   D. R. Bingham, B. M. de Silva, T. B. Swartz, Issues in Cricket and Golf, N. de Mestre, K. Kumar, Eds., 4th Australian Conference on Maths and Computers in Sport, Bond University, Queensland, Australia (1998).
3.   G. A. Christos, It's Just Not Cricket, N. de Mestre, K. Kumar, Eds., 4th Australian Conference on Maths and Computers in Sport, Bond University, Queensland, Australia (1998).
4.   B. M. de Silva, G. R. Pond, T. B. Swartz, Applications of the Duckworth-Lewis Method, G. Cohen, T. Langtry, Eds., 5th Australian Conference on Maths and Computers in Sport, University of Technology, Sydney, Australia (2000).
5.   K. Kumar, Is Cricket Really by Chance?, N. de Mestre, Ed., 3rd Australian Conference on Maths and Computers in Sport, Bond University, Queensland, Australia (1996).
6.   J. Armstrong, Masters Thesis, Monash University (1994).
7.   S. Bhogle. (Rediff On The NeT, 1999).
        http://www.cmmacs.ernet.in/nal/pages/dlm.htm

CAN A TENNIS PLAYER INCREASE THE PROBABILITY OF WINNING A POINT WHEN IT IS MORE IMPORTANT?

Graham Pollard

School of Information Sciences and Engineering

University of Canberra, ACT 2601, Australia

Abstract

The benefit a player can receive by "lifting his/her play" on the more important points is well known. Is there any evidence, however, that some players can lift their player on particular points? In this paper we analyse some situations in which a particular highly ranked player may wish to lift his play, and conclude that there is some evidence that he can do so.

1. Introduction

In a very elegant paper Morris (1977) demonstrated the benefit a player receives by increasing his/her probability of winning the more 'important' points, whilst correspondingly decreasing the probability of winning the less important points. In a related paper, Pollard (2002) considered the situation in which the server does not play a single type of point, but selects from a range of options on both the first and second service. He assumed that the player must use the full range of options in order to keep the opponent guessing as to which option is to be used next, and he demonstrated the benefits of playing the better options on the more important points.

In this paper we consider whether there is any evidence that a player can increase his/her probability of winning a point when he/she might wish to. Such evidence might be found for a highly ranked player in good form. We analysed the data for the seven matches played by Andre Agassi in the 2003 Australian Open. He was the second seed in the tournament, and was quite highly favoured to win the tournament, which he did.

2. The Analysis

Agassi's opponents and the outcomes in the seven matches were:

Round 1:	B. Vahaly	7/5, 6/3, 6/3
Round 2:	H-T. Lee	6/1, 6/0, 6/0
Round 3:	N. Escude	6/2, 3/6, 6/3, 6/4
Round 4:	G. Coria	6/3, 3/1 retired
Quarter-final:	S. Grosjean	6/3, 6/2, 6/2
Semi-final:	W. Ferreira	6/2, 6/2, 6/3
Final:	R. Scheuttler	6/2, 6/2, 6/1

2.1   Agassi Receiving

Given that breaking the opponent's serve (at least some of the time) is particularly important in order to win a match of men's tennis, we firstly considered all games in which Agassi received. Table 1 gives the relative frequencies of points won by Agassi's opponents when serving. As four of the seven matches had some games with quite a few deuces (two matches each had a game with nine deuces), and as we did not wish to have the potential to bias the conclusions as a result of such 'long' games, we decided to consider only the first six points (at most) in each service game.

Table 1 The relative frequencies of points won by Agassi's opponents on their services

Score	Relative Frequency of points won by Agassi's opponents on their service
0-0	48/87
15-0	26/48
0-15	18/39
30-0	14/26
15-15	21/40
0-30	11/21
40-0	8/14
30-15	10/33
15-30	18/30
0-40	4/10
40-15	9/16
30-30	19/41
15-40	10/16
40-30	13/26
30-40	12/32
Total	241/479

The following conclusions can be made from Table 1:

- The relative frequency of points won by Agassi's opponents on their service was 241/479=0.50.

- The relative frequency on the first or forehand court was 132/247=0.53, and was 109/232=0.47 on the second or backhand court. The Z-value for the difference between these two sample proportions is Z=1.41. Although not significant at the 5% level of significance, Agassi was somewhat more effective when receiving from the second court.

- The relative frequency of Agassi's opponents winning during the first two points was 92/174=0.53. For the 3rd and 4th points combined, this relative frequency was 86/174=0.49, and for the 5th and 6th points it was 63/131=0.48. Thus, Agassi was somewhat more effective on the later (and frequently more important) points in the game.

- The relative frequency of points won by Agassi's opponents at 0-0 was 48/87=0.55. At 15-15 this relative frequency was 21/40=0.53 and at 30-30 it was 19/41=0.46. Thus, for the situations in which the score was equal, Agassi was somewhat more effective on the later points (and frequently more important points) in the game.

- Considering the second or backhand court, the relative frequency of points won when Agassi was one point behind (ie. scores of 40-30, 30-15 and 15-0) was 49/107=0.46. When Agassi was one point ahead (ie. scores of 30-40, 15-30 and 0-15), it was 48/101=0.48, and on the other points (ie. 40-0 and 0-40, which are relatively unimportant), it was 12/24=0.50. Thus, Agassi was slightly more effective when the score was 'closer'.

Table 2 gives information on the number of times a point was won by Agassi's opponent following a point lost by Agassi's opponent (W/L), etc. Note that the first point in each game was omitted from the analysis as the previous point played on service was during the previous service game. Agassi's opponents' relative frequency of losing a point having lost the previous point (L/L) was 104/193=0.54, whereas their relative frequency of losing a point having won the previous point (L/W) was 95/199=0.48 (Z=1.22). Thus, although not significant at the 5% level of significance, Agassi was somewhat more likely to win a point when receiving having won the previous point (than having lost the previous point). This characteristic is useful for enhancing Agassi's chance of breaking service.

Table 2 One-step frequencies on Agassi's opponents' services (W represents won by Agassi's opponent)

1-step type	Frequency
W/W	104
W/L	89
L/W	95
L/L	104
Total	392

2.2   Agassi Serving

Table 3 gives the relative frequencies of points won by Agassi when serving.

Table 3 The relative frequencies of points won by Agassi on his service

Score	Relative Frequency of points won by Agassi on his service
0-0	50/82
15-0	31/50
0-15	24/32
30-0	21/31
15-15	21/43
0-30	7/8
40-0	15/21
30-15	26/31
15-30	21/29
0-40	1/1
40-15	25/32
30-30	17/26
15-40	6/9
40-30	17/25
30-40	12/15
Total	294/435

The following observations can be made:

- The relative frequency of points won by Agassi on service was 294/435=0.68.

- The relative frequency on the first court was 147/231=0.64, and on the second court was 147/204=0.72 (Z=1.87).Thus, Agassi was somewhat more effective when serving to the second court.

- The relative frequency of points won by Agassi on the first two, the second two and the 5th and 6th points in the game was 105/164=0.64, 112/164=0.68 and 77/107=0.72 respectively. Thus, Agassi was somewhat more effective on the later (and frequently more important) points in the game.

- The relative frequencies of points won by Agassi on service at 0-0, 15-15 and 30-30 was 50/82=0.61, 21/43=0.49 and 17/26=0.65 respectively. Although Agassi had a low relative frequency at 15-15, it was higher at the more important score of 30-30.

- The relative frequency of points won by Agassi when he was one point behind (ie. at scores of 0-15, 15-30 and 30-40) was 57/76=0.75, which is slightly greater than when he was one point ahead (74/106=0.70). This characteristic is useful for reducing Agassi's chance of losing service.

Table 4 One-step frequencies on Agassi's service (W represents won by Agassi)

1-Step Type	Frequency
W/W	156
W/L	88
L/W	70
L/L	39
Total	353

Table 4 gives information corresponding to that of Table 2 for the case in which Agassi was serving. In this case there was essentially no observed difference between Agassi's relative frequency of winning a point given that he lost the previous point (88/127=0.69) and the relative frequency of winning a point given that he won the previous point (156/226=0.69).

3. Conclusions

In this paper we have considered whether there is any evidence that a player can increase his probability of winning a point when he might wish to. The player considered was Agassi, and the data was his seven matches at the 2003 Australian Open.

When receiving, Agassi won a higher percentage of points towards the end of the game, he performed better at 30-30 than at 0-0 or 15-15, he performed possibly at his best when he was one point behind, and he was somewhat more likely to win a point having won the previous point (than having lost that point). All of these characteristics increase his likelihood of breaking his opponent's service.

When serving, Agassi won a higher percentage of points towards the end of the game, he performed better at 30-30 than at 0-0 or 15-15, and he performed possibly at his best when he was one point behind. These characteristics enhance Agassi's likelihood of holding his own service.

The overall conclusion is that there is some evidence that this player in particular can increase his probability of winning points that are more important. It would seem useful to repeat the type of analysis used in this paper to see whether some other top players have a similar capacity to lift their play.

Point-by-point data is available for Grand Slam Tournaments, and can be a useful source of data for statistical exercises for students. Such data can be useful for developing statistical concepts including independence, conditional probability and hypothesis testing.

4. References

[1]   Morris, C., "The most important points in tennis", *Optimal Strategies in Sports,* Ladany, S. P. and Machol, R. E. (Eds), North-Holland, Amsterdam (1977), 131-140.

[2] Pollard, G, "The effect of a variation to the assumption that the probability of winning a point in tennis is constant", *Proceedings of the sixth Australian conference on mathematics and computers in sport*, Cohen, G. and Langtry, T. (Eds), University of Technology, Sydney, (2002) 221-226.

SOME ATTRACTIVE PROPERTIES OF THE 16-POINT TIEBREAK GAME IN TENNIS

Graham Pollard and Ken Noble
School of Information Sciences and Engineering
University of Canberra, ACT 2601, Australia

Abstract

Factors such as first and second court effects when serving, end effects due to (say) the direction of the wind, and sun effects when serving can all affect the probability that a player wins a point in tennis. A method for designing fair lead by at least 2 and lead by at least 4 tiebreak games when such factors exist is described. The method includes a modification to the stopping rules presently in use or described in earlier research. Both singles and doubles are considered.

1. Introduction

A fair scoring system in tennis has the characteristic that the probability that each player (or team) wins is 0.5 when each player (or team) has the same probability of winning each type of point. In a recent paper Pollard and Noble (2002) showed that in some doubles situations the first to 7 points leading by at least 2 points tiebreak game (i.e. the 12 point tiebreak game) is unfair. They also noted that the first to 10 points leading by at least 2 points tiebreak game (i.e. the 18 point tiebreak game) is unfair in such doubles situations. The doubles situation considered was that in which the players on team A had probabilities on service p_{a1} and p_{a2} ($p_{a1} \neq p_{a2}$), and the players on team B had probabilities on service p_{b1} and p_{b2}, and with the two teams being "equal" in the sense that $p_{a1}=p_{b1}$ and $p_{a2}=p_{b2}$. They showed that fairness could be established for this case by making use of the 16-point tiebreak game. The advantage of considering multiples of 8 or 16 points was noted, as was a slight modification to the lead-by-2 rule if the number of points played is greater than 16 (after a score of 8-8 is reached).

In another paper Pollard and Noble (2003) showed that, in singles (and in doubles where $p_{a1}=p_{a2}$ and $p_{b1}=p_{b2}$), the first to 7 points leading by at least 4 points tiebreak game is fair. They noted the increased probability of the better player winning under the lead-by-4 rule (rather than the lead-by-2 rule). They also noted that lead-by-3 and lead-by-5 rules are unfair.

In this paper we show that it is possible to design a *fair* tiebreak game that handles additional parameters. In singles, we consider situations in which:

- Each player has the same advantage (or disadvantage) when serving to the forehand or first court (relative to serving to the backhand or second court).
- Each player has the same advantage (or disadvantage) when playing from one particular end of the court (relative to the other end). This situation can occur when there is a strong enough wind.
- Each player has difficulty with the position of the sun when serving from one particular end of the court. This can occur near the middle of the day, for example.

These various situations are also considered for the case of doubles.

Variations to the lead-by-2 and lead-by-4 rules, and how these variations contribute to fairness, are also considered.

2. Singles with first and second court effects, and with end and sun effects

2.1 Leading by at least 2 points and stopping rule, SR1

Table 1 gives probabilities for singles players A and B for the 16 point tiebreak game, with change of ends after points 4, 12, 20, … . Player A serves from "End 1" on points 1 and 4, and from "End 2" on points 5, 8, 9 and 12, and so on. The column labelled "End 1" shows an end effect of +e on *all* points for the player at that end, and a sun effect of +s for the player *serving* from that end . Note that these effects can be positive or negative, and cover all situations in which a player performs better from one end relative to the other, and where serving from one end is more challenging than from the other. Both players have an effect of +f when serving to the forehand court, and an effect of +b when serving to the backhand court. Thus, for example, player A's probability of winning point 4 is equal to $p_a+e+s+b$, and his probability of winning point 5 is equal to p_a-e+f. The receiver's probabilities are given in brackets. When $p_a= p_b$, the first eight point probabilities for player A: $p_a+e+s+f$, (q_b+e-b), (q_b+e-f), … , p_a-e+b can be seen to be equal to the first eight point probabilities for player B (note that $q_a=1-p_a$ and $q_b=1-p_b$), although the ordering of the eight probabilities is different for the 2 players. Correspondingly, when $p_a= p_b$, the second eight point probabilities for player A are equal to the second eight for player B. and also the third two sets of eight point probabilities are equal, and so on.

Table 1 - Singles with first and second court effects, and with end and sun effects
   Change Ends after points 4, 12, 20, …

Point	End 1		End 2	
1	$p_a + e + s + f$	[A serves]	$(q_a - e - s - f)$	
2	$(q_b + e - b)$		$p_b - e + b$	[B serves]
3	$(q_b + e - f)$		$p_b - e + f$	[B serves]
4	$p_a +e + s + b$	[A serves]	$(q_a - e -s - b)$	
5	$(q_a + e -f)$		$p_a - e + f$	[A serves]
6	$p_b + e + s + b$	[B serves]	$(q_b - e -s -b)$	
7	$p_b + e + s + f$	[B serves]	$(q_b - e -s -f)$	
8	$(q_a + e - b)$		$p_a - e +b$	[A serves]
9	$(q_a + e - f)$		$p_a - e + f$	[A serves]
10	$p_b + e + s + b$	[B serves]	$(q_b - e - s - b)$	
11	$p_b + e + s + f$	[B serves]	$(q_b - e - s - f)$	
12	$(q_a + e - b)$		$p_a - e + b$	[A serves]
13	$p_a + e +s + f$	[A serves]	$(q_a - e - s - f)$	
14	$(q_b + e -b)$		$p_b - e + b$	[B serves]
15	$(q_b + e - f)$		$p_b - e + f$	[B serves]
16	$p_a + e + s + b$	[A serves]	$(q_a - e - s - b)$	
17	$p_a + e + s + f$	[A serves]	$(q_a - e - s -f)$	
18	$(q_b + e - b)$		$p_b - e + b$	[B serves]

Here we note a result by Newton and Pollard (2004). They considered, for example, a conventional set of tennis and a fictitious set of exactly 10 games in which each player served exactly five times (in any order). They proved that the probability player A won 6-0, 6-1, 6-2, 6-3 or 6-4 using conventional scoring was equal to the probability player A won using the fictitious scoring, and the probability that the score reached 5-5 was the same under both conventional and fictitious scoring. Also, the probability player A lost 6-0, 6-1, 6-2, 6-3 or 6-4 using conventional scoring was the same as under the fictitious scoring. We note here that their proof carries over to the case in which player A has five different probabilities of winning each of his five service games, and player B also has five different probabilities of winning each of his service games. As in the

paper by Newton and Pollard (ibid) this result can be applied to points within the tiebreak game instead of games within the set. Thus, applying the result in this way, and considering sets of eight points for two equal players, it can be seen, by symmetry, that the probability player A wins equals the probability player B wins. Thus we can create a fair scoring system for the situation in Table 1 provided our 'stopping' rule make use of points eight at a time.

Thus, best of sixteen points (first to 9 points leading by at least 2 points), or, if the points score reaches 8-8, best of 24 points (first to 13 points leading by at least 2 points), or, if the points score reaches 12-12, best of 32 points (first to 17 points leading by at least 2 points), etc is a fair 'stopping' rule for Table 1, and is denoted by SR1. Possible winning scores for SR1 are 9-0, 9-1, 9-2, … , 9-7, 13-8, 13-9, 13-10, 13-11, 17-12, 17-13, 17-14, 17-15, 21-16, etc. Note that scores such as 10-8, 11-8,11-9, 12-8, 12-9, 12-10, …in which a player is leading by at least 2 points but has not reached 13 points, are not permissible (i.e. fair) finishing scores as the two players have not experienced equal point probabilities at that stage.

2.2   Leading by at least 4 points and stopping rules, SR2 and SR3

For the probabilities in Table 1, there are fair 'stopping' rules based on leading by at least 4 points, rather than leading by at least 2 points, again with the 'stopping' rule making use of points eight at a time. Thus, best of sixteen points (first to 10 points leading by at least 4 points), or if the points score reaches 8-8 or 9-7, best of 24 points (first to 14 points leading by at least 4 points), or if 12-12 or 13-11 is reached, best of 32 points (first to 18 points leading by at least 4 points), etc is a fair lead by at least 4 points 'stopping' rule for Table 1, and is denoted by SR2. The possible winning scores under SR2 are 10-0, 10-1, 10-2, … , 10-6, 14-7, 14-8, 14-9, 14-10, 18-11, 18-12, 18-13, 18-14, 22-15, etc. Note that scores such as 11-7, 12-7, 12-8, 13-7, 13-8, 13-9, …are not permissible fair finishing scores.

Lead by 4 points stopping rules can have large standard deviations in the number of points in the tiebreak game. This standard deviation can be reduced considerably whilst maintaining fairness by changing to a win by 2 points structure after (say) 16 points have been played. Such a fair stopping rule for Table 1, SR3, is the first to 10 points leading by at least 4 points, or if a score of 8-8 or 9-7 is reached, the first to 13 points leading by at least 2 points, or if the score reaches 12-12, the first to 17 points leading by at least 2 points, or if 16-16 is reached, the first to 21 leading by at least 2 points, etc.

(It is noted here as an aside that a variation of the scoring system in the above paragraph might be useful for the case in which the tiebreak game is used as a 'match deciding tiebreak' third or fifth set. In such a case it can be useful to use a 'longer' tiebreak game in which the better player has a higher probability of winning whilst maintaining an acceptable standard deviation of the number of points played. This might be achieved by playing (say) the best of 16 points leading by at least 6 points, or if the points score reaches 8-8, 9-7 or 10-6, the best of 24 points leading by at least 4 points, or if the points score reaches 12-12 or 13-11, the best of 32 points leading by at least 2 points, or if the points score reaches 16-16, best of 40 points leading by at least 2 points, …Possible winning scores would be 11-0, 11-1, 11-2, …11-5, 14-6, 14-7, …14-10, 17-11, 17-12, …17-15, 21-16, …)

2.3   Other comments and stopping rules, SR4 and SR5

The fairness of the above stopping rules for the probabilities in Table 1, *or indeed for any case in which probabilities for the players (or teams in the case of doubles) are equal in sets of 8 points*, can be demonstrated using exact probabilities or on a computer.

It is possible to shorten the (expected) duration of the tiebreak game in Table 1 (after (say) 16 points have been played) by changing ends after the 18th, 22nd, 26th, 30th, … points have been played, as shown in Table 2. We now have equal point probabilities four at a time (points 17,18,19 and 20, and then points 21,22,23 and 24, etc). Having equal point probabilities 4 at a time rather than 8 at a time, allows for additional fair stopping rules for Table 2. A fair stopping rule for Table 2, SR4, is the first to 9 points leading by at least 2 points, or if 8-8 is reached, the first to 11 points leading by at least 2 points, or if 10-10 is reached, the first to 13 points leading by at least 2 points, etc. Another fair stopping rule for Table 2, SR5, is the first to 10 points leading by at least 4 points, or if 8-8 or 9-7 is reached, the first to 12 points leading by at least 4 points, or if 10-10 or 11-9 is reached, the first to 14 points leading by at least 4 points, etc. It can be shown that SR4 and SR5 are fair stopping rules for any situation in which the probabilities are equal in sets of 4 at a time after the 16th point is played.

It is noted here that, although the present 12 point tiebreak game can be used when modelling singles with first and second court effects as well as with end and sun effects, this modelling involves considering points in sets of 12 points rather than in sets of just 8 points. This is a disadvantage as there are fewer finishing scores than for the 16 point tiebreak game described above. For example, the lead by at least two winning scores for this 12 point tiebreak game are 7-0, 7-1, 7-2, …, 7-5, 13-6, 13-7, …, 13-11, 19-12, …There is a possible advantage, however, in that an additional factor can be included in the model. It is a continuity effect, +c, which could possibly be of practical relevance when a server serves the second point (of each pair of points) from the same end as the first point (points 3, 5, 9, 11, 15, …).

Table 2 - Singles with first and second court effects, and with end and sun effects. First 16 points as in Table 1. Change Ends after points 18, 22, 26, …

Point	End 1		End 2	
17	$p_a + e + s + f$	[A serves]	$(q_a - e - s - f)$	
18	$(q_b + e - b)$		$p_b - e + b$	[B serves]
19	$p_b + e + s + f$	[B serves]	$(q_b - e - s - f)$	
20	$(q_a + e - b)$		$p_a - e + b$	[A serves]
21	$(q_a + e - f)$		$p_a - e + f$	[A serves]
22	$p_b + e + s + b$	[B serves]	$(q_b - e - s - b)$	
23	$(q_b + e - f)$		$p_b - e + f$	[B serves]
24	$p_a + e + s + b$	[A serves]	$(q_a - e - s - b)$	

3. Doubles with first and second court effects

Table 3 gives point probabilities for doubles in which players A1 and B1 each have a first court or forehand court effect of +f1 and a second court effect of +b1. Players A2 and B2 each have a first court effect of +f2 and a second court effect of +b2. It can be seen that the first two sets of eight probabilities are equal, the second two sets of eight are equal, the third two sets of eight are equal, etc and hence, as in Table 1, the stopping rules SR1, SR2 and SR3 are all fair.

Table 3 - Doubles with first and second court effects change ends after points 4, 12, 20, …

Point	End 1	End 2

1	$p_{a1} + f1$	[A1 serves]	$(q_{a1} - f1)$	
2	$(q_{b1} - b1)$		$p_{b1} + b1$	[B1 serves]
3	$(q_{b1} - f1)$		$p_{b1} + f1$	[B1 serves]
4	$p_{a2} + b2$	[A2 serves]	$(q_{a2} - b2)$	
5	$(q_{a2} - f2)$		$p_{a2} + f2$	[A2 serves]
6	$p_{b2} + b2$	[B2 serves]	$(q_{b2} - b2)$	
7	$p_{b2} + f2$	[B2 serves]	$(q_{b2} - f2)$	
8	$(q_{a1} - b1)$		$p_{a1} + b1$	[A1 serves]
9	$(q_{a1} - f1)$		$p_{a1} + f1$	[A1 serves]
10	$p_{b1} + b1$	[B1 serves]	$(q_{b1} - b1)$	
11	$p_{b1} + f1$	[B1 serves]	$(q_{b1} - f1)$	
12	$(q_{a2} - b2)$		$p_{a2} + b2$	[A2 serves]
13	$p_{a2} + f2$	[A2 serves]	$(q_{a2} - f2)$	
14	$(q_{b2} - b2)$		$p_{b2} + b2$	[B2 serves]
15	$(q_{b2} - f2)$		$p_{b2} + f2$	[B2 serves]
16	$p_{a1} + b1$	[A1 serves]	$(q_{a1} - b1)$	
17	$p_{a1} + f1$	[A1 serves]	$(q_{a1} - f1)$	
18	$(q_{b1} - b1)$		$p_{b1} + b1$	[B1 serves]

As a matter of possibly only theoretical interest, it can be shown that if end effects and sun effects are introduced into this model, we would need to work in multiples of 16 points in order to achieve sets of 16 equal probabilities. An appropriate 'lead by at least 2 points' stopping rule would be best of sixteen points (first to 9 points leading by at least 2 points), or if a score of 8-8 is reached, best of 32 points (first to 17 points leading by at least 2 points), or if 16-16 is reached, best of 48 points (first to 25 points leading by at least 2 points), etc. Alternatively an appropriate 'lead by at least 4 points' stopping rule would be best of sixteen points (first to 10 points leading by at least 4 points), or if a score of 8-8 or 9-7 is reached, best of 32 points (first to 18 points leading by at least 4 points), or if 16-16 or 17-15 is reached, best of 48 points (first to 26 points leading by at least 4 points), etc.

It is noted here by comparison that the 12 point and 18 point tiebreak games are unfair in some doubles situations, even without first and second court effects.

4. Conclusions

A method for designing fair lead by at least 2 and lead by at least 4 tiebreak games has been applied and results described. It has been shown how additional factors such as first court and second court effects, end effects and sun effects can be incorporated into the designs.

In particular, several fair '16 point' tiebreak games for singles with a first court/second court effect, as well as an end and sun effect have been demonstrated. Several fair '16 point' tiebreak games for doubles with additional factors have also been described.

It would appear that a tiebreak game that remains fair when one or more of the above-mentioned factors affects play, represents a useful practical and theoretical development in tiebreak scoring systems.

5. References

[1]   Newton, P. K. and Pollard, G. H., "Service neutral scoring strategies for tennis", prepared for the *7th Australasian Conference on Mathematics and Computers in Sport*, Palmerston North, New Zealand (August 30 – September 1, 2004).

[2] Pollard, G. and Noble, K., "A solution to the unfairness of the tiebreak game when used in tennis doubles", *Proceedings of the Sixth Australian Conference on Mathematics and Computers in Sport*, Cohen, G. and Langtry, T. (eds), University of Technology, Sydney, (2002), 221-226.

[3] Pollard, G. and Noble, K., "A new tiebreak game with four proposed applications", *Tennis Science and Technology 2,* Miller, S. (Ed), International Tennis Federation, London, (2003), 341-346.

THE BENEFITS OF A NEW GAME SCORING SYSTEM IN TENNIS, THE 50-40 GAME

Graham Pollard and Ken Noble
School of Information Sciences and Engineering
University of Canberra, ACT 2601, Australia

Abstract

With a view to reducing the likelihood of matches that last a long time, the no-add game, in which only

one point is played if deuce if reached, has been approved recently by the International Tennis

Federation as an option within the rules of tennis. In this paper we propose a new type of game as an

alternative to the no-add game for reducing the likelihood of long matches. This alternative is shown

to have several advantages over the no-add game.

1.   Introduction

Tennis matches that last a long time can have several disadvantages. For example, a tournament schedule can be affected considerably by long matches. Players, spectators and television programming can be inconvenienced. The winner of a long match can be too tired to play well in the next match, giving rise to an unfairness in the tournament setting. Also, it would seem reasonable to believe that long matches can produce a disproportionate number of injuries. Indeed, the very need to train for the possibility of long matches can cause additional injuries.

Thus, in the last few years there have been tennis scoring modifications introduced to reduce the likelihood of matches that last a long time. For example, the 'no-add' game in which only one point is played if deuce is reached is now an approved option within the International Tennis Federation's Rules of Tennis 2003. Another recently approved option is the use of the tiebreak *game* as the third set in a best of three sets match, or as the fifth set in a best of five sets match. These options were approved formally following a two-year trial period, which included surveys of players, coaches, spectators, officials and media personnel.

Recently, Pollard and Noble (2003) reported on the use of drawn games and drawn sets to remove long matches and increase fairness in a tournament setting. In this paper we define a new type of game, the '50-40 game' involving a simple modification of the present game scoring methods. In a 50-40 game the server is required to reach 50 (one point more than 40) before the receiver reaches 40 in order to win the game. The receiver is required to reach only 40 in order to win the game. Such games require at most 6 points.

2. A Comparison of the 50-40 Game with the No-add Game

The aim of this paper is to compare two types of games: the no-add game and the 50-40 game. For completeness however we have included results for standard deuce games. All results are for the best of 3 tiebreak sets matches.

As in earlier work (Pollard and Noble (ibid)) we carried out simulations of 1,000,000 matches for each parameter set, as this number of matches is ample to achieve appropriate accuracy for the results. Table 1 gives results for the best of three tiebreak sets using no-add games. Rows 1 and 2 give values for the parameters p_a=the probability player A wins a point on service, and p_b=the probability player B wins a

point on service. Row 3 gives the average number of points played in a match, row 4 gives the standard deviation of the number of points played, and row 5 gives the probability player A wins the match. The efficiency of the scoring system is given in row 6. The 98% point, that is the number of points such that 2% of matches have a greater number of points in total, is given in row 7. Values for the 2% point and for the difference, 98% point-mean, are also given.

We note here that two scoring systems can be compared for their efficiency (at correctly identifying the better player). Thus, given two scoring systems, scoring system 1 and scoring system 2 with the same expected number of points played in a match, scoring system 1 is said to be more efficient than scoring system 2 if scoring system 1 has a higher value for the probability that the better player wins the match. The efficiencies of scoring systems with differing values for the expected number of points played can also be evaluated (Miles (1984)). It is noted here that scoring systems with high efficiencies (that is, efficiencies close to 1) typically have particularly large standard deviations, and are not appropriate in the sporting context.

Columns 1 and 2 in Table 1 have parameter values for (p_a, p_b) of (0.5,0.5) and (0.52,0.48). These values are used to represent a group of players who typically do not have an advantage when serving. Such players are young juniors and some social players. We call them Group 50. Columns 3 and 4 with parameters (0.55,0.55) and (0.57,0.53) are used to represent the group of players who have a slight advantage when serving. This might include boys in their middle teens, many amateurs, some professional women, and some professional men on very slow surfaces. We call them Group 55. Columns 5 and 6 are used to represent the group of players who have an advantage when serving. This group includes stronger professional women, and average professional men on slower to average surfaces, and is called Group 60. Columns 7 and 8 are used to represent the group of players who have a significant advantage when serving. This group, called Group 65, includes stronger professional men on average to fast surfaces, and men's doubles on slow to average surfaces. Group 70 (columns 9 and 10) includes strong serving professional men on fast grass court surfaces, and men's doubles on average to fast surfaces. Group 75 (the last two columns) is for players with a highly significant advantage when serving, and includes men's doubles on fast grass court surfaces.

Table 1: Results for the best of three tiebreak sets using no-add games.

		1	2	3	4	5	6	7	8	9	10	11	12
1	Pa	0.50	0.52	0.55	0.57	0.60	0.62	0.65	0.67	0.70	0.72	0.75	0.77
2	Pb	0.50	0.48	0.55	0.53	0.60	0.58	0.65	0.63	0.70	0.68	0.75	0.73
3	mean	142.3	139.3	142.5	139.6	143.5	140.7	145.4	142.7	148.9	146.3	154.8	152.2
4	Std-dev	35.1	35.2	35.1	35.2	35.1	35.3	35.4	35.5	35.9	36.1	36.8	37.0
5	P(A wins match)	0.500	0.682	0.500	0.682	0.499	0.681	0.501	0.679	0.500	0.676	0.500	0.673
6	efficiency	na	0.624	na	0.618	na	0.584	na	0.537	na	0.464	na	0.383
7	98% point	213	211	213	212	214	213	216	215	220	219	225	224
8	2% point	84	82	85	83	86	84	88	86	91	89	93	92
9	98% point - mean	71	72	70	72	70	72	71	72	71	73	70	72

Table 2 gives corresponding results for the best of three tiebreak sets using 50-40 games.

Table 2: Results for the best of three tiebreak sets using 50-40 games.

		1	2	3	4	5	6	7	8	9	10	11	12
1	pa	0.50	0.52	0.55	0.57	0.60	0.62	0.65	0.67	0.70	0.72	0.75	0.77
2	pb	0.50	0.48	0.55	0.53	0.60	0.58	0.65	0.63	0.70	0.68	0.75	0.73
3	mean	120.2	118.1	121.0	118.8	122.4	120.1	124.4	122.0	127.1	124.6	131.3	128.6
4	std-dev	29.9	30.1	30.2	30.4	30.6	30.7	31.0	31.1	31.6	31.7	32.5	32.6
5	P(A wins match)	0.500	0.667	0.501	0.671	0.500	0.675	0.500	0.680	0.500	0.682	0.499	0.686
6	efficiency	na	0.616	na	0.633	na	0.640	na	0.629	na	0.585	na	0.526
7	98% point	182	181	183	182	185	184	188	187	192	191	197	196
8	2% point	71	70	71	70	72	70	73	72	76	74	79	77
9	98% point - mean	62	63	62	63	63	64	64	65	65	66	66	67

The following observations can be made from tables 1 and 2:

- For the parameters tabled the mean or average number of points played in a match ranges from 118.1 to 131.3 in table 2, compared with 139.3 to 154.8 in table 1. Thus, on average, a match using 50-40 games requires about 20 fewer points than one using no-add games.
- The standard deviation of the number of points played ranges from 29.9 to 32.6 in Table 2, compared to 35.1 to 37.0 in table 1. Thus, the standard deviation using 50-40 games is about 5 points less than that when using no-add games.
- For Groups 60 and 65 the probability player A wins is comparable in tables 1 and 2. For Groups 70 and 75 the probability player A wins is greater using 50-40 games, even though at least 20 fewer points are required on average.
- For Groups 55, 60, 65, 70 and 75, 50-40 games are more efficient than no-add games. The greater efficiency is quite marked for Groups 65, 70 and 75. For example, when (p_a, p_b) equals (0.77,0.73) the efficiency using 50-40 games is 0.526 compared to only 0.383 using no-add games.
- The 98% point ranges from 211 to 225 points in table 1, whereas it ranges from only 181 to 197 points in table 2. Thus, the 98% point using 50-40 games is about 30 points less than the 98% points for the no-add case. Hence, 50-40 games are somewhat fairer in the tournament setting as they reduce the occurrence of long matches, which can affect the winning player's performance in the following round.

We noted above that 50-40 games require at most six points. The no-add game, on the contrary, sometimes requires a seventh point. When a seventh point is required in singles (and in men's and women's doubles), the receiver elects the side to which the server will serve, whilst in mixed doubles the woman serves to the woman and the man to the man. It can be argued that this seventh point creates a lack of symmetry, and that this is less than ideal for tennis. The use of the 50-40 game instead of the no-add game removes the need for this seventh point.

For completeness we have given in table 3 the corresponding results for the case when standard deuce games are used. The means, standard deviations and 98% points in this table can be seen to be considerably greater than those in tables 1 and 2. The efficiencies for Groups 65, 70 and 75 are even lower for table 3 than for table 1. It is interesting to note that for Groups 70 and 75 the probability player A wins is greater using 50-40 games than it is using standard deuce games, despite the substantial difference in the number of points played under the two systems.

Table 3: Results for the best of three tiebreak sets using standard deuce games.

		1	2	3	4	5	6	7	8	9	10	11	12
1	pa	0.50	0.52	0.55	0.57	0.60	0.62	0.65	0.67	0.70	0.72	0.75	0.77
2	pb	0.50	0.48	0.55	0.53	0.60	0.58	0.65	0.63	0.70	0.68	0.75	0.73
3	mean	164.6	160.2	164.4	160.1	164.0	160.0	164.0	160.4	165.6	162.4	168.9	166.2
4	std-dev	42.0	42.2	41.8	42.0	41.2	41.5	40.7	40.8	40.4	40.6	40.0	40.3
5	P(A wins match)	0.500	0.704	0.500	0.703	0.500	0.697	0.499	0.689	0.500	0.680	0.499	0.669
6	efficiency	na	0.689	na	0.676	na	0.616	na	0.534	na	0.438	na	0.335
7	98% point	251	248	250	248	248	246	246	245	245	244	243	242
8	2% point	94	91	94	91	95	93	96	94	98	96	100	98
9	98% point - mean	86	88	86	88	84	86	82	85	79	82	74	76

3. Conclusions

In this paper we have defined a new type of game, the 50-40 game and compared it with the no-add game as a mechanism for reducing matches that take a long time. By using 50-40 games instead of no-add games:

- The longest matches (as measured by the 98% point in the distribution of duration) can be reduced by about 30 points. Thus, the 50-40 game is a better mechanism for reducing long matches, and hence is fairer in the tournament setting.
- For matches where serving is a moderate advantage, the probability player A (the better player) wins using 50-40 games is comparable to that when using no-add games, even though about 20 fewer points are required on average.
- For matches where serving is even more of an advantage, the probability player A wins is greater when using 50-40 games than it is when using no-add games.
- Matches involving strong servers are shown to be considerably more efficient when using 50-40 games instead of no-add games.
- A seventh point, creating an unattractive lack of symmetry in the game is never required.

As with the recent development of options within the tennis scoring system, it would seem reasonable to trial the 50-40 game over a period of 12 to 24 months, including the surveying of all relevant stakeholders.

4. References

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

[2]   Pollard, G. and Noble, K. (2003). Scoring to remove long matches, increase tournament fairness and reduce injuries. Medicine and Science in Tennis 2003; 8(3): 12-13.

THE EFFECT OF HAVING CORRELATED POINT OUTCOMES IN TENNIS

Graham Pollard and Ken Noble
School of Information Sciences and Engineering
University of Canberra, ACT 2601, Australia

Abstract

In many papers on modelling tennis, it is assumed that the probability a player wins a point on service is constant. In this paper we consider a model in which the probability a player wins a point on service is a function of 'previous performance' and 'performance on the day'. The factor 'performance on the day' is shown to have no great effect on the mean and standard deviation of the number of points played in a match. The effect of even a moderate weighting on the 'performance on the day' is, however, to increase substantially the likelihood of the weaker player winning the match.

1. Introduction

In many papers on modelling tennis, it is typically assumed that player A has a constant probability, p_a, of winning a point when serving, and that player B has a constant probability, p_b, of winning a point when serving.

It is clear, however, that in practice a player will sometimes have a 'good' (or 'bad') run of points by winning (or losing) a disproportionate number of points over several games or a set or more. This situation was considered by Pollard and Noble (2003). For example, they considered the situation in which player A had, at random and equally likely, service games in which his/her point probability was $p_a+\delta$ for every point in the game or was $p_a-\delta$ for every point in the game. Corresponding assumptions were made for player B. They also considered situations in which the point probability was fixed at the higher or lower value for the whole set, or even the whole match. They concluded that these modifications in modelling tennis do not have substantial effects on the mean and standard deviation of the number of points played in a match of tennis. They did conclude, however, that such changes have greater proportional effect on the probability that the weaker player, player B, wins, and that the increase in the weaker player's probability of winning can be quite substantial when his/her point variation, δ, is maintained over longer periods of play such as a set or a match.

In the model previously considered by Pollard and Noble, points have higher (or lower) values for every point in a game, a set, or a match. This model is somewhat 'discrete' in nature as points have better (or worse) likely outcomes for the 'whole unit' such as a game, a set, or a match. In this paper we consider a more continuous approach to formalizing a player's probability of winning a point at any stage in the match. The effect on characteristics such as the mean and standard deviation of the number of points played in a match, and the probability that each player wins is reported.

2. The Analysis

In the model considered in this paper, the probability that player A wins a point on service is assumed to be made up of two components:

> one component based on the relative frequency, f_a, of points won on service in the match to date, weighted by θ
> the other component (based on information prior to the match) is p_a, weighted by $(1-\theta)$.

Thus, Probability (player A wins the next point on service)$= \theta f_a+(1-\theta)p_a$

A corresponding approach is taken for Player B. We note here that our analysis has shown that this approach is 'fair', as the probability each player wins the match is 0.5 when $p_a=p_b$.

It can be seen that if player A is 'having a good run of points' on service, his/her probability of winning the next point on service is greater than p_a. Correspondingly, if player A is 'having a bad run of points' on service, his/her probability of winning the next point on service is less than p_a. Thus, point outcomes on service for player A are positively correlated. There is a corresponding positive correlation of point outcomes on player B's service.

It is frequently observed that a player who is 'having a good day' will continue to 'have a good day' for the entire match. Thus, if $f_a> p_a$ by (say) about halfway through a match, it is likely in practice that $f_a>p_a$ for the remainder of the match. The above model has this characteristic. Given the importance of 'performance on the day' in determining the outcome of a match, it is suggested that a value for θ in the vicinity of 0.5 is appropriate for many tennis matches.

In this paper we consider best of three tiebreak sets matches, and use simulations of 1,000,000 matches for each set of parameters, as simulations of this size provide appropriate accuracy. In order to be consistent with earlier work (Pollard and Noble (ibid, 2002) we assume $p_a=0.652$ and $p_b=0.572$ (values agreed on and used in earlier work carried out for the International Tennis Federation on men's singles tennis). We consider the effect of the model described above on various characteristics of a tennis match for $\theta>0$.

We also consider the effect of 'bounding' the point probabilities by various amounts. For example, a 20% bound gives
$$0.8p_a<\theta f_a+(1-\theta)p_a<1.2p_a$$
$$0.8p_b<\theta f_b+(1-\theta)p_b<1.2p_b$$
Also, in order to remove any 'initialising problems', we consider the effect of using p_a rather than $\theta f_a+(1-\theta)p_a$ as the point probability (early) in the match until player A has won at least (say) n=2 points and has lost at least (say) n=2 points on service. We treat player B on service correspondingly.

Table 1 gives the results for $\theta = 0$ (0.1) 1 with 20% bounds and n=2. It can be seen that, as may have been expected, the mean decreases somewhat, and the standard deviation increases a little as θ increases from 0 to 1. The probability player B wins, however, increases substantially from 0.153 when $\theta = 0$, to 0.253 when $\theta = 0.4$, and to 0.347 when $\theta = 1$. These observations are similar to those under the alternative models of Pollard and Noble (ibid).

Table 1 – Results for $p_a=0.652$ and $p_b=0.572$ with 20% bounds and n=2

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=1.0$
Mean	149.9	148.6	146.8	144.5	141.7	139.0	136.8	135.2	134.2	133.5	133.0
Standard deviation	40.7	41.2	41.8	42.5	43.0	43.3	43.4	43.4	43.5	43.5	43.5
P(B wins)	0.153	0.173	0.196	0.223	0.253	0.281	0.306	0.325	0.339	0.344	0.347

Given the relatively small effect on the mean and standard deviation of the number of points played, Table 2 focuses on just the probability player B wins, and displays this statistic for differing percentage bounds and n values, for $\theta = 0$ (0.1) 1. In all cases, the probability player B wins increases substantially as θ increases from 0 to 1. For values of θ up to about 0.5, it is clear that the most important determinant of the probability player B wins is the value for θ, with the bounds and n values being of less importance. For example, when $\theta = 0.2$, the values

for this statistic range from 0.192 to 0.199. Even when θ = 0.4, the range of values for this statistic (from 0.235 to 0.267) is moderately narrow.

Table 2 – P(B wins) for p_a=0.652 and p_b=0.572 for differing bounds and n values

	θ=0	θ=0.1	θ=0.2	θ=0.3	θ=0.4	θ=0.5	θ=0.6	θ=0.7	θ=0.8	θ=0.9	θ=1.0
10% bounds n = 0	0.153	0.173	0.197	0.222	0.241	0.255	0.263	0.267	0.270	0.272	0.275
10% bounds n = 2	0.153	0.173	0.196	0.219	0.239	0.254	0.263	0.268	0.271	0.274	0.276
10% bounds n = 4	0.153	0.171	0.192	0.215	0.235	0.250	0.261	0.266	0.270	0.272	0.275
20% bounds n = 0	0.153	0.173	0.199	0.229	0.260	0.288	0.311	0.327	0.337	0.342	0.342
20% bounds n = 2	0.153	0.173	0.196	0.223	0.253	0.281	0.306	0.325	0.339	0.344	0.347
20% bounds n = 4	0.153	0.171	0.192	0.218	0.244	0.272	0.296	0.315	0.329	0.338	0.341
30% bounds n = 0	0.153	0.173	0.199	0.231	0.265	0.300	0.332	0.356	0.372	0.377	0.377
30% bounds n = 2	0.153	0.173	0.196	0.223	0.254	0.287	0.321	0.349	0.371	0.382	0.387
30% bounds n = 4	0.153	0.171	0.192	0.218	0.244	0.273	0.303	0.331	0.354	0.369	0.379
no bounds n = 0	0.153	0.173	0.199	0.231	0.267	0.309	0.350	0.387	0.418	0.446	0.463
no bounds n = 2	0.153	0.173	0.196	0.223	0.254	0.287	0.323	0.360	0.393	0.423	0.449
no bounds n = 4	0.153	0.171	0.192	0.218	0.244	0.273	0.303	0.333	0.363	0.390	0.414
Range of values	0.153, 0.153	0.171, 0.173	0.192, 0.199	0.215, 0.231	0.235, 0.267	0.250, 0.309	0.261, 0.350	0.266, 0.387	0.270, 0.418	0.272, 0.446	0.275, 0.463

3. Conclusions

A model in which the probability a player wins a point is a linear combination of his/her 'performance (so far) on the day', with weighting θ, and his/her 'previous performance', with weighting 1- θ, has been considered. The mean and standard deviation of the number of points played in the match are not greatly affected by the value of θ. The probability that the weaker player wins, however, does depend on θ, and increases substantially as θ increases from 0 to 1. As θ in the vicinity of 0.5 is probably quite appropriate for many tennis matches, the value for the probability that the weaker player wins when θ equals 0 considerably underestimates this probability for such matches.

A player who is playing a more highly ranked player should be considerably encouraged by the results of this study in that a strong 'performance on the day' increases substantially his/her chances of

winning. Correspondingly, the more highly ranked player in a match should be on his/her guard against a poor 'performance on the day'.

4. References

[1] Pollard, G. and Noble, K., "The characteristics of some new scoring systems in tennis", *Proceedings of the sixth Australian conference on mathematics and computers in sport*, Cohen, G. and Langtry, T. (Eds), University of Technology, Sydney, (2002), 221-226.

[2] Pollard, G. and Noble, K., "The effect of changing the assumption that the probability of winning a point in a tennis match is constant", *Tennis Science and Technology 2,* S. Miller, (Ed), International Tennis Federation, London, (2003), 341-346.

THE FALLACY OF THE USE OF PURE STATISTICS IN SPORT

Dr George Serrallach
Managing Director, Verusco Technologies Ltd
Palmerston North, New Zealand

Dr Andy Martin
Department of Management, Massey University
Palmerston North, New Zealand

Abstract

This paper reviews data gathered using the GoalSeeker coding system, which provides comprehensive game analysis focusing on the quality of skill execution and shows Key Performance Indicators (KPIs), which highlight 'player and team profiles' and 'factors influencing games outcomes' identified from the analysis of a database of over 700 games. Nowadays statistical analysis of games is being used more and more by teams. However, many of the statistics provide meaningless information related to the outcome of a game. The development of KPI's shows which team is likely to win the next game taking into account the quality of technical execution. The probability of this happening is around 93%, if a team has KPIs better than those of their next opponents. These KPI's allow coaches to plot trends and see if their team performance is getting better or worse.

FORECASTING THE 2003 RUGBY WORLD CUP

Stefan Yelas and Stephen Clarke
School of Mathematical Sciences
Swinburne University
PO Box 218 Hawthorn, Vic, 3122
Australia

Abstract

A simple forecasting model was built to predict the results of each game and the tournament in the 2003 Rugby World Cup. An exponential smoothing technique was optimised on all 566 games played between the 20 World Cup teams from 1996. The model predicted the winning team, the winning margin and the probability of a win. A simulator used these predicted probabilities to calculate a team's overall chances of winning or placing in the tournament. Predictions were regularly updated on our web site www.swin.edu.au/sport. The model selected the correct winner in 46 of the 48 games, and the predicted margins were used for profitable gambling.

1. Introduction

Swinburne Sports Statistics has a 25 year history of sporting predictions in the media. Several years ago we began to use the internet to publish predictions on our web site www.swin.edu.au/sport. The aim is to demonstrate the use of statistical modeling to add value to data, and show the general public that predictions based on mathematical models can be as accurate as media experts. The site has built up a reputation for timely and accurate forecasts of sporting events, resulting in over 4000 visits per week in 2002,. From the beginning, we have not just predicted winners, but given more detail such as predicted scores or margins, chances of winning, and chances in the competition as a whole. So for example our predictions for the F1 motor race in Melbourne consist of four tables of the chances of both drivers and constructors finishing in any points position in both the race and the World Championship.

The forecasts produced for our site often attract media publicity. In addition to regular weekly predictions of national Australian rules football, rugby league, basketball and netball competitions, we like to focus on other major events such as the Olympics, Grand Prix racing and the Brownlow medal count. Successful examples discussed in the literature include the Major tennis tournaments[7] and the World Cup of soccer[8].

In 2003 we were keen to produce predictions for the upcoming World Cup in Rugby. The subject Sports Performance Modeling, undertaken by Stefan and taught by Steve in 2003, is part of the Graduate Diploma/Masters in Applied Statistics at Swinburne, and includes a project component worth 40%. Since the timing was right, Stefan chose predicting the Rugby World Cup as his project topic. The aim was to produce individual match predictions, and also via a simulation, predictions for the tournament as a whole. We wished to build a forecasting model to predict not only the winner of each game of the 2003 Rugby World Cup, but also each team's chance of a place finish (first to eighth) at any given time in the tournament. Several models were to be explored based on the forecasting method outlined by Clarke[6], which was used to predict the outcome of AFL matches. The best performing model would be used for publishing predictions on the Swinburne Sports Statistics website.

Rugby is a winter game played between two teams of fifteen players on a rectangular ground. Points are scored by grounding the oval shaped ball behind the opposition goal line for a try (5 points), kicking the ball through upright posts located on the opposition's goal line for a conversion (2 points), a drop goal (3 points) or a penalty goal (3 points). A match lasts for two halves of 40 minutes each and the rules and play are complex. The gulf between skill levels of national teams means that 70 plus point margins are frequent in a World Cup, while evenly matched teams usually result in margins less than 20.

The first Rugby World Cup was played in New Zealand and Australia in 1987 between 16 countries. The 2003 World Cup was played in Australia between the top 20 teams in the world, competing in four pools (A to D) of five teams each. The pool phase consisted of a round robin with each team playing each other team

within the pool once. At the completion of the pool phase, the teams in a pool were ranked one through five based on their cumulative match points. The first two teams in each pool progressed to a knockout tournament, with fixed order of play. The two Semi Final losers played off for third and fourth. This structure resulted in 40 pool matches and 8 finals.

2. Prediction Method

Since the late 1970s various authors [9, 10, 13, 14] have suggested using linear models to predict sporting events. The models usually select winners via a prediction of team scores, or margins. The predictors may include attack and defence or combined team ratings, and distinct or common home advantages. The models are optimised by minimising a function of the errors, such as least squares or average absolute error. The distribution of the errors can then be used to convert predicted margins into chances of winning. Clarke [5, 6] had used these methods to produce predictions of AFL football as good as the media experts, and Morton [11] had demonstrated they could be used to provide predictions of rugby good enough to produce profitable betting.

Three models, with varying degrees of complexity, were eventually built to predict the signed margin. Each model used team ratings and a common home advantage (H). An exponential smoothing constant α was used to update team ratings depending on how the predicted margin (P) differed from the actual margin (M). The models used exactly the same input data and only varied by either the prediction equation or measure of error (E) used to update the ratings.

Model 1 used standard exponential smoothing of team ratings to produce a margin prediction, similar to the first model outlined in Clarke [6]. The ratings were updated using simple exponential smoothing of the actual error.

$P = Team1\ rating + H - Team\ 2\ rating$ .... (1)
$E = M - P$ .... (2)
$Updated\ Team\ 1\ rating = Team\ 1\ rating + \alpha\ E$ .... (3)
$Updated\ Team\ 2\ rating = Team\ 2\ rating - \alpha\ E$ .... (4)

Model 2 used the square root of the predicted and actual margins to reduce the relative importance of large errors and magnify the relative effect of errors near the win/loss boundary. Thus the importance of the error in predicting a margin of –1 when the actual margin was +1 became {sqrt(1) + sqrt(1) =2}, equivalent to that when predicting a margin of 49 when it actually was 81. This was achieved by replacing (2) above with

$E = sgn(M)\ \sqrt{|M|}\ -\ sgn(P)\ \sqrt{|P|}$ .... (5)

Model 3, similar to the third model of Morton [11], used an attack and defence rating for each team. This model would not only predict margins but also the score $S1$ and $S2$ for each team, and so had two associated errors $E1$ and $E2$. The predicted margin was obtained by subtraction, and the ratings again were updated by simple exponential smoothing. Thus we had

$P1\ = Team\ 1\ Attack\ rating + H - Team\ 2\ Defence\ rating$ .... (6)
$P2 = Team\ 2\ Attack\ rating - Team\ 1\ Defence\ rating$ .... (7)
$P = P1 - P2$ .... (8)

$E1 = S1 - P1$ .... (9)
$E2 = S2 - P2$ .... (10)

$Updated\ Team\ 1\ Attack\ rating = Team\ 1\ Attack\ rating + \alpha\ E1$ .... (11)
$Updated\ Team\ 2\ Defence\ rating = Team\ 2\ Defence\ rating - \alpha\ E1$ .... (12)
$Updated\ Team\ 2\ Attack\ rating = Team\ 2\ Attack\ rating + \alpha\ E2$ .... (13)
$Updated\ Team\ 1\ Defence\ rating = Team\ 1\ Defence\ rating - \alpha\ E2$ .... (14)

The methods were compared on the optimal value of three key indicators, namely the percentage of games correctly predicted, the standard deviation and the average absolute value of the actual errors.

Although the general form of the prediction equations was known, the value of the smoothing parameters and home advantage was yet to be determined. The data used to fit the model were the historical results from all 566 games matches between any of the 20 competing teams from 1996 (when rugby became professional). The data used were the team names, score margin and which team was playing at home.

Unlike least squares methods, exponential smoothing requires initial values for the ratings. The Zurich World Team Rankings, developed in 1998 by www.planetrugby.com were used as initial ratings for the prediction models. While this would introduce some element of feedback that could be expected to produce better than usual predictions, the importance of the chosen initial values gradually diminishes as the predictions continue. The accuracy of the models were compared over the last 50 predictions as well as all the data. All ratings were divided by 10 so that a 1 point different in ratings equated to a 1 point difference on the field.

Excel was used to build the models. Once the historical data and the equations were entered, solver was used to find the values of the home ground advantage and the exponential smoothing constants, which minimised the average absolute error (AAE). Table 1 compares the AAE, the standard deviation of errors, and the proportion correctly predicted for all 566 games and for the last 50 games.

Clearly there is little difference in the forecasting accuracy of the three models. Interestingly, for all data the first and simplest of the models was optimal on all measures, with the smallest AAE, smallest standard deviation of errors, and largest proportion of correctly tipped winners. For the last 50 games differences are also small.

It was decided to use Model 1 for the live predictions. This had a common home advantage (to be enjoyed only by Australia) of 5 points.

Table 1: Model fit

	Statistic	**Method 1**	**Method 2**	**Method 3**
All 566 games	Number of winners	446	445	444
	Percentage correct	78.8%	78.6%	78.4%
	Average absolute error	13.9	14.0	14.0
	Stand deviation of errors	18.3	18.6	18.5
Last 50 Games	Number of winners	43	42	41
	Percentage correct	86.0%	84.0%	82.0%
	Average absolute error	11.8	11.2	12.1
	Stand deviation of errors	14.1	14.4	14.5

The model was now ready to predict the outcome (margin and winner) of each game in the World Cup. Since the values given in Table 1 come from optimising, and not from a holdout sample, the accuracy is overstated, and we were expecting live forecasts to be somewhat less accurate. However, to make the predictions more useful for potential betting a probability of a team winning or losing was also added. In Excel a standard deviation of the error was calculated, and used as an input to a function that returned the normal cumulative distribution for this standard deviation and specific mean (the predicted margin). This gave the probability that a team would lose. Subtracting this probability from one gave the probability that a team would win. The probabilities were also used as inputs into a tournament simulator to give the probabilities of a team finishing in a particular place.

The tournament simulator was also built in Excel, using as inputs the probabilities of the outcomes of unplayed games and the actual outcomes of all games played. The tournament structure was programmed as closely as possible (the bonus point system was not employed due to the prediction model chosen only predicting the margin and not the score or make-up of the score). 10,000 simulations were performed using a

data table. The simulations were repeated each time the model was updated with actual results, and care was taken to 'hard code' the results of games that had been played. As the model was dynamic and team ratings could change during the course of the World Cup, so too could the probabilities of teams finishing in a particular place. At regular stages (when each set of matches were completed), updated predictions were placed on our Web site http://www.swin.edu.au/sport/rugbywc.htm. and previous predictions were archived.

3. Results

Table 2 shows the probabilities of each team finishing in a particular place prior to the start of the tournament generated from the forecasting model and the tournament simulator. At this stage England was a clear favourite, followed by New Zealand, then Australia.

Table 2. Predicted percentage chances of team placing, as generated before the first game of the World Cup

Pool	Team	Pool Matches		Tournament			
		Winning Pool	Runner up in Pool	1st	2nd	3rd	4th
A	Australia	64%	25%	25%	18%	24%	10%
A	Ireland	15%	33%	2%	6%	6%	13%
A	Argentina	22%	42%	4%	10%	11%	15%
A	Romania	0%	0%	0%	0%	0%	0%
A	Namibia	0%	0%	0%	0%	0%	0%
B	France	83%	14%	6%	12%	13%	16%
B	Scotland	10%	44%	0%	1%	1%	5%
B	Fiji	5%	27%	0%	0%	0%	2%
B	United States	2%	14%	0%	0%	0%	0%
B	Japan	0%	1%	0%	0%	0%	0%
C	England	87%	12%	34%	26%	18%	9%
C	South Africa	12%	68%	1%	5%	4%	11%
C	Samoa	2%	16%	0%	0%	0%	1%
C	Georgia	0%	1%	0%	0%	0%	0%
C	Uruguay	0%	3%	0%	0%	0%	0%
D	New Zealand	92%	8%	28%	22%	21%	10%
D	Wales	7%	55%	0%	2%	1%	6%
D	Canada	0%	11%	0%	0%	0%	0%
D	Italy	1%	21%	0%	0%	0%	1%
D	Tonga	0%	5%	0%	0%	0%	0%

Table 3 shows the predictions and actual results for each match, as produced prior to each match.

The model performed extremely well in that the correct winner was forecast in 46 out of 48 matches (96% correct). Furthermore the AAE was 15 points. As expected, this is higher than the 12 points achieved during the optimisation phase. However this is affected by a few very large values such as Australia's 142 point win over Namibia. The median absolute error was 12 points, and the first quartile 5 points.

Table 3. Predictions and actual results for each match.

Team 1	Team 2	Predicted Winner	Predicted Margin	Predicted Chance	Actual Winner	Actual Margin	Correct
Australia	Argentina	Australia	14	78%	Australia	16	Yes
NZ	Italy	NZ	54	100%	NZ	63	Yes
Ireland	Romania	Ireland	53	100%	Ireland	28	Yes
France	Fiji	France	40	99%	France	43	Yes
S. Africa	Uruguay	S. Africa	44	99%	S. Africa	66	Yes
Wales	Canada	Wales	26	92%	Wales	31	Yes
Scotland	Japan	Scotland	42	99%	Scotland	21	Yes
England	Georgia	England	79	100%	England	78	Yes
Argentina	Namibia	Argentina	64	100%	Argentina	53	Yes
Fiji	USA	Fiji	7	64%	Fiji	1	Yes
Italy	Tonga	Italy	8	68%	Italy	24	Yes
Samoa	Uruguay	Samoa	26	92%	Samoa	47	Yes
NZ	Canada	NZ	53	100%	NZ	62	Yes
Australia	Romania	Australia	56	100%	Australia	82	Yes
France	Japan	France	55	100%	France	22	Yes
S. Africa	England	England	-22	88%	England	-19	Yes
Wales	Tonga	Wales	27	93%	Wales	7	Yes
Ireland	Namibia	Ireland	60	100%	Ireland	57	Yes
Georgia	Samoa	Samoa	-32	96%	Samoa	-37	Yes
Scotland	USA	Scotland	15	79%	Scotland	24	Yes
Italy	Canada	Italy	5	60%	Italy	5	Yes
Argentina	Romania	Argentina	43	99%	Argentina	47	Yes
Fiji	Japan	Fiji	15	79%	Fiji	28	Yes
NZ	Tonga	NZ	55	100%	NZ	84	Yes
S. Africa	Georgia	S. Africa	54	100%	S. Africa	27	Yes
Australia	Namibia	Australia	70	100%	Australia	142	Yes
Italy	Wales	Wales	-17	82%	Wales	-12	Yes
France	Scotland	France	17	83%	France	42	Yes
Argentina	Ireland	Ireland	-2	55%	Ireland	-1	Yes
England	Samoa	England	42	99%	England	13	Yes
Japan	USA	USA	-15	79%	USA	-13	Yes
Georgia	Uruguay	Uruguay	-2	54%	Uruguay	-12	Yes
Canada	Tonga	Canada	4	58%	Tonga	17	Yes
Namibia	Romania	Romania	-21	87%	Romania	-30	Yes
France	USA	France	37	98%	France	27	Yes
Scotland	Fiji	Scotland	11	72%	Scotland	2	Yes
S. Africa	Samoa	S. Africa	14	78%	S. Africa	50	Yes
Australia	Ireland	Australia	24	91%	Australia	1	Yes
England	Uruguay	England	65	100%	England	98	Yes
NZ	Wales	NZ	37	98%	NZ	16	Yes
NZ	S. Africa	NZ	20	86%	NZ	20	Yes
Australia	Scotland	Australia	39	97%	Australia	17	Yes
France	Ireland	France	2	55%	France	22	Yes
England	Wales	England	33	96%	England	11	Yes
Australia	NZ	NZ	-3	45%	Australia	12	No
England	France	England	12	74%	England	17	Yes
NZ	France	NZ	13	76%	NZ	27	Yes
Australia	England	Australia	1	51%	England	3	No

4. Discussion

While the model's correct prediction record is surprising, it should be noted that the one-sidedness of many matches would have meant that over half of these would have been pretty easy to pick for even a non-rugby follower. To quote one newspaper article in The Herald Sun headlined "Easy Pickings"

> Much was made this week of the fact Swinburne Uni's Sports Statistics Unit had picked every winner so far in the Rugby World Cup. Using a tipping system based on the way international chess players are rated, the stats rats were happy to announce they had selected the winners of all the matches. What, like everybody else didn't? Forty-four matches won by 44 favourites"

While this article is incorrect in stating all winners were favourites, the general thrust is true – the large differences in standards result in a higher success rate than is normally the case. However it is of interest that such large margins are still useful in judging the relative strengths of the stronger teams, and the model can achieve a high level of accuracy using information which many human tipsters might dismiss as being irrelevant.

In addition, the strength of the model has been in predicting margins. This quote from an email received during the tournament from a professional punter in the UK highlights this point

> I was actually going to mention your student and the rugby in my last e-mail as I have been following his selections. So far based on his hcaps when compared to the bookies he has managed to find 18/22 winners and of scratch been 100% correct including the two underdogs in the last two days (Uruguay and Canada). The problem is he has Romania as 21 pt favs today which is too close to the bookies line for a bet.

As the above email indicates, an obvious application for the model is sports betting and as the pool games are normally one sided the logical application is margin betting. After the first week of the competition (once the model had been seen to work) a $50 pool was placed with website www.betfair.com. Betfair is a betting market where members can lay or take bets. Bet Fair only takes a percentage on winning bets. Using the predicted probabilities and margins, bets were placed on games where a 20% or higher overlay was apparent. A fraction of the pool of 40% of the probability of the predicted margin range was bet. At the end of the competition the pool was $638. Note that even the semi final and final matches in which the model predicted the incorrect winner resulted in a significant contribution to this profit of $160 and $121 respectively. In both cases the model suggested Australia would do better than the line indicated. As these matches were late in the tournament, the bets were a reasonable proportion of what was, by then, a relatively sizeable pool.

The project also achieved its publicity aims. Exposure included a Campus review article (Figure 1), published predictions on the front page of The Age sports section, short mentions in The Herald Sun, The Sydney Morning Herald and the Wellington times, and interviews on Bathurst and Melbourne radio. Stefan achieved his objectives with a HD in the subject, and a bonus with his winnings.

5. Conclusion

Simple forecasting models can be effective and profitable. The evidence is mounting that mathematical models can be used to create profitable betting. While more traditional markets such as horse racing [4] and dog racing [12] may be more efficient and require complicated models, there is evidence that the newer forms of sports betting may be less efficient with positive gains coming from simpler models to predict soccer [7, 15], tennis [3], Australian rules football [1, 16], cricket [2], and rugby [11].

Figure 1: An example of media coverage in Campus Review

Swinburne statisticians predict England to win

SWINBURNE sports statisticians believe England is a firm favourite to take out the 2003 Rugby World Cup. After examining statistical data, experts from Swinburne's Sports Statistics Unit say England has a 49 per cent chance of winning the prestigious event that kicks off in Sydney this Friday, October 10.

New Zealand has the next greatest chance of taking out the title, with a 29 per cent probability of winning the competition, according to the statisticians.

Australia's home ground advantage may not be enough as they are only ranked third with a 14 percent chance of winning the tournament. Without the home ground advantage, the sports statisticians believe Australia would only have a seven percent chance of winning.

Other teams ranked in the top eight by the Swinburne statisticians are: France (fourth), Ireland (fifth), South Africa (sixth), Argentina (seventh) and Wales (eighth).

The predictions were compiled by Stefan Yelas as a part of his masters in applied statistics course in conjunction with Professor Stephen Clarke, who is head of the sports statistics unit at Swinburne.

Stefan used the results of some 600 matches played from 1996 until this year between the teams competing in the Rugby World Cup as a basis for rating each team and predicting each match.

He also simulated the 48 games being played in the tournament 10,000 times before coming up with his final predictions. The Swinburne predictions for the 2003 World Cup winner will change during the tournament and will regularly be updated on the Swinburne Sports Statistics web site at www.swin.edu.au/sport.

6. References

1. Bailey, M., Identifying arbitrage opportunities in AFL betting markets through mathematical modelling, in *Proceedings of the Fifth Australian conference on Mathematics and Computers in Sport*, G. Cohen and T. Langtry, Editors. 2000, University of Technology Sydney: Sydney. p. 37-42.

2. Bailey, M. and S.R. Clarke, Market Inefficiencies in Player Head to Head Betting on the 2003 Cricket World Cup, in *Economics, Management and Optimisation in Sports*, P. Pardalos, S. Butenko, and J.G. Lafuente, Editors. 2004, Springer. p. 185-201.

3. Barnett, T.J. and S.R. Clarke, Using Microsoft Excel to model a tennis match, in *Proceedings of the Sixth Australian conference on Mathematics and Computers in Sport*, G. Cohen and T. Langtry, Editors. 2002, University of Technology Sydney: Sydney. p. 63-68.

4. Benter, W., G. Miel, and P.D. Turnburgh, Modelling distance preference in thoroughbred racehorses, in *Proceedings of Third Conference on Mathematics and Computers in Sport*, N. de Mestre, Editor. 1996, Bond University: Queensland. p. 207-218.

5.   Clarke, S.R., Computer and human tipping of AFL football - a comparison of 1991 results., in *Mathematics and Computers in Sport*, N. de Mestre, Editor. 1992, Bond University: Gold Coast, Qld. p. 81-93.

6.   Clarke, S.R., Computer forecasting of Australian Rules football for a daily newspaper. *Journal of the Operational Research Society*, 1993. **44**(8): p. 753-759.

7.   Dixon, M.J. and S.G. Coles, Modelling Association football scores and inefficiencies in the UK football betting market. *Applied Statistics*, 1997. **46**: p. 265-280.

8.   Dyte, D. and S.R. Clarke, A ratings based model for World Cup soccer simulation. *Journal of the Operational Research society*, 2000. **51**: p. 993-998.

9.   Harville, D.A., Predictions for National Football League games via linear-model methodology. *Journal of the American Statistical Association*, 1980. **75**: p. 516-524.

10.   Leake, R.J., A method for ranking teams with an application to college football, in *Management Science in sports*, R.E. Machol, S.P. Ladany, and D.G. Morrison, Editors. 1976, North Holland. p. 27-46.

11.   Morton, R.H., Team rating and home advantage in SANZAR rugby union, 2000-2001, G. Cohen and T. Langtry, Editors. 2002, University of Technology Sydney: Sydney. p. 182-187.

12.   Shepherd, C., Discrete choice modelling in a small betting market, Unpublished PhD Thesis, Department of Econometrics and Business Statistics. 2001, Monash University: Melbourne. p. 442.

13.   Stefani, R.T., Football and basketball predictions using least squares. *IEEE Transactions on systems, man, and cybernetics*, 1977. **7**: p. 117-121.

14.   Stefani, R.T., Improved least squares football, basketball and soccer predictions. *IEEE Transactions on Systems, Man and Cybernetics*, 1980. **SMC -10**(2): p. 116-123.

15.   Stefani, R.T., Observed betting tendencies and suggested betting strategies for European football pools. *The Statistician*, 1983. **32**: p. 319-329.

16.   Yelas, S., Analysis of the Swinburne Computer performance in identifying inefficiencies in the AFL betting market, Unpublished MSc Thesis, School of Mathematical Sciences. 2003, Swinburne University of Technology: Hawthorn. p. 45.