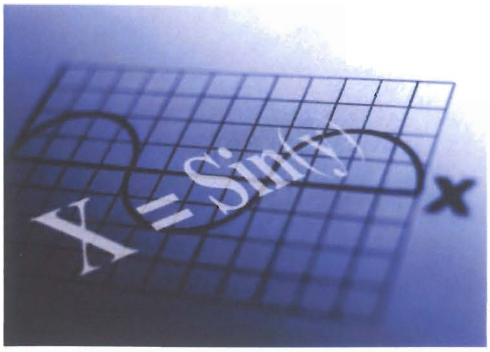
MATHEMATICS AND COMPUTERS IN SPORT



Edited by
John Hammond

MathSport

PROCEEDINGS OF THE NINTH AUSTRALASIAN CONFERENCE ON

MATHEMATICS AND COMPUTERS IN SPORT

edited by

John Hammond

9M&CS

August 30th to September 3rd 2008

Tweed Heads, New South Wales, Australia

Published by MathSport (ANZIAM)

ISBN: 978-0-9578623-4-0

The papers in these proceedings have gone through a full peer review refereeing process.

CONFERENCE DIRECTORS' REPORT

Welcome to the ninth Australasian Conference on Mathematics and Computers in Sport. The venue at the Twin Towns Resort is 400 metres from the venue for the 2006 (8^{th}) conference but is now in New South Wales instead of Queensland. However, it will still be warm in September in northern NSW.

One of the previous co-directors of the conference Professor John Hammond has now moved to the United Kingdom but has acted as Scientific Director and Proceedings Editor, efficiently handling all of your papers and putting them through the refereeing process. He had considerable assistance in editing the proceedings from Mrs. Joan Hammond (editorial assistant) and Emeritus Professor John Norman (editorial consultant) from Sheffield University. The latter unfortunately cannot be with us this year because of other commitments.

In addition to wide representation from Australia, we also have delegates from Austria, Germany, New Zealand and the United Kingdom. Our keynote principal speakers Professor Arnold Baca (University of Vienna) and Dr Ian Renshaw (Queensland University of Technology) should provide us with interesting insights into their research activities in sport.

The conference will be opened by Geoff Pollard, President of Tennis Australia, who is also presenting papers. The papers are many and varied and I'm sure you will enjoy the conference. I hope that you gain great benefit from the programme and associated discussions.

Neville de Mestre

Professor Emeritus and Conference Director

LIST OF CONFERENCE PARTICIPANTS[1]

Arnold Baca	Vienna, Austria	arnold.baca@univie.ac.at
James Baglin	RMIT	james.baglin@rmit.edu.au
Anthony Bedford	RMIT	anthony.bedford@ems.rmit.edu.au
Michael Bailey	Monash	Michael.Bailey@med.monash.edu.au
Tristan Barnett	Swinburne	strategicgames@hotmail.com
Alan Brown	Swinburne	abrown@labyrinth.net.au
Tania Churchill	Canberra	tania@fitsense.com.au
Neville de Mestre	Bond	margnev@omcs.com.au
Philip Fink	Massey, NZ	
Tim Gabbett	Broncos	timg@broncos.com.au
John Hammond	Lincoln	jhammond@lincoln.ac.uk
John Haynes	New England	jhaynes2@une.edu.au
Tim Heazlewood	ACU	
Karl Jackson	Swinburne	cjackson@swin.edu.au
Ina Janssen	AIS	ina.janssen@ausport.gov.au
H. Koehler	Karlsuhe, Germany	
Philipp Kornfeind	Vienna, Austria	philipp.kornfreind@univie.ac.at
Monique Ladds	RMIT	monique.ladds@rmit.edu.au
Ian Lisle	Canberra	ian.lisle@canberra.edu.au
Ken Louie	Waikato, NZ	kenlouie@xtra.co.nz
Tony Lewis	Oxford Brookes, UK	aj-lewis@tiscali.co.uk
John McCullagh	La Trobe	j.mccullagh@latrobe.edu.au
Denny Meyer	Swinburne	dmeyer@swin.edu.au
Geoff Pollard	Swinburne	gpollard@tennisaustralia.com.au
Graham Pollard	Canberra	graham@foulsham.com.au
Ian Renshaw	QUT	i.renshaw@qut.edu.au
Richard Ryall	RMIT	r.ryall@ems.rmit.edu.au
Jonathan Sargent	RMIT	j.sargent@ems.rmit.edu.au
Steven Stern	ANU	Steven.Stern@anu.edu.au
Elliot Tonkes	Energy Edge	etonkes@energyedge.com.au
Niven Winchester	Otago, NZ	nwinchester@business.otago.ac.nz

[1] At time of going to press

TABLE OF CONTENTS

TRACKING MOTION IN SPORT – TRENDS AND LIMITATIONS

Baca, Arnold

Department of Biomechanics, Kinesiology and Applied Computer Science, ZSU, University of Vienna, Austria

KEYNOTE ADDRESS

Abstract. A survey of recent developments in hard- and software for tracking the motion of athletes and/or objects used in sport is presented. Computer vision-based systems are considered first, putting specific focus on markerless methods for 3D-pose estimation in order to identify how the position of a human body and the configuration of its segments change in time. Potentials and limitations will be discussed from a biomechanics perspective mainly. The usefulness of methods from computer vision is also outlined presenting applications utilizing feature tracking methods such as template matching to find corresponding objects in consecutive frames. Examples from biathlon shooting and archery are given. Object tracking methods are then discussed as very powerful tools for tracking players in game sports. These methods are compared with such based on wireless technologies, using (electronic) transmitters and/or receivers attached to the athlete. By applying triangulation algorithms (GPS and related systems) or by perceiving tags on passing subjects through receivers placed on fixed positions – e. g. along a running route – (RFID-based systems) positions may be estimated with more or less accuracy. Applications of tracking systems based on accelerometers and gyroscopes will finally be discussed. In addition to virtual sports, examples on their usefulness for controlling motion tasks in prevention/rehabilitation programs will be shown. Concluding, expectations on future developments will be given.

Keywords: markerless motion analysis, GPS, position data

INTRODUCTION

Throughout the last decades numerous systems have been developed for capturing human motion. Kinematic information describing the motion of the centre of mass and the change of the body configuration of athletes in time is a prerequisite for biomechanical analyses; information on the change of positions of players in time is required for individual notational analysis and tactics analysis in game sports.

Wireless technologies are beneficial for obtaining valid motion data, since they support non-invasive monitoring of both kinematic and positional information without affecting the athletes in executing their motion (Armstrong, 2007). Active sensing systems operating wirelessly are therefore very popular in sports applications. Devices are mounted on the athletes (small tags) or in the environment and transmit or receive signals, respectively (Moeslund & Granum, 2001). On the other hand, methods from computer vision, which do not require attaching any device or markers to the athlete and/or equipment, are particularly promising. Both, active (wireless) sensing methods and markerless video based methods provide powerful means in tracking motion in sport, but have, however, although some drawbacks. These potentials and limitations will be discussed in the sequel and illustrated by exemplary applications.

COMPUTER VISION BASED SYSTEMS

Video-based motion analysis systems are used in biomechanics in order to identify how a human body and its individual limbs are configured during a motion. Using markers, the pose is defined by the 3D-position of marker points attached to the human body. If the motion is recorded first from several cameras and the digital videos are post-processed afterwards, a more or less long delay between motion execution and the availability of the results must be put up with. Optoelectronic real-time motion capture systems, which identify the image coordinates of the marker points by processors being part of the cameras, overcome this problem.

Model based 3D-motion tracking of athletes

Using image sequences acquired simultaneously from multiple views, 3D joint data at each instant may also be reconstructed without the use of markers. Moeslund et al. (2006) report on significant advances from 2000-2006 in reconstructing human motion either from monocular or multiple view image sequences. They identify model based analysis-by-synthesis of human motion as a dominant methodology for human pose estimation in these (Moeslund et al., 2006) as well as in earlier years (Moeslund & Granum, 2001). Those kinematic pose parameters of a human body model are estimated, which result in a most similar appearance of its synthesized shapes to the actual shapes (edges, silhouettes, contours, etc.) of the real subject in the multi-view camera images. Figure 1 illustrates the principle. Rosenhahn et al. (2001), for example, use silhouettes for model fitting with 21 degrees of freedom (upper part of the body only) in a 4 camera setup and report promising results when comparing their approach with a commercial marker-based tracking system. The geometric models for the body segments used either constructed of simple volumetric primitives (Gavrila & Davis, 1996) or of freeform surface patches (Rosenhahn et al., 2006). Complex environments, noise, occlusions and shading complicate tracking of human segments yielding less reliable results. Hence most methods are applicable only under laboratory conditions.

Although a development towards almost real-time systems achieving results comparable to those from marker-based systems can be observed, their application for analysing loads in biomechanical studies is still limited. In order to estimate joint loads and joint torques, exact locations of joint centres and accurate angular accelerations of the body segments are required. In marker-based methods markers are attached according to certain protocols with regard to certain palpable parts of the skeleton, thus indirectly defining the location of the joint centre. Additional (redundant) markers may be used to improve accuracy. In analysis-by-synthesis methods there are no such means. However, there are some promising results (Corazza et al., 2007), which give hope that this problem might be overcome in the not too far future. Even though the estimation of angular accelerations (involving the calculation of second derivatives from angular displacement data) is also a severe problem in marker-base methods, this may be more crucial using analysis-by-synthesis-methods, since even more noise in the angular displacement data can be expected.

Nevertheless, such systems provide an interesting alternative for obtaining kinematics parameters (linear and angular displacements). This is particularly the case during competitions, where no markers may be attached. It requires, however, that the athletes are wearing tight fitting clothes.

Figure 1: Analysis-by-synthesis. Volumetric primitives representing the body segments are optimally fitted into views from multiple cameras. Schematic view.

Feature tracking

Another approach for tracking objects in time is to identify certain features (patterns, edges, contours, etc.) and to detect these features in consecutive frames. Marker based analysis techniques are typical representatives of such systems. If the colour, intensity or appearance of an object (e. g. a certain body segment of the athlete) differs significantly from the environment, it may easily be tracked using methods from pattern recognition such as template matching.

A colour based tracker is, for example, utilized by Mauthner, Koch, Tilp and Bischof (2007). Athletes are tracked during a beach volleyball competition, using a single camera only. The authors report sufficient accuracy compared to hand annotated position estimation.

Another sports-specific example from biathlon shooting (Baca & Kornfeind, 2006) shall be described in some more detail. Coaches and athletes are interested in the motion of the barrel of the rifle just before shooting. This is a crucial factor because of the preceding high exertions of the athletes. Commercial laser based systems, which measure and store the hit point of the shot and the on-target trajectory of the alignment of the weapon, require to attach a (laser) device to the rifle. A (low-cost) video-based system was therefore developed, which does not necessitate fixing any such device or sensor onto the rifle. A video camera is set up in a distance of about 5-7 m in front of the athlete in a laterally displaced position and records the muzzle. The digital video is directly stored on computer disk. An application program enables the user to start and stop the video acquisition, to specify the template containing the image of the muzzle and to track this muzzle automatically using 2D-normalized cross correlation for template matching. Tracking is performed in a user-selectable time interval before the shot. From the sound track recorded the instants of shooting are estimated. Applying the results of a calibration procedure, the image coordinates obtained by the tracking algorithm are converted to object space coordinates. Figure 2 shows an image of the barrel and the reconstructed trajectories of the muzzle.

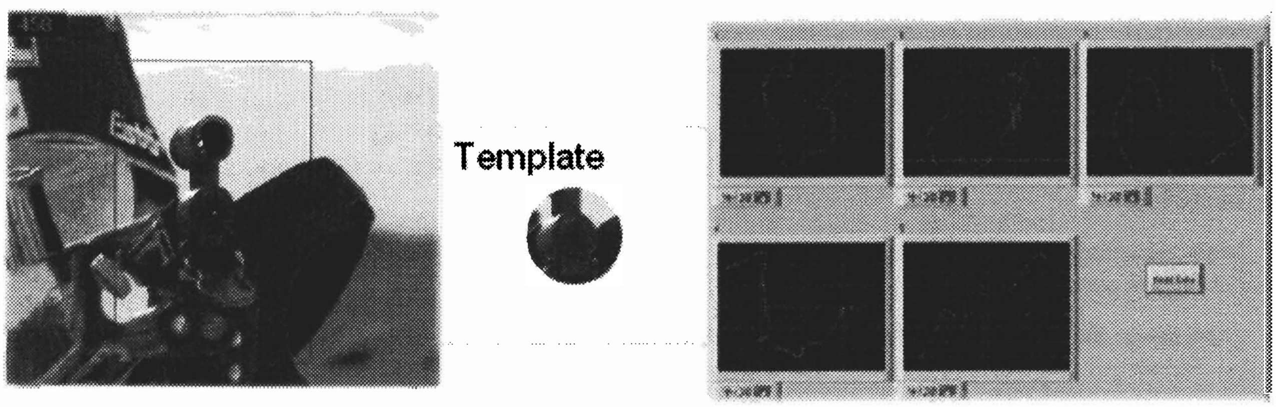


Figure 2: Tracking the muzzle in biathlon shooting. Left: Frame from video recording. Middle: Template to be tracked. Right: Reconstructed trajectories (5 shots).

The system has practically been used for analysing biathletes from the Austrian Junior team (Heller et al., 2006). Trajectories of the muzzle, which moved in a range of less than 1 mm in many cases, could well be reconstructed and visualized graphically.

The method has successfully also been adapted for analysing the aiming process in archery.

Object tracking

Video-based observation systems have also been developed for overall tracking of subjects in game sports (particularly in soccer). Players are viewed as single objects neglecting the motion of their body segments (Beetz et al., 2006; Perš & Kovačič, 2000). Match activities, distances covered during a game and running speeds may be collected. Players and ball need not be equipped by any tag.

There exist a few commercial, semi-automatic systems using video camera systems (normally 8). A sophisticated system for automatic analysis of a soccer game is proposed by Beetz et al. (2006). The video streams of a set of TV cameras are evaluated in order to track players and ball.

However, for identifying strengths and weaknesses of players and for supplying suggestions for game strategies, automatic processing of positional data alone is not sufficient. Moreover, methods from image understanding have to be exploited (Beetz et al., 2005; Lames, 2008).

Both feature and object tracking methods may form the basis for automatically segmenting motions (from individuals or teams) into (classified) actions or for categorizing or judging the motions subsequently.

GPS, RFID AND SIMILAR WIRELESS LOCATION DETECTING TECHNOLOGIES

Video based systems are limited in the observation area to be covered. GPS-based systems do not have these restrictions and are well suited for the determination of speed and position in sport's activities, where athletes move along rather simple trajectories, such as in rowing, cycling, running (Townshend et al., 2008). They provide a cheaper and easier to operate alternative to (more accurate) differential GPS (DGPS) systems, which require a fixed reference station. Both techniques may, however, only be used outdoor under good weather conditions and require small antennas onto the objects to be tracked.

Actual technical developments like the Leica GS20 Professional Data Mapper (total mass with battery: 0.652 kg) combined with a high precision GPS antenna (AT501) provide accuracy in position of typically 5 - 10 mm + 2 ppm (rms) and data collection of up to 20 Hz which is quite challenging for a lot of sports applications. (cf. www.leica-geosystems.com, accessed March 25th, 2008).

Active (battery powered) or passive tags transmitting radio or microwaves are also increasingly used to compute the position of freely mobile objects within a local bounded 3D area. In contrary to GPS-based systems, this calculation is not done on the object tracked, but in a central control unit. 3D positions are estimated from the times an electromagnetic wave needs from light weighted tags (transponders) attached to the objects to be tracked to a certain number of receiving stations using a triangulation method. These stations surround the environment under investigation and are linked to the central station. Figure 3 illustrates a schematic view of a typical system configuration. It is possible to refer to several objects (tags) independently and to receive their specific position in real time. Objects may thus continuously be tracked and identified.

Main advantages of this technique are a high spatial accuracy and a recording frequency of up to 1000 measurements per second.

Athletes or players are, however, obliged to wear special tags. Transponder-based systems are therefore well suited for in- and outdoor training, but difficult to use in regular matches.

RFID (Radio Frequency Identification) based applications shall conclude this section. Upon activation, small RFID-chips transmit a unique signal, which is identifiable by short distance readers. This technology is, for example, used in marathon races. Whenever a runner wearing such a chip crosses fixed mats throughout the race, the time is recorded.

3- AND 6-DEGREE OF FREEDOM (DOF) MOVEMENT MEASURING DEVICES

The number of degrees of freedom (DOF) refers to the different kinds of motion that a sensor is capable of measuring. A head tracking system used in a head mounted display, for example, is a 3-DOF system with degrees of freedom to measure roll and pitch of the head as well as to measure head rotation. A 6-DOF system considers 3 axes of rotation and 3 axes of lateral motion. 3D-position and orientation of objects can thereby be determined.

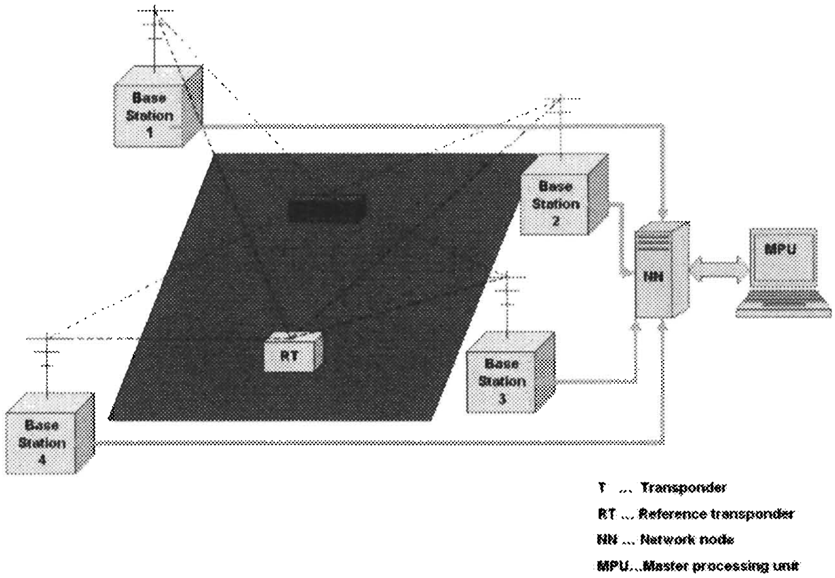


Figure 3: Setup of a transponder based tracking system, consisting of one reference transponder, k measurement transponders (tags), four receiver stations, a fiber optical network with one network node and the master processing unit. Adapted from Fischer at al. (2003).

Such devices require several sensors, for example 3 accelerometers and 3 rotation sensors, (gyroscopes). Changes in position are calculated from all signals measured. The devices may be small in weight and size. Measured data may be transmitted using wireless technologies.

Virtual Sports

One very popular application of such devices lies in the area of virtual sports. Computer-Human-Interaction systems have already been designed for different sports. Users' motions are captured and used as input for controlling the virtual run in order to improve the realism and the enjoyment during the game experience. Head, hand and body tracking devices are designed for this purpose. Nintendo's Wii video game console is a typical representative of systems of that kind. Accelerations in 3 dimensions of a wireless controller are detected enabling to control the game.

Beside of using such devices for traditional 3D-motion analyses by attaching them on body segments, they may also be applied in real-time feedback systems such as is the case in the following example.

Example: Cervical spine training

In order to objectively assess cyclic motion patterns in the region of the cervical spine different measuring methods are applied. Heller et al. (2008) report on the development of an audio-visual feedback system for training this region, allowing an objective estimation of the range of motion as well as a practical application in prevention and rehabilitation.

A 3D wearable virtual reality device (see Figure 4) is used to record the motion of the head. This device enables to track the motion of the head in 3 dimensions using gyroscopes and accelerometers. It is therefore possible to control the motion of a mouse cursor on a monitor by moving the head (see Figure 5).

The total range of motion of the cervical spine in 3 dimensions is captured in order to individualize the dimensions of geometrical shapes to be traced. General accepted exercises for training the cervical spine (Dreher-Edelmann, 1997) have been integrated into the software developed. Coloured graphics of geometrical figures (see Figure 5) are drawn on a white background for this purpose.

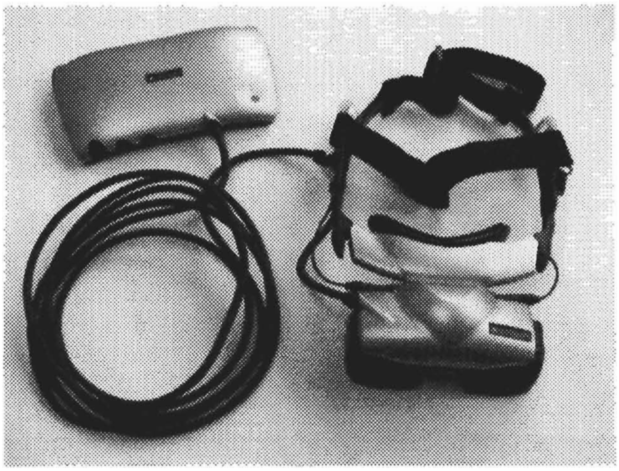


Figure 4: 3D wearable virtual reality device (eMagin Z800 Visor with controller) with 2 SVGA micro displays

During the motion of the head the current mouse position is checked. If it is outside the target area an acoustic signal is given (Figure 5).

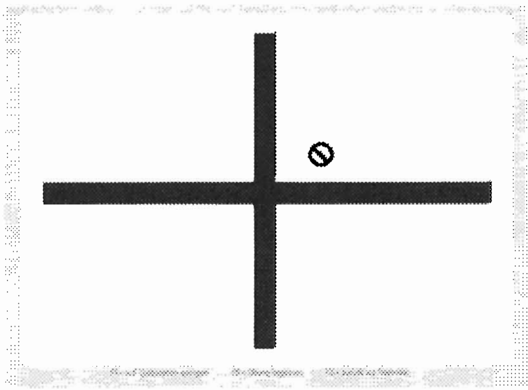


Figure 5: Screenshot showing a simple tracking exercise. The mouse cursor has to be moved along the cross. If it is outside the valid (black) area, as is the case here, a prohibit-sign is displayed.

The audio-visual feedback system is currently still under development. Upon successful completion and verification of the test quality investigations will be performed in order to check its suitability for use in practical preventive and rehabilitative measures.

CONCLUSION

Model based methods from computer vision provide a promising alternative for marker based methods. In order to be applicable in outdoor sports activities, more research on their application in complex environments is required. First steps towards biomechanical applications have also already been taken.

Video based object tracking systems and wireless technologies using active sensing systems as well as 3- or 6-DOF sensing devices provide manifold means for tracking athletes. There is no doubt that they have the potential to supply data for an automatic subsequent computerized detection and analysis of actions and intentions from individual athletes and from teams.

Acknowledgements

I wish to thank Mario Heller, Philipp Kornfeind and Martin Böcskör for their assistance with this paper.

References

Armstrong, S. (2007) Wireless connectivity for health and sports monitoring: a review. *British Journal of Sports Medicine*. **41:** 285-289.

Baca, A., Kornfeind, P. (2006) Rapid feedback systems for elite sports training. *IEEE Pervasive Computing*. **5 (3):** 70-76.

Beetz, M., Kirchlechner, B. and Lames, M. (2005) Computerized real-time analysis of football games. *IEEE Pervasive Computing*. **4 (3):** 33-39.

Beetz, M., v. Hoyningen-Huene, N., Bandouch, J., Kirchlechner, B., Gedikli, S. and Maldonado, A. (2006) Camera-based observation of football games for analyzing multi-agent activities. *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 42-49.

Corazza, S., Mündermann, L., and Andriacchi, T. (2007) A framework for the functional identification of joint centers using markerless motion capture, validation for the hip joint. *Journal of Biomechanics*. **40:** 3510-3515.

Dreher-Edelmann, G. (1997) *Gymnastik für die Hals- und Brustwirbelsäule*. [*Gymnastics for the cervical and thoracic spine*.] München: Urban & Fischer.

Fischer, A., Pracherstorfer, G., Stelzer, A. and Söser, A. (2003) Local position measurement system for fast and accurate 3D-monitoring. *Proceedings of SPIE, vol. 5048*, SPIE Press, Billingham, 128-135.

Gavrila, D. M. and Davis, L. S. (1996) 3-D model based tracking of humans in action: a multi-view approach. *Proceedings IEEE Computer Vision and Pattern Recognition*. IEEE Computer Society, Washington DC, 73-80.

Heller, M., Baca, A., Kornfeind, P. and Baron, R. (2006) Analysis of methods for assessing the aiming process in biathlon shooting. In H. Schwameder, G. Strutzenberger, V. Fastenbauer, S. Lindinger and E. Müller (eds.) *Proceedings of the 24th International Symposium on Biomechanics in Sports, vol. 2*. University Press, Salzburg, 817-820.

Heller, M., Karall, E., Schleimer, C., Baca, A. and Baron, R. (2008) Entwicklung eines audio-visuellen Feedbacksystems für den Einsatz in der Prävention und Rehabilitation im Bereich der Halswirbelsäule [Development of an audio-visual feedback system for application in prevention and rehabilitation in the cervical spine region]. *Proceedings of the 7th Symposium of the Section "Computer Science in Sport" of the German Association of Sport Science* (in press).

Lames, M. (2008) Coaching and computer science. In P. Dabnichki and A. Baca (eds.) *Computers in Sport*, WIT Press, Southampton, 99-120.

Mauthner, T., Koch C., Tilp, M. and Bischof, H. (2007) Visual tracking of athletes in beach volleyball using a single camera. *International Journal of Computer Science in Sport*. **6 (2):** 21-34.

Moeslund, T. B., and Granum, E. (2001) A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*. **81:** 231-268.

Moeslund, T. B., Hilton, A., and Krüger, V. (2006) A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*. **104:** 90-126.

Perš, J. and Kovačič, S. (2000) A system for tracking players in sports games by computer vision. Electrotechnical Review, Ljubljana, Slovenija. **67 (5):** 281-288.

Rosenhahn, B., Brox, T., Kersting, G., Smith, A.W., Gurney, J. K. and Klette, K. (2006) A system for marker-less motion capture. *Künstliche Intelligenz*. **6:** 45-51.

Townshend, A. D., Worringham, C. J., and Stewart, I. B. (2008) Assessment of speed and position during human locomotion using nondifferential GPS. *Medicine & Science in Sports & Exercise*, **40:** 124-132.

USE OF THE DUCKWORTH/LEWIS METHODOLOGY TO PROVIDE ADDITIONAL BENEFITS FOR ONE-DAY CRICKET

Lewis, Tony

The Business School, Oxford Brookes University, OX33 1HX, UK..

Paper Submitted for Review: 24 March 2008
Revision submitted and accepted: 22 May 2008

Abstract. The Duckworth /Lewis (D/L) method for resetting targets in rain-interrupted one-day cricket matches has been in use for well over eleven years and has been the world standard method for most of this time. The formula upon which it is based, giving the further runs that are obtainable on average as a function of overs remaining and wickets down, can also be used for secondary purposes. Such usage has the potential for enhancing interest in games and for better understanding of the progress of the two teams. This paper discusses two facilities that would add to the value of media coverage of a game; projection of the final score of the side batting first and monitoring the performance of the side batting second towards its target. During the first innings, current practice of the TV broadcasting companies is to provide a range of possible final scores based entirely on run-rates, which ignore the effects of wickets remaining. During the second innings comparisons are made with the progress of the first-innings at similar points, which is of interest but not, it is argued, the most relevant measure of comparison. The paper suggests that use of the D/L methodology would provide more meaningful and valid measures of the progress of the two teams in their respective goals, which can be presented to viewers in easy-to-grasp graphical form. Data from several one-day-international matches are used to illustrate the suggestions.

Keywords: one-day cricket, media interest

INTRODUCTION

The Duckworth/Lewis (D/L) method of resetting target scores in rain-interrupted one-day cricket matches was first used in 1997. In 2001, it was formally adopted by the International Cricket Council (ICC) as the world standard and it has remained so ever since.

The principle of the method is that when a match is shortened after it has started, the target score for the side batting second (referred to here as Team 2) is adjusted according to the run scoring resources that the two sides have for their innings. These resources are a combination of overs to face and wickets to lose. The methodology is fully described in Duckworth and Lewis (1998).

For the purposes of this paper we use the D/L model of average runs scored $Z(u,w)$ in u overs remaining when w wickets have been lost, and the model of resource percentages remaining $P(u,w) = Z(u,w)/Z(50,0)$ which is an expression of the percentage of the average runs scored in a standard one-day international (ODI) innings of 50 overs and 10 wickets.

Tables provided in Duckworth and Lewis (2004) summarise the percentages in-use until 2002. In 2002, following a review of the data from more recent one-day matches, the tables were updated to reflect changes in the way the game was played. These tables are current and can be seen on the Internet. [http://www.icc-cricket.com/rules/d-l_method.pdf].

To operate better in very high scoring matches, the method was upgraded in 2003. This upgrade, for which a computer is essential for its operation, is referred to as the Professional Edition and is explained in Duckworth and Lewis (2004). The previous manually operable version, now known as the Standard Edition, has been retained for use where computers are not generally available, principally at lower levels of the game.

As well as its standard uses for setting revised targets and deciding the results of prematurely terminated matches, the D/L methodology can also be used for other purposes which have the potential for enhancing interest in an individual game and in one-day cricket in general. The capability for implementing some of these additional uses is already incorporated in the software for the Professional Edition. This paper describes these additional potential uses of the D/L methodology.

PROJECTION OF TEAM 1'S FINAL SCORE

Coverage by the broadcasting media, principally TV, of one-day matches often includes projections of Team 1's score from around the mid-point of their innings. This is merely a linear extrapolation of the score based on the number of overs still to be faced and is given for various assumed run rates, usually four, six or eight runs per over, and it is left to the commentators to guess which rate is the most appropriate given the match situation.

The projections given add very little to what the commentators, or indeed the viewers, can guess for themselves. Their main weakness is that no account is taken of the wickets lost and hence of the real value of the remaining overs. If many wickets are still in hand, clearly the team can take greater risks in the later overs and attempt to achieve a higher run rate. If, however, many wickets have been lost, they will need to exercise greater caution in order to avoid forfeiting the opportunity to make the best use of the remaining overs available.

A single, more meaningful, projection may be made by taking account of the total run-scoring resources remaining for the innings, and the D/L methodology provides the way of combining the contributions to these resources of both overs remaining and wickets in hand. As the actual parameters of the formula in the Professional Edition (Duckworth & Lewis, 2004) are not known until the innings has been completed, the resource percentage tables used to operate the Standard Edition, which represent the average situation over many matches, are used for this purpose. O'Riley and Ovens (2006) used the Standard Edition tables in assessing the D/L method, and several other methods, for their ability to predict the final score.

What we propose is that we take a weighted average of the projected further runs based on the runs per resource already experienced, and the projected further runs based on applying the remaining resources to the prior estimate of the total score for a 50-over innings, the weighting factors being the resources used and to come respectively.

If Team 1 starts with its full 50 overs and have scored S runs for w wickets lost, and now have u overs left, a projection of their final total, based on their pro-rata scoring rate in relation to resources consumed would be $F_1 = S/[1 - P(u,w)]$. For instance, if after 28 overs have been bowled of a 50-over innings Team 1 have scored 110 runs for the loss of 3 wkts, the current resource table cited tells us that on average 52% of their run scoring resources remain (a combination of 22 overs to be received and 7 wickets to lose), and an extrapolation of this run rate per unit resource would give a projected final total of 110 x 100/52 = 212, i.e. a further 102 runs in the remaining 22 overs.

This does not, however, take into account the variable nature of a cricket innings; nor does it allow a reliable projection to be made from an early stage of the innings as the start may well not be an indicator of how the innings will eventually build. It merely assumes that the innings will carry on at the same rate of runs per unit of resource as it has achieved so far.

An alternative extrapolation, which is guided only by what is seen on average in the remaining resources and takes no account of the rate achieved so far, would be to apply the remaining resources to the prior 50-over expectation (in the same way that enhanced targets are calculated in the Standard Edition) by applying Team 2's excess resources to the quantity called G50, this being the average 50-over total for the relevant standard of cricket (Duckworth & Lewis, 1998 and 2004). On average $Z(u,w)$ runs would be scored to add to the S already scored to find the projection. $Z(u,w)$ is easily obtained for a particular match, by using the D/L tables, as $P(u,w)Z(50,0)$. However, the value $Z(50,0)$, currently 235 for ODIs, does not have to be the basis for the average for a particularly match. It could be a different figure which takes into account team strengths and ground conditions – what commentators may regard before the match as 'par for the course'. If we denote this prior value by G then the projection would be $F_2 = S + P(u,w) * G$. In our example, taking G = 235, this second method of projection would give a predicted further runs of 235 x 0.52 = 122 and a projected total of 232 since there are already 110 runs on the board.

Both of these approaches are combined within our computer software by taking the two alternative values F_1 and F_2 for the further runs to be scored and weighting them relative to the resources consumed and those remaining respectively. The rationale is that F_1 would be a less reliable as a predictor in the earlier stages of an innings, with F_2 , a priori, a more reliable predictor in the earlier parts of an innings. As the innings progresses, what has happened in the match will have more and more bearing on the final total than average performance and so the weightings would reverse in importance. The weights used are the resource percentages consumed and still remaining respectively, rather than the proportions of the overs available, as this is consistent with the D/L's methodology of assessing an innings according to its resources left. The projected total is then $F = [1 - P(u,w)]F_1 + P(u,w)F_2$. For our example, this gives that the further runs expected in the remaining 22 overs would be 102 with a weight of 0.48 and 122 with a weight of 0.52, giving a combined prediction of 112, when rounding to the nearest integer. With 110 runs already on the board, the projected 50-over total would therefore be 222.

It can easily be appreciated that this method gives a continuous projection from the start to the finish of the innings. Before the first ball is bowled, the projection is the value chosen for G, and as the innings progresses, it gradually takes more account of what has happened, and when the innings closes the projection converges with the final score. This estimate of the further runs is then added to the runs already scored to give a single projected total score. This procedure has already been incorporated into the D/L software.

We have monitored the projection during many important matches and found that it very often gives reliable results. Of course, no method can allow for the possibility that Shahid Afridi (Pakistan) or Jacob Oram (New Zealand) may hit 40 runs off the final two overs, but in the long run, we have found that it works well.

Figure 1 gives a typical example taken from Pakistan's innings against Australia on 19th June 2001. The projection as the innings proceeded is shown alongside the runs scored and wickets lost, which are represented by the small dark discs. [http://uk.cricinfo.com/db/ARCHIVE/2001/OD_TOURNEYS/NWS/SCORECARDS/AUS_PAK_NWS_ODI8_19JUN2001_RR-COMPARISON.html.]

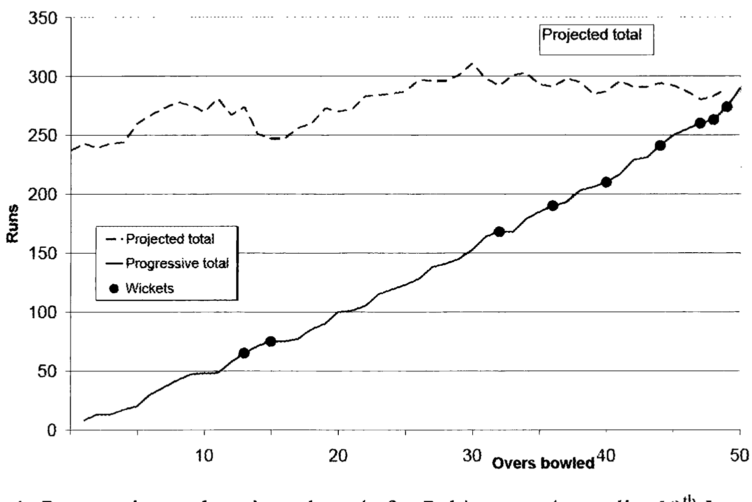


Figure 1: Progressive and projected totals for Pakistan, v Australia, 19th June 2001

Figure 1 illustrates that a score of about 300 was 'expected' from around the 20th over although the deviation up above 300 at around the 30th over and back again showed how Australia checked Pakistan's progress by the taking of several wickets in the last 20 overs.

MONITORING TEAM 2'S PROGRESS TOWARDS THEIR TARGET

TV graphics are a fine art and have become extremely sophisticated over the years. Nevertheless, we believe that in developing their sophistication they have sacrificed usefulness by not displaying the most relevant quantities. The classic examples are the 'worm' and 'Manhattan' plots designed to illustrate how the progress of Team 2's innings compares with the way Team 1's total accrued.

A statistic that is constantly displayed, with the same intended purpose, is the required run rate. Commentators continually refer to this as though 'staying up with the run rate' were the prime objective of the chasing side. Again, this statistic takes no account of the number of wickets that are down. For instance, to take an extreme example to illustrate the point, it would be of little comfort for a side chasing 200 in 50 overs, who were on 120/9 after 25 overs, to be told that they were well ahead of the required run rate.

The 'Manhattan' plot is a three-dimensional [perspective] block diagram showing the runs per over. It is visually appealing but provides very little useful guidance on how Team 2 are performing with respect to their target.

One relevant statistic, that is displayed on the scoreboard, but is very seldom shown on TV screens, is the D/L par score. The par score is displayed primarily to inform players and spectators alike. It is the score that Team 2 would have to exceed to win the match in the event of the match being terminated at that point. Team 2 would win or lose (or tie) by the number of runs they are ahead or behind (or level with) par upon premature termination of the match. Par score schedules are available to teams and match officials, so that teams should be aware of exactly how they stand after each ball.

But as well as indicating what would be the result if the match were to be terminated, the par score schedule also provides a reliable indication of how Team 2 are progressing relative to their target. If they are ahead of par, they are winning and if they are behind, they are losing. This is a much more relevant guide to their progress than run rate compared with the rate required or than can be displayed by worms or Manhattan plots.

To illustrate what we would propose, Figure 2 gives an example drawn from the match between South Africa and West Indies in the 2003 World Cup on 9 February 2003. [http://uk.cricinfo.com/db/ARCHIVE/WORLD_CUPS/WC2003/SCORECARDS/POOL-B/RSA_WI_WC2003_ODI1_09FEB2003_RR-COMPARISON.html]

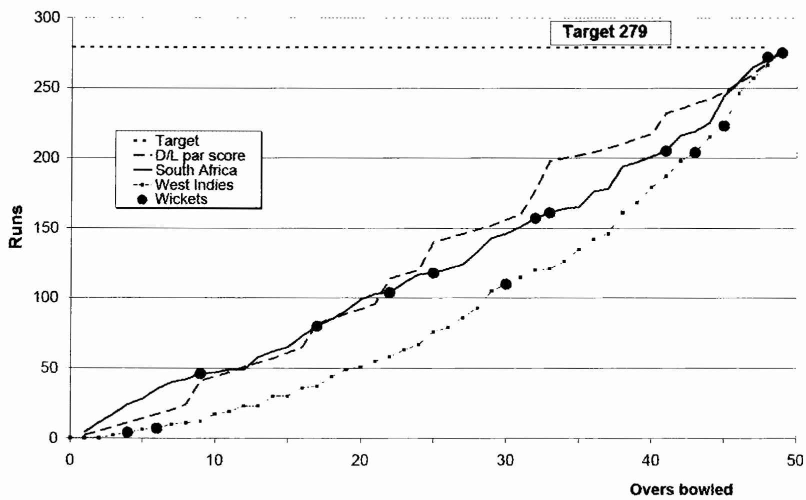


Figure 2: South Africa's progress relative to D/L par and West Indies' score: World Cup 9 Feb 2003

The West Indies (WI) scored an abnormally large percentage of their 278 runs in the latter part of the innings. Consequently, comparison of South Africa's (SA) progressive score at a more 'normal' pace was meaningless in assessing South Africa's progress in chasing the 279 to win. It shows SA always ahead of the WI graph. A much more informative indicator would have been the comparison of SA's progress relative to the D/L par score, which shows SA generally behind their requirement to win and they finally lost by three runs.

Interruptions to Team 2's inning

When matches are interrupted then the traditional worms are totally meaningless as a means of comparison of Team 2's performance relative to its revised target. This will be true whether there are interruptions in either innings. An example of this is given in Figure 3 and relates to the match New Zealand (NZ) v India, Jan 1999. In this match India scored 257 in their 50 overs. NZ were 168/3 in 30.4 overs when 11 overs were lost due to failed floodlights. After the restart, the D/L method reduced the target to 200 so that the comparison of the worms was then of little meaning and the only sensible comparison was NZ's progress relative to the D/L par. In this particular match NZ were ahead of par at the stoppage and went on to win the match with an over to spare.

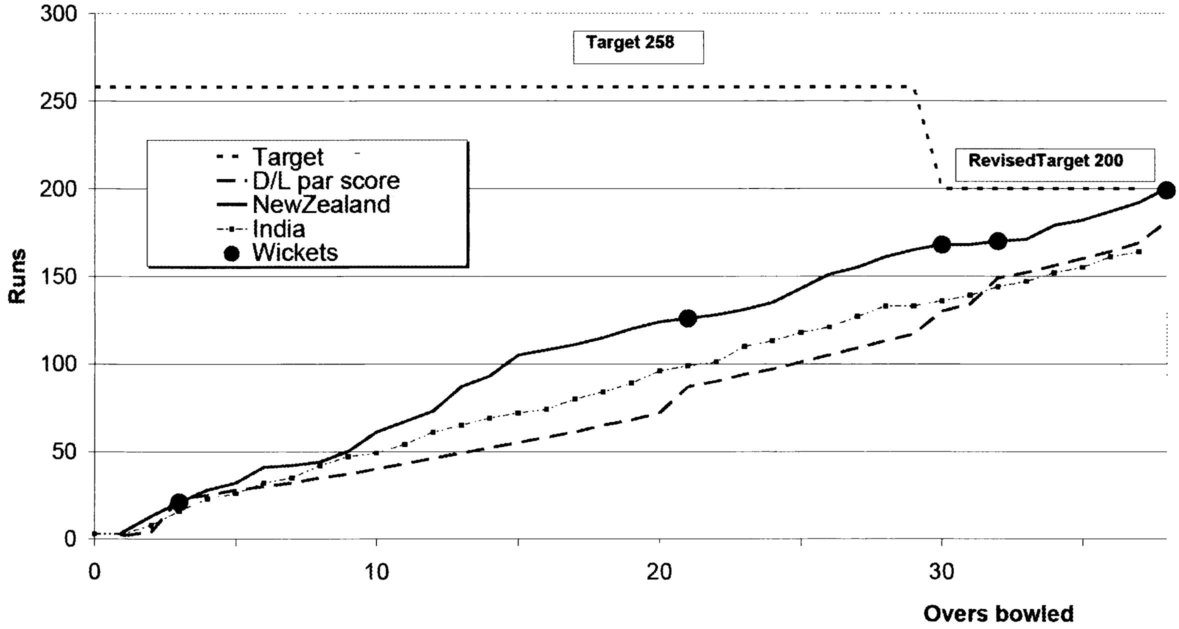


Figure 3: NZ's progress relative to D/L par and India's score: 9 Jan 1999

Clearly, if there are interruptions in Team 1's innings then comparing worms is pointless even from the start of Team 2's innings.

SUMMARY AND CONCLUSION

We have illustrated that there is considerable potential for the use of the D/L methodology beyond that of its original purpose for resetting targets in interrupted one-day cricket matches. The methodology could be used for graphical displays of the projected score for the team batting first and for the progress of the team batting second towards its target. The challenge is to persuade the TV and marketing companies of the benefits that can be realised from the additional uses of the Duckworth/Lewis methodology.

Acknowledgment

The author acknowledges the collaborative work undertaken by his colleague, Dr Frank Duckworth, that contributes to the analysis and conclusions summarised in this paper.

References

Duckworth, F. and Lewis, A. (1998) A fair method for resetting the target in interrupted one-day cricket. matches. *Journal of Operational Research Society*. **29**: 220-227.

Duckworth, F. and Lewis, A. (2004) A successful operational research intervention in one-day cricket. *Journal of Operational Research Society*. **55**: 749-759.

O'Riley, B. and Ovens, M. (2006) Impress your friends and predict the final score: An analysis of the psychic ability of four target resetting methods. *Journal of Sports Science and Medicine*. **5**: 488-494.

RANKING INTERNATIONAL LIMITED-OVERS CRICKET TEAMS USING A WEIGHTED, HETEROSKEDASTIC LOGISTIC REGRESSION WITH BETA DISTRIBUTED OUTCOMES

Stern, Steven

School of Finance and Applied Statistics, The Australian National University, Canberra, Australia

Paper Submitted for Review: 17 March 2008
Revision submitted and accepted: 11 June 2008

Abstract. A regression-based ranking method is developed and applied to international limited-overs cricket, using a database of matches played between September 1999 and December 2007. The structure employed is a generalised linear model with logistic link function and beta distributed outcomes and is used to estimate team strength parameters which in turn yield ranking scores. The outcome variable for the regression is a newly proposed measure of the margin of victory based on the Duckworth-Lewis methodology. The model uses Weibull weighting to discount the impact of matches played in the past and incorporates a heteroskedastic structure to account for the potentially skewing effects of uncommonly large victories. Finally, the model is flexible enough to allow examination of the effects of other factors such as home ground advantage.

Keywords: Duckworth-Lewis Method, Generalised Linear Model, Relative Margin of Victory.

INTRODUCTION

Who's Number One? Sports pundits, participants and enthusiasts alike are obsessed with answering this question and, more generally, with rankings of all kinds. Often, the methods used to arrive at the answers are based on nothing more than "expert opinion" or simple statistics such as team win-loss records. While such subjective approaches lead to enthusiastic and revealing debate, they typically reveal more about the debaters than the actual answer to the question of accurate rankings. Of course, the appropriateness of any ranking system depends on the use to which the resultant ranks will be put. If the intent, as it often is, is simply to create a kind of on-going or annual competition, the winner of which will be the best performed team in the year, then detailed, objective methodology is perhaps less crucial and methods which are simple and intuitive may be the best approach. However, more and more in international sports, "official" rankings are being used for activities which involve monetary outcomes, such as seeding international tournaments, and in these circumstances it seems important to ensure an objective ranking which is based on true team strengths.

In this paper, an objective methodology for ranking international limited-overs cricket teams is developed. The methodology will be general enough that it may be modified to apply to other sports; however, limited-overs cricket is a nice starting point as there are relatively few nations which play the sport at international standard. In addition, as will be discussed in more detail subsequently, the relative margin of victory for any match can be meaningfully defined. This latter issue is of paramount importance if an objective ranking methodology is to be developed which incorporates all the information available in the history of results of any given sporting competition. In particular, an objective ranking methodology should incorporate all the information inherent in:

- The results of all matches in the competition, suitably discounted according to how long ago they occurred;
- The inter-relationships between head-to-head results and results between common opponents when determining relative rankings of individual teams; and,
- The relative margins of victory in matches.

The first two of these criteria lend themselves quite nicely to a weighted regression approach, with match results as the outcome variable, team strengths as the model parameters and weights determined according to the amount of time elapsed since each match has been played. The final of the three criteria listed above is crucially important from the perspective of determining rankings for the purpose of assessing true team strength. A method based only on results would always yield the same change in rankings after incorporating a new result regardless of the margin of victory, implying that should a very low ranked team lose to a very high ranked team, but only by a small margin, this would have the same effect as if the lower rank team had been defeated convincingly. While such a focus purely on result may be sensible from the perspective of a "ranking competition", it seems clear that if a low ranked team looses narrowly to a high ranked opponent, this should be taken to indicate an improvement in the lower ranked team, and ought to be reflected in an increase in standing. A methodology based solely on win-loss outcomes would, by its nature, indicate that any loss would result in a decrease in standing.

In the remainder of the paper, a measure of relative margin of victory suitable for limited-overs cricket is introduced, based on the famous Duckworth-Lewis methodology (1998; 2004), and a ranking based on applying a weighted generalised linear model incorporating beta distributed errors to the victory margins is developed and applied to a database of the results of the 12 major limited-overs cricketing nations: Australia, Bangladesh, England, India, Ireland, Kenya, New Zealand, Pakistan, South Africa, Sri Lanka, the West Indies and Zimbabwe. Modifications to the regression structure to incorporate various desirable features into the ranking mechanism are also discussed. Similar methods have been investigated by de Silva et al. (2001) as well as Clarke and Allsopp (2001) and, where appropriate, comparisons with their work are made.

A MEASURE OF VICTORY MARGIN IN LIMITED-OVERS CRICKET

Unlike many international sports, the pattern of play in limited-overs, or one-day international (ODI) cricket does not entail each team undertaking their offensive and defensive activities in a dynamic flow throughout the game. Instead, ODI cricket playing structure consists of two batting innings, one for each team, played consecutively and during which the batting team compiles its score of runs. Each batting innings continues until either the completion of a fixed number of overs, usually fifty, or the loss of ten wickets, or (in the case of the team batting second) the number of runs scored is sufficient to ensure victory, whichever occurs first. This structure, whereby each team undertakes its entire offensive activity contiguously, makes it particularly unique among high profile international sports and also makes a determination of the margin of victory for a match complex. If the team batting first wins the match then, assuming that the match was uninterrupted, the margin of victory is typically determined by the difference in the number of runs scored by the two teams. However, if the team batting second wins, then their innings ends as soon as they have scored enough runs, and thus their margin of victory is typically stated in terms of either the number of wickets or the number of overs (or both) still remaining when they achieved victory.

This asymmetry in the reporting of victory margin is further compounded by the fact that the outcomes of matches interrupted by weather or other circumstances are determined using a method developed and then further improved by Duckworth and Lewis (1998; 2004). Fortunately, the added complication of the Duckworth-Lewis (D/L) methodology also allows for a sensible way to develop a symmetric and practical definition for the margin of victory in ODI matches. The essence of the D/L method is "scoring resources". At any stage of a batting innings, the D/L method uses both the number of overs and the number of wickets remaining to determine the proportion of scoring resources still available. The primary use of the D/L method is to determine the proportion of scoring resources which are lost due to interruptions, so that appropriate comparison of scores from the two innings can be made to determine a winner. However, the resource calculations can also help determine a margin of victory. In particular, one sensible method of determining the margin of victory is to calculate the proportion of available resources which the winning team did not need. In this way, the size of a victory can be determined in a symmetric fashion regardless of whether the team batting first or the team batting second wins.

To implement this margin of victory calculation, define S_1 and S_2 to be the runs scored by the team batting first and the team batting second, respectively. Similarly, let U_1 and U_2 be the amount of their available scoring resources actually utilised by each team, and let M_2 be the total resources available to the team batting second (the total resources available to the team batting first is always equal to U_1), as determined by the D/L methodology. For details of calculating resources using the D/L method, see Duckworth and Lewis (2004); however, for clarity, note that D/L resources are calculated on a proportional scale where unity is equivalent to the resources associated with a fifty over innings and ten available wickets, so that in an uninterrupted match U_1 and M_2 will always be 1 and U_2 will be 1 whenever the team batting first wins (or the game is tied) and less than 1 whenever the team batting second wins. If the team batting second wins the match, their margin of victory, V, can then be calculated as:

$$V = \frac{M_2 - U_2}{M_2},$$

which is the proportionate amount of their unused resources. Alternatively, if the team batting first wins, then the proportion of unnecessarily used resources (recall that the team batting first will always use all of its allotted resources, as it does not know beforehand how much it will ultimately need) can be calculated as:

$$V = \frac{U_1 - R_1}{U_1},$$

where R_1 is the amount of resources actually needed for the team batting first to have achieved victory. To achieve victory, the team batting first needs only to have scored more runs than the team batting second would have scored given an equivalent amount of resources. Thus, R_1 is the solution to the equation:

$$\left(\frac{S_1}{U_1}\right) R_1 = \left(\frac{S_2}{U_2}\right) U_1,$$

which implies $R_1=(S_2U_1^2)/(S_1U_2)$.

To simplify the calculations, note that the D/L method "par score" at any point in the second innings is the number of runs the team batting second would need to have scored to make the match a tie were it terminated at that point. The value of the par score at the end of the second innings is readily calculated as $P=S_1U_2/U_1$ (though, typically, P is rounded to the nearest whole number). Using this relationship, the victory margin for a team batting first can be re-written as:

$$V = \frac{U_1 - R_1}{U_1} = \frac{S_1U_1U_2 - S_2U_1^2}{S_1U_1U_2} = \frac{P - S_2}{P}.$$

Alternatively, if the team batting second wins, their score is equivalent to the victory target (at least approximately, as the actual score of the team batting second may be a few runs larger than the victory target, depending on the number of runs scored on the winning scoring stroke), which itself is essentially the par score associated with the maximum resources available to the team batting second, S_1M_2/U_1. Thus, it is seen that $S_2 \approx S_1M_2/U_1$ and the victory margin in the case that the team batting second wins is then given by:

$$V = \frac{M_2 - U_2}{M_2} = \frac{S_1M_2 - S_1U_2}{S_1M_2} = \frac{S_2 - P}{S_2}.$$

Combining the two cases yields:

$$V = \frac{|P - S_2|}{\max\{P, S_2\}}.$$

Finally, for the sake of using a single value for all matches, the signed victory margin is defined as:

$$D_1 = \frac{P - S_2}{\max\{P, S_2\}},$$

and will be referred to as the *Relative Resource Differential* (RRD). Note that D_1 is positive if the team batting first wins and negative otherwise; in other words, the RRD is a margin of victory for the team batting first, a negative value indicating the margin of their loss. For the sake of completeness, note that an alternative measure of victory margin defined in terms of the "effective" runs differential

$$\frac{S_1}{U_1} - \frac{S_2}{U_2} \approx D_1 \max\left\{\frac{S_1}{U_1}, \frac{S_2}{U_2}\right\}$$

has been previously proposed, debated and even employed to assess relative team strength (Clarke & Allsopp, 2001, 2002; de Silva et al., 2001; Duckworth & Lewis, 2002). The RRD is preferred here for modelling team strengths since it inherently adjusts for the overall scoring rate in the match. In other words, the RRD recognises that a 50 run victory is less substantial (at least in terms of resources saved) when the final score was 350 to 300 than when the final score was 250 to 200.

RANKING INTERNATIONAL LIMITED-OVERS CRICKET TEAMS

To estimate the team strength of the 12 major ODI teams, a database of all results of matches between these teams from the start of the 1999 Cricket World Cup until the end of the 2007 calendar year is investigated. This database consists of 1066 matches and the breakdown of head-to-head games, including the tabulation of which team batted first and which batted second, is given in Table 1.

Table 1: Number of Head-to-Head ODI Matches by Team and Batting Order
(September 1999 – December 2007)

Batted First	Batted Second											
	AUS	BAN	ENG	IND	IRE	KEN	NZL	PAK	SAF	SRL	WIN	ZIM
AUS	–	3	10	16	0	0	24	11	18	13	15	8
BAN	9	–	7	5	0	2	6	4	5	8	4	16
ENG	15	1	–	17	2	0	8	10	7	11	6	9
IND	17	5	12	–	0	5	11	17	15	12	17	15
IRE	1	1	0	1	–	1	0	0	1	1	1	1
KEN	3	6	2	4	0	–	0	3	4	1	2	7
NZL	13	4	4	12	1	1	–	12	16	12	12	5
PAK	13	8	9	19	1	1	19	–	18	15	13	10
SAF	12	3	10	14	1	3	13	12	–	12	10	9
SRL	11	9	13	20	0	2	14	21	18	–	8	12
WIN	7	7	9	15	0	3	7	13	11	6	–	15
ZIM	7	13	14	10	0	3	5	5	11	9	15	–

The small numbers of matches between some of the teams means using the inter-relationship information contained in the outcomes of matches between common opponents is critical in accurately assessing team strengths, and this information is directly used in a regression approach. A beta regression is employed as described in the following, applied to the relative resource differential (RRD) values, suitably transformed to

$$Y = \tfrac{1}{2}(D_1 + 1),$$

so that the outcomes take values in the unit interval. Note that Y values less than 0.5 correspond to losses for the team batting first and Y values greater than 0.5 to their wins.

Weighted Logistic Regression with Beta Distributed Outcomes

Consider a random quantity, Y, whose outcomes are values within the unit interval. A convenient model for the distribution of Y is given by the beta family. Specifically, the beta family consists of a collection of distributions with support on the unit interval and probability density functions of the form:

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\phi\mu)\Gamma\{\phi(1-\mu)\}}\, y^{\phi\mu-1}(1-y)^{\phi(1-\mu)-1},$$

where $\Gamma()$ is the gamma function, μ is the expectation of the distribution and the variance is $\mu(1-\mu)/(1+\phi)$. As such, ϕ is a measure of dispersion of the distribution, small values of ϕ corresponding to large dispersions.

Regressions based on a generalised linear model with the beta distribution as the error structure have been used recently in various areas (Ferrari & Cribari-Neto, 2004; Smithson & Verkuilen, 2006). To model team strengths, a beta regression model for the RRD values, D_1, is employed with mean structure of the form:

$$E(Y_{ij}) = E\{\tfrac{1}{2}(D_{1,ij} + 1)\} = \mu_{ij} = g^{-1}(\beta_i - \beta_j)$$

where $D_{1,ij}$ is the victory margin for a match between teams i and j in which team i bats first and $g()$ is a suitable *link* function. There are many possibilities for the link, the only requirements being that the function map the unit interval to the entire real line and that it is invertible; however, the common choices are the probit function based on the inverse of the cumulative distribution function of the standard normal distribution, the complementary log-log function, $g(x)=\ln\{-\ln(1-x)\}$, and the logistic function $g(x)=\ln\{x/(1-x)\}$. The latter is chosen for what follows for reasons discussed later. Note that, if the link is the identity function, the margin of victory used is the "effective" runs difference noted above and the beta error structure is replaced by a normal error structure, the model reduces to that of de Silva et al. (2001) as well as Clarke and Allsopp (2001) (with their "first innings advantage" parameter, h, set to 0). As for those models, a parameter constraint is needed here, and so the β_i's will be required to sum to zero to ensure identifiability.

Once a link function is chosen, the estimation of the parameters, $\beta_1,\ldots,\beta_{12}$, is accomplished using maximum likelihood methodology. However, for the application here, the information associated with matches must be discounted according to their age. In general, this can be accomplished quite simply by defining the parameter estimates to be the maximising values of the weighted log-likelihood function:

$$l(\beta_1,\ldots,\beta_{12},\phi)=\sum_{k=1}^{n} w_k \ln\left[f\left\{Y_{i_k j_k}; g^{-1}\left(\beta_{i_k}-\beta_{j_k}\right),\phi\right\}\right]$$

where i_k is the team batting first in the k^{th} match of the dataset, j_k is the team batting second and the w_k's are suitably defined weights. While there are many possible choices for weights, for what follows a choice is required that is based on the age of a match, A_k, and takes values of essentially unity for matches less than a certain age and then decreases steadily until matches beyond a certain age have essentially no contribution. One choice of weights with these features is based on the survival function of the Weibull distribution, w_k=exp(cA_k^d), for some choice of positive constants c and d. Table 2 shows the values of the Weibull survival function weights for different choices of the constants and matches of various ages.

Table 2: Weibull weights for matches of various ages

Constants	Age of match (in years)							
	1	2	3	4	5	6	7	8
c = 0.01, d=4	0.990	0.852	0.445	0.077	0.002	0.000	0.000	0.000
c = 0.005, d=6	0.995	0.726	0.026	0.000	0.000	0.000	0.000	0.000
c = 0.015, d=3	0.985	0.887	0.667	0.383	0.153	0.039	0.006	0.000

In the analysis that follows, the values c=0.01 and d=4 are used, which indicate that matches played within 1 year of the date on which the model is fit are given a weight of essentially unity and matches which are 5 years old or more are given essentially no weight at all. While other choices are possible, the effect on ranking scores of varying values of c and d was investigated and found to be minimal (results not presented).

A Ranking of International Limited-Overs Cricket Teams

The logistic link structure is used for the beta regression model employed here. The expected outcome for a match in which team i bats first and team j bats second is then defined as:

$$\mu_{ij}=\frac{\exp(\beta_i-\beta_j)}{1+\exp(\beta_i-\beta_j)}.$$

This mean structure yields values less than 0.5 if the team batting second has a larger strength parameter and values greater than 0.5 if the team batting first has the larger strength parameter. Also, some algebra shows: $\mu_{ji}=1-\mu_{ij}$. Thus, the chosen mean structure is symmetric about 0.5, as it should be in this case, as the only difference between μ_{ij} and μ_{ji} is the order in which the teams bat, and the expected outcome should therefore reflect that the expected RRD should simply change sign. This symmetric structure is not sustained by the other common link choices, either the probit or the complementary log-log functions.

Once estimates of the β's are obtained, they may be used to derive ranking scores and a standings table. To do so, the expected transformed victory margin, Y, for team i against a generic opponent is used:

$$R_i=\frac{\exp(\beta_i)}{1+\exp(\beta_i)}.$$

For the sake of simplicity, interpretability and comparison with other methods, these values are scaled by a factor of 200 and then round to the nearest tenth of an integer. This means that a team whose average result is a tie (which does not necessarily imply that they win half their games) has a ranking score of 100.

Using this ranking scheme (and the Weibull weighting according to the age of matches described previously, with constants c=0.01 and d=4), the standings as of the end of the 2007 calendar year are:

Table 3: ODI Team Rankings using Basic Beta Regression Model (as of 1/1/2008)

TEAM	RANKING SCORE	TEAM	RANKING SCORE	TEAM	RANKING SCORE
AUSTRALIA	125.5	INDIA	108.4	BANGLADESH	83.4
SOUTH AFRICA	118.1	ENGLAND	106.4	IRELAND	79.6
NEW ZEALAND	117.2	WEST INDIES	105.2	ZIMBABWE	74.6
SRI LANKA	115.0	PAKISTAN	105.1	KENYA	63.2

Over-dispersion and the Effect of Uncommonly Large Victory Margins

A common concern raised over using margins of victory as the basis for rankings is that either a single or relatively few unusually large wins or losses will unduly affect the team strength parameters. One way to counteract this effect is to employ a model structure which deals not just with the mean structure, but with the variation structure as well. If each team is assigned both a strength parameter and a volatility parameter, then a few large victories will tend to increase the volatility parameter estimate, and thus insulate to some degree the strength parameter from the adverse effects of uncommonly large margins.

For beta regression models, heteroskedastic structure may be incorporated using volatility parameters to model the relationship between individual outcomes and the dispersion parameter, ϕ. Specifically, the dispersion structure for the team strength model is defined as: $\phi_{ij}=h^{-1}(\gamma_0+\gamma_i+\gamma_j)$, where γ_i and γ_j are the volatility parameters for teams i and j, respectively, and $h()$ is a suitable link function. As for the mean structure, the volatility parameterisation requires a constraint to ensure identifiablility, so the γ_i's will be required to sum to zero. The estimated strength and volatility parameters are then the values which maximise:

$$l_{\text{het}}(\beta_1,\ldots,\beta_{12},\gamma_0,\gamma_1,\ldots,\gamma_{12})=\sum_{k=1}^{n} w_k \ln\left[f\left\{D_{1,i_k j_k};g^{-1}\left(\beta_{i_k}-\beta_{j_k}\right),h^{-1}\left(\gamma_0+\gamma_{i_k}+\gamma_{j_k}\right)\right\}\right]$$

The ranking scores are determined as before, using the estimated team strength parameters, $\beta_1,\ldots,\beta_{12}$. For the analysis here, the volatility link structure $h(x) = \ln(x)$ is used. While all that is required of the volatility link function is that it be invertible and map the positive half-line to the entire real line, the logarithmic choice is the simplest and most common. The resultant ranking scores from fitting this model are given in Table 4.

Table 4: ODI Team Rankings using Heteroskedastic Beta Regression Model (as of 1/1/2008)

TEAM	RANKING SCORE	TEAM	RANKING SCORE	TEAM	RANKING SCORE
AUSTRALIA	124.9	INDIA	108.1	BANGLADESH	83.0
SOUTH AFRICA	118.4	ENGLAND	106.7	IRELAND	80.8
NEW ZEALAND	116.9	PAKISTAN	105.6	ZIMBABWE	73.8
SRI LANKA	113.5	WEST INDIES	105.4	KENYA	64.6

The differences between these ranking scores and those derived from the model without heteroskedastic structure are small; however, there is one reversal in ranking order with Pakistan moving ahead of the West Indies. A likelihood ratio test indicates that the inclusion of heteroskedastic structure does not significantly improve the model fit (LRS = 12.81 on 11 degrees of freedom, p-value of 0.306). Nevertheless, maintaining it in the model is recommended so that the potential for adverse effects due to extremely large victories is addressed. Indeed, it so happens that the last match in the dataset used for this analysis was played on December 31, 2007 between New Zealand and Bangladesh. New Zealand won the match by bowling out Bangladesh for 93 runs and then scoring 95 runs in 6 overs, resulting in a victory margin of $D_1 = -0.937$, the largest (in absolute value) of any match in the database. The changes in ranking scores under the two models, homoskedastic and heteroskedastic, as a result of this match are shown in Table 5.

Table 5: Change in Ranking Scores as a Result of the Match on December 31, 2007

	RANKING SCORE FROM BASIC MODEL		RANKING SCORE FROM HETEROSKEDASTIC MODEL	
	WITHOUT 31/12/07 MATCH	*WITH 31/12/07 MATCH*	*WITHOUT 31/12/07 MATCH*	*WITH 31/12/07 MATCH*
NEW ZEALAND	115.3	117.2	115.8	116.9
BANGLADESH	85.4	83.4	84.5	83.0

The effect of the massive victory is moderated by the heteroskedastic volatility structure. In this sense, the heteroskedastic structure down-weights large victories or, equivalently, gives more weight to any victory. In other words, this new model gives a "bonus" for simply achieving victory.

DISCUSSION

The regression-based methodology described here was designed for international limited-overs cricket. However, the structure may be applied to other sports. All that is required is a sensible definition of a relative margin of victory. Of course, this may not be easy. Cricket is unique in many ways, and the Duckworth-Lewis methodology made the definition of the RRD possible. Typically, cricket scores are large enough that relative margins of victory are meaningful. By comparison, an appropriate parallel concept may be more difficult to define in other sports. For example, in rugby, it is difficult to decide whether a 6-3 victory is more comparable to a 60-30 victory or a 33-30 victory. Nevertheless, given an appropriate definition of relative margin of victory, the model structure defined here will provide objective ranking scores.

The mean and volatility model structures are also flexible enough to allow inclusion of components to address other aspects of matches which may affect outcomes and should be accounted for in assessing team strength. For example, a factor for home ground advantage may be incorporated into the mean structure as:

$$\mu_{ij} = g^{-1}\left(\beta_i - \beta_j + \kappa\delta_{ij}\right)$$

where κ is the home ground effect and $\delta_{ij} = 1$ if team i is the home team, $\delta_{ij} = -1$ if team j is the home team and $\delta_{ij} = 0$ if the match is played at a neutral site. Table 6 shows the ranking scores from the heteroskedastic beta regression model including this home ground advantage factor.

Table 6: ODI Team Rankings using Heteroskedastic Model with Home Advantage (as of 1/1/2008)

TEAM	RANKING SCORE	TEAM	RANKING SCORE	TEAM	RANKING SCORE
AUSTRALIA	125.5	INDIA	107.5	BANGLADESH	83.2
SOUTH AFRICA	118.0	ENGLAND	106.6	IRELAND	80.3
NEW ZEALAND	117.0	PAKISTAN	105.6	ZIMBABWE	73.3
SRI LANKA	114.4	WEST INDIES	105.3	KENYA	64.8

The estimated home ground advantage factor is κ=0.101 (which is statistically significant, p=0.002). Clearly, teams are aided by playing in familiar surroundings, and this ought to be accounted for when estimating team strengths. For example, India is far more successful at home, and thus their ranking score under this new model has decreased. Other effects of interest are readily examined in a similar way; such as, factors for the potential effects of batting first, as in Clarke and Allsopp (2001), or winning the initial coin toss.

As noted previously, one crucial aspect of a sensible ranking system is that it is not overly affected by a few very large victories. In other words, there should be something sacrosanct about a victory, so that the effect on ranking of the scoring play which achieved victory is larger than for any other scoring play. To a certain extent, the heteroskedastic structure provides this aspect. However, it may be the case that more is needed. One way of including such an aspect into the model is the use of modified or penalised likelihood, which could give increased likelihood to the larger parameters even in the case of a relatively narrow victory.

Finally, in closing, a comparison is made between the rankings developed here and the official ICC ODI rankings, based on a method developed by David Kendix. Details of the Kendix method can be found on the ICC's web-site (www.icc-cricket.com); essentially, though, the method is based on awarding "ranking points" to each team involved in a given match, and then creating a standings table based on teams' average ranking points per match. Relative team strengths are incorporated in the Kendix method by allowing the ranking points awarded for any match to depend on the current ranking of the opponents, so that defeating a lower ranked opponent provides fewer ranking points than defeating a higher ranked opponent. The weighting of matches based on age is accomplished by giving full weight to any match which occurs between the ranking date and the preceding August 1st, half weight to any match which occurs in the 12 months prior to the preceding August 1st and one-quarter weight to matches occurring in the 12 months prior to that. The Kendix method does not account for margins of victory in its ranking procedure. The major benefit of the Kendix method is its relative simplicity. Its calculation scheme makes it simple to see how the outcome of any match will affect the rankings. By comparison, the method developed here makes the effect of the outcome of a single match less obvious (though, simply fitting the model with and without the result of the match in question will clearly indicate its ultimate effect on the rankings). Moreover, for the Kendix method, the result of any match only affects the ranking points of the two teams involved, whereas the regression-based rankings allow the outcome of any match to potentially affect the entire ranking table, as it fully incorporates the "common opponent" information that each result entails. Also, the smoothly varying weightings provided by the Weibull survival function structure avoids the discontinuity associated with the method employed by the Kendix rankings, which may lead to notable shifts in team rankings each August. The ICC ODI ranking table at the end of the 2007 calendar year is given in Table 7 below.

Table 7: Official ICC ODI Team Rankings (as of 1/1/2008)

TEAM	RANKING SCORE	TEAM	RANKING SCORE	TEAM	RANKING SCORE
AUSTRALIA	130	SRI LANKA	108	BANGLADESH	47
SOUTH AFRICA	124	PAKISTAN	107	IRELAND	28
NEW ZEALAND	112	ENGLAND	107	ZIMBABWE	20
INDIA	110	WEST INDIES	100	KENYA	0

The team ordering is in close agreement with the results presented here. However, the gaps between the ranking scores are noticeably different. New Zealand is nearer South Africa in the rankings presented here and Ireland is nearer Bangladesh, and both are nearer the West Indies. The difference in rankings at the lower end of the table is partly a reflection of the fact that these teams play less frequently and the Kendix method uses only matches which are no more than three years old. In addition, investigation of the effect of individual matches indicates that the Kendix method is more volatile in its ranking scores, and the outcome of a few matches can make dramatic changes to the ranking scores, which is ironic given that margins of victory were not included in the method in part because of their perceived potential for just such an effect.

Acknowledgements

This work owes a great deal to an ongoing discourse with Frank Duckworth and Tony Lewis. I gratefully acknowledge their willingness to continue that discourse and provide considered advice on my ideas.

References

Clarke, S.R. and Allsopp, P. (2001) Fair measures of performance – the world cup of cricket. *Journal of the Operational Research Society*. **52:** 471-479.

Clarke, S.R. and Allsopp, P. (2002) 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*. **53:** 1160-1161.

de Silva, B.M., Pond, G. and Swartz, T. (2001) Estimation of the magnitude of victory in one-day cricket. *Australian and New Zealand Journal of Statistics*. **43:** 259-268.

Duckworth, F.C. and Lewis, A.J. (1998) A fair method of resetting the target in interrupted one-day cricket matches. *Journal of the Operational Research Society*. **49:** 220-227.

Duckworth, F.C. and Lewis, A.J. (2002) Comment on Clarke and Allsopp (2001) Fair measures of performance – the world cup of cricket. *Journal of the Operational Research Society*. **53:** 1159-1160.

Duckworth, F.C. and Lewis, A.J. (2004) A successful operational research intervention in one-day cricket. *Journal of the Operational Research Society.* **55:** 749-759.

Ferrari, S. and Cribari-Neto, F. (2004) Beta regression for modeling rates and proportions. *Journal of Applied Statistics.* **31:** 799-815.

Smithson, M. and Verkuilen, J. (2006) A better lemon squeezer? Maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods.* **11:** 54-71.

PREDICTING THE NUMBER OF RUNS SCORED PER OVER IN ONE-DAY INTERNATIONAL CRICKET MATCHES

Bailey, Michael [1] and Clarke, Stephen [2]

[1] Department of Epidemiology & Preventive Medicine, Monash University, Melbourne, Australia

[2] Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia

Paper Submitted for Review: 21February 2008
Revision submitted and accepted: 16 May 2008

Abstract. One-Day International cricket (ODI) is a popular sport worldwide. The advent of the Internet has increased opportunities for punters to wager on differing outcomes associated with each match, with one outcome of interest being the number of runs scored per over. Using information gathered from 627 past ODI matches (55000 overs), specific 'within match' and 'between match' variables were identified and explored to determine the statistical significance of their relationship with the number of runs scored. Separate models were constructed for both first and second innings with ten first-order variables and four second-order variables found to be highly significantly related to runs per over in both innings with a p-value less than 0.0001. These variables include the number of overs bowled, the stage of the game, the number of wickets fallen, recent wickets fallen, the magnitude of the existing partnership, current run rates, host country, team batting strength and opposition bowling strength. A multivariate linear regression model was used to combine and weight the relative importance of all variables with goodness of fit determined using the absolute average error between predicted and actual results, the maximum log-likelihood and an R-square statistic. This process serves to ascertain and rank the relative importance of various predictors and shed some light on the predictability of runs per over in ODI cricket.

Keywords: Predictors of runs scored, ODI Cricket, Linear Regression

INTRODUCTION

The first official one-day international (ODI) match was played in 1971 between Australia and England at the Melbourne Cricket Ground. Whilst ODI cricket has developed over the past 37 years, the general principles have remained the same. Both sides bat once for a limited time (maximum 50 overs) with the aim in the first innings to score as many runs as possible, and in the second innings to score more than the target set in the first innings.

As an international sport that has been in existence for over 100 years, it is of little surprise that much has been written on the topic of cricket. Clarke (1998) contains a summary of the work done on cricket up to that time, including that on distribution of scores. The first published work on cricket by Elderton and Elderton (1909) was also one of the first published bodies of work on statistical methods. Wood (1941, 1945) and Elderton (1945) investigated if the distribution of batsman scores in test cricket was geometric. They found an excess of very small and very large scores. Reep et al. (1971) and Pollard et al. (1977) explored whether a negative binomial distribution may in fact be a more appropriate fit to the distribution of batsmen scores. Kimber and Hansford (1993) found that apart from a greater risk in the early part of the innings, the chance of dismissal is reasonably constant throughout the innings.

Statisticians have more recently turned their attention to one-day cricket. Using batting records of only a few batsmen, Clarke (1991) found that the observed frequency of very small and very large scores were actually less in ODI cricket than expected by a geometric distribution. Increasing availability of data has resulted in several studies for different purposes. With a view to developing fair rain interruption rules, Duckworth and Lewis (1998, 2004) investigate the expected team scores in the remainder of a ODI innings as a function of overs left and wickets remaining. Alternative approaches to the same problem are taken by de Silva et al. (2001) and Gray and Le (2002). With the aim of forecasting individual batsman's scores, Bailey and Clarke (2004) found that the fielding restrictions and time constraints of ODI matches combined to ensure that the distribution of batsmen scores in ODI cricket could be reasonably well approximated by a log-Normal distribution, whilst comparisons made between batsmen could still be achieved using a geometric approach. Whilst much has been done to investigate the scores made by batsmen and teams there is no work to date relating to the prediction of runs scored per over (RPO).

BACKGROUND

Prior to February 2004, 2100 official ODI matches had been played between 20 competing countries. Although match results and player information is available for all matches played, information at an individual over level has only become available in recent years, and has been collected for 627 matches, creating a database of 55,000 overs. Although the actual number of runs scored per over clearly does not follow a Normal distribution (see Figure 1A), the enormity of the database used allows for some practical benefit to be gained by using a parametric approach to determine prediction variables.

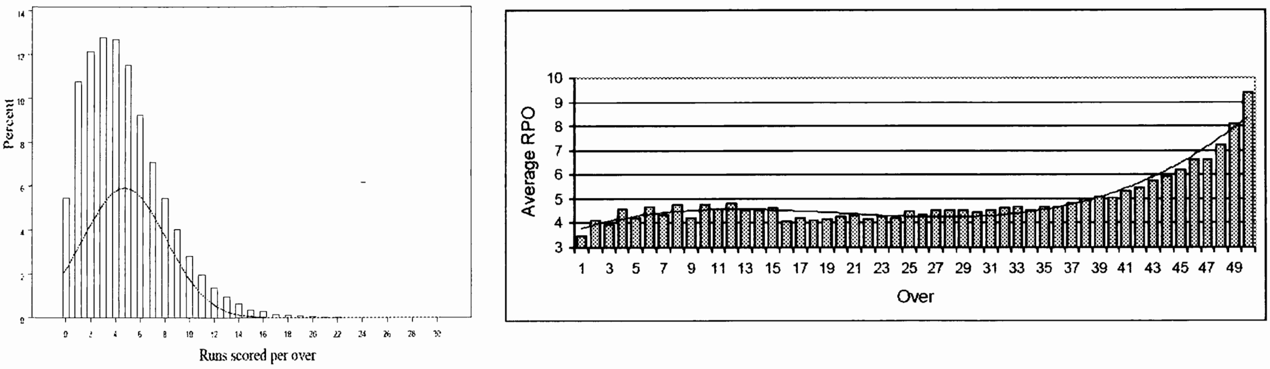


Figure 1A and 1B: Histograms of runs and plot of average runs scored for each over

PREDICTIVE FACTORS FOR RUNS PER OVER (RPO)

Prediction variables can be readily divided into 'within match' and 'between match' variables. Not surprisingly, significantly more variation occurs within games than between games. Within game factors that impact on RPO include innings number, over number, wickets fallen, duration of current partnership, scoring rate for the previous five overs, wicket fallen in the last over, run rate from the same end and best bowler. Between match variables include the location of the ground and the quality of both the competing teams.

Innings Number

More runs are scored per over by the team batting first than by the team batting second (1^{st} inning 4.80 ± 0.02 vs. 2^{nd} inning 4.66 ± 0.02, $p<0.0001$). This may not necessarily mean that runs are easier to score in the first innings. The constraints imposed upon batsmen in the second inning are different to that of the first. In the first inning a batsmen has the aim to score as many runs as possible, whereas in the second innings, batsmen face a specific target to win. There exists a trade off in ODI cricket between risk and reward, whereby to score more runs generally requires the batsmen to increase the risk of losing his wicket. In the second innings, if a target is relatively small, a batsman may opt to score at a slower rate thus reducing risk. The impact of having a target to chase (2^{nd} innings) in comparison to maximising the number of runs scored (1^{st} innings) has a dramatic impact on the predictability of RPO. Over twice as much variation can be explained in the first innings compared to the second, suggesting that external factors increase the variability associated with the second inning. Using linear regression, significant interactions could be found between the batting sequence and several other prediction variables, suggesting the need to model the first and second innings separately.

Overs

A linear predictor applied to RPO would suggest that in the first innings of a game, the expected number of RPO would increase at a rate 0.057±0.001 for each additional over. Interestingly, a linear predictor applied to the second innings would suggest that the expected number of RPO would increase at a rate of only 0.024±0.001 runs for each additional over. In reality, the effect of the new ball and fielding constraints ensure that a linear relationship between runs and overs is overly simplistic.

From Figure 1B it could be hypothesised that the relationship between overs bowled and runs scored during the course of a match follows a polynomial distribution with three degrees. Although perhaps a little simplistic there is merit in assuming that the course of ODI innings goes through three distinct stages. With fielding restrictions in place for the first 15 overs[2], a clear distinction can be seen between the fifteenth and sixteenth overs. Further scrutiny reveals that the scoring rate increases more dramatically as the match approaches its conclusion, suggesting a need to identify the turning point in which teams begin to accelerate the scoring rate towards the end of the match. To do this, a series of indicator variables were created, categorising the data as either above or below a given cut-off of overs. By then maximising the likelihood, the cut-off that produces the best fit to the data was identified. By considering a generalised linear model of the following form

$$Runs = A+B(First15)+C(Overs)+D(first15*Overs)+E(Cutoff) +F(Cutoff*Overs) \quad (1)$$

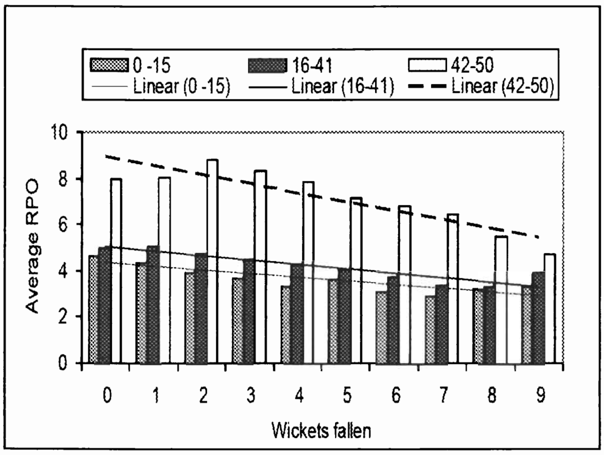
each match could be divided into three sections, with a linear model fitted to each section. A cut-off at the 41 over mark produced the best fit to both the first and second innings. Thus each innings could now be viewed as having three distinct phases, start (overs 1-15), middle (overs 16-41) and end (overs 42-50).

Wickets Fallen

Cricket is played between two teams of 11 players – of which two are required to be at the batting crease at any given time. As a general rule, better players bat higher in the batting order so as to optimise the time available from which to score runs, although players batting down the order can often score at a faster rate. Not only is each batting team constrained by the maximum number of overs that they can receive, they are also limited by the number of batsmen that they have available. Theoretically, one could expect that as wickets fall, so too would the scoring rate.

Using a generalised linear model, with wickets fallen as a continuous variable, the fall of each wicket was found to have significantly more impact towards the end of a game with each additional wicket reducing the average run rate by about a half a run per over. There was no significant difference between the rate of decline between the start and middle stages of the game. This is reflected in Figures 3A and 3B, with the relationship between wickets fallen and RPO differing significantly between the three stages of the first inning. (start: -0.13±0.03 runs per wicket fallen, middle: -0.19±0.03, end: -0.59±0.03).

[2] See Discussion section for information concerning 'Power Plays'



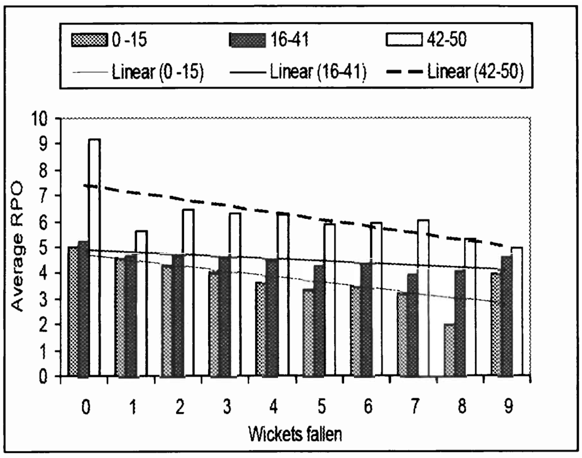


Figure 3A and 3B: RPO for wickets fallen for the three stages of the first inning and second innings respectively

From Figure 3B it can be seen that in the second innings, the fall of a wicket will reduce the run rate by the greatest amount during the start of the innings (start: -0.18±0.04, middle: -0.06±0.05, end: -0.12±0.04).

Wicket last over

Factors that occur within the game can have enormous effect on RPO. A good example is if a wicket has fallen in the previous over. As seen in Bailey and Clarke (2004), a study of all ODI batting performances revealed that as in test cricket, batsmen are most vulnerable when they first come to the crease. This reflects a brief "training" period where batsmen adjust to the conditions and the way the opposition are bowling. If a wicket has fallen in the previous over, a team will score on average one run less in the following over in the first innings and 1.3 runs less in the second innings (1st innings 0.98±0.05 vs. 2nd innings 1.32±0.06, $p<0.0001$).

Partnerships

Another strong predictor for RPO is the magnitude of the partnership for the batsmen at the crease. This equates to the number of runs scored since the fall of the last wicket. It is generally accepted that the longer a batsman spends at the crease, the more comfortable he becomes, thus making it easier to score runs.

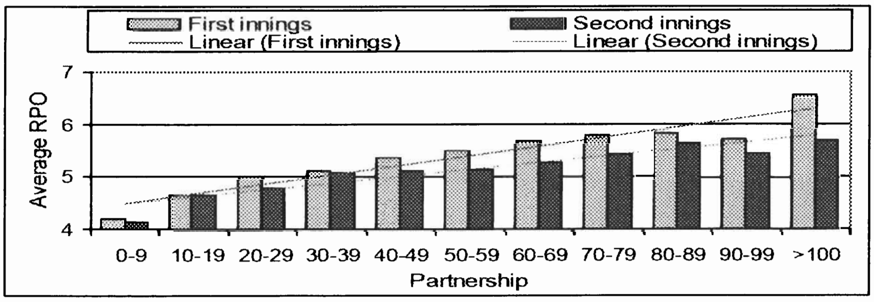


Figure 4: Relationship between partnership and runs per over

From Figure 4 it is possible to see a clear linear trend between partnership and RPO. Although partnership appears to have a slightly stronger relationship with RPO in the first innings than in the second, this difference was not statistically significant. Figure 4 shows the greatest difference between innings occurs for partnerships in excess of 100 runs. This is not a surprising result, as in the second innings batsmen are chasing a specific target and are thus not required to maximise the scoring rate to its optimal potential. From the multivariate models, the parameter estimate for partnership in the first innings is equivalent to 0.006±0.001 RPO whilst for the second innings the parameter estimate is equivalent to 0.005±0.001 RPO.

Runs previously scored in the match

Significant auto-correlation exists between consecutive overs in ODI matches. Because bowlers will often bowl consecutive overs from the same end, it is of no surprise that runs scored in the previous over from the same end provide an even better predictor of RPO. By measuring the Average Absolute Error (AAE) between actual runs scored and predicted runs scored, it is possible to compare the explanatory capacity of various within game predictors. From Figure 5 it can be seen that the more information that can be used from within the match, the lower the AAE, thus the better the prediction.

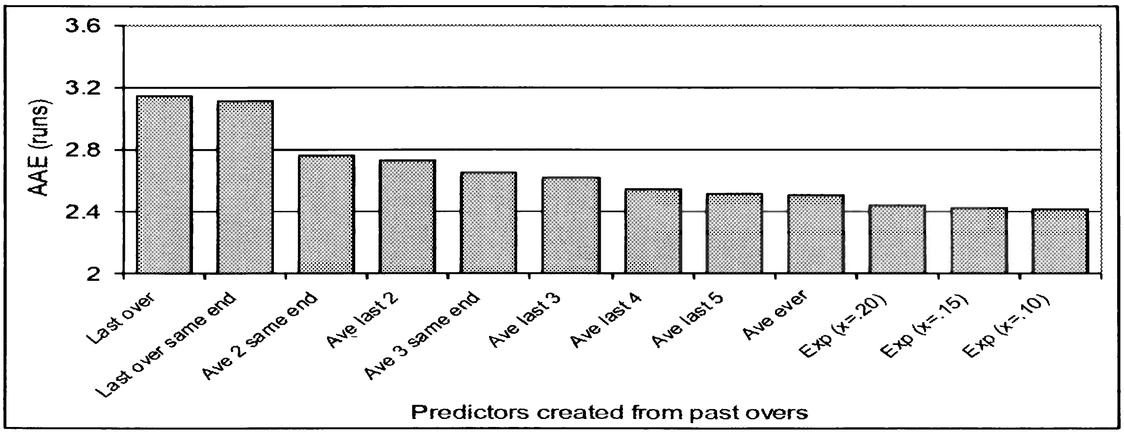


Figure 5: AAE for predictors created from previous overs bowled

Arithmetic averages weight each over equally, irrespective of when it occurred. An alternative approach is to give more weight to more recent overs by exponentially smoothing past results. It is possible to derive an unbiased prediction for future data by using the formula

$$\textit{Smoothed score} = \alpha \textit{ actual runs} + (1-\alpha) \textit{ Previous smoothed score} \quad (2)$$

where α is the smoothing parameter. Both rolling averages and exponentially smoothed predictors were compared. Whilst Figure 5 would suggest that exponential smoothing produced the lowest AAE, the run rate for the previous five overs and the average runs scored from the last three overs from the same end were also highly significant predictors in the multivariate model.

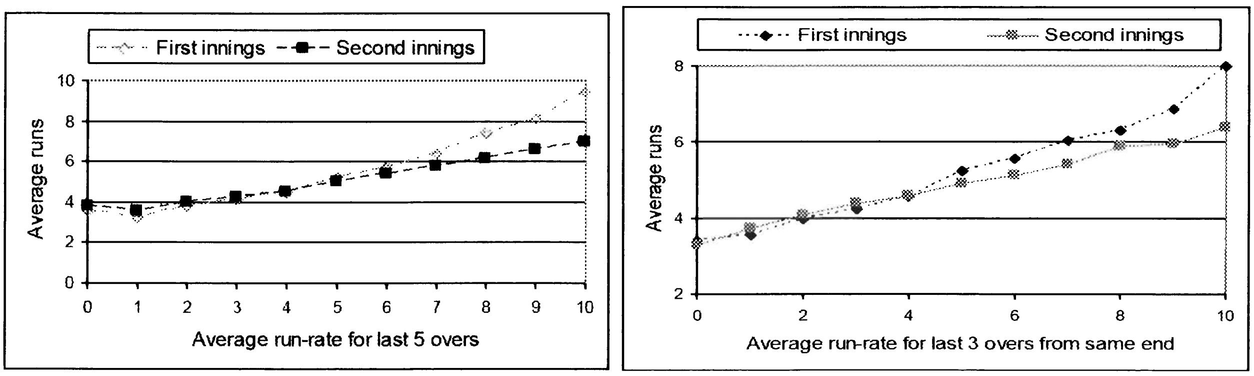


Figure 6A and 6B: RPO vs run-rate for the previous five overs and run-rate for the last 3 overs from the same end

For every one run increase in the run rate for the previous five overs, RPO was found to increase in both the first and second innings at a rate of 0.15±0.03. While Figure 6A might suggest the rate to be slightly different between the first and the second innings, this was not statistically significant. Similarly, for every one run increase in run rate from last three overs from the same end, RPO was found to increase at a rate of 0.11±0.02. There was no significant interaction with innings for either of these two predictors (See Figure 6B).

Best Bowler

A closer examination of the first 15 overs reveals some interesting trends. At the commencement of each game, the bowling side is given the use of a new cricket ball. This new ball combined with a joint training effect experienced by both batsmen as they "get their eye in" ensures that significantly less runs are scored on average in the first over than all others (first over 3.44±0.08 vs. all other 4.76±0.01, $p<0.0001$). When considering first inning performance versus second inning performance, the first over effect is greater in the first innings (first over 3.20±0.11 vs. all other 4.83±0.02) than in the second, (first over 3.69±0.11 vs. all others 4.68±0.02), with an interaction term between innings and first over statistically significant ($p<0.0001$).

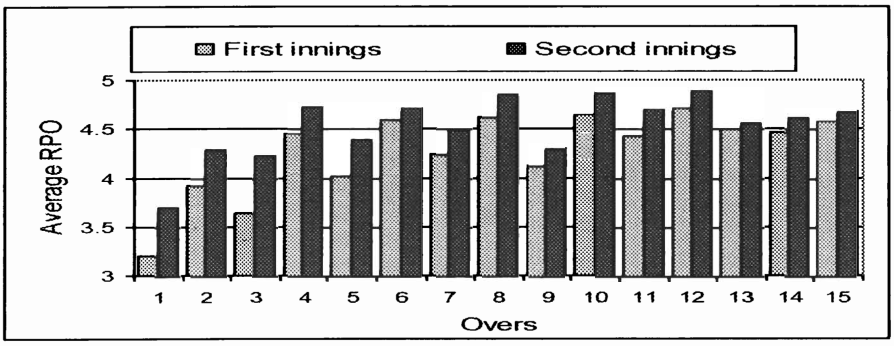


Figure 7: Average runs scored in the first fifteen overs

It is normal for the best fast bowler in the side to have first use of the new ball. Each bowler can bowl up to 10 overs per game, but it is unusual for an opening bowler to bowl all 10 overs consecutively. Although dependent upon the performance of the bowler, it would be usual for the best fast bowler in the side to have a spell of bowling that would last about 5-7 overs. This means that it is not unrealistic to expect that the best fast bowler in each side would bowl the first five odd numbered overs (1,3,5,7 & 9) and the second best fast bowler would bowl the first five even numbered overs (2,4,6,8 &10). The fact that bowlers alternate ends can clearly be seen from Figure 7 with a significant difference existing between the best and second best fast bowlers from each country. This result can be further confirmed by averaging runs scored for the first five odd numbered overs in comparison to the first five even numbered overs (odd 4.04±0.04 RPO vs. even 4.57±0.04 RPO $p<0.0001$). To account for the best bowler from each country, an indicator variable was created to identify the first five odd numbered overs from each innings. There was no significant interaction between the best bowler and innings.

Host Country

Twenty different host countries are represented in the database, although only 12 host countries have more than 1000 overs of available information. From these 12 countries, runs are primarily scored at the greatest rate in sub-continental countries, where pitches are perceived to be more batsmen friendly.

Team

Significant differences can be seen in the relative strength of the batting teams of competing nations. Interestingly, there is a strong correlation between the average number of runs scored per team and the number of overs each team has faced over the past six years. From Figure 8 a clear distinction can be drawn between the more established cricketing nations and those who play international cricket on a less frequent basis. This may also reflect the fact that the weaker cricketing countries are less likely to bat and bowl for their full 50 overs. Australian batsmen have scored at the fastest rate in the past six years, followed by India and South African batsmen. Second tier teams such Scotland, Canada and the Netherlands, score at the lowest run-rate and have faced the fewest number of overs at an international level in the last six years.

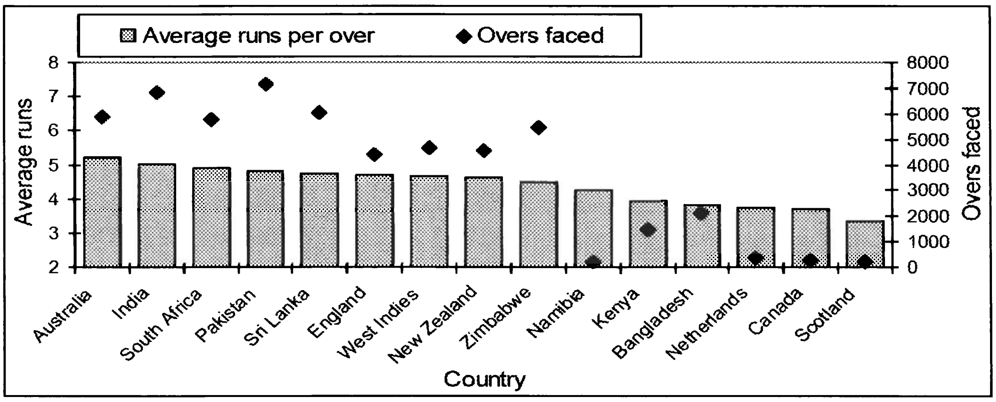


Figure 8: RPO and overs faced for the batting team

Team by stage

In addition to differences that exist between scoring rates for different countries, there are also significant differences that exist between countries for each of the three stages in the match. This may well reflect a difference in coaching strategies with some countries such as Australia, opting to follow the advice of Clarke (1988) and score faster earlier in the match, whilst countries such as Pakistan may have a preference to be more conservative earlier in the game, but score faster in the later stages of the match. An alternative explanation could be that Australia has faster scoring batsmen at the start of their batting order, whilst Pakistan has faster scoring batsmen lower down their batting order.

Opposition by Stage

An important contributing factor to the number of runs scored per over is the quality of the opposition bowlers and fielders. From the 627 matches played, there were 15 countries represented, with 11 of these countries having more than 1000 overs of data available. Overall, South African and Australian bowlers have been the hardest to score runs from, although different countries appear to perform better at differing stages of the match. Once again, this may well reflect the coaching strategies adopted by each country, or may simply be a legacy of team composition.

Multivariate linear models

A parametric approach to exploring variation associated with RPO enables the use of generalised linear modelling to determine the statistical significance of potential predictive variables. An additional benefit of the linear approach is that it allows practical comparisons to be made through the use of the R-square statistic. 18.4% of the variation in first innings score can be explained by a multivariate model, whereas only 7.5% of the variation in the second innings could be explained. This confirms the need to model innings separately. Multivariate models were constructed using stepwise selection and backwards elimination procedures before undergoing validation for plausibility using standard diagnostic procedures.

Ten first-order variables and four second-order variables were found to be significantly related to RPO in both innings with a p-value less than 0.0001. The corresponding reduction in AAE and -2 log likelihood for each stage of development can be seen in Table 1.

Table 1: Goodness of fit for stages of model development for both innings

Stage	Variable	First innings			Second innings		
		AAE	-2LL	R^2	AAE	-2LL	R^2
1	Over Number	2.57±.01	157200	5.7%	2.58±.01	132233	0.9%
2	+ Stage	2.53±.01	156187	8.9%	2.57±.01	132012	1.8%
3	+Over X Stage	2.51±.01	155712	10.3%	2.57±.01	131997	1.9%
4	+Wicket	2.46±.01	154204	14.1%	2.54±.01	128976	3.3%
5	+Wicket X Stage	2.44±.01	153911	14.9%	2.54±.01	128950	3.4%
6	+Partnership	2.43±.01	153652	15.7%	2.53±.01	128732	4.3%
7	+Run rate last 5	2.42±.01	153235	16.7%	2.51±.01	128310	5.7%
8	+Wicket last over	2.41±.01	153144	17.0%	2.51±.01	128272	5.9%
9	+Ave. last 3 same end	2.41±.01	152997	17.1%	2.50±.01	128076	6.2%
10	+Best Bowler	2.41±.01	152984	17.2%	2.50±.01	128071	6.2%
11	+Team	2.41±.01	152900	17.5%	2.50±.01	128040	6.4%
12	+Home Country	2.40±.01	152854	17.7%	2.49±.01	127999	6.7%
13	+Opponent X Stage	2.40±.01	152744	18.1%	2.49±.01	127935	7.0%
14	+Team X Stage	2.39±.01	152693	18.4%	2.48±.01	127851	7.5%

Key: AAE = Average absolute error, -2LL = -2 * log likelihood
R^2 = Percentage of variation explained by model

Although the host country and the quality of both teams can be shown to be independently significant, their contributions towards improving the model are quite small. Whilst parameter estimates differ significantly, the key components for the first innings model, namely, overs, wickets, partnerships and run rates are also the most important predictors for the second inning models. By comparing the R-square statistic for 14 stages of model development between the two innings it can be seen that although the models develop in a similar fashion, far more unexplained variation exists in the second innings.

DISCUSSION

In July 2005 the International Cricket Council (ICC) announced two new rules for ODI matches. The first of these rules, the introduction of a super-sub, was found to bias against the team batting second and was subsequently dropped after 12 months. The second rule, an increase in field restrictions from 15 to 20 overs is still in operation. Previously, fielding restrictions were in place for the first 15 overs of each innings only, whereas now, fielding restrictions are mandatory for the first 10 overs only, with the fielding captain now responsible for the timing of two additional blocks of five overs referred to as 'Power Plays.' Whilst this new rule has served to increase the average number of runs scored per over (4.98±0.02 vs 4.78±0.01 $p<0.0001$) the fact that the first Power Play has been taken from overs 11-15 in over 90% of matches ensures that the three stage models incorporated in this paper are still applicable. With this in mind, future models should seek to further explore the effect of Power Plays.

CONCLUSION

Using past data it is possible to identify features of the match that can aid in the prediction of the number of runs scored per over. The use of mathematical models to predict RPO during the course of the match opens new doors for bookmakers. The speed and objectivity of a mathematical approach supersedes traditional bookmaking methods. A mathematical model, incorporating simple past features of the game enables a dynamic price setting process and creates a much greater scope of betting opportunities for the punter.

References

Bailey, M. J. and Clarke, S. R. (2004) Market inefficiencies in player head to head betting on the 2003 cricket world cup. In S. Butenko, J. Gil-Lafuente and P. Pardalos (eds.) *Economics, Management and Optimization in sport*. Heidelberg, Springer-Verlag, 185-201.

Clarke, S. R. (1988) Dynamic programming in one-day cricket - optimal scoring rates. *Journal of the Operational Research Society* **39(4):** 331-337.

Clarke, S. R. (1991) Consistency in sport - with particular reference to cricket. *Proceedings 27th NZ OR Conference. Wellington, New Zealand Operations Research Society,* 30-35.

Clarke, S. R. (1998) Test Statistics. In J. Bennett (ed.) *Statistics in Sport*. London, Arnold, 83-101.

de Silva, B. M., Pond G. R. and Swartz T. B. (2001) Estimation of the magnitude of victory in one-day cricket. *The Australian & New Zealand Journal of Statistics* **43(3):** 259-268.

Duckworth, F. and Lewis, T. (1998) A fair method for resetting the target in interrupted one-day cricket matches. *Journal of the Operational Research Society* **49:** 220-227.

Duckworth, F. C. and Lewis, A. J. (2004) A successful operational research intervention in one-day cricket. *Journal of the Operational Research Society* **55:** 749-759.

Elderton, W. E. (1945) Cricket scores and some skew correlation distributions. *Journal of the Royal Statistical Society (Series A)* **108:** 1-11.

Elderton, W. P. and Elderton, E. M. (1909) *Primer of Statistics*. London, Black.

Gray, S. and Le, T. A. (2002) How to fix a one-day international cricket match. In G. Cohen and T. Langtry (eds.) *Proceedings of the sixth Australian conference on mathematics and computers in sport*. University of Technology, Sydney: Gold Coast. pp. 14-31.

Kimber, A. C. and Hansford, A. R. (1993) A statistical analysis of batting in cricket. *Journal of the Royal Statistical Society (Series A)* **156:** 443-445.

Pollard, R. and Benjamin, B. (1977) Sport and the negative binomial distribution. In S. P. Ladany and R. E. Machol (eds.) *Optimal Strategies in Sports*. Amsterdam, North Holland: 188-195.

Reep, C., Pollard, R. and Benjamin, B. (1971) Skill and chance in ball games. *Journal of the Royal Statistical Society (Series A)* **134:** 623-629.

Wood, G. H. (1941) What do we mean by consistency? *The Cricketer Annual*: 22-28.

Wood, G. H. (1945) Cricket scores and geometrical progression. *Journal of the Royal Statistical Society (Series A)* **108:** 12-22.

A COMPARISON OF DISTRIBUTIONS FOR THE RUNS SCORED PER OVER IN ONE-DAY INTERNATIONAL CRICKET MATCHES

Bailey, Michael [1] and Clarke, Stephen [2]

[1] Department of Epidemiology & Preventive Medicine, Monash University, Melbourne, Australia

[2] Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia

Paper Submitted for Review: 21 February 2008
Revision submitted and accepted: 16 May 2008

Abstract. The distribution of runs per over (RPO) scored in one-day international cricket does not follow a known distribution. Bailey and Clarke (2008) used multiple linear regression to identify and weight 14 highly significant variables found to be independently predictive of RPO. This paper expands on this work by comparing a range of possible distributions for modeling RPO using both standard and novel approaches. Predictive capacity was determined by developing parameter estimates using 75% of the available 55000 overs and applying the subsequent prediction models to the remaining 25% of the data. Goodness of fit was determined by averaging the log of the predicted probabilities for the actual runs scored with the model producing the highest average indicative of the best fit to the data. Of the 10 approaches examined, a slivered binomial approach appeared to produce the best fit to the data, although due to the increased number of models required, considerable care is needed to ensure that the data is not over-fitted. Both the ordinal logistic regression and the negative binomial approach produced good fits to the data and were more simplistic in implementation. Although not significantly different from each other, these three approaches were significantly better than the remaining seven models that were considered ($p<0.0001$).

Keywords: Runs Per Over, ODI cricket, Comparison of distributions

INTRODUCTION

The advent of the internet has dramatically increased world-wide exposure to cricket and has paved the way for a wealth of betting opportunities. In particular, punters can now wager on the number of runs per over (RPO). Whilst one-day international (ODI) cricket matches have been played internationally for 37 years, the literature relating to ODI cricket is sparse. Duckworth and Lewis (1998) have investigated expected team innings scores to develop an unbiased approach to resolving ODI matches when rain delays occur. Similarly, Bailey and Clarke (2004) have compared distributions for batsman scores in ODI cricket. Neither have modeled data at an individual over level.

Each over in ODI cricket comprises of six legitimate deliveries, with the batsmen capable of scoring anywhere from zero to six runs from each delivery. Given that the probability for each run outcome differs markedly, it is of little surprise that the number of runs scored per over will not follow a unique distribution. Bailey and Clarke (2008) applied a multiple linear regression to 627 ODI matches to identify 'within' and 'between' match features that could aid in the prediction of RPO. Whilst the magnitude of the database used (55000 overs) allows for some practical benefit to be gained, it remains that a basic assumption for the use of linear regression is that the underlying distribution of RPO is approximately normal. From Figure 1, it can be seen that this is not the case.

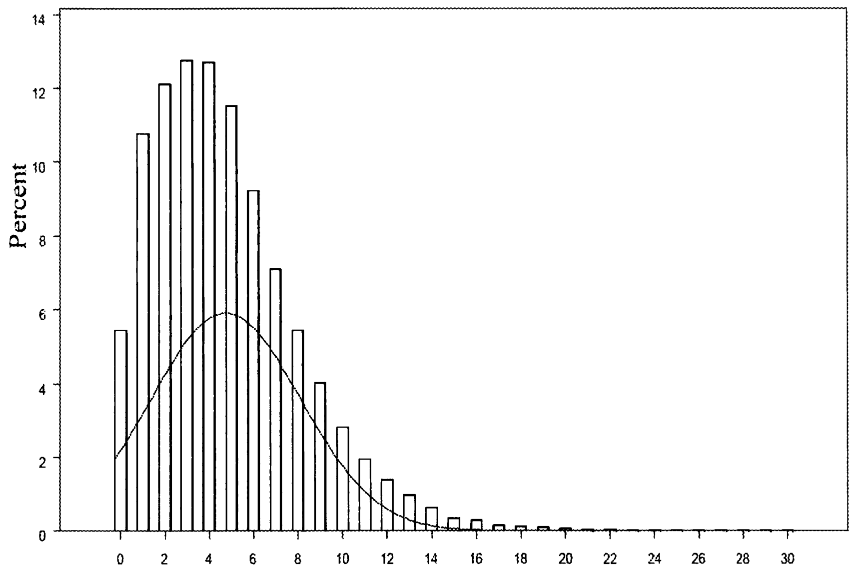


Figure 1: Histogram of runs scored per over in ODI cricket

This paper seeks to build on the work of Bailey and Clarke (2008), by utilising the same prediction variables that were identified by linear regression and applying them to a range of prediction approaches that consider differing underlying distributions.

METHODS

ODI cricket is played world wide with more that 2600 matches played. Although match and player information is available for all matches played, information at an individual over level has only become available in recent years. Over by over information was gathered from 627 matches played between 1998 and 2004. Although in theory, 100 overs could be bowled in each ODI, due to the nature of the game, this has only occurred about 12% of the time. There are various reasons why a game would not go for the full 100 overs. Rain delays, one or both of the sides being dismissed before using their full resource of 50 overs, the second side completing the required target within the 50 overs, or penalties imposed upon teams for slow play are all reasons why 100 overs would not be bowled. On average, 88 overs were bowled per match, creating a database of 55000 overs. Using multiple linear regression, ten first-order variables and four second-order variables were found to be significantly related to RPO in both innings with a p-value less than 0.0001. In order of importance these variables were:

1. Over number
2. Stage of match: start (overs 1-15), middle (overs 16-41) and end (overs 42-50)
3. Interaction (Over x Stage)
4. Wickets fallen
5. Interaction (Wickets x Stage)
6. Size of current batting partnership
7. Run rate for the last 5 overs
8. Wicket fell in last over
9. Average runs scored from the last 3 overs of the same end
10. Best bowler: the first five odd numbered overs in the innings
11. Team strength
12. Home Country
13. Interaction (Opponent x Stage)
14. Interaction (Team x Stage)

To determine what underlying distribution might best approximate RPO, five distributions namely Normal, Log-normal, Gamma, Poisson and Negative-binomial were modelled to the data using these 14 previously identified prediction variables. These five distributions were initially compared using the Absolute Average Error (AAE) between the predicted mean and the actual number of runs scored.

By collapsing all scores greater than nine into the same category, two additional modelling approaches were incorporated using logistic regression. Firstly, an ordinal logistic regression was applied to the 11 possible outcomes (0, 1, 2...9, >9). This model incorporates a different intercept term for each additional run, but has fixed parameter estimates. To further enable variation in the parameter estimates for each run category, a series of 10 binomial models were constructed, each with a different cut-off for success. The first model predicted whether the number of runs scored would be greater than or equal to one, whilst the second model predicted whether the number of runs would be greater than or equal to two, and so on. By subtracting sequential models the probability for each run category could be determined.

A more robust 'Reduced Binomial' model was also created by removing the four 'between match' variables (Team, Home country, Team X Stage and Opposition X Stage) from the training models. For further comparison, two additional approaches were incorporated, one that assigned an equal probability of 0.091 to the 11 run categories, and a final comparison that assigned the actual probabilities derived from the training data set.

To avoid the bias associated with over-fitting, predictive capacity must be assessed in a sample of data separate from that which parameter estimates were derived. This was achieved by developing parameter estimates from all data prior to 2003 and applying to matches played in 2003 and 2004 (169 matches). Goodness of fit was determined by averaging a function of the log of the predicted probabilities for the actual runs scored, with the model producing the highest average, indicative of the best fit to the data. This average is the information content of the predictions, and is related to the likelihood of the outcome, but gives a minimum of zero for the case when all possible outcomes are allocated equal probabilities. Statistical significance between models was determined using Wilcoxon rank sum tests.

RESULTS

Using generalised linear modelling, five multivariate models were constructed, each assuming a differing underlying distribution (Normal, Poisson, Gamma, Negative binomial and Log-normal). The 14 variables that were found to be highly significant ($p<0.0001$) using the Normal distribution model were applied to the other four distributions, and found to be equally significant ($p<0.0001$) regardless of what underlying distribution was fitted to the data.

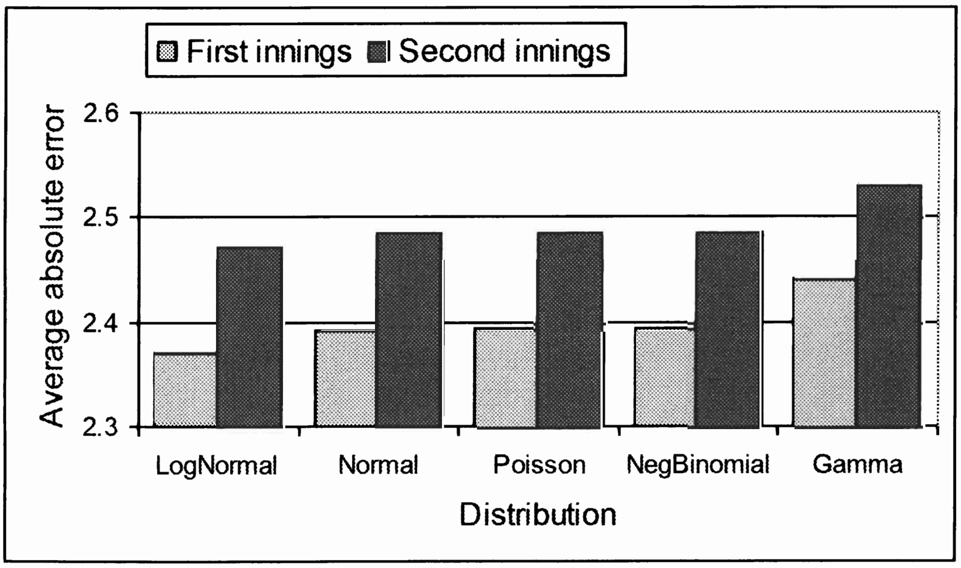


Figure 2: Comparison of AAE for the five distributions

The AAE between the predicted and actual scores was compared for the five distributions and can be seen in Figure 2. Clear differences exist between the first and second innings with the first innings being more than twice as predictable as the second innings. Overall, the log-normal approach was found to produce the lowest AAE for both the first and second innings.

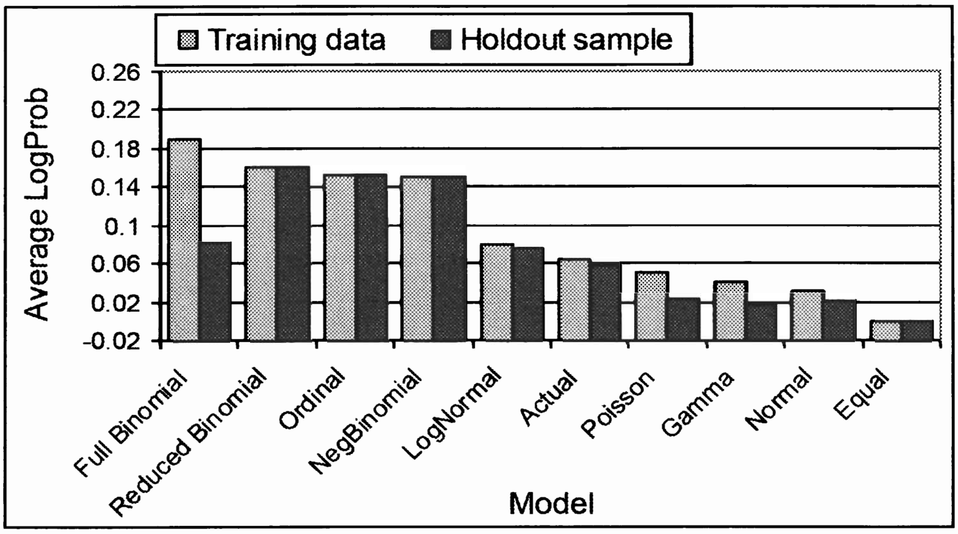


Figure 3: Comparison of average LogProb between models for first innings

Predictive capacity for first and second innings models can be seen respectively in Figures 3 & 4. Although the 14-parameter multivariate binomial approach was found to produce the best fit to the training data, the large difference between the training data and the holdout data in both innings indicated over-fitting. To alleviate the bias of over-fitting, the four 'between match' variables (Team, Home country, Team X Stage and Opposition X Stage) were removed, creating the 'Reduced Binomial'. The similarity between the average log of the probabilities in the training and holdout samples for the Reduced Binomial model, suggest that the source of bias due to over-fitting had been removed. From Figures 3 & 4 it is possible to see that the Reduced Binomial model appeared the best performed model in both the first and second innings. Although no statistically significant differences existed between the Reduced Binomial, Ordinal Logistic and Negative Binomial models, all three of these approaches were significantly better than all others ($p<0.0001$) for both first innings and second innings.

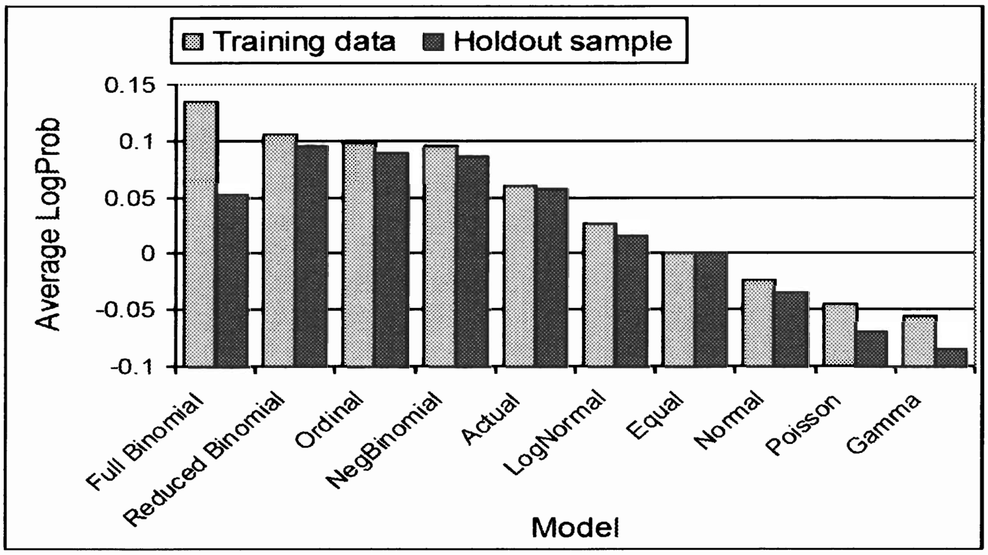


Figure 4: Comparison of average LogProb between models for the second innings

DISCUSSION

Whilst some loss in predictability between training and holdout samples is expected, with datasets for both first and second inning models in excess of 20000 data points it is realistic to expect that bias associated with over-fitting would be small. Interestingly, it was only the full-binomial model which had differing intercepts and differing parameter estimates for each run category that was significantly biased. The primary source of this bias could be traced to 'between match' variables such as team, country and opposition that do not remain consistent between training and holdout samples. Over fitting is best avoided by choosing highly significant variables with lower degrees of freedom.

CONCLUSION

The wealth of internet data available provides an excellent opportunity to build robust prediction models for RPO. Using past data it is possible to identify features of the match that can aid in the prediction of the number of runs scored per over. Whilst the use of linear regression provides an easy way to identify and rank predictors of RPO, when constructing prediction models for RPO, significantly better approaches can be employed. While the use of a binomial model with differing intercepts and differing parameter estimates provides the most accurate prediction approach, this was not significantly better that a binomial model with fixed parameter estimates, or a negative binomial model. In addition, the increased complexity of the slivered binomial approach dramatically increases the chance of over-fitting the data. While the bias of over-fitting can be alleviated by choosing a simple prediction model constructed from within match parameters only, the ease of use of the ordinal or negative binomial models suggest a greater usability.

The use of mathematical models to predict RPO during the course of the match opens new doors for bookmakers. The speed and objectivity of a mathematical approach supersedes traditional bookmaking methods. A mathematical model, incorporating simple past features of the game enables a dynamic price setting process and creates a much greater scope of betting opportunities for the punter.

References

Bailey, M. J. and Clarke, S. R. (2008) Predicting the number of runs scored per over in one-day international cricket matches. In J. Hammond (ed.) *Proceedings of the Ninth Australiasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM) Coolangatta, Queensland, Australia (in press).

Bailey, M. J. and Clarke, S. R. (2004) Market Inefficiencies in player head to head betting on the 2003 cricket world cup. In S. Butenko, J. Gil-Lafuente and P. Pardalos *(eds.) Economics, Management and Optimization in sport.* Heidelberg, Springer-Verlag: 185-201.

Duckworth, F. and Lewis, T. (1998) A fair method for resetting the target in interrupted one-day cricket matches. *Journal of the Operational Research Society* **49:** 220-227.

IMPROVED SCORING SYSTEMS FOR A CONTEST BETWEEN TWO TENNIS TEAMS

Lisle, Ian [1], Pollard, Geoff [2], and Pollard, Graham [1]

[1] Faculty of Information Sciences and Engineering, University of Canberra, Australia

[2] Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia

Paper Submitted for Review: 28 March 2008
Revision submitted and accepted: 11 June 2008

Abstract. There are several scoring system variations for contests between tennis teams in operation throughout the world. Here we analyse the major team tennis competition in Sydney ('Badge'), as it would appear that any improvements possible in it could be applicable elsewhere. In 'Badge', Team A and Team B each have two doubles pairs. Pair A1 plays two tie-break sets against pair B1 on court 1, whilst A2 plays two tie-break sets against B2 on court 2. When both of these matches are finished, A2 plays B1 on (say) court 1 whilst A1 plays B2 on court 2 to complete the fixture, making 8 tie-break sets in total. Advantage games are used, and the tie-break game is played at 5-5 (if necessary), not 6-6. The winner is the team that wins at least five sets, or if each team wins 4 sets, the winner is the team that wins the most number of games (on a 'countback'). Otherwise, the match is a draw (or is unfinished), unless one team must win despite losing all unplayed games. Note that considerable court time can be wasted if the first two sets finish very quickly on one court (e.g. 6-2, 6-1) relative to the other court (e.g. 6-4, 6-5). A maximum time of 2 hours and 30 minutes is allocated to each fixture. The aim of this study was to use simulation methods to analyse and report on whether scoring system changes such as no-ad games, '50-40' games, 'first-to-7-games' sets, drawn sets,... could be useful in decreasing court-time wasted at the cross-over, increasing the number of fixtures completed, and increasing the likelihood of the better team winning. An alternative scoring system with these properties was found. Similar approaches may be applicable to other sporting contests between teams, tennis in particular.

Keywords: improved scoring systems for team tennis, drawn sets and first to seven games sets of tennis.

INTRODUCTION

The major teams' tennis competition held across Sydney on Saturday afternoons is the 'Badge Competition' involving men's (and women's) doubles. There are 8 teams in each grade, and each team plays each other team twice, on a home-and-away basis. Each team has two doubles pairs. In an afternoon when team A plays team B, the pair A1 plays two tie-break sets against pair B1, and also plays two tie-break sets against pair B2. Correspondingly, pair A2 plays two tie-break sets against pair B2 and two tie-break sets against pair B1, making 8 tie-break sets in total. Advantage games are used throughout. The tie-break game is played at 5-5, not 6-6. This is very interesting as it is a rare case in which a scoring system that is not an option within the ITF Rules of Tennis 2008 is used in an important competition.

Prior to 2006, the order of play could be represented by A1 vs B1, A2 vs B2, A2 vs B1, and finally A1 vs B2, the matches being played sequentially on one court. Under this structure half of the players were off the court at any point in time. In order to reduce substantially the amount of time spent off the court by the players, a major change was made to this structure in 2006. Two courts are now used for approximately half the time previously taken. Thus, the order of play can now be represented by A1 plays B1 on court 1, whilst A2 plays B2 on court 2. When both of these matches are finished, A2 plays B1 on (say) court 1 and A1 plays B2 on (say) court 2 to complete the contest (i.e. there is a cross-over). The winner is the team that wins at least five sets; or if each team wins 4 sets, the winner is the team that wins the most number of games (on a 'countback'). If the teams win an equal number of sets and games, the match is a draw. If one team (say team A) has won (say) 4 sets, and the other team (team B) has won (say) 3 sets but cannot possibly win on a countback because they have not won enough games in total, then team A is the overall winner.

The new format was established so that the players spent much less time off the court. As two courts are now necessary for a fixture, there are two sessions for the competition during the afternoon: the early afternoon session from 12:00 to 14:30, and the later session from 14:30 to 17:00. The new format can work very well when the first two matches in a contest finish at about the same time. If this happens, very little time is wasted by the players. However, some court-time can be wasted (i.e. one of the two courts is not used) if the first two sets finish relatively quickly on one court (eg 6-2, 6-1), but the two sets take much longer on the other court (eg 6-4, 6-5). More fixtures might be completed if less court-time was wasted.

The recent changes lead to two questions.

1. Is there an alternative scoring system (to the present two tie-break sets with a tie-break game at 5-5) that would co-ordinate better the completion of the first two parallel matches, whilst keeping such things as the likelihoods of each team winning the match much the same as they are now?
2. Why play the tie-break game at 5-5, given there is often a real likelihood of delaying the starting time of the following two matches? Why not award half a set to each team when the score reaches 5-5? Any one set is presumably not of sufficient importance to require a winner to be determined.

The aims of any change to the scoring system presently being used would be to reduce as much as possible the court-time wasted at the cross-over, and to increase number of matches completed (and won by the better team) rather than drawn.

In addressing these questions and aims, we can make use of some recently published research by Pollard et al. (2007) on alternative scoring sytems for doubles. They studied the mathematical/statistical properties of the new doubles scoring system adopted by the ATP Tour in 2006, and compared this system with the system previously in use. They also studied the properties of five other alternative scoring systems for men's doubles, and concluded that 'all of the five alternative scoring systems would appear to be just as good as, or even a little better than the (recently adopted) scoring system, from a statistical point of view. They are all at least slightly more efficient, and have higher values for P, the probability that the better pair wins'. They also stated that 'the system using 'draw-sets' is perhaps the most interesting one. It has a reduced mean, a small standard deviation, a low 98% point, and a good efficiency with an improved value of P'.

The five scoring systems mentioned above made use of several scoring system structures including:

- '50-40 games' (and '60-50 games')
- 'first-to-seven-games sets' (and 'first-to-five-games sets'), which can be fairer than the first-to-six-games sets that are presently used (Pollard, 2005)
- 'draw-sets' that result in a win, a draw or a loss in the set to either pair, and
- 'first-to-nine-points tie-break games' (which can be fairer than the first-to-seven-points and first-to-ten-points tie-break games that are presently used (Pollard, 2005)).

Given the statistical characteristics reported in the above study for the four scoring system structures above, it would appear that the first three of them could be quite useful in addressing the situation in Sydney Badge doubles. For this reason these three structures are discussed in the next few paragraphs.

Firstly, the usefulness of the '50-40 game', particularly in men's doubles has been noted (Pollard & Noble, 2004). In a 50-40 game, the server is required to reach 50 (one more point than 40) before the receiver reaches 40, in order to win the game, whilst the receiver wins the game by reaching 40. Pollard & Noble (2004) noted that, for parameter values relevant to professional men's doubles, 'the longest best-of-three tie-break sets matches (as measured by the 98% point in the distribution of the number of points played) can be reduced by about 30 points by using '50-40 games' instead of no-ad games (an option within the ITF Rules of Tennis 2008 in which only one point is played if deuce is reached)'. They also noted that for the same parameter values 'the probability that the better player wins a best-of-three tie-break sets match using '50-40 games' is comparable to when no-ad games are used, even though about 20 less points are required on average in the match'. The classical or advantage games used in Badge clearly take longer than no-ad games, and so for the situation being considered in this study, there would appear to be a moderate-sized savings in points played by using 50-40 games. Also, it is noted that advantage games can sometimes cause quite long delays on (say) one of the two courts being used.

The results in the paper by Pollard et al. (2007) indicate that a 'first-to-five-games (tie-break game at 4-4) set' using '60-50 games' (a natural extension or variation of the 50-40 game) has about 2–3 less points on average than a standard tie-break set using no-ad games. Thus, the 60-50 game when used with 'first-to-five-games sets' and a drawn set at 4-4, might be a useful scoring system for the teams' situation being considered in this paper. Note that with such a scoring system, it is only necessary to change ends after games 2 and 6 in the set, and not after games 1, 3, 5 and 7, leading to a further savings in time.

Secondly, it has been noted that the present first-to-six-games sets can be unfair in doubles (Pollard, 2005). This unfairness results from the fact that one player in a doubles pair may have three service games in a set whilst his partner may have just two. This unfairness can be removed by playing (say) 'first-to-seven-games sets', which operate probabilistically like 'best-of-twelve-games sets' in which each player has three service games each. Alternatively, the unfairness can be removed by playing 'first-to-five-games sets', which operate probabilistically like 'best-of-eight-games sets' in which each player has just two service games each. The results in the paper by Pollard et al. (2007) indicate that 'first-to-seven-games (with tie-break game at 6-6) sets' when used with 50-40 games have about the same expected number of points as the standard tiebreak set when used with no-ad games. Thus, since at present classical or advantage games are used in Badge tennis, this 'first-to-seven games sets' system is quite possibly a reasonable alternative for this study, especially if the set is declared a draw at 6-6 creating some further decrease in expected duration (and a decrease in court-time wasted). Note that for such a system it is only necessary to change ends after games 2, 6 and 10, rather than after games 1, 3, 5, 7, 9 and 11.

Thirdly, the idea of declaring a set a draw if the games' score reaches 6-6 has been considered by Pollard and Noble (2003). Note that the rationale behind declaring a draw at 6-6 is that the tie-break game, with an expected duration of about 12 points, is too often lost by the better player; and this is after a considerable investment in time (say about 60 points) has been made in getting to 6-6. A loss in the tie-break game by the better player is not good for the efficiency of the scoring system in total or for its variance of duration. For a draw-set structure, the set score is incremented by 2, 1 or 0 for a win, draw or loss respectively. Pollard and Noble (2003) considered 'best-of-two-sets' matches with the winner of the match being the first player to reach a set score at least 3. If a set score of 2-2 was reached, a (long) tie-break game was played to determine the winner. Interestingly, they noted that 'when playing no-ad games and 'best-of-two sets' with a deciding match tie-break (7), it is (perhaps surprisingly) more efficient not to play the tie-break game at 6-6 (in each of the first two sets) than it is to play it'. We note here that declaring a set to be a draw at 6-6, although fair for the players, can be seen as an unattractive rule by those spectators who like the excitement of the tie-break game. Then again, for most non-professional matches (and even for some professional matches), and certainly in our present context, there are typically few or no spectators. Thus, a change to declaring a draw at 5-5 would appear to have its merits in our present considerations. Note that with the present scoring system used for Badge, the two sets between a two pairs are quite often drawn (1-1), so the concept of a drawn match in Badge is not new. In fact there could well be a reduction in the number of drawn matches with some of these 1-1 draws being replaced by 1.5 to 0.5 set wins to one team or the other. Further, a drawn set at 5-5 in a set would appear to be quite a reasonable outcome as it is the outcome of the overall contest that is of interest, rather than the outcome of any particular set or match. Further, playing a tie-break game at any stage of the contest often extends the duration of the contest and the time wasted.

METHOD

Aims of this study

The aims of this study were to identify a scoring system which, compared to the current Badge system:

1. Finishes the match fully, and/or gives a win/loss outcome for the fixture more often.
2. Wastes less court-time (i.e. at the cross-over, has a fewer expected number of points on one of the courts whilst the match has finished on the other court).
3. Plays more tennis in total.
4. Achieves a similar or higher value for the probability that the better team wins.

Perhaps it could be possible to find a system that was better on all of these four aims.

The data available were the results from rounds 15 to 21 of the 2007 Sydney Badge competition as published in a Sydney newspaper on the mornings following each round. Not all results of the previous day were available, as some did not meet the cutoff time for the newspaper. In total there were 420 results from all men's and women's grades each showing the number of sets and games won by each team, the winning team or whether there was a draw or no result. On average 69.2 games were played per match. Only 7 sets were completed 18.1 % of the time, and only 6 sets were completed 6.7 % of the time.

Simulation Study

A computer simulation program was devised which simulates individual sets, and then combines the sets into pairs of sets representing a single match between two doubles pairs. This aspect of the program was checked using convolution methods for the two sets played. Two such two-set matches are then considered in parallel. The theory of the distribution of the maximum, and the maximum minus the minimum of two random variables was used to check the program logic at the cross-over. Further, the program was checked by point-by-point monitoring of play to see that scoring and logging of match statistics were correct. The results were also checked against other published work and found to agree (Pollard, 1983; Pollard & Noble, 2004; Pollard et al., 2007; Brown et al., 2008). Although the program has the capacity to handle four different servers (and different rules as to who serves at the beginning of each set), only two different types of servers were considered in this study, as this was sufficient to study the issues under consideration. The authors believe that the software developed for this study is a useful resource for studying other scoring system issues for individuals and for teams, and even for some other sports.

For each of the scoring systems under consideration, simulations of 1,000,000 fixtures were carried out. This was a sufficient sample size to achieve appropriate accuracy. Team A was assumed to have serving point probabilies of pa, and team B was assumed to have serving point probabilities of pb. The team that served first at the beginning of every match was selected at random. Limiting the number N of points played on each court operated as a proxy for the court-time limit of 2½ hours.

The scoring systems structures considered

A range of scoring system structures was considered. It included four game types (advantage games, no-ad games, '50-40 games' and '60-50 games'); sets played as first-to-six, 'first-to-seven', and 'first-to-five games'; sets declared tied at 5-5 and at 6-6; tied sets resolved by tie-break games, or drawn sets (no tie-break).

RESULTS

An alternative system

Rather than report on the range of specific scoring systems that were equal or similar to the existing one, we report in detail on just one system that can be argued to be better than the present one. It was possible to disregard some of the scoring system options under consideration. For example, it was noted that the 50-40 game, whilst seen to be very useful in some other scoring system applications (e.g. Pollard & Noble, 2004; Brown et al., 2008), produced sets that were typically too short, and could not be used in this Badge application without making sets the 'first-to-eight games'. Such a large number of games in a set seemed an unattractive characteristic, particularly given variance and fairness considerations. Correspondingly, 60-50 games produced (first-to-six games) sets that were typically too long.

The recent modification of playing the tie-break game at 5-5 rather than 6-6 was observed in the simulations to be a successful one in achieving the various aims of Badge tennis. However, it was observed in the various simulations that it was consistently better not play the tie-break game at all. Why play the tie-break game when the set is already long, having reached 5-5 (or 6-6), and when the real interest is in which team wins the whole *fixture* rather than on which pair wins a particular *set*? By removing the tie-break game and declaring the set a draw, the number of fixtures requiring a countback to games decreased, and the percentage of fixtures won by the better team increased. Thus, whilst playing the tie-break game resolves the outcome of a particular set, it does not assist in resolving the outcome of the overall fixture.

The alternative system with greatest potential appeared to be one using no-ad games, 'first-to-seven games sets', and drawn sets if 6-6 is reached. It is noted here that 'first-to-seven games sets', in which each player has the potential to have three service-games in a set, can be fairer than present first-to-six games sets (Pollard, 2005).

It can be seen from Table 1 that when there are no time constraints (N = ∞) the alternative system has a similar mean to the present system, but a smaller standard deviation (25.8 points rather than 39.2) and a smaller court-wastage (14.7 points rather than 22.2). These two characteristics are very useful when time

Table 1: Statistical characteristics (i.e. probabilities, number of points, number of games or percentages) of the present and alternative systems when (pa, pb) = (0.6, 0.5) and N = ∞, 290, 280 and 270, where N is the maximum number of points playable on one court.

	Present	Present	Present	Present	Alt	Alt	Alt	Alt
N	∞	290	280	270	∞	290	280	270
6 sets	0	0.019	0.046	0.101	0	0.000	0.002	0.024
7 sets	0	0.111	0.185	0.269	0	0.002	0.033	0.148
8 sets	1	0.870	0.768	0.629	1	0.998	0.965	0.828
Points: mean	482.1	480.1	478.0	474.2	484.9	484.8	484.7	483.6
s.d.	39.2	36.3	34.1	30.9	25.8	25.8	25.5	24.3
Games: mean	71.7	71.4	71.1	70.5	84.3	84.2	84.2	83.9
s.d.	4.2	3.9	3.8	3.6	4.1	4.1	4.0	3.8
Win	0.982	0.973	0.965	0.951	0.976	0.976	0.973	0.962
Loss	0.016	0.013	0.012	0.010	0.022	0.022	0.021	0.017
Draw	0.003	0.013	0.023	0.039	0.003	0.003	0.006	0.021
Sets	0.957	0.940	0.926	0.906	0.974	0.973	0.971	0.960
C/back	0.043	0.060	0.074	0.094	0.026	0.027	0.029	0.041
Wastage (points)	22.2	22.2	22.2	22.2	14.7	14.7	14.7	14.7
Usage	n.a.	0.828	0.854	0.878	n.a.	0.836	0.866	0.896

constraints are involved. When N = 290, 280 and 270 points, it is more likely that all eight sets are completed under the alternative system than under the present system. Under the alternative system the probability of a draw is smaller, the probability of a win by the better team is larger, fewer fixtures are determined by countbacks, and there is less wastage at the cross-over and greater court-usage. The likelihood of all eight sets being completed is quite sensitive to the value of N. It is noted that when N = 280 and (pa, pb) = (0.6, 0.5), the distribution of the number of complete sets (6, 7 or 8 sets) for simulations of the present system (0.046, 0.185, 0.768) is similar to the observed distribution (0.067, 0.181, 0.752), although the average number of games played (71.1) is greater than the observed value (69.2).

Table 2 with N = 275 and (pa, pb) = (0.61, 0.49) gives a reasonable overall fit to the data values for 6/7/8 sets completed (0.067, 0.181, 0.752) and the average number of games played (69.2). These two statistics are the appropriate ones for fitting the model of the present system to the available data. For these parameter values, more fixtures are completed under the alternative system, there are fewer draws and unfinished fixtures, more points are played on average, more fixtures are won by the better team, there is less court-wastage at the cross-over, and there is higher court usage overall. These results indicate that the alternative system is better than the present system.

The performance of the present and the alternative system is now considered when the time constraint is more severe than those estimated in Table 2. This can occur as a result of slow play, a brief shower of rain, etc. Four situations for the case when N = 260 are considered. The first is for a close match between two moderately strong servers (columns 2 and 3), and the second is for a not-so-close match between such servers (columns 4 and 5). Experience suggests that the (pa, pb) values in these 4 columns are representative of the higher men's grades. The third situation is for close matches between weaker servers (columns 6 and 7), and the fourth situation is for not-so-close matches between such servers (columns 8 and 9). The (pa, pb) values in these 4 columns are assumed to be representative of the lower men's grades and

Table 2: Statistical characteristics (i.e. probabilities, number of points, number of games or percentages) of the present and the alternative systems when (pa, pb) = (0.61, 0.49) when N = 275.

	Present	Alt
6 sets	0.0453	0.0043
7 sets	0.1806	0.0554
8 sets	0.7739	0.9403
Points: mean	467.3	477.0
s.d.	33.98	26.1
Games: mean	69.9	83.0
s.d.	3.8	4.1
Win	0.984	0.988
Loss	0.003	0.007
Draw	0.013	0.005
Sets	0.959	0.984
C/back	0.041	0.016
Wastage (points)	22.1	15.1
Usage	0.850	0.867

women's grades. It follows that even though these results with N = 260 represent a more time-constrained situation than at present, the alternative system still outperforms the present system for all of these situations.

Table 3: Statistical characteristics (i.e. probabilities, number of points, number of games or percentages) of the present and alternative systems when N = 260, for various values of pa and pb.

pa, pb	**0.62, 0.58**	**0.62, 0.58**	**0.65, 0.55**	**0.65, 0.55**	**0.52, 0.48**	**0.52, 0.48**	**0.55, 0.45**	**0.55, 0.45**
System	**Present**	**Alt**	**Present**	**Alt**	**Present**	**Alt**	**Present**	**Alt**
6 sets	0.316	0.204	0.178	0.096	0.362	0.250	0.198	0.118
7 sets	0.363	0.371	0.329	0.290	0.360	0.383	0.339	0.314
8 sets	0.318	0.424	0.491	0.614	0.269	0.367	0.460	0.567
Points: mean	480.9	492.0	469.1	481.0	481.0	491.3	467.7	479.5
s.d.	22.1	16.7	26.2	20.4	22.5	17.5	27.3	21.6
Games:mean	72.4	86.1	71.4	84.6	69.8	84.2	68.8	82.7
s.d.	3.4	2.8	3.4	3.3	3.6	2.9	3.6	3.4
Win	0.628	0.657	0.928	0.937	0.621	0.646	0.931	0.934
Loss	0.116	0.138	0.010	0.014	0.103	0.129	0.007	0.012
Draw	0.257	0.205	0.063	0.050	0.276	0.226	0.062	0.054
Sets	0.640	0.809	0.877	0.939	0.635	0.806	0.881	0.937
C/back	0.360	0.191	0.123	0.061	0.366	0.194	0.119	0.063
Wastage	20.8	12.5	21.0	13.6	22.4	13.8	22.6	15.0
Usage	0.925	0.946	0.902	0.925	0.925	0.945	0.899	0.922

FURTHER CONSIDERATIONS

It is known that if a (slightly) better pair serves in the first game of a set, the most likely games score is 6-3, whereas if the (slightly) better pair serves in the second game of a set, the most likely game score is 6-4.

Given that the number of games won by each team is used in a countback situation, it would be fairest for one team to serve first on one court and the other team to serve first on the parallel court. Note that two tosses are then required for the whole match. At present there are 4 tosses, which can result in one team winning 3 or even 4 tosses, with the resultant unfairness.

The situation in which each team has a first or stronger pair and a second or weaker pair may be relevant in some team situations played elsewhere. In such situations the two matches played before the cross-over could be between the two stronger pairs and between the two weaker pairs, or each of the two matches could be between a stronger and a weaker pair. These two scenarios would lead to (slightly) different court-wastage statistics, different fixture completion rates, etc. but this has not been studied in this paper. The software used in this study could be used for such an analysis.

The method for determining the winning team is to count sets firstly, and then games if necessary to determine the winning team. It can be shown that, if point probabilities are constant, the better team is more likely to be selected if games are used firstly, and sets secondly. However, the reason that in practice sets are used first and games second is that this gives each team an opportunity to overcome a period of poor play, and this is considered an appropriate characteristic for a tennis scoring system. The reader who is interested in countback methods is referred to the paper by Pollard and Noble (2006).

Using the methods of Pollard (2006), it can be shown that, even in the presence of sun and wind effects, changing ends of the court after games 2, 6 and 10 is sufficient for fairness in doubles. There is no need to change after every two games. This could save a little time on present practices.

The results on the number of completed sets in Tables 1–3 are quite sensitive to the time limit N points. The figures in Table 2 represent best-fit of N to observed match statistics, but it would certainly be useful to directly collect data such as the number of points played on each court, and the proportion of points won by the server, etc., for the various grades. This would provide useful additional information on this topic.

CONCLUSIONS

This study considered whether the circumstances of team tennis are such that alternative scoring systems to those used in tournament play can be useful. The objectives of team tennis at the local competitive level can differ from those in the tournament setting. For example, the requirements that the players complete a good percentage of matches, that a high percentage of court-time is used, that a small percentage of court-time is wasted, etc. have very practical relevance in a time-constrained fixture.

Badge tennis in Sydney was considered as an example of team tennis. Drawn sets at 6-6 (rather than playing a tie-break game) were seen to be useful in this situation. A scoring system was identified that resulted in more matches being completed, more matches being won by the better team, a higher usage of the available court-time, and more matches being won by the better team without the need for countbacks. The scoring system used no-ad games, first to seven games sets, and drawn sets at 6-6. Otherwise the system was the same as at present.

Performance of the new system appears to be robust to variations in server ability, gaps in ability between teams and time limits. The new system could thus be considered for use in some or all grades. However, it is recommended that more detailed statistics on the present system be collected before any change is implemented, especially of the number N of points that can be played in 2½ hours.

The methodology and results of this study could be applicable to other team tennis situations around the world, and even to other sports.

Acknowledgement

The authors wish to thank Dr John Best for supplying the data.

References

Brown, A., Barnett, T., Pollard, G. N., Lisle. I. and Pollard, G. H. (2008) The characteristics of various men's tennis doubles scoring systems. In J. Hammond (ed.) *Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM), Coolangatta, Queensland, Australia. (in press).

Pollard, G. (1983) An analysis of classical and tiebreaker tennis. *Australian Journal of Statistics*. **25(3)**, 496-505.

Pollard, G. H., Barnett, T., Brown, A. and Pollard, G. N. (2007) Some alternative men's doubles scoring systems. In S. Miller and J. Capel-Davies (eds.) *Tennis Science and Technology 3*, International Tennis Federation Licensing (UK) Ltd, Roehampton, London SW15 5XZ. 301-309.

Pollard, G. (2005) On solving an aspect of unfairness in the tennis doubles scoring system. *Chance*, **18(4)**, 17-19.

Pollard, G. (2006) A fair tennis scoring system for doubles in the presence of sun and wind effects-an application of probability. In P. Brown, S. Liu and D. Sharma (eds.) *Proceedings of the International Statistics Workshop, 2005: Contributions to Probability and Statistics, Applications and Challenges*. World Scientific Publishing, Singapore, 89-96.

Pollard, G. and Noble. K. (2003) Scoring to remove long matches, increase tournament fairness and reduce injuries. *Medicine and Science in Tennis*, **8(3)**, 12-13.

Pollard, G. and Noble, K. (2004) The benefits of a new game scoring system in tennis: the 50-40 game. In R. Morton and S. Ganesalingam (eds.) *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport.* Massey University, Palmerston North, New Zealand, 262-265.

Pollard, G. and Noble, K. (2006) Countback methods and auxiliary scoring systems in tennis. In J. Hammond and N. de Mestre (eds.) *Proceedings of the Eighth Australasian Conference on Mathematics and Computers in Sport*, MathSport (ANZIAM), Coolangatta, Queensland, Australia, 170-177.

THE EFFICIENCY OF DOUBLES SCORING SYSTEMS

Pollard, Geoff [1] and **Pollard, Graham** [2]

1 Faculty of life and Social Sciences, Swinburne University of Technology, Melbourne, Australia

2 Faculty of Information Sciences and Engineering, University of Canberra, Australia

Paper Submitted for Review: 27 March 2008
Accepted Without Revision

Abstract. The standard model for tennis singles is to assume player A has a probability pa of winning a point on service, and player B has a probability pb of winning a point on service. Player A is the better player if pa > pb. Miles (1984) noted that a tennis singles scoring system can be seen as a binomial sequential statistical test of the hypothesis pa > pb against the alternative hypothesis pa < pb. Noting Wald (1947) had recommended alternating statistical trials (a, b, a, ...), AL, for this two-sample binomial hypothesis test, he proposed the scoring systems W1(Wn(ALa), Wn(ALb)) as the standard scoring systems against which any singles scoring system could be assessed. He called such systems 'biformats' because of their two-part structure. In doubles there are four players. Pair A has probabilities pa1 and pa2 of winning a point on service when players A1 and A2 serve respectively, and correspondingly pair B has probabilities pb1 and pb2. In this paper a doubles biformat structure for testing the hypothesis that pair A is better than B versus the alternative hypothesis that pair B is better than A, is established. As in singles, this scoring system can be used as a benchmark against which the efficiency of any doubles scoring system can be assessed. Thus, the relative efficiency of the various scoring systems presently used in doubles can be assessed. Further, Further, the methods of this paper can be applied to a match between two teams of 2, 4, 8, ... doubles pairs.

Keywords: efficiency of tennis scoring systems, play-the-loser, volleyball, team tennis.

INTRODUCTION

In a very elegant paper, Miles (1984) noted the link between sports scoring systems and sequential statistical hypothesis testing, and the following few paragraphs outline the essential features of his contribution to scoring systems in that paper. One important characteristic of a scoring system is its efficiency. (Other important characteristics include the mean, the variance, and the skewness of the number of points played, and the probability that the better player or better team wins when that scoring system is used.) If two scoring systems SS1 and SS2 have the same probability of correctly identifying the better player, SS1 is said to be more efficient than SS2 if it has a smaller expected number of points played. Here we identify how the efficiency (a simple numerical value) of a scoring system can be determined.

Unipoints

Many sports consist of playing a sequence of points each of which is won by either player A or player B. If there is only one type of point in the match, we have unipoints in which p (q) is the probability player A wins (loses) each point ($p + q = 1$). Player A is the better player if $p > 0.5$. Miles (1984) considered fair scoring systems possessing sensible regularity conditions and only one type of point, which he called uniformats.

For testing the hypothesis H_0: Player A better than player B, versus

the alternative hypothesis H_1: Player B better than player A,

he applied the result of Wald and Wolfowitz (1948) to conclude that there is a unique class of optimal uniformats, given by the Sequential Probability Ratio Test, and that it is $\{W_n\}$ ($n = 1,2,3,\ldots$) where W_n is the uniformat in which the winner is the first player to achieve a lead of n points over his opponent. The key characteristics of W_n are P_n, the probability that player A wins, and μ_n, the expected number of points played. They are given by

$$P_n = p^n/(p^n + q^n)$$
$$\mu_n = n(P_n - Q_n)/(p - q)$$

where $Q_n = 1 - P_n$, and the efficiency of a general uniformat with key characteristics P and μ is given by

$$\rho = \frac{(P-Q)\ln(P/Q)}{\mu(p-q)\ln(p/q)}$$

Bipoints

If there are two types of points (a-points when player A serves, and b-points when player B serves), we have bipoints. The probability player A wins a point on service is p_a, and the probability player B wins a point on service is p_b. Player A is the better player if $p_a > p_b$. For the bipoints situation, Miles (1984) considered fair scoring systems with sensible regularity conditions, which he called biformats.

Noting that Wald (1947) had recommended the use of paired trials for comparing two binomial probabilities (H_0: $p_a > p_b$ Vs H_1: $p_b > p_a$), Miles (1984) set up {W_n(point-pairs)}, where n = 1, 2, 3, ..., as the standard biformat family of scoring systems (with unit efficiency) against which the efficiency of any tennis singles scoring system could be measured (e.g. Pollard & Noble, 2004). Here, point-pairs represent the playing of pairs of points consisting of an a-point and a b-point.

He showed that the efficiency of a general bipoints scoring system with mean μ and probability P that the better player wins, is given by

$$\rho = \frac{2(P-Q)\ln(P/Q)}{\mu(p_a - p_b)\ln(p_a q_b / p_b q_a)}$$

where Q = 1 – P, $q_a = 1 - p_a$ and $q_b = 1 - p_b$. He also set up what he called the reversal biformat $W_1(W_nAL^a, W_nAL^b)$, where the superscript represents the first type of point played, and AL represents alternating point types played (e.g. ababab...), and (W_nAL^a, W_nAL^b), operating like a point-pair, results in a win by player A, a draw, or a loss by player A. Pollard (1986, 1992) showed that $W_1(W_nAL^a, W_nAL^b)$ was stochastically equivalent to the scoring system W_n(point-pairs). Note that in both of these scoring systems, P and μ do not depend on the type of the point first played.

Using the play-the-loser mechanism (PL) in which a win by player A (B) is followed by a b- (a-) point, Miles (1984) showed that the family of reversal biformats {$W_1(W_nPL^a,W_nPL^b)$} (n = 2, 3, 4, ...) was slightly more efficient than {W_n(point-pairs)} (n = 1, 2, 3, ...) when $p_a + p_b > 1$, which is the tennis context. Pollard (1986, 1992) showed that this PL family of scoring systems was not only more efficient than the AL or point-pairs family of systems, but that it (and its stochastic equivalent, for example, one based on PL generalized point-pairs) was the most efficient possible. (Note that when using PL generalized point-pairs, a win to A (AA) is followed by a (b, b) point-pair, a draw (AB) is followed by an (a, b) point-pair, and a loss to A (BB) is followed by an (a, a) point-pair.) He showed correspondingly that the family of play-the-winner reversal biformats {$W_1(W_nPW^a,W_nPW^b)$} (n = 2, 3, 4, ...) is optimally efficient when $p_a + p_b < 1$ (which is the volleyball situation).

If P_k (Q_k) denotes the probability that player A wins (loses) in k points, then a scoring system is said to have the constant probability ratio property (c.p.r.) if P_k/Q_k is constant for all k for which $Q_k > 0$. Pollard (1992) showed that the above families of AL, PL and PW scoring systems possessed the c.p.r. property.

Quadpoints

In tennis doubles, we have four probabilities, p_{a1}, p_{a2}, p_{b1} and p_{b2} (using an obvious notation), so it is natural to call this situation quadpoints. Using these four probabilities, Pollard (2005) showed that a first to six games rule for a set of tennis could be unfair in some doubles situations, as could the present 'first to seven' and 'first to ten' tie-break games rules. Also, Pollard (1986, 1990) carried out research on the asymptotic efficiencies of some quadpoints W_n systems (for large n). He noted that some complex W_n systems could be decomposed into smaller independent components called modules, which could in turn be analyzed to produce approximate *asymptotic* values for P, μ and ρ for the whole system. In a natural extension of point-pairs, he set up a basic module consisting of the 4 points {a1, b1, a2, b2}, which we call *'point-quads'*. He also considered a second module, called PL_{teams}, using PL and generalized point-pairs in the following way. The first section of the PL_{teams} module starts with an (a1, b1) point-pair, and the PL mechanism operates in the following way. A point-pair win to pair A, (AA), is followed by a (b1, b2) point-pair whilst a point-pair win to pair B, (BB), is followed by an (a1, a2) point-pair. The first section of the module finishes as soon as a draw, (AB), occurs, and the second section of the module begins with a point-pair (a2, b2), and proceeds in the same manner as above until a draw occurs, thus completing the whole module. He showed that, provided the modulus of $(p_{a1} - p_{a2})$ equals the modulus of $(p_{b1} - p_{b2})$, PL_{teams} (PW_{teams}) modules used in conjunction with W_n systems (n large) are *asymptotically* more efficient than the {W_n(point-quads)} (n large) system when $((p_{a1}+p_{a2})/2 + (p_{b1}+p_{b2})/2)$ is greater (less) than 1.

In this paper some quadpoints results are established for *non-asymptotic* cases.

METHODS

Recall that in doubles we have four service probabilities p_{a1}, p_{a2}, p_{b1} and p_{b2}, using an obvious notation. In order to establish the efficiency of a doubles scoring system we need to set up an appropriate structure so that P and μ do not depend on the order of the four points being played. This suggests setting up the family of scoring systems {$W_n(W_1(W_2AL(a1, b1), W_2AL(a2, b2)), W_1(W_2AL(a1, b2), W_2AL(a2, b1)))$} (n = 1, 2, 3, ...) as the standard scoring system against which the efficiency of any doubles scoring system can be measured. Here AL(a1, b1), AL(a2, b2),...have obvious meanings. Note that AL(a1, b1) is stochastically equivalent to (a1, b1) point-pairs.

An expression for the efficiency of a general doubles scoring system is now derived. Firstly we note that the above scoring system has four components, each with an AL structure. The component listed first, $W_2AL(a1, b1)$ has characteristics P_1, μ_1 and Q_1 given by

$$P_1 = (p_{a1}q_{b1})/(p_{a1}q_{b1} + p_{b1}q_{a1})$$
$$\mu_1 = (2(P_1 - Q_1))/(p_{a1} - p_{b1})$$
$$Q_1 = 1 - P_1$$

where P_1 is the probability pair A wins this first component, and μ_1 is the mean duration of (or the expected number of points played in) this first component.

It is noted that $W_2AL(a1, b1)$ has the c.p.r. property, and that μ_1 is equal to the mean number of points conditional on pair A winning, and is also equal to the mean number of points conditional on pair A losing (Pollard, 1992). This fact is used in the following analysis.

Corresponding expressions for P_2, Q_2 and μ_2 for the second component can be written down.

The first half of the above scoring system, $W_1(W_2AL(a1, b1), W_2AL(a2, b2))$, is now analysed. This first half of the scoring system amounts to playing $W_2AL(a1, b1)$ and $W_2AL(a2, b2)$ until one pair wins both components. If these two components are won by different pairs, the process is repeated until one pair wins both. It can be shown that $W_1(W_2AL(a1, b1), W_2AL(a2, b2))$ has the c.p.r. property. The probability pair A wins this half of the above scoring system is given by

$$P_{1,2} = (P_1P_2)/(P_1P_2 + Q_1Q_2)$$

and the expected duration of this first half is given by

$$\mu_{1,2} = (\mu_1 + \mu_2)/(1 - R_{1,2})$$

where $R_{1,2}$ is the probability that the pairs each win one component of this half, and is given by

$$R_{1,2} = 1-(P_1P_2+Q_1Q_2)$$

The second half of the above scoring system can be analyzed similarly giving corresponding results. Combining the results for the two halves of the scoring system, we have, using an obvious notation,

$$P_{1,2,3,4} = (P_1P_2P_3P_4)/(P_1P_2P_3P_4 + Q_1Q_2Q_3Q_4)$$
$$\mu_{1,2,3,4} = (\mu_{1,2} + \mu_{3,4})/(1-R_{1,2,3,4})$$
$$1-R_{1,2,3,4} = (P_1P_2P_3P_4 + Q_1Q_2Q_3Q_4)/((P_1P_2+Q_1Q_2)(P_3P_4+Q_3Q_4))$$

Thus, for the complete scoring system (general n), which also possesses the c.p.r. property, the probability pair A wins, P_n, and the mean number of points played, μ_n, satisfy the equations

$$P_n/Q_n = (P_{1,2,3,4}/Q_{1,2,3,4})^n$$
$$\mu_n = (n(P_n-Q_n)/(P_{1,2,3,4}-Q_{1,2,3,4}))\mu_{1,2,3,4}$$

The efficiency of a doubles scoring system with mean μ and probability pair A wins equal to P, is now considered. The efficiency is equal to μ_n/μ where μ_n is the mean of the above standard scoring system with the same P value (i.e. $P = P_n$). Noting

$$2n = \ln(P_n/Q_n)/\ln(p_{a1}p_{a2}q_{b1}q_{b2}/p_{b1}p_{b2}q_{a1}q_{a2})$$

and the above expression for μ_n, the efficiency ρ of a general doubles scoring system with probability P and mean μ is given by

$$\rho = ((P-Q)\ln(P/Q)\mu_{1,2,3,4})/(2\mu(P_{1,2,3,4}-Q_{1,2,3,4})\ln(p_{a1}p_{a2}q_{b1}q_{b2}/p_{b1}p_{b2}q_{a1}q_{a2}))$$

Expressing $\mu_{1,2,3,4}$ and $(P_{1,2,3,4} - Q_{1,2,3,4})$ as functions of p_{a1}, p_{a2}, p_{b1} and p_{b2}, we have, after some algebra,

$$\rho = \frac{f(p_{a1},p_{a2},p_{b1},p_{b2})(P-Q)\ln(P/Q)}{\mu\ln(p_{a1}p_{a2}q_{b1}q_{b2}/p_{b1}p_{b2}q_{a1}q_{a2})}$$

where

$$f(p_{a1},p_{a2},p_{b1},p_{b2}) = ((p_{a1}+p_{a2})(q_{b1}+q_{b2})+(p_{b1}+p_{b2})(q_{a1}+q_{a2}))/(p_{a1}p_{a2}q_{b1}q_{b2}-p_{b1}p_{b2}q_{a1}q_{a2})$$

This is the general expression for the efficiency of a doubles scoring system. Note that when $p_{a1} = p_{a2} = p_a$ and $p_{b1} = p_{b2} = p_b$, this expression is equal to expression given in the introduction for the efficiency of tennis singles scoring systems.

The efficiency of $\{W_n(a1, a2, b1, b2)\}$

The efficiency of the W_n systems using sets of the four points at a time, called *point-quads*, is considered. (Note that this system involves playing sets of 4 points, and then making a decision as to whether pair A has won or lost, or the match continues). This system is a natural extension of point-pairs to the quadpoints case, and we might expect it to have unit efficiency, as in the point-pairs case. As in the paper by Pollard (1990), this set of four points is called a module. The independent modules effectively become the steps of a general one-dimensional walk in discrete time. Using the approach and notation of Cox and Miller (1965, p. 46-58), the steps in the random walk, Z_i, are mutually independent random variables on the integers ..., -2, -1, 0, 1, 2,... and the moment generating function (m.g.f.) of Z_i is defined by

$$f^*(\theta) = \sum_{j=-\infty}^{\infty}\exp(-j\theta)P(Z_i = j)$$

If P(Q) represent the probability of absorption in states [a, infinity) ((-infinity, -b]), and E(N) is the expected number of steps to absorption, then, neglecting the excess over the barriers,

$P/Q = (1-\exp(\theta_0 b))/(\exp(-\theta_0 a)-1)$ when $E(Z_i) \neq 0$ and

$$E(N) = (a + b - a\exp(\theta_0 b) - b\exp(-\theta_0 a))/(E(Z_i)(\exp(-\theta_0 a) - \exp(\theta_0 b)) \quad \text{when } E(Z_i) \neq 0,$$

where θ_0 is the non-zero solution of the equation $f^*(\theta) = 1$. For W_n we set a and b equal to n. Also, the random variable S (= Z_i) is used to represent the increase in pair A's score (i.e. the better pair's score) during the play of one module, and D is used to represent the expected number of points in the play of one module. It can be seen from above that the efficiency of system 1 relative to system 2 is given by the expression

$$\rho_1 / \rho_2 = ((P_1 - Q_1)\ln(P_1/Q_1)/\mu_1)/((P_2 - Q_2)\ln(P_2/Q_2)/\mu_2).$$

Using the above expressions for the ratio P/Q and E(N), and representing W_n(point-quads) as system 1, and the above standard system as system 2, we have

$$\rho_1 / \rho_2 = (\exp(\exp(-n\theta_2) - \exp(-n\theta_1)))(E(D_2/\theta_2 E(S_2))/(E(D_1/\theta_1 E(S_1)).$$

where θ_1 and θ_2 are the non-zero solutions of their respective equations.

It turns out that the efficiency of W_n(point-quads) (n = 1, 2, 3, ...) is slightly less than unity. For example, when $p_{a1} = 0.9$, $p_{a2} = 0.8$, $p_{b1} = 0.7$, and $p_{b2} = 0.6$, it can be shown that $E(D_1) = 4$, $E(S_1) = 0.8$, and $\exp(\theta_1) = 1.7650$, and $E(D_2) = 39.7541$, $E(S_2) = 7.8502$, and $\exp(\theta_2) = 1.7908$ (note that $\exp(4\theta_2) = (p_{a1}p_{a2}q_{b1}q_{b2})/(p_{b1}p_{b2}q_{a1}q_{a2})$), and it follows that, when n = 30 say, $\rho_1/\rho_2 = 0.9876$, which is slightly less than 1. A few differences between systems 1 and 2 (or their modules) are noted here. A module of system 2 has the same expected number of a- and b- points in total, but the expected number of a1-points is not equal to the expected number of a2-points, and the expected number of b1-points is not equal to the expected number of b2-points (For example, these are 10.3159, 9.5611, 10.8192 and 9.0579 respectively in the above example). Also, system 1 can have excesses of 1, 2 or 3 over the boundaries, whereas system 2 cannot have any excesses.

The efficiency of some PL and PW quadpoints systems

In this section we consider whether the PL and PW service exchange mechanisms can be used to find super-efficient scoring systems for quadpoints, as has been noted was possible in the case of bipoints. We firstly consider $\{W_n(W_1(W_{2m}PL(a1, b1), W_{2m}PL(a2, b2)), W_1(W_{2m}PL(a1, b2), W_{2m}PL(a2, b1)))\}$ (n = 1, 2, 3, ...; m = 2, 3, 4, ...). Note that $W_{2m}PL(a1, b1)$, for example, is the generalized PL point-pair structure described earlier that makes use of the W_{2m} stopping rule.

For the first of the four components, $W_{2m}PL(a1, b1)$, we have, using a similar notation to above,

$$P_1 = p_{a1}q_{b1}^{2m-1}/(p_{a1}q_{b1}^{2m-1} + p_{b1}q_{a1}^{2m-1})$$
$$\mu_1 = (2(P_1 - Q_1)(1 + (m-1)(p_{a1} + p_{b1})))/(p_{a1} - p_{b1})$$

Similar expressions for the other three components can be written down. Firstly, the case when n = 1 is considered. Using similar methods to above, expressions for $P_{1,2,3,4}$ and $\mu_{1,2,3,4}$ can be determined for this PL scoring system, and it can be shown by substituting these expressions into the formula for ρ above that the efficiency of this PL system is less than unity for all m = 2, 3, 4, ... Table 1 has values for the efficiency of this PL system when n = 1 and m = 2, 3, 4 and 5, when $p_{a1} = 0.9$, $p_{a2} = 0.8$, $p_{b1} = 0.7$ and $p_{b2} = 0.6$. This table also gives values for the efficiency of the corresponding alternating and play-the-winner scoring systems $\{W_n(W_1(W_{2m}AL(a1, b1), W_{2m}AL(a2, b2)), W_1(W_{2m}AL(a1, b2), W_{2m}AL(a2, b1)))\}$ (n = 1, 2, 3, ...; m = 2, 3, 4, ...) and $\{W_n(W_1(W_{2m}PW(a1, b1), W_{2m}PW(a2, b2)), W_1(W_{2m}PW(a1, b2), W_{2m}PW(a2, b1)))\}$ (n = 1, 2, 3, ...; m = 2, 3, 4, ...), for the cases when n = 1 and m = 2, 3, 4 and 5.

It was observed that the efficiency of all of these systems (when n = 1) is less than unity as soon as m is greater than unity. It is clear that the efficiency of these systems when n is greater than 1 (and m is not unity) must also be less than unity. Thus, it was concluded that it is not possible to use these PL and PW mechanisms in the above manner to increase efficiency above 1. The fact that the efficiencies of these systems are less than 1 (when m is greater or equal to 2), appear to be somewhat akin to the fact that efficiencies drop below 1 when scoring systems are nested using nests that are not entirely W systems.

The efficiency of some other PL and PW quadpoints systems

For the non-asymptotic situation, can the PL mechanism be used in another way so as to achieve efficiencies greater than 1? Here we consider the PL_{teams} module approach mentioned in the introduction. For example, for this module as defined, when $p_{a1} = 0.9$, $p_{a2} = 0.8$, $p_{b1} = 0.7$ and $p_{b2} = 0.6$, it was shown numerically that θ was approximately 0.8642 and E(D)/E(S) was 7.5 for each 'half' of the module. (It was verified numerically that E(D)/E(S) for this module is in general equal to $(p_{a1} + p_{a2} + p_{b1} + p_{b2})/(p_{a1} + p_{a2} - p_{b1} - p_{b2})$. The value for θ, however, needed to be calculated numerically. Thus, for these parameter values, $E(D_2)/(\theta_2 E(S_2))$ is equal to 8.6909 (from above), and $E(D_1)/(\theta_1 E(S_1))$ is equal to 8.6785, so the ratio is 1.0014. When n = 30, $\exp(\exp(-n\theta_2)-\exp(-n\theta_1)) = 1.0000$, so the efficiency of this $W_{30}(PL_{teams})$ relative to the standard system is equal to 1.0014 when $p_{a1} = 0.9$, $p_{a2} = 0.8$, $p_{b1} = 0.7$ and $p_{b2} = 0.6$. Thus, in the tennis context, for quadpoints just as for bipoints, it appears that it is possible to use the PL mechanism to achieve efficiencies greater than 1. However, it is noted that the PL_{teams} module defined in the introduction should be extended so that it includes two additional components beginning with (a1, b2) and (a2, b1). This has not been done, and so cannot be reported on. However, as in the bipoints case, any increases in efficiency above 1 are expected to be miniscule. Correspondingly, when the underlying p-values are less than 0.5 as in volleyball, the PW mechanism should be able to be used to achieve efficiencies marginally greater than 1.

The above module, PL_{teams}, made use of point-pairs. However, a corresponding module making use of points rather than point-pairs can be formed. This is done by using the PL mechanism between teams but alternating the points within each team. It was verified that the value of E(D)/E(S) was again given by $(p_{a1} + p_{a2} + p_{b1} + p_{b2})/(p_{a1} + p_{a2} - p_{b1} - p_{b2})$, and that $\exp(\theta_0)$ was the square root (since points rather than point-pairs are involved) of its value for the PL_{teams} module. Thus, these two modules are equivalent. The advantage of this formulation, however, is that it can be extended to multi-points (a1, a2,..., an; b1, b2,..., bn) (n = 3, 4, 5, ...) by 'rotating' or 'cycling' through the relevant points within each team rather than alternating them. This result (when n = 4) is relevant to the situation where there are two teams of two doubles pairs (i.e. a1, a2, a3, a4; b1, b2, b3, b4), even though 'rotating' or 'cycling' would be impractical as it would involve (amongst other things) players going on and off the court continuously. Nevertheless, a relative measure of efficiency in this team situation could be evaluated. It can be seen that this paragraph is related to the next section.

Table 1: The efficiency of the PL, AL and PW systems when $p_{a1} = 0.9$, $p_{a2} = 0.8$, $p_{b1} = 0.7$ and $p_{b2} = 0.6$

ρ when n = 1	PL	AL	PW
m = 1	1	1	1
m = 2	0.9477	0.7622	0.5701
m = 3	0.9139	0.6888	0.4585
m = 4	0.8989	0.6526	0.4055
m = 5	0.8926	0.6322	0.3742

An extension of the quadpoints systems

Suppose team A has two doubles pairs represented by [(a1, a2), (a3, a4)] and team B has two doubles pairs represented by [(b1, b2), (b3, b4)]. Then, denoting the above quadpoints standard scoring system $\{W_1(W_1(W_2AL(a1, b1), W_2AL(a2, b2)), W_1(W_2AL(a1, b2), W_2AL(a2, b1)))\}$ by SS(a1, a2; b1,b2), it follows that $W_n(W_1(SS(a1, a2; b1,b2), SS(a3, a4; b3,b4)), W1(SS(a1, a2; b3,b4), SS(a3, a4; b1,b2)))$ (n = 1, 2, 3,...) has the c.p.r. property, and is the corresponding standard family of scoring systems for two teams with two doubles pairs per team. Thus, it can be used as the standard family of scoring systems against which to assess the efficiency of team doubles with two pairs per team. In the same way, this process can be extended to teams of 4 doubles pairs, 8 doubles pairs, etcetera.

CONCLUSIONS

In a very elegant paper, Miles (1984) established a method for determining the efficiency of a singles scoring system in tennis, which is typically modeled with two probability parameters. In tennis doubles there are four players. Pair A has probabilities p_{a1} and p_{a2} of winning a point on service when players A_1 and A_2

serve respectively, and correspondingly pair B has probabilities p_{b1} and p_{b2}. In this paper a family of standard scoring systems for testing the hypothesis that pair A is better than B versus the alternative hypothesis that pair B is better than A, is established. As the characteristics of this family of scoring systems do not depend on the order of the four types of points played, it is possible to develop the efficiency of any doubles scoring system relative to it. Thus, this scoring system can be used as a benchmark against which the efficiency of any doubles scoring system can be evaluated. This is particularly useful as there is a range of scoring systems presently used for doubles. It has been shown that the efficiency of a doubles scoring system can be expressed as a single formula, similar to the case of singles tennis. The standard family of scoring systems that has been set up is very efficient, and has been given an efficiency of unity. It has been shown that, as in singles tennis with two parameters rather than four, there appears to be a family of ever-so-slightly more efficient systems that make use of the play-the-loser (or play-the-winner) service exchange mechanism. As in singles, this super-efficient family of scoring systems is of theoretical rather than sporting relevance.

The methods of this paper can be applied to a match between two teams of 2, 4, 8, etc. doubles pairs. That is, it is possible to establish a yardstick for the efficiency of such team contests.

The standard family and the super-efficient family of scoring systems for singles were seen to have applications in the munitions-testing and drug-testing situations where large variations in the number of trials required for the test are not necessarily the problem that they can be in tennis. In a similar way, it would appear that there may be use for this new family of tennis doubles scoring systems in such contexts.

References

Cox, D. and Miller, H. (1965) *The theory of stochastic processes*. Chapman and Hall, London, UK.

Miles, R. (1984) Symmetric sequential analysis: the efficiencies of sports scoring systems (with particular reference to those of tennis). *Journal of the Royal Statistical Society B*. **46(1):** 93-108.

Pollard, G. (1986) *A stochastic analysis of scoring systems*. PhD thesis. Australian National University.

Pollard, G. (1990) A method for determining the asymptotic efficiency of some sequential probability ratio tests. *Australian Journal of Statistics*. **32(1):** 191-204.

Pollard, G. (1992) The optimal test for selecting the greater of two binomial probabilities. *Australian Journal of Statistics*. **34(2):** 273-284.

Pollard, G. (2005) On solving an aspect of unfairness in the tennis doubles scoring system. *Chance*. **18(4):** 17-19.

Pollard, G. and Noble, K. (2004) The benefits of a new game scoring system in tennis: the 50-40 game. In R. Morton and S. Ganesalingam (eds.) *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport*, 30 August-1 September, 2004, Massey University, New Zealand, 262-265.

Wald, A. (1947) *Sequential analysis*. Wiley, New York.

Wald, A. and Wolfowitz, J. (1948) Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*. **19:** 326-339.

THE CHARACTERISTICS OF VARIOUS MEN'S TENNIS DOUBLES SCORING SYSTEMS

Brown, Alan [1], Barnett, Tristan [2], **Pollard, Geoff** [2], Lisle, Ian [3] and Pollard, Graham [3]

[1] Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Melbourne, Australia

[2] Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia

3 Faculty of Information Sciences and Engineering, University of Canberra, Australia

Paper Submitted for Review: 27 March 2008
Revision submitted and accepted: 11 June 2008

Abstract. In recent times a range of best of five and best of three sets tennis scoring systems has been used for elite men's doubles events. These scoring system structures include advantage sets, tie-break sets, match tie-breaks, tie-break games, advantage games and no-ad games. Several tournament organizers, tennis administrators, players and ATP Tour representatives are interested in comparing the characteristics of these various scoring systems. These characteristics include such things as the likelihood of each pair winning, and the mean, the variance, and the 'tails' of the distribution of the number of points played under the various systems. In this paper these characteristics and the distribution of the number of points in a match are determined for these various doubles scoring systems at parameter values that are relevant for elite men's doubles. Advantage games and no-ad games, both approved alternatives within the rules of tennis, are considered, as is the '50-40' game.

Keywords: new scoring systems for men's tennis doubles, no-ad and '50-40' games of tennis.

INTRODUCTION

In a match of tennis, points are played to determine the winner of a game, games are played to determine the winner of a set, and sets are played to determine the winner of a match. Traditionally, a game is the best of six points, but if the score reaches 3-3, play continues until one player leads by two points. A traditional advantage set is the best of 10 games, but if the games score reaches 5-5, play continues until one player leads by two games. A match is the best of 5 sets, or the best of 3 sets.

This scoring system survived unchanged throughout the amateur era until 1968 when tennis was opened up to professional players, and tournaments became major television events. The first significant change was the introduction of the tie-break game at six games all in all sets (except in some cases the last set) to determine the winner of a set. Also, all ATP Tour tournaments gradually reduced men's doubles and then men's singles to best of 3 sets. More recently, men's and women's doubles on the ATP and WTA Tours have introduced sudden death or no-advantage scoring at deuce so that whoever wins the next point (instead of leading by two points) wins the game. They have also replaced the last set with an extended tie-break game, known as a match tie-break (10 points).

These and other recent modifications to the traditional scoring system were primarily designed to address the fact that the requirement to lead by two points to win a game, or lead by two games to win a set, produces a match of unknown and quite variable length that leads to considerable scheduling difficulties for tournament organizers and television coverage in particular. It was also hoped that these changes would encourage top singles players to compete in the doubles as well, but the principal objective, however, was to decrease both the mean and the variance of the length of a tennis match.

Very little research has been carried out on the effect of the various scoring system options on the mean and standard deviation of the length of tennis doubles matches. Pollard et al. (2007) studied the effect of the most recent changes to the scoring system used for men's doubles for ATP Tour events.

In this paper the traditional and all of the commonly used scoring systems in professional tennis are analyzed as well as other possible systems in order to advance the prime objective of reducing the length and variance of a tennis match without significantly affecting the overall probability that the better pair would have won if the traditional scoring system was still in place.

The following eight scoring systems have been identified and analysed using advantage games. These systems are called systems 1(a) to 8(a).

1. Best of 5 advantage sets. The traditional men's scoring system, allowable under the International Tennis Federation's 'Rules of Tennis 2008', but no longer in use.
2. Best of 5 sets, first four tie-break and fifth set an advantage set. Used in men's singles Grand Slams (except the US Open) and Davis Cup singles and doubles, as well as men's doubles at Wimbledon.
3. Best of 5 tie-break sets. Used in men's singles at US Open and final of Tennis Masters Cup.
4. Best of 5 sets, first four tie-break sets and fifth set a match tie-break (10 Points). Not currently used but a possibility under the present rules of tennis, and consistent with system 8. below.
5. Best of 3 advantage sets. The traditional scoring system for women and shorter men's matches, allowable under the Rules of Tennis 2008, but not presently in use.
6. Best of 3 sets, first two tie-break and third set an advantage set. Used in women's matches at Grand Slams (except the US Open) and at the Federation Cup.
7. Best of 3 tie-break sets. Used in doubles at most Grand Slams (not Wimbledon), and also used for women's singles at US Open and for men's and women's singles on ATP and WTA Tours.
8. Best of 3 sets, first two tie-break sets and third set a match tie-break (10 Points). This is used for mixed doubles at most Grand Slams, and for doubles on the ATP and WTA Tours.

Given that the ATP and WTA have now introduced no-ad scoring at deuce for doubles on their Tours, the above analysis was repeated using no-ad games. These scoring systems are called systems 1(b) to 8(b).

Also considered is the '50-40' game scoring system (Pollard & Noble, 2004), whereby the server is required to reach 50 (i.e. one more point than 40) before the receiver reaches 40 in order to win the game, whilst the receiver wins the game by reaching 40. The above eight systems using this '50-40' game are called systems 1(c) to 8(c). Thus, in total, 24 scoring systems are considered in this paper.

In tennis the server has an advantage over the receiver and, therefore, a greater than 50% chance of winning the point. This chance is generally greater for men than for women, greater on grass than on clay, and greater in doubles than in singles. The focus of this research is on professional men's doubles, where 'long matches' can be an issue and where alternative systems are being trialed. For this reason, the analysis in this paper considers situations where the probability of the server winning the point varies from near 0 60 to near 0.75. Given that the relevant point-probability values for the servers depend on a range of things such as serving form on the day, the receivers' form on the day, the rankings of each pair, the court surface, the weather conditions, etc., it was considered that this range from 0.6 to 0.75 represented an appropriate range for covering men's professional doubles matches, at the present time and into the future.

For this analysis, the length of a match is measured by the number of points played under the various scoring systems. Obviously the actual time to play the match (in hours and minutes) also depends on other factors such as the court surface and the style of play. For example, the average time taken to play a point on a clay surface is typically longer than the time taken on a grass surface. Also, some players are known to play 'longer' points than others, and/or take more time between points. These factors have not been considered in this paper, but could be 'factored-in' by using additional information.

Following the changes in 2006 to the best of 3 sets scoring system used for men's doubles in a range of professional tournaments, there are now several scoring systems commonly used for professional men's doubles. In this paper the characteristics of the various scoring systems presently used in men's doubles or ones currently available under the Rules of Tennis 2008, are determined, and displayed in a quantitative manner. The various scoring systems are then compared with the view to determining how well they achieve the types of objectives mentioned above. The primary objective of this paper was to evaluate the characteristics of the present doubles systems, used or allowable under the rules. Given the research nature of this publication, however, it was considered appropriate to consider the merits of a new game scoring system, not in the Rules of Tennis 2008, but one that looked particularly interesting with regard to its possible use in men's doubles. This new game scoring system, the '50-40' game, is considered in the next paragraph. It is noted that its possible use in best of 5 sets matches has not previously been studied.

The usefulness of the '50-40' game, particularly in men's doubles, has been noted (Pollard & Noble, 2004). They noted that, for the p-values relevant to professional men's doubles matches (values typically greater than 0.60), 'the longest [best of 3 tie-break sets] matches (as measured by the 98% point in the distribution of the number of points played) can be reduced by about 30 points [by using '50-40' games instead of no-ad games]'. They also noted that for such p-values 'the probability player A wins [a best of 3 tie-break sets match] using '50-40' games is comparable to when using no-ad games, even though about 20 less points are required on average'. Note that the seventh point in the no-ad game creates a lack of symmetry with respect to serving to the right half or the left half of the court, and also creates a potential for unfairness, but these two issues are removed by using the '50-40' game. Thus, it would appear that the '50-40' game has some merit relative to the no-ad game for men's doubles.

What are the desirable (statistical) characteristics of a good tennis scoring system? The 'three-nesting' aspect of tennis (points, games and sets) is taken as a given or fixed part of tennis scoring systems, as it allows either player to overcome a period of poor play. Games and sets might be made longer or shorter than at present if there is an advantage in doing so. It is fundamental, however, that a scoring system should have an appropriate average number of points played, and an appropriate value for the probability that the better player wins. Also, the standard deviation of the number of points played should not be too large, so that matches have a reasonably predictable duration. Strongly positively skewed distributions of duration are to be avoided as, under such scoring systems, very long matches can result, and this can delay the matches that follow, can lead to unfairness in the tournament setting, and is obviously an issue for television. Scoring systems with good efficiency at correctly identifying the better player are preferable to ones with not-so-good efficiency. Given that a men's doubles tournament involves matches between very strong servers as well as matches between weaker servers, what is needed is a scoring system that works satisfactorily at both ends of this spectrum. This is an important consideration in selecting a tournament scoring system.

Given that the above characteristics need to be considered before adopting a tournament scoring system, compromises may need to be made between them when choosing one particular scoring system over another.

ANALYSIS OF VARIOUS SCORING SYSTEMS

For the 24 scoring systems under consideration, the probability that the better pair wins the match, and the mean and higher moments of the number of points played, were evaluated using recursion methods. The details of this (exact) recursion method are omitted here, but are available (Brown et al., 2008).

The process used to estimate the distribution of the number of points in a match required some care since, for a best of 3 sets match, it may be bi-modal, depending upon the design of the third set. Furthermore, a distribution is not uniquely determined by its moments. To overcome these difficulties, the distributions for matches requiring 2 sets or 3 sets to complete were estimated separately. This was done using the Normal Power approximation, and the first four moments of each distribution. The Normal Power approximation uses a basic assumption that the distribution is uni-modal, and it would be inappropriate to use it in conjunction with the statistics for an overall match. The probability weighted sum of the two distributions (or three in best of 5 sets matches) was used to estimate the distribution of the number of points in the overall match, from which the 98% point was obtained by interpolation. This point in the distribution is used as a measure of 'long' matches. This method was used in earlier work (Pollard et al., 2007). Further, the results agreed with those in the paper by Pollard & Noble (2004), who used simulations of 1,000,000 matches. With the exception of some cases of the 98% statistic, all statistics agreed with simulations of 4,000,000 carried out by the authors in 2008, and with other exact results (Pollard, 1983). If a difference of more than 2 points existed in the two estimated 98% statistics, the simulated value was reported in the tables.

We note here that two scoring systems can be compared for their efficiencies (at correctly identifying the better pair). Thus, given 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 pair wins the match. The efficiencies of scoring systems with differing values for the expected number of points played can also be evaluated, as in the elegant paper by

Miles (1984). It is noted here that scoring systems with high efficiencies (i.e. efficiencies close to 1) typically have particularly large standard deviations, and hence are not appropriate in the sporting context.

The best of 5 sets systems 1, 2, 3 and 4 above using (a) advantage games, and (b) no-ad games, being approved tennis scoring systems, are considered firstly, and then these systems using (c) '50-40' games are considered. The following questions were considered. What are the characteristics of these various scoring systems? What are the differences between them? Do some of them have good/not-so-good characteristics?

Five characteristics are sufficient to make relevant comparisons of the above scoring systems. They are P(A wins) where A is the better pair, the mean and the standard deviation of the number of points played, the efficiency of the scoring system, and the 98% point of the cumulative distribution of the number of points.

These five statistical measures were evaluated for a range of p_a (the probability pair A wins a point when serving) and p_b (the probability pair B wins a point when serving) values. This range covered p-values of 0.60, 0.65, 0.70 and 0.75, with pair-deviations of 0, plus and minus 0.02, and plus and minus 0.04. A close inspection of all of these results indicated that it was sufficient to report on just the results for p-values near 0.60 and for p-values near 0.75, as the results for the intermediary parameter values lay between the results at these values. Thus, the statistical measures reported in table 1 are for matches at the 'strong-serving' end ($p_a = 0.77$, $p_b = 0.73$), and at the 'weaker-serving' end ($p_a = 0.62$, $p_b = 0.58$) of the above range.

It is useful to introduce a definition at this stage. We say a scoring system X is 'a better scoring system' than scoring system Y if it has an equal or larger P(A wins) value, an equal or smaller mean, an equal or smaller standard deviation, an equal or lower 98% point, and an equal or greater efficiency (provided there is at least one inequality). It is noted that if scoring system X is better than scoring system Y, and if scoring system Y is better than scoring system Z, then scoring system X must be better than scoring system Z.

A comparison of the best of 5 sets systems 1, 2, 3 and 4, using advantage and no-ad games

Firstly, comparing scoring systems 1(a) and 1(b), the means of systems 1(a) and 1(b) for a match between strong servers are very large (484.3 and 366.7), as are the standard deviations (218.2 and 149.6), and the 98% points (1049 and 755). Even for matches between weaker servers, the standard deviations and the 98% points are large ((73.8 and 61.6) and (442 and 377) respectively). It is these characteristics that caused system 1(a), that survived the amateur era, to be replaced by others in the professional era.

Secondly, comparing scoring systems 2(a) and 2(b), the standard deviations of both systems 2(a) and 2(b) for a match between strong servers, are large (99.5 and 76.7), as are the 98% points (582 and 467). Even for a match between weaker servers, these statistics are still quite large ((63.6 and 54.2) and (395 and 341)). Given the size of these characteristics, systems 2(a) and 2(b) (although accepted in Grand Slam singles) are only really applicable to the finals of important events, where the winners are not required to play another match on the following day, and in situations in which there is no match scheduled to follow. For matches between strong servers, scoring system 2(b) is very close indeed to being a better system than system 2(a).

Thirdly, comparing scoring systems 3(a) and 3(b), there is not a huge difference in the means, in the standard deviations and in the 98% points for the matches between strong servers and those between weaker servers ((272.0 and 248.9), (261.0 and 229.9); (60.7 and 55.7), (61.6 and 52.7); and (385 and 354), (383 and 333) respectively). Thus, under scoring systems 3(a) and (3b), the characteristics of 'moderately long' matches are not hugely different for matches between weaker servers and for matches between strong servers. It is interesting to note that the P(A wins) values for the matches between weaker servers under these scoring systems are much the same as their values under scoring systems 1(a) and 1(b), whereas, for matches between stronger servers, the P(A wins) values are less than those under systems 1(a) and 1(b) (as a result of the decreased means under systems 3(a) and 3(b)). It can be seen that system 3(b) is a better scoring system than system 3(a) for the stronger servers. However, for matches between weaker servers, although the value of P(A wins) for system 3(b) (i.e. 0.721) is less than its value under system 3(a) (i.e. 0.741), its value is nevertheless greater than the associated value for the stronger servers (i.e. 0.712). The other characteristics of system 3(b) (except efficiency) for matches between weaker servers are 'better' than those of system 3(a). Thus, from a statistical point of view, system 3(b) could reasonably be preferred to system 3(a).

Fourthly, considering scoring systems 4(a) and 4(b), the means, the standard deviations and the 98% points of scoring systems 4(a) and 4(b) are, as expected, less than those of systems 3(a) and 3(b). For the matches between stronger servers, scoring system 4(b) is better than scoring system 4(a) (whilst, for the weaker servers, it 'goes somewhat close' to being better). It is also noted that, for matches between weaker servers, although the value of P(A wins) for 4(b) (i.e. 0.706) is less than its value for system 4(a) (i.e. 0.722), its value is nevertheless greater than the associated value for the matches between strong servers (i.e. 0.702). Thus, it can be argued that system 4(b) could reasonably be preferred to system 4(a). It can also reasonably be argued that, if the fifth set is simply a match tie-break, it would seem appropriate to use no-ad games rather than advantage games during the first four sets.

Scoring systems 1(a) and 1(b) are not considered any further in this paper as they have large values for the mean, the standard deviation and the 98% point, particularly for the strong servers, and also because they are no longer in use.

For matches between strong servers, the 98% point for system 3(a) (385) is substantially less than the 98% point for system 2(a) (582), whilst the value for P(A wins) is a little less for system 3(a) (0.708 compared to 0.723). For matches between weaker servers, the decrease in P(A wins) under system 3(a) rather than under system 2(a) is miniscule (0.741 down from 0.743). It follows that system 3(a) is a very reasonable alternative to system 2(a).

We have noted above that, from a statistical point of view, system 3(b) might reasonably be preferred to system 3(a). On the other hand, system 3(b) involves the use of no-ad games, which some players, spectators and viewers might find less attractive than advantage games. Nevertheless, if a lower mean and a lower 98% point than those of system 3(a) were required, system 3(b) could reasonably be used.

It can be seen from table 1 that scoring system 4(a) has no overall advantage over system 3(b) either for matches between strong servers or for matches between weaker servers. Also, we have noted above that system 4(b) could reasonably be preferred to system 4(a). Thus, if a smaller average match duration is required to that under system 3(b), system 4(b) could be considered. Judgements about the appropriate average length (and variation) required for a particular tennis match or tournament, however, are best made by the tennis enthusiasts, and not by mathematicians.

Summarizing, the conclusions about the best of 5 sets scoring systems presently available, are

1. Systems 2(a) and 2(b) are only really applicable to the finals of important doubles events, where the winners are not required to play another match within the next few days, and in situations in which there is no match scheduled to follow.
2. Across all matches with both strong and weaker servers, the characteristics of 'moderately long' matches are not hugely different for the two scoring systems 3(a) and 3(b).
3. From a statistical point of view, all things considered, system 3(b) could reasonably be preferred to system 3(a), and system 4(b) could reasonably be preferred to system 4(a).
4. If the fifth set is simply a match tie-break, it would be appropriate to use no-ad games rather than advantage games throughout the first four sets.
5. If a lower mean and a lower 98% point than those of system 3(a) are required, system 3(b) could reasonably be used.
6. For a shorter match than under system 3(b), system 4(b) (rather than 4(a)) could be used.

A comparison with the best of 5 sets scoring systems using '50-40' games

The four best of 5 sets systems using '50-40' games (see columns (c) in table 1) are now considered, and the following observations made.

1. For matches between strong servers, scoring system 2(c) is a better scoring system than both 2(a) and 2(b), scoring system 3(c) is a better scoring system than both 3(a) and 3(b), and scoring system 4(c) is a better scoring system than both 4(a) and 4(b), and, perhaps surprisingly given the number of points played, scoring system 4(c) is even a better scoring system than both 3(a) and 3(b).

2. Across all matches with both strong and weaker servers, scoring system 4(c) goes very close to being a better scoring system than system 4(b), and scoring system 3(c) goes very close to being a better scoring system than system 3(b).
3. The characteristics of system 3(c) are reasonably similar for the matches between strong servers and those between weaker servers, as are the characteristics of system 4(c). Thus, any match under system 3(c) is likely to have quite similar duration characteristics regardless of the serving strength of the players, and likewise, any match under system 4(c) is also likely to have quite similar duration characteristics.
4. Overall, systems 2, 3 and 4 using '50-40' games appear to be at least as good as these three systems when using no-ad games. Thus, the '50-40' game provides a real practical alternative to the no-ad game, when used for best of five sets matches.

Table 1: The five statistics for the twelve best of 5 sets scoring systems (* are simulated values)

5.

(i) P(A wins) (ii) Mean (iii) Stand Dev (iv) Efficiency (v) 98% Point		**(a)** $p_a = 0.77; p_b = 0.73$	**(b)** $p_a = 0.77; p_b = 0.73$	**(c)** $p_a = 0.77; p_b = 0.73$	**(a)** $p_a = 0.62; p_b = 0.58$	**(b)** $p_a = 0.62; p_b = 0.58$	**(c)** $p_a = 0.62; p_b = 0.58$
1	(i)	0.778	0.759	0.741	0.752	0.728	0.718
	(ii)	484.3	366.7	230.4	274.2	239.3	200.1
	(iii)	218.2	149.6	68.2	73.8	61.6	49.4
	(iv)	0.34	0.38	0.52	0.61	0.57	0.61
	(v)	1049*	755*	395*	442*	377*	306*
2	(i)	0.723	0.721	0.730	0.743	0.723	0.715
	(ii)	290.3	259.0	211.7	262.1	230.7	196.6
	(iii)	99.5	76.7	52.0	63.6	54.2	46.3
	(iv)	0.34	0.38	0.51	0.59	0.55	0.61
	(v)	582*	467*	323	395	341	290
3	(i)	0.708	0.712	0.727	0.741	0.721	0.715
	(ii)	272.0	248.9	210.0	261.0	229.9	196.3
	(iii)	60.7	55.7	48.6	61.6	52.7	45.7
	(iv)	0.32	0.36	0.50	0.58	0.55	0.60
	(v)	385	354	306	383	333	286
4	(i)	0.699	0.702	0.714	0.722	0.706	0.700
	(ii)	255.4	234.2	198.7	245.5	216.7	185.8
	(iii)	44.8	41.6	37.7	46.9	40.0	35.5
	(iv)	0.31	0.35	0.46	0.52	0.50	0.55
	(v)	333	308	270	335	291	253

A comparison of best of 3 sets systems 5, 6, 7 and 8, using deuce, no-ad and '50-40' games

The corresponding results for the best of 3 sets scoring systems are given in table 2, and the following observations are made:

- Many of the comments and conclusions on the best of 5 sets systems apply to the corresponding best of 3 sets systems.
- For matches between strong servers, system 6(b) goes very close to being a better system than 6(a), whilst scoring system 6(a) has a large standard deviation (93.1), and a large 98% point (480). Thus, it could be argued that system 6(a) is only really applicable to the finals of important doubles events, where the winners are not required to play another match within the next few days, and in situations in which there is no following match. It is noted that the corresponding 98% point for system 6(b) is 366, which is a much better value (than 480) for tournament play.

- For matches between strong servers, system 6(c) is better than 6(a) and 6(b), system 7(c) is better than 7(b) (which in turn is better than 7(a)), and system 8(c) is better than 8(b) (which in turn is better than 8(a)). System 8(c) even goes close to being better than system 7(a).
- As with the best of 5 sets systems, the efficiency of all of the (a) and (b) systems is low for matches between strong servers, whereas it is around 0.5 for the (c) systems. The best of 3 sets systems are, as expected, slightly more efficient than the corresponding best of 5 sets systems.
- With the exception of efficiency (and the value of P(A wins)), the characteristics of system 7(a) are quite similar for matches between strong and weaker servers, so the length of matches under system 7(a) are not particularly dependent on the strength of the players' serves.
- If it was felt that matches under system 7(c) and/or 8(c) were 'too short', then playing first to 7 games sets would increase the duration whilst maintaining efficiency and increasing the value of P(A wins). (It is noted that first to 6 games sets can be unfair (Pollard, 2005), and that the associated unfairness can be removed by playing first to 7 games sets.)

Table 2: The five statistics for the twelve best of 3 sets scoring systems (* simulated values)

(i) P(A wins) (ii) Mean (iii) Stand Dev (iv) Efficiency (v) 98% Point		**(a) $p_a = 0.77$; $p_b = 0.73$**	**(b) $p_a = 0.77$; $p_b = 0.73$**	**(c) $p_a = 0.77$; $p_b = 0.73$**	**(a) $p_a = 0.62$; $p_b = 0.58$**	**(b) $p_a = 0.62$; $p_b = 0.58$**	**(c) $p_a = 0.62$; $p_b = 0.58$**
5	(i)	0.730	0.713	0.698	0.707	0.687	0.678
	(ii)	298.2	225.3	141.2	168.3	146.5	122.4
	(iii)	165.0	112.1	48.9	51.6	42.8	33.9
	(iv)	0.36	0.40	0.55	0.65	0.60	0.65
	(v)	750*	533*	269*	295*	250*	200*
6	(i)	0.690	0.686	0.690	0.701	0.683	0.676
	(ii)	192.1	166.4	131.0	161.6	141.8	120.5
	(iii)	93.1	66.8	37.9	44.6	37.6	31.7
	(iv)	0.37	0.41	0.54	0.63	0.59	0.65
	(v)	480*	366*	225*	261	226*	188
7	(i)	0.669	0.672	0.686	0.697	0.681	0.675
	(ii)	166.3	155.2	128.6	160.0	140.7	120.1
	(iii)	40.3	37.0	32.6	41.4	35.3	30.7
	(iv)	0.34	0.38	0.53	0.62	0.58	0.64
	(v)	243	224	196	246	212	183
8	(i)	0.656	0.658	0.667	0.670	0.658	0.655
	(ii)	142.8	131.5	112.5	137.8	122.0	105.2
	(iii)	21.8	20.5	19.9	24.5	20.5	18.8
	(iv)	0.33	0.37	0.48	0.52	0.51	0.56
	(v)	187	174	158*	191	166	146

CONCLUSIONS

This paper provides answers for those persons wanting to do something about the undefined length of tennis matches, where the scoring systems used lead to matches of unpredictable length, with considerable variation and exposure to excessive length, all of which affect the scheduling, television coverage and players' health. Experimentation with modified scoring systems is taking place in professional tennis, especially in men's doubles where eight different scoring systems have been used to date. All of these eight systems have been analysed in this paper, using the usual advantage scoring in each game, no-ad scoring, and the more efficient '50-40' game scoring system. Thus, twenty four scoring systems in total have been analysed and discussed.

Five statistics are sufficient to describe and compare tennis scoring systems. By using these statistics, one can gain an insight into the effects of selecting one scoring system for a tournament rather than another.

For best of 5 tie-break sets matches, the characteristics of 'long' matches are not hugely affected by whether advantage or no-ad games are used. If tournament organizers require fewer points in such 'long' (best of 5 tie-break sets) matches, a match tie-break fifth set and no-ad games throughout, might be used instead of five tie-break sets. Indeed, the results in this paper indicate that if a match tie-break is to be used for the fifth set, no-ad games are better (than advantage games) to use throughout the first four sets. For all of the best of 5 sets systems presently in use, the results show that the use of the '50-40' game would be at least as good as using the no-ad game, making the 50-40 game a practical alternative to the no-ad game. For the best of 3 sets matches, similar comparisons exist between the various systems.

At the Grand Slam level, Wimbledon retains best of 5 sets for men's doubles using scoring system 2(a) (4 tie-break sets and fifth set advantage). The other Grand Slams have all moved progressively to system 7(a) (best of 3 tie-break sets) in order to achieve a considerable reduction in the average length of matches, the standard deviation of the length, and the 98% point for the length of matches. Whilst Wimbledon may not particularly wish to achieve savings in average length and variation in length for doubles matches, the analysis in this paper shows that Wimbledon could achieve similar reductions to those in the other Grand Slam events but retain best of 5 sets structure by using system 4(c) (i.e. 4 tie-break sets and the fifth set a match tie-break, using '50-40' scoring in the first 4 sets).

The ATP and WTA have sought even greater reductions in length and variation in length of matches, and are currently using system 8(b) (2 tie-break sets and third set a match tie-break, with no-ad scoring). If they used system 8(c) (same as 8(b) but using '50-40' scoring), they would achieve further reductions in mean length, standard deviation, and in the 98% point, while efficiency would actually improve. Alternatively, they could use system 6(c) (which uses 50-40 games) and thus retain the 3 tie-break sets structure, and avoid the match tie-break.

In general, tournaments would like much greater control over the length of doubles matches, but different tournaments may want different average lengths of matches. This paper presents 24 scoring systems options that tournaments might use to increase the certainty that matches approximate that desired length. It is argued that it is not appropriate in this paper to be prescriptive about which scoring system to use in one tournament or another. However, tournament organizers who are interested in two or more scoring systems for possible use in a tournament can refer to the results in the body of this paper as a guide to making that decision.

References

Brown, A., Barnett, T. and Pollard, G. (2008) A recursion method for evaluating the moments of a nested scoring system. In J. Hammond (ed.) *Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM), Coolangatta, Queensland, Australia. (in press).

Miles, R. (1984) Symmetric sequential analysis: the efficiencies of sports scoring systems (with particular reference to those of tennis). *Journal of the Royal Statistical Society, B.* **46(1):** 93-108.

Pollard, G. (1983) An analysis of classical and tiebreaker tennis. *Australian Journal of Statistics.* **25(3):** 496-505.

Pollard, G. (2005) On solving an aspect of unfairness in the tennis doubles scoring system. *Chance.* **18(4):** 17-19.

Pollard, G. H, Barnett, T., Brown, A. and Pollard, G. N. (2007) Some alternative men's doubles scoring systems. In S. Miller and J. Capel-Davies (eds.) Tennis Science and Technology 3, International Tennis Federation Licensing (UK) Ltd, Roehampton, London SW15 5XZ. 301-309.

Pollard, G. and Noble, K. (2004) The benefits of a new game scoring system in tennis: the 50-40 game. In R. Morton and S. Ganesalingam (eds.) *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport.* Massey University, Massey, New Zealand. 262-265.

A PLAYER RATING MODEL IN MEN'S BADMINTON

Ladds, Monique and Bedford, Anthony

School of Mathematical and Geospatial Sciences, RMIT University, Australia

Paper Submitted for Review: 29 January 2008
Revision submitted and accepted: 16 June 2008

Abstract. The World Badminton Federation currently rates players based on a moving twelve month window utilising tournament points. For a players rating to improve, they simply need to win as many games as possible in as high a quality tournament as possible. The existing system allows for some graded change of rating beyond a simple win game to increase rating, given that the higher quality tournament reward players with greater points when successful. Whilst this method goes some way to grade player ratings, this system has a number of flaws. Firstly it does not account for the quality opponent, as depth of tournament success is the sole rating determinant. Secondly, being a twelve month window, players' ratings from a year ago hold the same weight as a tournament completed last week, thereby not allowing for current form. To overcome these perceived shortcomings, we have created a performance rating model based on the Elo system that considers players score at the end of each round. Players will begin with a predicted score based on their rating in comparison to the opposition. The predicted score will then be compared to the actual score obtained by the player in that round. A new rating will be acquired based on whether a player performs above or below expectations. The new ratings will then be discussed in comparison to the current BWF rating system as a descriptive and predictive tool.

Keywords: Rating, ranking, badminton, Elo

INTRODUCTION

Badminton is the world's fastest racquet sport. It is played across all continents and by many countries. Badminton is extremely popular in Europe and Asia. It made its Olympic debut at the Barcelona Olympics in 1992. Competitive badminton is predominantly played under a tournament structure. The Badminton World Federation (BWF) is the sports governing body. The BWF ranks players using a rating method akin to that used by the ATP (Association of Tennis Professionals) in tennis. These rankings and ratings are utilised for tournament seeding and Olympic qualification. In this research we improve upon the existing BWF ranking method using both optimization and smoothing techniques on an Elo-based model which we call for simplicity the *SORTED* model (Smoothed Optimized Ratings with Tournament Estimated Depth). Intrinsic to this approach are games scores, which capture both a players' current form, the strength of the opponent, and the tournament depth. We will show dramatic differences in *SORTED* and BWF men's rankings, and demonstrate that BWF rankings remain a very poor predictor of game outcome at the elite level.

A number of seminal works focus on racquet and tennis-type sport ratings which account for differences in player rating based on the nuances of sport scoring. Joe (1991) devised a rating system based on paired comparison models for table tennis and chess. Player performance ratings are adjusted based on comparison of actual and predicted outcome. This allows lowly ranked players the opportunity to move up in rankings far quicker than the current system. For example, if a player was predicted to lose by 10 points but actually loses by five, then they have exceeded the predicted outcome and should therefore move up in their ratings. This follows the logic of Elo (1978), whose method of ratings chess players is intrinsic to this work.

Strauss and Arnold (1987) looked at rating players in racquetball utilising a markov chain approach. The authors consider the probability of winning a point based on the strength of server. This is necessary given that in racquetball to win a point you must first serve. Marcus (2001) develops a ratings model for table-tennis. The game is similar to badminton in that some players can play over a dozen tournaments a year whereas some will only play one a year. They develop an algorithmic approach to solving each players rating. They differentiate new players from existing players by assigning different distributions. They also consider time between tournaments, and finalise their work via a reporting system.

Barrie (2003) suggests that ratings be calculated based on an old rating, uncertainty of old rating, tournament rating and estimated error of tournament to create a new rating. Uncertainties would be smaller based on the level of experience of the player. This allows for the experience of a player to become a factor. It also not only takes into account players and opponents scores, but it allows for estimated error in tournament rating.

All these papers incorporate the ideas of Elo (1978), which were also applied to men's tennis in Bedford and Clarke (2000). In this paper we have created a rating system which builds upon this work, including a tournament strength variable, and multiple optimisation processes.

METHOD AND RESULTS

At present ratings for badminton are calculated by the BWF (www.internationalbadminton.org) and are based on an accumulation of points from various tournaments within the current 12 month period. Only points acquired from their best 10 tournament scores make up their overall points tally which forms the player rating. Players are awarded points based on which point they are eliminated from a tournament, scaled by the quality of the tournament. Points from any tournament are only valid for 12 months for ratings. At the end of 12 months the points are lost and the next best points from another tournament are added. However, if a player has not accrued new points then they lose their points and possibly drop in ratings. A breakdown of tournament points earned is given in Table 1, the values which we shall call $D_{*,j}$.

Table 1: BWF points by tournament and tournament depth

$D_{*,j}$	World Series	BWF Super Series	BWF Grand Prix Gold	Grand Prix	International Challange	International Series	Future Series
Winner	12000	9200	7000	5000	4000	2500	1700
Run Up	10200	7800	5950	4250	3400	2130	1420
SF	8400	6420	4900	3500	2800	1750	1170
QF	6600	5040	3850	2750	2200	1370	920
1\16	4800	3600	2720	1920	1520	920	600
1\32	3000	2220	1670	1170	920	550	350
1\64	1200	880	660	460	360	210	130
1\128	600	430	320	220	170	100	60
1\256	240	170	130	90	70	40	20
1\512	120	80	60	40	30	20	10
1\1024	60	40	30	25	20	10	5

The data required for this model was obtained from the BWF who provided a full excel file of each tournament down to each players individual result. However the data was not in a suitable format for our ratings and as a result four separate excel files had to be created to transform the data. Three of the excel files were different only for the initial number of players in the tournament. The raw data extracted from these was player number, name, opponent and games score. Added to that was win or lose so that the players outcome in the game was always known, and tournament type, tournament date and depth of tournament. The flowchart of the process is shown in Figure 1.

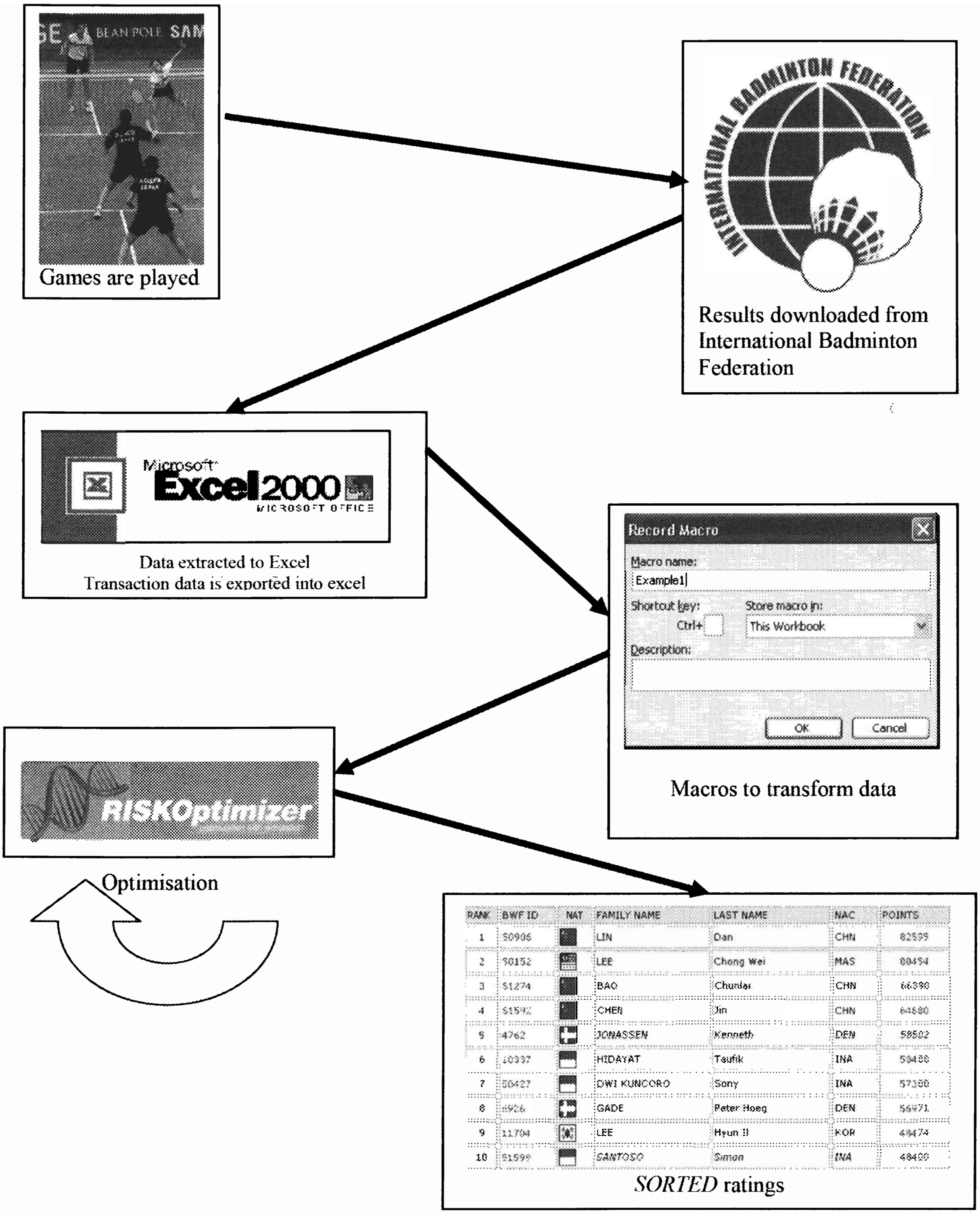


Figure 1: Flowchart of data import and application of new rating system.

As seen in Figure 1, the process concludes with optimisation before the *SORTED* ratings are obtained. Prior to optimisation we developed a number of formulae to encompass the aspects of individual game data, and tournament depth. As in Bedford and Clarke (2000), we first determine the value of winning a game, and set, to a player i, or the observed result, through the following:

$$O_i = \gamma S + G \tag{1}$$

where S = total sets won, G = total games won, and γ = set multiplier. The *SORTED* model consists of a smoothed rating, utilised through the generic formula

$$R_{i,t} = R_{i,t-1} + \alpha(O_i - E_i) + \tau(D_{i,j}) \tag{2}$$

where $R_{i,t}$ = ratings of player i at time t, α = smoothing coefficient; E =expected outcome for player i, and τ = smoothing coefficient.

The initial trial encompassed tournament variation, as in Barrie (2003), in the hope that by including the variation in quality of players commencing a stage of a tournament would improve the final prediction percentage. Barrie (2003) utilised the standard deviation of players in a tournament, and we followed a similar route using the standard deviation of pre-game ratings as a value of $D_{i,j}$, that is

$$D_{i,j} = SD(\{R_{i,t}\}) \tag{3}$$

where $\{R_{i,t}\}$ is the set of players pre-game ratings at stage j of a tournament.

To our logic, a high value of D_j as in (3) implies a tournament of low volatility, that is, a theoretically more predictable result. Inversely, a low value in (3) yields a high volatility, given players ratings are quite close together. The questioned remained as to what to do with the values in (3). If (3) reflected a volatile tournament value, then should the ratings have greater or lesser meaning, here realised by weighting into (2), than a tournament with low volatility? Should an upset in a non-volatile tournament count for more than in a volatile tournament? Without any presumption, we utilised RiskOptimizer to gain improvements on a basic model as in (2). For all values within the optimization process, using (3) yielded no improvement and therefore was excluded from the final model. We later trial another form of D_j using the existing BWF points as detailed in Table 1. Hence the final model includes

$$D_{i,j} = BWF_j \tag{4}$$

where BWF_j = BWF points for player i at stage j as given by Table 1.

The final model, combining (1), (2) and (4), and our modified expectation on pre game ratings, yielded

$$R_{i,t} = R_{i,t-1} + \alpha\left(\left([\gamma S + G]_i - [\gamma S + G]_{opp}\right) - \left([R_{i,t-1} - R_{opp,t-1}]\beta^{-1}\right)\right) + \tau(D_{i,j}) \tag{5}$$

The process of optimization was stepwise, using a method developed in Bedford (2004), where the predicted percentage correct was maximized first, modifying α, β, γ and τ, before minimizing error. The process of optimisation was as follows:

(a) Determine integer-wise best value of set multiplier, γ.
 Maximise % correct by changing α for fixed $\gamma = 0,1,\ldots,21; \tau = 0, \beta = 5$.

(b) Maximise % correct by changing τ and α for fixed $\gamma = 3; \beta = 5$ (based on (a) optimal solution).

(c) Minimise MSE by changing τ and α for γ fixed $\beta = 5$.

The final *SORTED* model yielded the values $\alpha = 0.78, \beta = 5; \gamma = 2.96; \tau = 0.001$ for a predicted correct 75.33% or 1878 out of 2493 games considered. The model was run on 3199 games, however games whereby either player had plated less than three matches were modeled but not considered, to allow for form catch-up. Interestingly, the inclusion of tournament depth makes marginal improvements to the results. Figure 2 shows the changes to the set multiplier and MSE on % correct, and Figure 3 the values of α.

Bedford and Clarke (2000) found a set multiplier of six yielded both practical significance as well as optimal prediction. For tennis, they show that by winning a set, a player is rewarded an additional six SPARKS (Set Point mARKS), equivalent to (1). Here a value of $\gamma = 2.96$ seems rather low, tending to suggest that given badmintons high point count to win a set contributes nearly enough information to determine an appropriate margin of victory. A detailed example of how the *SORTED* model works is now shown (Table 2) where Chong Wei Lee defeats Peter Gade.

Table 2: Semi Final: BWF Super Series; Dec-02-2007

	Chong Wei Lee	Peter Gade
R_{t-1}	1361.62	1322.21
E	7.882129	-7.88213
Result	21 – 17 ; 22 – 20	
O	11.92273	-11.9227
R_t	1369.81	1324.10

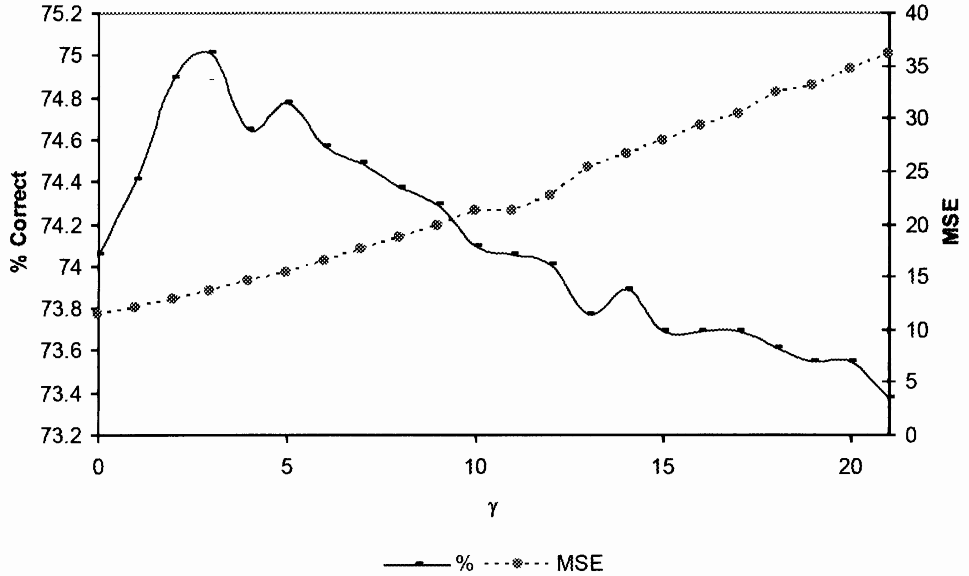


Figure 2: Comparison of set multiplier on MSE and % predicted correctly.

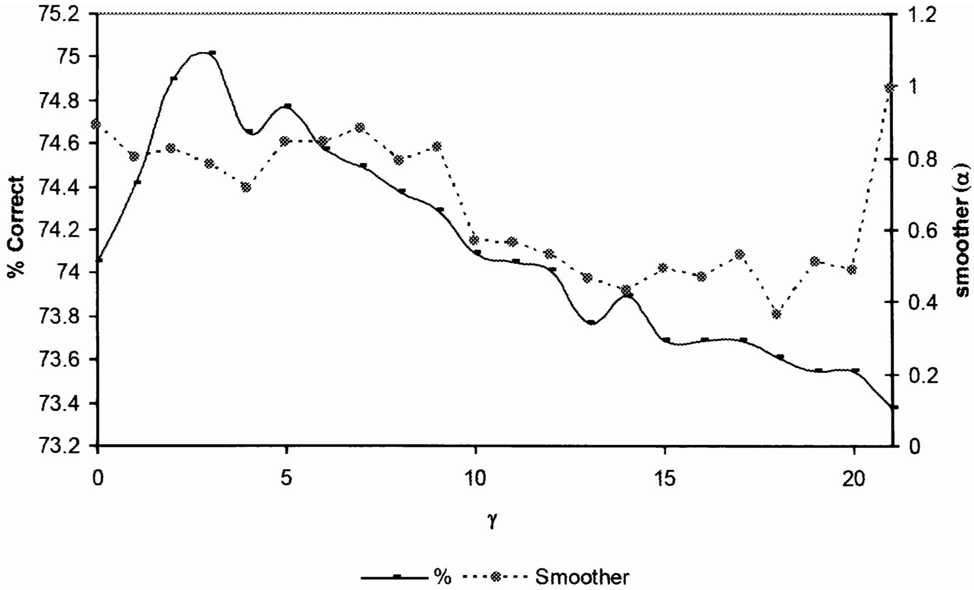


Figure 3: Comparison of set multiplier on smoother and % predicted correctly.

We found the optimal value of $\alpha = 0.77$ to be of some interest, given our past research found values of 0.2 or below. This indicates a rapidly reducing contribution of past performances to the current rating. Consider Figure 4, the top two players in men's badminton. The BWF have ranked Dan Lin as world number one for the entire 2007 season, however our model shows a differing result, with Chong Wei Lee overtaking Dan Lin towards the end of the year. Overall the percentage of correctly predicted games by our rating system is 75.33%. Compared to the BWF ranking system of only 61.43% predicted correctly, our rating system far surpasses the BWF in terms of accuracy of prediction.

Digging deeper, we investigated the performance of the BWF ratings to the SORTED model in terms of a threshold level. To have a game considered both players must be ranked higher than the threshold level, for example a threshold of 10 implies both players must be ranked within the top 10 according to the BWF rankings. Figure 5 indicates the distinct gap between the BWF ratings as a predictive tool too that developed in this research.

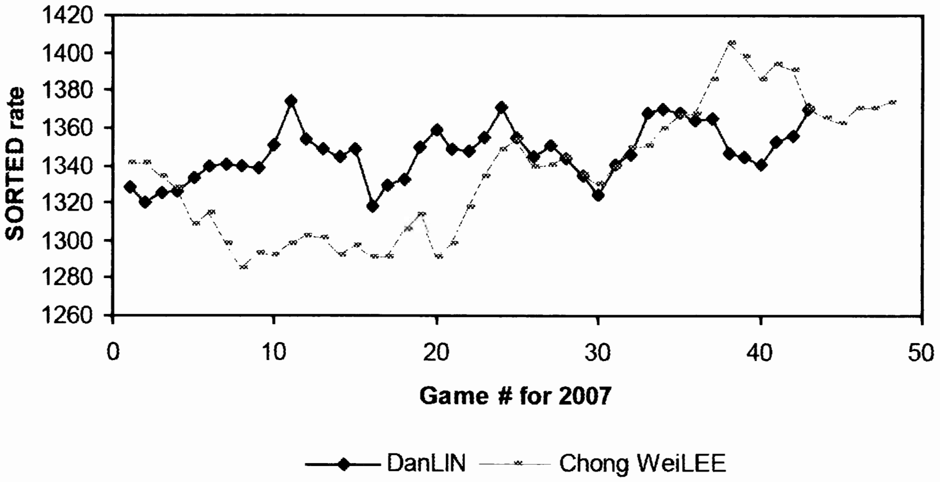


Figure 4: Top two players season track, 2007.

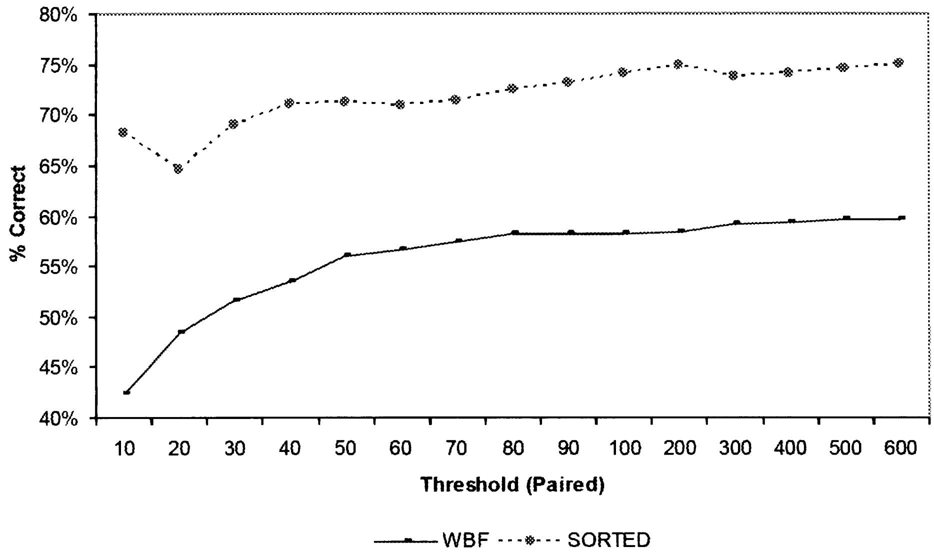


Figure 5: % predicted correctly, WBF vs SORTED, based on threshold

We then looked at the final top 10 for 2007, by both the BWF and *SORTED* model, as shown in Table 3. Notably, Chong Wei Lee is number one, quite different to the BWF, given that Dan Lin is well in front using the BWF ratings. The BWF ratings in this instance have not accounted for recent form, and it is near impossible for Chong Wei Lee to make number one until Dan Lin's best past performance moves outside the 12-month window.

Table 3: Top 10 comparison at seasons end, 2007

	Player	Country	*SORTED*	Player	Country	BWF
1	Chong Wei LEE	MAS	1373.39	Dan LIN	CHN	85394.51
2	Dan LIN	CHN	1369.89	Chong Wei LEE	MAS	74828.67
3	Sony DWI KUNCORO	KOR	1354.87	Chunlai BAO	CHN	70530
4	Chunlai BAO	CHN	1337.51	Jin CHEN	CHN	64800
5	Peter Høeg GADE	DEN	1324.10	Peter Høeg GADE	DEN	59262.51
6	Muhd Roslin HASHIM	MAS	1309.12	Sony DWI KUNCORO	KOR	57380
7	Jin CHEN	CHN	1303.05	Taufik HIDAYAT	INA	57047.8
8	Taufik HIDAYAT	INA	1301.76	Kenneth JONASSEN	DEN	56264.13
9	Sung Hwan PARK	KOR	1291.06	Yu CHEN	CHN	54620
10	Kenneth JONASSEN	DEN	1287.95	Hong CHEN	CHN	48920

Whilst the *SORTED* model provides a far more accurate method of prediction, its utility as the new ratings method requires much further investigation. The biggest obstruction to the models implementation is the points attraction associated with tournaments. Olympic selection and tournament seeding is based upon the BWF ratings, hence there is strategy involved in selecting which tournaments to attend. A player may well choose to enter an International Series event in the knowledge of a greater likelihood of high points (with weaker opponents) than an International Challenge, where the quality of opponent in likely to be higher. The reward for winning the International Series event is 2500, whereas to gain the equivalent score at the Challenge event requires at least a semi-final place (see Table 1). If the *SORTED* model were to be implemented as a replacement for the BWF series points, then players are likely to enter events solely on the prize money available, and the quality of opponent. Hence we contended the existing system remains and the BWF points are utilised within (5), given there is reward in winning at a better quality tournament to a players rating.

CONCLUSION

Through optimisation techniques we have developed a ratings model with is vastly superior to the existing BWF ratings system. We argue that a rating model also provides a greater incentive for players to chase down every shuttle, and attempt to win as many points as possible, given the point based nature of the model. We aim to further extend this research to women's and doubles formats.

Acknowledgements

We thank Badminton Australia, and the Australian Olympic players, for their feedback on the ratings. We also acknowledge the invaluable assistance of ex-CEO Damian Kelly and the World Badminton Federation in obtaining the data.

References

Barrie, P. J. (2003) A new sports ratings system: the tiddlywinks world ratings. *Journal of Applied Statistics.* **30**: 361-372.

Bedford, A. (2004) Predicting women's World Cup handball match outcomes using optimised ratings models. In H. Morton and S. Ganesalingam (eds.) *Proceedings of 7th Australasian Conference on Mathematics and Computers in Sport.* Massey University, New Zealand, 66-74.

Bedford, A. and Clarke, S.R. (2000) A comparison of the ATP ratings with a smoothing method for match prediction. *In Proceedings of 5th Australasian Conference on Mathematics and Computers in Sport.* UTS, Sydney, 43-51.

Elo, A.E. (1978) The Rating of Chessplayers, Past and Present. Batsford, London.

Joe, H. (1991) Rating systems based on paired comparison models. *Statistics & Probability letters.* **11**: 343-347.

Marcus, D. J. (2001) "New table-tennis rating system". *The Statistician.* **50**: 191-208.

Strauss, D. and Arnold, B. C. (1987) The rating of players in racquetball tournaments. *Applied Statistics.* **36:**163-173.

A PLAYER RATING SYSTEM FOR AUSTRALIAN RULES FOOTBALL USING FIELD EQUITY MEASURES

Jackson, Karl

Faculty of Life and Social Sciences, Swinburne University of Technology, VIC Australia

Paper Submitted for Review: 25 March 2008
Revision submitted and accepted: 18 June 2008

Abstract. In this paper, we will introduce a new player rating system that takes advantage of recent developments in the collection of Australian Football League statistics. Since 2004, Champion Data have collected the physical location of all possessions in the competition in addition to information about the type and quality of possession. Previous work that used these locations to establish a measure of Field Equity – the scoreboard value of a possession – will form the basis of the player rating system introduced in this paper. This new system will be an improvement over current systems, in that it removes the emphasis on subjective classification of possessions and replaces it with an objective value for the worth of each possession. Players who consistently win the ball and progress it towards goal will be rewarded more than players who build a high number of possession without improving the position of their team. Correlations between game-by-game Equity Ratings and Champion Data player ratings are presented for each of eight different playing positions. While early results suggest bias towards forward line players, methods of reducing this bias are presented in order to improve the player rating system so that is fair for all playing positions.

Keywords: AFL, Player Ratings, Equity.

INTRODUCTION

As public interest and media expenditure in sport increases, so does the need for a more in-depth analysis of team and player performance. Supporters and media commentators alike want to know the standout players of the game without having to dig through raw statistics. Effective measurement of player performance could also have applications to the presentation of end of year awards (such as club best and fairest and the Brownlow Medal) and the evaluation of a player's true value for use in decisions on trading and renewal of contracts.

Since player performance and team performance are often hard to separate, rating individual players in team sports can be problematic. Some recent developments on individual ratings in team sports can be seen in Bracewell (2003), Bennett (2005) and Schatz (2005). Bracewell uses multivariate techniques and data mining to reduce the dimensionality of the available data from Rugby matches to find key performance indicators and quantify player performance. Bennett (2005) uses 'player game percentages' to evaluate the true value of the contributions a player makes within a game, based on the effect a player has on his team's probability of winning the game. Schatz (2005) uses a system called defence-adjusted value over average (DVOA) to rate player performance in American Football. DVOA compares the performance of offensive players in each successive game situation with the average performance of all teams in the same situation throughout the season and adjusts the value according to the strength of the defensive team.

One of the most visible uses of player rating systems to the general public is online fantasy football competitions. With more households connected to the internet every year, these competitions are becoming increasingly popular, with over 206,000 teams entered in the Australian Football League's (AFL's) official fantasy competition, Dream Team, in 2007. The aim of these competitions is to select a squad of 30 players subject to salary cap and positional restraints, of which 22 are selected to "play" each week. The goal is to field a team that will perform the best based on a pre-defined player rating system. Ideally, player ratings systems should be completely objective and be fair for all positions across the field, but one criticism of current player rating systems for AFL is their bias in favour of midfield and half-back players who dominate possession. AFL Dream Team uses a simplistic player rating system that considers only the most basic of statistics, such as the number of kicks, marks, handballs and goals. While this system is easily interpreted, it fails to take into account the quality of a player's possessions and disposals, relying purely on the number of such events.

The AFL's second largest online fantasy competition, the Herald Sun's Supercoach (with over 135,000 entries in 2007), uses a more complex method of player ratings, developed by the official statisticians for the AFL, Champion Data. Whilst this system still relies on volume of possessions/disposals to evaluate ratings, it also takes into account the quality of disposal, a more accurate assessment of the type of possession and the current state of the game. For example, the Dream Team rating system allocates 4 points for a kick, whereas the Champion Data ratings classify the same event as either an effective kick (4 points), ineffective kick (0 points) or a clanger (uncontested turnover) kick (-8 points). Similar breakdowns of point allocation occur for handballs, marks and possessions from disputed balls.

While the Champion Data ratings are a significant improvement over the simplistic Dream Team system, the subjective and discrete classification of disposals and possessions is not ideal. For example, the definition of an effective kick is either a short kick (less than 40m) to an unopposed teammate or a long kick (greater than 40m) to an evenly matched (or 50/50) contest or better. So, a kick to a 50/50 contest 39 metres away is considered an ineffective kick, but a kick that travels 41 metres to the same contest is considered effective.

In order to remove this subjective classification of events, this paper will make the first progress towards a new player rating system will be introduced using Field Equity measures. O'Shaughnessy (2006) first introduced the concept of Field Equity to Australian Rules football, a method of evaluating the scoreboard value of individual possessions. A similar concept was used by Romer (2002) to decide in which situations an American football team should "go for it" on the 4th down.

AVAILABLE DATA

The locations on the ground of all possessions and stoppages in the AFL competition have been recorded by Champion Data since 2004, resulting in over 500,000 data points up to the end of the 2007 season. Since the dimensions of AFL grounds vary, the physical location of all possessions were mapped to standard dimensions, those of the MCG, using a method that preserves ground markings and angles from goal. These locations are time-coded and linked to information about the type of the possession that occurred, so further investigation of specific situations within a game is possible. Noting that each quarter is started by a Centre Bounce and that no further possessions/stoppages can occur after the siren at the end of each quarter, the 14 different types of possession/stoppage that can occur within a game can be classified within the following seven phases:

- Set – Set possession, with no physical pressure.
    - Contested/Uncontested Mark, Free Kick
- Uncontested – Minimal physical pressure from opponents.
    - Gather, Handball Receive, Kick-In to Self
- Loose – Significant physical pressure from opponents.
    - Loose Ball Get
- Hard – Direct physical pressure from opponents.
    - Hard Ball Get, Ground Kick
- Stoppage – Umpire possession to restart the ball into general play.
    - Ball-Up Bounce, Out of Bounds, Centre Bounce
- Kick-In – Restarts play from the defensive goal square, following a behind.
- Boundary Restart – Restarts play following an opposition "out on the full" at the location of the infringement.

Each of the above phases of possession will be analysed separately in order to establish Field Equity values continuously across the ground.

FIELD EQUITY

O'Shaughnessy (2006) defines Match Equity, $E_M(m,t,\mathbf{x},\varphi)$, as the probability that a team will win from the current match situation, dependent on the current score margin (m), the time left in the match (t), the location of the ball ($\mathbf{x}$) and the current phase of possession(φ). It is then suggested that the match Equity can be decoupled to give two measures, Field Equity $E_F(\mathbf{x},\varphi)$, and a pressure factor $\Pi(m,t)$. The Field Equity is the expected value of the next score, based on the current location of the ball and the current phase of possession. The pressure factor is a weighting that is applied to the Field Equity, essentially the effect that next score will have on the outcome of the game. Since this paper is a first attempt at developing an Equity-based rating system, for simplicity it is assumed that Field Equity is independent of the current margin and the time left in the game. We now define:

X_i as the observed location of the i^{th} possession

$X_i \in$ an ellipse 160m x 138m (the dimensions of the MCG, used as the standard field size);

S_i as the next score that occurs after the i^{th} possession (with possessions at the end of each quarter where no further scoring occurs being removed)

$S_i \in \{-6, -1, 1, 6\}$; and

φ_i as the phase of possession i

$\varphi_i \in$ {Set, Uncontested, Loose, Hard, Stoppage, Kick-In, Boundary Restart}.

We can then estimate the Field Equity for a particular phase of possession φ, at location x, as the average of all observed "next score" events that occur within 6m of x.

$$E_F(x,\varphi) = \sum_i S_i \times I_{|X_i - x| \le 6} \times I_{\varphi_i = \varphi}$$

where I is the indicator function. Note that data points where no further score occurred in a game were removed from the analysis, so the next score cannot take on a value of zero. Of the seven different phases of play, only the first four need to have Field Equity values calculated in this manner. This paper is primarily an investigation of the effect of possessions and disposals on player ratings, so in order to keep the rating system as simple as possible, we will make the assumption that neither team has an advantage at stoppages, regardless of field position, meaning stoppages will be assigned zero Equity. O'Shaughnessy (2006) proposed that the negative Equity associated with goal square Kick-Ins was due to the observed inequality between two competing teams, and not due to the location/phase of the Kick-In itself. For the purposes of this paper, however, the expected value of the next score will be used for the Equity of Kick-Ins. From the observed data, it was found that the average next score from Kick-Ins was –0.12, meaning that the team taking the Kick-In was at a disadvantage, so kicking a behind gives your team an average scoreboard advantage of 1.12 points.

Table 1 below contains summaries of the Field Equity values calculated for the seven different phases of possession for all matches in the 2004-2007 seasons. Note that in all areas of the field shown, set possession is worth more than uncontested possession, which in turn is more valuable than loose and hard possession. This is as expected because players with set possession have more time with the football to make an optimal decision. Players gaining hard possession have very limited time and often have to take the first option that presents itself. This difference becomes more obvious as the ball gets closer to goal. In the defensive 50m zone (Def.50), set possession is worth on average 0.19 and 0.67 points higher than uncontested and hard possession, respectively. In the forward 50m zone (Att.50), set possession is worth on average 1.00 and 1.76 points higher than uncontested and hard possession, respectively.

Table 1: Summary of Field Equities for possession phases, with overall range and average values across the whole ground, in the defensive and attacking 50m zones and in the midfield

Play Possession Phase	Count for Possessions	Range in Field Equity Values	Mean Field Equity Values by Region			
			All	Def. 50	Midfield	Att. 50
Set	159148	(-0.92, 5.78)	1.53	0.18	1.29	3.73
Uncontested	177004	(-0.75, 5.70)	1.23	-0.01	1.16	2.73
Loose	64302	(-0.89, 5.42)	1.09	-0.11	1.01	2.59
Hard	54731	(-1.59, 4.48)	0.69	-0.49	0.66	1.97
Stoppage	54573	N/A	0	0	0	0
Kick-In	16624	N/A	-0.12	N/A	N/A	N/A
Boundary Restart	3368	(-0.12, 0.40)	0.12	-0.07	0.08	0.34

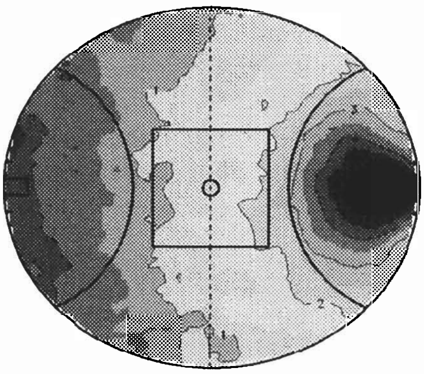


Fig 1: Equity values - Set Possessions.

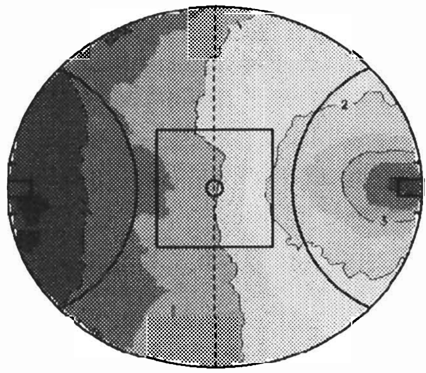


Fig 2: Equity values - Loose Possessions.

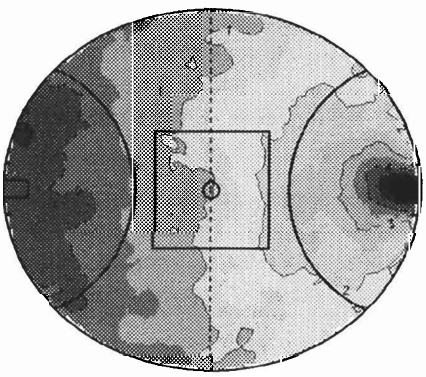


Fig 3: Equity values - Uncontested Possessions.

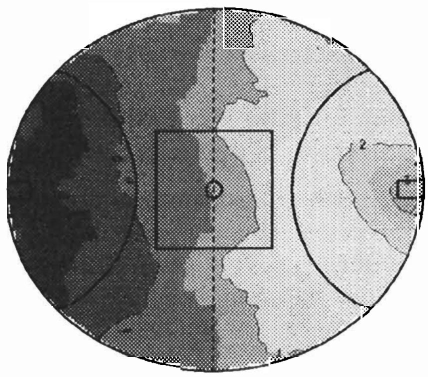
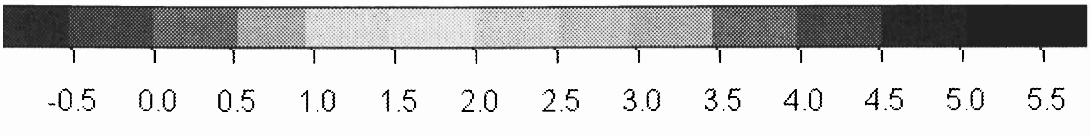


Fig 4: Equity values - Hard Possession



The spatial patterns of average Field Equity values for all four general play possession phases (set, uncontested, loose and hard) can be seen in Figures 1-4 above. All graphs are shown with the team in possession running left to right. That is, the left side of the graph is the defensive zone and the right side is the attacking zone. Note the steep gradient of Field Equity in the forward 50m zone, as compared to the defensive 50m zone, especially for the set and uncontested possession phases. This steep gradient is due to the scoring range of most players in the competition being 40-60m from goal. It was decided at the commencement of this paper that previous assumptions of equality between the left and right sides of the ground would be tested. No significant differences appear to be present, so future analysis will be made using spatial data that has been folded along the centre of the ground, leading to roughly twice as many observations for each location on the ground.

EQUITY RANKINGS FOR PLAYERS

In order to develop a player rating system that is unbiased with respect to player position, it was decided that information about a player's disposal and possession needed to be taken into account. As seen in Figures 1-4, there is a much higher rate of change for Equity in the forward half, and more specifically in the forward 50m zone, than in the defensive half and defensive 50m zone. If only disposals are taken into account, forwards and midfielders would benefit greatly from this increased gradient, to the detriment of defenders. In order to balance this bias, in addition to the result of a player's disposal, the player's possession also needs to be taken into account to calculate their rating. In doing this, we reward defenders for turning over the ball in a position where their opposition had a high Field Equity.

In view of the above discussion we can now define the Equity Value (V_i) of possession i as the difference between the Field Equity resulting from a player's disposal and the Field Equity of the previous possession. i.e.

$$V_i = E_F(X_{i+1}, \varphi_{i+1}) - E_F(X_{i-1}, \varphi_{i-1}).$$

Several methods of evaluating player performance within a game based on these Equity Values are possible, depending on the aim of the ratings system. Three possible systems are:

- Equity per minute, $\Sigma(V_i)$/(Time on Ground)
- Equity per possession, $\Sigma(V_i)$/(Number of Possessions)
- Total Equity per game, $\Sigma(V_i)$

These would be used to evaluate a player's efficiency, disposal/possession quality and total scoreboard value, respectively. Since pre-existing rating systems are based on the performance throughout a single game, Equity per game was used in Table 2. In a concurrent paper, Meyer, et al. (2008) used Equity per minute as a measure of a player's quality and investigated possible influencing factors.

Table 2 below is a summary of the results by playing position of the Equity ratings for games in 2007, sorted by median Equity per game. The correlations presented are between the individual game ratings for the Equity ratings and for Champion Data's (CD's) player rating system. All measures indicate that forward line players are benefiting more from this equity system than defensive players. Reasons behind this and suggested corrections are presented in the Discussion section, below.

Table 2: Total Equity per Game Statistics for Individual Players

Player Position	No. Games	Total Equity per Game for Individual Players					Correlation with CD's Ratings
		Mean	St. Dev	Approx. 95% CI For Mean	Median	Maximum	
Key Forward	676	12.13	8.87	(11.45,12.81)	11.16	54.94	0.910
Gen. Forward	836	11.28	7.80	(10.74,11.82)	10.25	50.19	0.804
Midfielder	2945	10.20	7.07	(9.93,10.46)	9.56	43.82	0.850
Gen. Utility	293	9.27	6.82	(8.47,10.06)	8.43	33.82	0.877
Gen. Defender	1214	8.89	6.06	(8.54,9.24)	8.36	31.76	0.773
Tall Utility	366	8.94	6.77	(8.23,9.65)	8.15	32.29	0.868
Key Defender	795	8.20	5.38	(7.82,8.58)	7.83	32.60	0.771
Ruckman	619	6.64	5.36	(6.21,7.07)	5.91	30.16	0.834
All Positions	**7744**	**9.69**	**7.02**	**(9.53,9.85)**	**8.84**	**54.94**	**0.805**

DISCUSSION

One of the main aims of this paper was to produce a rating system that was unbiased with respect to player position. However, due to the steep Equity gradient in the forward 50m zone and the relatively flat gradient in the defensive 50m zone, as seen in Figures 1-4, players who spend the majority of their time in the forward half are likely to get higher Equity ratings. Table 2 above contains evidence of this, with forwards having a higher Equity rating than midfielders, who in turn rated higher than defenders and ruckmen. General utilities are players who are able to play in the midfield and either forward or back and tall utilities are players who generally play as either forwards or defenders, resulting in both classes of utilities having close to average ratings.

By examining the typical possession patterns of the specific positions, the source of this inequality can be seen more easily. For example, a player that kicks a goal gets a positive increase to their Equity rating of 4.31 points on average, but a missed shot on goal that results in a behind only decreases the player's Equity value by 0.34 points. It is worth noting that the season's maximum Equity per game, Jonathan Brown's 54.94 points in round 16, was the result of kicking 10 goals, 1 behind. In Round 21, Lance Franklin kicked 2 goals, 11 behinds and still managed an above average Equity rating of 16.04 points.

In order to even out the inequalities between forwards and defenders, more thought and planning will need to go into the definition of the Equity value. The current system only takes into account where the ball came from and where it went, not where the actual possession was taken. Using information about the location of the possession could help to even out the ratings. Currently, a missed goal from the goal square has the same negative effect to players' rating as a missed goal from outside the 50m arc, as long as the previous possessions had come from the same location and phase. Further penalising missed goals will reduce the positive bias towards players in the forward line. Key and general defenders should also benefit from using the location of the possession. One possibility is to take into account what might have happened if the player did not get that possession. For example, taking a mark in the defensive goal square could be further rewarded since a failure to do so may have resulted in an opposition mark and a near certain opposition goal.

Since no information was used about their presence at stoppages, ruckmen did not fare well with this ratings system. Their primary role in the game is to win the ball out of a stoppage, playing as either an extra man in defence or attack during general play. If however, stoppage information could be incorporated, this would result in an increase to ruckmen's ratings. Using the raw "next score" data, a first possession from a centre bounce is worth 0.66 points and a first possession from a ball up in the forward 50 is worth 1.86 points. Champion data have a pre-existing statistic called "Hit-out to advantage", which occurs when the ruckman knocks the ball from the stoppage to an unopposed teammate. By combining this information with the location of stoppages and the ensuing possessions, ratings for ruckmen should get closer to the average. In 2007, teams averaged over 7 hit-outs to advantage from centre bounces, so with two ruckmen per team, this should result in an increase of 2.3 points per player from centre bounces alone.

As the Equity rating system gets developed further, more information from the game statistics that are already being recorded by Champion Data could be incorporated. Spoils, smothers, tackles and 50m penalties, among others could all be used to influence a player's rating. Even though these statistics aren't strictly possessions, they have the ability to change the phase of possession and/or the location of the next possession, which both heavily affect the equity.

CONCLUSION

In this paper, it was shown that Field Equity values, as defined by O'Shaughnessy (2006) are approximately the same for the left and right sides of a football field. This is an important result because it means that further analysis can be conducted with data folded across the centre of the ground, essentially doubling the amount of available data. It was also shown that a rating system for players in Australian Rules football could be developed through the use of Field Equity. It is hoped that a different method of combining Field equities to produce ratings will remove the current bias in the favour of forward line players. Early analysis shows promising results for within position comparisons with current rating systems. Further development of the rating system would provide a more accurate representation of the effect of a possession/disposal. This could be achieved with the addition of a weighting based on the state of the game – the "pressure factor". Since Champion Data's existing player rating system incorporates this, integration between the two ratings systems should also achieve this goal. Other measures that were not considered in this paper, such as the Equity per minute and the Equity per possession, may prove to be better estimates of player quality.

Acknowledgements

The author wishes to thank Champion Data for providing the data used in the development of this paper.

References

Bennett, J. (2005) World series player game percentages. Statistics in sports section of the American Statistical Association

Bracewell, P. (2003) Quantifying Individual Rugby Player Performance through Multivariate Analysis and Data Mining. Ph.D. Thesis, Massey University, Albany, NZ

Meyer, D., Jackson, K. and Clarke, S.R. (2008) Multi-level models for player performance in AFL football. In J. Hammond (ed.) *Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM), Coolangatta, Queensland, Australia. (in press).

O'Shaughnessy, D. (2006) Possession versus position: Strategic evaluation in AFL. Journal of Sports Science and Medicine. **5:** 533-540.

Romer, D.H., (2002) It's Fourth Down and What Does the Bellman Equation Say? A Dynamic Programming Analysis of Football Strategy. NBER Working Paper No. W9024

Schatz, A. (2005) Methods to out madness: DVOA explained. Football Outsiders.Available from URL: https://www.footballoutsiders.com/methods.php

WHAT IS THE OPTIMAL ALLOCATION OF SUPER RUGBY COMPETITION POINTS?

Winchester, Niven

Department of Economics, University of Otago, New Zealand

Paper Submitted for Review: 6 April 2008
Revision submitted and accepted: 24 June 2008

Abstract. Competition points are awarded in sports events to determine which participants qualify for the playoffs or to identify the champion. In Super Rugby, bonus points are awarded for scoring four or more tries and/or losing by seven points or less. As a result, there are 17 possible pairwise allocations of competitions points per match. We use competition points to measure strength in a prediction model and choose competition points to maximise prediction accuracy. This allows us to determine the allocation of competition points that most appropriately rewards strong teams. Our prediction model relates the winning margin of the home team to home advantage and travel variables, and indices of competition points earned by the home and away teams using non-linear least squares. The model determines both optimal values (relative to the number of points awarded for a win) and optimal thresholds (in terms of the number of tries or minimum losing margin required to earn a bonus point) for bonus points. We find that the current allocation of competition points is not optimal. Specifically, we argue that the try bonus point should be scrapped and a narrow loss bonus awarded if a team losses by five or fewer points. Our findings also have implications for other competitions.

Keywords: Non-linear least squares; Sports predictions

INTRODUCTION

Administrators of sports competitions involving round-robin or group stages typically award competition points in order to rank participants. These rankings are used to determine which competitors advance to the playoffs or to identify the overall winner. It is, therefore, important that organisers employ allocation criteria that accurately reflect the strength of participants. Bonus points are awarded in a number of sports, including rugby, ice hockey and cricket. We concentrate on Super Rugby as the allocation of points in this competition is more complicated than most others and our modelling framework is well suited to this event.

The inclusion of bonus points often produces a different hierarchy of teams relative to that obtained if bonus points were not included. Despite the influence bonus points can have on the ordering of teams and ultimately the selection of semi-finalists, the appropriateness of the allocation of Super Rugby competition points has not been evaluated. Furthermore, most other rugby competitions have adopted a similar system for allocating competition points, including the Rugby World Cup – reputably the world's third largest sporting event.

We determine the allocation of Super Rugby points that is best at revealing strong teams by using competition points to construct strength measures and choosing competition points to maximise prediction accuracy. Intuitively, maximising prediction accuracy allows us to determine the optimal allocation of competition points, as predictions using strength indices built on an allocation that is not good at revealing strong teams will be less accurate than predictions based on an allocation that is good at identifying strong teams.

This paper has three further sections. Section 2 outlines the salient features of Super Rugby. Our modelling framework and results are set out in Section 3. Section 4 concludes.

SUPER RUGBY

Super Rugby is played under rules set out by the International Rugby Board. Teams earn points during a game by scoring tries (placing the ball over their opponent's goal line) and kicking goals. A try is worth five points and grants an opportunity to kick a conversion, which, if successful, is worth an additional two points (so seven points can be scored in a single scoring play). Teams can also attempt to kick a goal when they are awarded a penalty or attempt a drop goal in general play. A goal is worth three points.

The Super Rugby competition has been played annually since 1996 by provincial/state sides from Australia, New Zealand and South Africa. Between 1996 and 2005, there were 12 Super Rugby teams (five from New Zealand, four from South Africa and three from Australia) and the competition was known as the Super 12. Two extra teams, one each from Australia and South Africa, were added in 2006 and the tournament was renamed the Super 14. The name 'Super Rugby' encompasses both competitions.

Each tournament begins with a round-robin phase where each team plays every other team once. Each team has one bye, so from 1996-2005 each team played 11 games over 12 rounds. The top four teams from this stage qualify for the semi-finals and the two winning semi-finalists contest the final. Hosting rights for the semi-finals and the final are awarded to the team in each contest that gained the highest round-robin ranking. Competition points are awarded at the completion of each match according to several decision rules. A winning team is awarded four points, a losing team zero points and each team earns two points if a match is tied. In addition, bonus points may be awarded for (i) scoring four or more tries, and/or (ii) losing by seven or fewer points. So, a winning team may earn five or four competition points, a team that ties a match may be awarded three or two points, and a losing team may earn two, one or zero points. In total, there are 16 possible pairwise allocations of points per match (5-2, 5-1, 5-0, 4-2, 4-1, 4-0, 3-3, 3-2, 2-2, 2-3, 0-4, 1-4, 2-4, 0-5, 1-5, 2-5).

Geographically, Super Rugby franchises are diverse. The time difference between New Zealand and South Africa is 10 hours, travelling between New Zealand and South African cities can take up to 30 hours, and teams based on South Africa's Highveld are around 2000 meters above sea level while most other teams have coastal headquarters. We distinguish four regions – Australia, New Zealand, South Africa-coastal and South Africa-Highveld – to capture geographic diversity.

Evidence of strong home advantage is that between 1998 and 2005 home teams won 61.3% of matches, away teams won 36.6%, and 2.1% of matches were tied. Further evidence of home advantage is that, on average, home teams score 6.4 more points and nearly one more try per match than away teams.

There are also large variations in average home winning margins across teams, not all of which are due to differences in team ability. Average home winning margins for the Brumbies, Crusaders and Highlanders are 17.0, 14.8 and 11.6 respectively, while corresponding figures for the Bulls and Cats are -1.5 and 0.4 respectively. The data also suggest that home advantage is also likely to depend on the distance travelled by the away team.

MODELLING FRAMEWORK AND RESULTS

The details of many sports ranking/prediction systems are not available to the public because of their application to sports gambling and/or for proprietary reasons. Prediction models in the public domain include Stefani (1980), Zuber et al. (1985), Clarke and Stefani (1992), Clarke (1993), Glickman and Stern (1998), and Bailey and Clarke (2004). To our knowledge, no existing system evaluates the appropriateness of the allocation of competition points.

We use the home team's net score (points scored by the home team minus points scored by the away team) to characterise the outcome of a match and predict match outcomes by regressing net scores on location and net strength variables. Let i denote the home team, j denote the away team, n and s index regions (Australia, New Zealand, South Africa-coastal and South Africa-Highveld), r index rounds and y index years. We specify the following regression equation

$$NSC_{ij,r,y} = \alpha_0 + \boldsymbol{\alpha}_1 \mathbf{D}_1 + \boldsymbol{\alpha}_2 \mathbf{D}_2 + \beta.NSTRN_{ij,r,y} + \varepsilon_{ij,r,y} \quad (1)$$

where $NSC_{ij,r,y}$ is i's net score against team j in round r in year y; α_0 measures base home advantage (applicable to all teams); $\boldsymbol{\alpha}_1$ is a 1x12 vector of team-specific additional home advantage parameters $\{\boldsymbol{\alpha}_1 = (\alpha_1^{Blues},...,\alpha_1^{Waratahs})\}$ $\mathbf{D}_1$ is a 12x1 vector of binary variables equal to 1 if a team played a home match, zero otherwise $\{\mathbf{D}_1 = (D_{ij,r,y}^{Blues},...,D_{ij,r,y}^{Waratahs})\}$; $\boldsymbol{\alpha}_2$ is a 1x16 vector capturing additional home advantage when i is located in region n and j is located in region s $\{\boldsymbol{\alpha}_2 = (\alpha_2^{NZ,NZ},...,\alpha_2^{SA-Highveld,SA-Highveld})\}$; $\mathbf{D}_2$ is a 16x1 vector of binary variables equal to 1 if a team from region n hosted a team from region s, zero otherwise $\{\mathbf{D}_2 = (D_{ij,r,y}^{NZ,NZ},...,D_{ij,r,y}^{SA-Highveld,SA-Highveld})\}$; *NSTRN* is the net strength of team i; β captures the influence of *NSTRN* on *NSC*, and ε is an error term.

In our estimations described below, we drop one team-specific home advantage parameter (α_1^{Blues}) and the four regional home advantage parameters for which $n = s$ to avoid introducing perfect multicolinearity. *NSTRN* is defined as

$$NSTRN_{ij,r,y} = STRN_{i,r,y} - STRN_{j,r,y} \quad (2)$$

where *STRN* measures team strength and is a time-varying weighted average of competition points earned per-game in years y and y-1. In the first match of each year, the weight on competition points earned in the current season is zero and this weight increases by a constant amount after each game. Specifically, *STRN* is calculated as

$$STRN_{i,r,y} = \lambda_{i,r,y} POINTS_{i,12,y-1} + (1 - \lambda_{i,r,y}) POINTS_{i,r-1,y} \quad (3a)$$

where $POINTS_{i,r,y}$ denotes competition points earned per-match by team i in year y at the completion of round r, and $\lambda_{i,r,y}$ is equal to $(11 - g_{i,r,y})/11$ where g denotes the number of games played by team i prior to round r in year y. (Even though there is an even number of teams in the competition, there is not a direct correspondence between the number of games played by a team and the round number as each team has one bye each year.)

Noting that competition points are awarded for winning, tieing and losing by seven points or less yields the following expression

$$\begin{aligned} STRN_{i,r,y} = {} & \lambda_{i,r,y}(\theta^{WIN} WIN_{i,12,y-1} + \theta^{TIE} TIE_{i,12,y-1} \\ & + \theta^{LOSS} LOSS_{i,12,y-1}) + (1 - \lambda_{i,r,y})(\theta^{WIN} WIN_{i,r-1,y} \\ & + \theta^{TIE} TIE_{i,r-1,y} + \theta^{TRY} TRY_{i,12,y-1} + \theta^{LOSS} LOSS_{i,r-1,y}) \end{aligned} \quad (3b)$$

where θ^{WIN}, θ^{TIE}, θ^{TRY} and θ^{LOSS} are competition points awarded for, respectively, winning, tieing, scoring four tries or more, and losing by seven points or less; and $WIN_{i,r,y}$, $TIE_{i,r,y}$, $TRY_{i,r,y}$ and $LOSS_{i,r,y}$ are the average number of matches team i has, respectively, won, tied, scored four or more tries in, and lost by seven points or less in year y at the completion of round r. We replace r with 12 when referring to the average number of wins etc in the previous year as there are 12 rounds in each season. Substituting (2) and (3b) into (1) gives the equation to be estimated.

As the optimal allocation of competition points is invariant to multiplication by any positive scalar, we normalise points with respect to θ^{WIN}. That is, we set θ^{WIN} equal to one and express values for other events attracting competition points relative to the number of competition points awarded for a win.

Our strength measure is well suited to Super Rugby on three grounds. First, there is no relegation or promotion so the same teams play each other each year. Second, probably because rugby has only been professional for just over a decade, most players are local to the province/state they represent and there is not a well-developed transfer market. Indeed, most movements between franchises are by players on the fringe of selection for their 'home' team seeking an opportunity at another franchise.

Third, as each competition begins with a round-robin, each team plays a balanced schedule each year and there is a higher probability that a team has played a schedule of average difficulty as the season progresses. Combined, the three characteristics indicate that it is appropriate to measure a team's strength at the start of each year as average competition points earned in the previous year and increase the weight on current year competition points as the season progresses.

We estimate (1) using 528 round-robin Super Rugby matches played during the years for which there were no changes in Super Rugby teams, 1998-2005. We fix all competition points at (normalised) values currently used in Super Rugby ($\theta^{WIN} = 1$, $\theta^{TIE} = 0.5$, and $\theta^{TRY} = \theta^{LOSS} = 0.25$) in our first regression exercise. In this case all components of *NSTRN* are exogenous. We eliminate insignificant home advantage parameters using a general-to-specific methodology with a single search path. Specifically, we start by including all home advantage parameters (except those omitted to avoid introducing perfect multicolinearity) and eliminate the parameter with the lowest t-statistic in each subsequent estimation until the highest p-value is less than 0.05. We also search for structural breaks in home advantage parameters by allowing values for these parameters to differ pre and post 2002.

We find that all regional home advantage parameters are not significantly different from zero expect those for Australian teams hosting teams from the Highveld ($\alpha_2^{Australia,SA-Highveld}$) and New Zealand teams hosting sides from the same region ($\alpha_2^{NZ,SA-Highveld}$). The results also indicate that additional home advantage parameters are only significant throughout the sample period for the Brumbies ($\alpha_1^{Brumbies}$) and Crusaders ($\alpha_1^{Crusaders}$). The additional home advantage parameter for the Highlanders ($\alpha_1^{Highlanders}$) was also significant but only until the end of 2002. All other home advantage parameters are not significantly different from zero. (The Wald test for the joint significance of omitted home advantage parameters has a p-value of 0.924.)

Results from estimating (1) using ordinary least squares (OLS) and only including significant home advantage coefficients are reported in column (a) of Table 1. The estimates reveal that most teams experience an advantage from playing at home equal to 3.42 game points. Home advantage for the Brumbies and the Crusaders, however, against most teams is equal to 12.14 (3.42 + 8.99) and 10.97 (3.42 + 7.55) respectively. Prior to 2003 the Highlanders enjoyed the largest home advantage of all Super Rugby teams (13.34). Turning to the regional home advantage variables, when an Australian team (except the Brumbies) hosts a team from the Highveld home advantage equates to 19.34 (3.42 + 15.92). Meanwhile, New Zealand teams (except the Crusaders and the Highlanders prior to 2003) entertaining Highveld sides benefit by 11.64 (3.42 + 8.22) points. The impact of location is largest when the Brumbies host a Highveld team and assists the Brumbies by 28.33 (3.42 + 8.99 + 15.92) points. Interestingly, the Brumbies win-loss record against Highveld teams in our sample period is 8-0. Given the geographic dispersion of Super Rugby regions, the large impact of location on match outcome is not surprising. As the influence of home advantage is consistent across specifications, we do not discuss these parameters for other estimations.

Table 1: Regression results

	(a)	(b)	(c)	(d)	(e)
Estimation	OLS	NLS	NLS	OLS	NLS
Try partition	4	4	8	8	-
Narrow-loss partition	7	7	5	5	5
Constrained	Yes	No	No	Yes	No
α_0	3.42***	3.31***	3.47***	3.42***	3.34***
	(0.84)	(0.85)	(0.85)	(0.83)	(0.85)
$\alpha_1^{Brumbies}$	8.99***	9.49***	8.32***	9.09***	9.24***
	(2.64)	(2.65)	(2.66)	(2.61)	(2.62)
$\alpha_1^{Crusaders}$	7.55***	7.76***	7.33***	7.41***	7.52***
	(2.76)	(2.81)	(2.85)	(2.78)	(2.81)
$\alpha_1^{Highlanders(1)}$	9.92***	10.37***	9.10***	9.63***	10.21***
	(2.74)	(2.84)	(2.85)	(2.76)	(2.83)
$\alpha_2^{Australia,SA-Highveld}$	15.92***	15.90***	16.58***	15.88***	15.90***
	(3.59)	(3.60)	(3.62)	(3.60)	(3.60)
$\alpha_2^{NZ,SA-Highveld}$	8.22***	8.31***	8.87***	8.27***	8.47***
	(2.46)	(2.48)	(2.49)	(2.45)	(2.48)
β	13.49***	16.26***	13.12***	15.23***	15.78***
	(2.56)	(3.32)	(3.10)	(2.65)	(2.79)
θ^{TIE}	0.50	0.88	1.26	0.67	0.88
	-	(0.77)	(0.99)	-	(0.78)
θ^{TRY}	0.25	-0.02	1.64	0.33	-
	-	(0.26)	(1.06)	-	-
θ^{LOSS}	0.25	0.57*	0.89*	0.33	0.71**
	-	(0.30)	(0.45)	-	(0.36)
R^2	0.20	0.21	0.22	0.21	0.21
Correct predictions	347	352	356	349	343

Note: ***, **, and * denote significance at the 1%, 5% and 10% significance level respectively. Robust standard errors are reported in parentheses. (1) only significant between 1998 and 2002.

The positive and significant coefficient on *NSTRN* indicates that our strength measure is a significant determinant of match outcomes. The value for β implies that, in the absence of home advantage, a team that wins every match and a collects a try bonus will beat a team that loses every match without earning any bonus points by 16.86 (1.25*13.49) points. Relative to the impact of where the match is played, the strength of the two opponents appears to have a moderate impact on match results. Overall, the model is able to explain about 20% of the variation in the sum of squared net scores and correctly selects the winning team in around two-thirds of matches.

Results from estimating (1) using nonlinear least squares (NLS) are presented in column (b). The estimate for β indicates that the net average number of wins by the home team is a significant determinant of net scores. Estimates for θ^{TIE} and θ^{TRY} are not significantly different from zero, so the average number of ties and the average number of times four or more tries are scored are not significantly correlated with team strength. The estimate for θ^{LOSS} is only different from zero at a 10% significance level, indicating that the average number of losses by seven or fewer points is a weak determinant of team strength.

The average number of ties may be an insignificant determinant of strength as there has not been enough tied matches to accurately gauge the impact of this event on team strength (only 2.1% of matches were tied). The appropriateness of the try bonus can be questioned on the grounds that it is not uncommon for teams that lose by a large margin to earn a try bonus. For example, in round nine of the 1998 competition the Stormers earned a try bonus even though they lost 24-74 to the Blues. This could be because whether or not a losing team earns such a bonus is largely determined by the attitude of the winning team. For instance, a dominant team may decide to bring on bench players and/or play with less aggression/enthusiasm. Support for this hypothesis is that the average losing margin when defeated teams are awarded a try bonus (13.0) is similar to the average losing margin when beaten sides do not earn a try bonus (14.7). We also regress the losing margin on a binary variable equal to one if the losing team scored four or more tries (and zero otherwise). The p-value on the coefficient for the binary variable is 0.420.

Regarding the narrow-loss bonus, perhaps a seven-point margin is not indicative of a close game. After all, such a margin implies that the losing team could earn a narrow-loss bonus if an additional maximum scoring play (a converted try) by this team would have tied the game. Between 1992 and 1995 in New Zealand's National Provincial Championship, teams could only earn a narrow-loss bonus if an additional maximum score by the losing team would have reversed the outcome of the match. So, history suggests that administrators are unsure how to define a narrow loss.

We examine the appropriateness of cut-offs or partitions used for bonus points by estimating (1) for a range of alternative combinations of partitions for try and narrow-loss bonuses. The sum of squared errors is minimised when a try bonus is awarded for scoring eight or more tries and a narrow-loss bonus granted for losing by five or fewer points. The cut-off for the try bonus makes it very unlikely that a losing team will earn such a bonus. The cut-off for the narrow-loss bonus indicates that a defeated team should be awarded such a bonus if at most two additional penalty goals or a converted try by this team are required to reverse the outcome of the match.

Column (c) in Table 1 reports results when these partitions are used. There is a slight improvement in the R^2 and the number of correct predictions. The point estimate for θ^{TIE} suggests that a tie should attract more points than a win, but this estimate is not significantly different from zero. The estimate relating to the try bonus indicates that scoring eight or more tries should attract 1.64 points. Although, unlike the estimate for θ^{TIE}, the estimate for the try bonus relative to number of points awarded for a win is not illogical, such an allocation may be unpalatable to rugby administrators and supporters. In any case, the estimate for θ^{TRY} is not significantly different from zero, although the p-value for this estimate (0.120) is much smaller than the corresponding p-value (0.952) in (b).

Turning to the narrow-loss bonus, the results suggest that losing by five or less points should attract almost 90% of the points awarded for a win. Like in (b), θ^{LOSS} is only different from zero at a 10% significance level but the p-value for this estimate improves from 0.059 in (b) to 0.051. The p-value for joint significance of θ^{LOSS} and θ^{TRY} is 0.098.

Overall, point estimates for competition points awarded for a tie and the two bonuses are higher than interested parties would find agreeable. Consequently, we impose the following constraints when estimating (1)

$$\theta^{TIE} \leq \frac{1}{2}(\theta^{WIN} + \theta^{LOSS}) \tag{4.1}$$

$$\theta^{TRY} \leq \frac{1}{2}\theta^{TIE} \tag{4.2}$$

$$\theta^{LOSS} \leq \frac{1}{2}\theta^{TIE} \tag{4.3}$$

The first constraint places an upper limit on the amount of points awarded for a tie. We allow a tie to attract more than half the number of points awarded for a win on the grounds that the most likely alternative outcome to a tie is a narrow loss for one team, and our specification allows teams that tie to share the competition points awarded for a win and a narrow loss. Constraints (4.2) and (4.3) stipulate that bonuses should be less than or equal to half the number of points awarded for a tie, as in the allocation currently used in Super Rugby.

Results from estimating (1) with the appropriate constraints imposed and determining optimal thresholds for bonus points are reported in column (d) of Table 1. As expected, a lower R^2 and fewer correct predictions are associated with the constrained model than the unconstrained model, but the model does better on both of these criterion than (a). The p-value for the joint Wald test of the appropriateness of the constraints (0.525) indicates that the data cannot reject the restrictions. Overall, the results suggest that an allocation that awards three points for a win, two points for a tie, one point for scoring eight or more tries, and one point for losing by five or fewer points is marginally better at identifying strong teams than the current allocation.

Our results suggest that, after controlling for the number of wins and narrow losses in previous matches, the number of try bonuses earned by a team does not increase that team's predicted net score. In other words, offering a try bonus encourages teams to play a style of rugby that does not increase the probability of winning. Anecdotal evidence also supports our assertion that the try bonus is not correlated with team strength. In the quarterfinals of the 2007 Rugby World Cup, New Zealand and Australia, two teams heavily favoured to advance to the next round, lost to France and England respectively. Several rugby experts, including Australian coach John Connolly, suggested that the attacking style adopted by the two favourites was partly responsible for the unexpected results. In turn, the incentive structures in place in the competitions that these teams regularly participate in may influence playing styles. Specifically, New Zealand and Australia compete in the Tri-Nations competition where a try bonus point is offered, and France and England participate in the Six Nation tournament, where a try bonus is not offered.

We report results from estimating (1) when the try bonus is dropped in column (e) of Table 1. The larger estimate for β reveals that the average number of wins has a greater influence on net score than in all other specifications except (c). Additionally, unlike in other specifications, the point estimate for θ^{TIE} is not irrational and the p-value for the significance of θ^{LOSS} is less than 0.05. We drop the try bonus and impose constraints (4.1) and (4.3) in an unreported specification. The data cannot reject the constraints (the p-value for the joint test of the constraints is 0.584). Interestingly, the data can also not reject the joint test $\theta^{TIE} = 0.5$ and $\theta^{LOSS} = 0.25$ (the p-value for this test is 0.437). This suggests that dropping the try bonuses and changing the losing margin required to earn a narrow loss bonus to five points will improve strength accuracy while requiring minimum changes to the current allocation.

CONCLUSIONS

We determined the allocation of competition points that most appropriately rewards strong teams. We focused on Super Rugby as the allocation of points in this competition is relatively complicated and several features of this tournament fit our modelling strategy.

We found that the bonus awarded for scoring four or more tries is not significantly correlated with team strength and that the bonus for losing by seven or fewer points is only a weak determinant of strength. If competition points are allocated solely to reward strong teams, a try bonus should not be awarded but a bonus should be awarded for losing by five points or less. In addition to distorting league tables, it could be argued that the try bonus encourages teams to play in manner that does not increase the probability of winning. The finding that the (modified) narrow-loss bonus is a significant determinant of team strength suggests that it may be beneficial for other sports competitions to adopt such a bonus, although the different nature of alternative sports may give quite different results.

Before closing we note that an alternative allocation of competition points may encourage teams to behave differently to that observed in our sample. Although, given the competitive nature of most professional athletes, it seems reasonable to assume that teams wish to win and prefer a narrow loss to a large loss, there are several cases where shifting the 'goal posts' may alter teams' actions. First, a team awarded a kickable penalty near the end of a match and behind by seven points will be more likely to attempt to score a try rather than opting for a shot at goal if a bonus is awarded for losing by seven or fewer points than if a loss by five or less is required. Second, a coach whose team has scored four tries and has a comfortable lead will be less likely to substitute key players if eight or more tries are needed for a try bonus than if only four tries are required.

Acknowledgements

I would like to thank Ray Stefani, Dorian Owen, Stephen Dobson, Liam Lenten, an anonymous referee and seminar participants at the Universities of Colorado (at Boulder) and Wyoming, and La Trobe University for helpful comments and suggestions. Remaining errors are my responsibility.

References

Bailey, M. J. and Clarke, S. R. (2004) Deriving a profit from Australian Rules Football: A statistical approach. In R. H. Morton and S. Ganesalingam (eds.) *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport.*, Palmerston North, Massey University, 48-55.

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

Clarke, S. R. and Stefani, R. T. (1992) Predictions and home advantage for Australian Rules Football. *Journal of Applied Statistics*. **19(2)**: 251-259.

Glickman, M. E. and Stern, H. S. (1998) A state-space model for National Football league scores. *Journal of the American Statistics Association*. **93(441)**: 25-35.

Stefani, R. T. (1980) Improved least squares football, basketball and soccer predictions. *IEEE Transactions on Systems, Man and Cybernetics*. **10**: 116-123.

Zuber, R. A., Gandar, J. M. and Bowers, B. D. (1985) Beating the spread: testing the efficiency of the gambling market for National Football League games. *Journal of Political Economy*. **93(4)**: 800-806.

PERFORMANCE AND LEARNING OF MOTOR SKILLS: A CONSTRAINTS-LED PERSPECTIVE FOR STUDYING HUMAN MOVEMENT SYSTEMS

Renshaw, Ian

School of Human Movement Studies, Queensland University of Technology, Queensland.

KEYNOTE ADDRESS

Abstract. Given the challenges in increasing activity levels of the population, an important factor in enhancing lifelong adherence to sport and physical activity are the early sporting experiences of children. To this end, coaching practice needs to be based on a principled application of motor learning theory. Current research using interceptive cricket actions as the task vehicle has demonstrated that movements are tightly coupled to information. These findings provide support for the constraint-led model proposed by Newell (1986) and highlight the need to develop learning approaches that adopt a more hand-off approach than traditional coaching.

Keywords: constraints-led; motor learning

INTRODUCTION

Around the world there is growing concern over levels of physical activity and their impact on population health. To arrest this 'epidemic' there has been increased recognition of the role of physical activity in exercise and sport. In Australia, the importance of physical activity appears to be highly valued with many children participating in organised sport and being "athletic" is closely associated with personal growth and development. However, in 1995 the proportion of overweight or obese children and adolescents in Australia was 23% and 6% respectively (Booth et al., 2001). These figures demonstrate that the prevalence of overweight children had almost doubled, and the prevalence of obese children more than tripled over the previous decade (Waters & Baur, 2003). Given these alarming statistics and the likely knock on effect on health as these children grow up, how then do we encourage children to be more active? Although millions of dollars are being spent on well meaning exercise programmes, perhaps the key to lifelong activity is to enable children to become skilful sports performers?

An important factor in enhancing participation in sport and physical activity are the early experiences of children in sport. The importance of high quality experiences from the 'first coach' is well known in the sport literature and it appears that these early experiences act to facilitate the motivation and interest needed to continue participation in sports and physical activities throughout the lifespan (Martens, 2004). To this end, interaction between movement scientists and pedagogists is important for the development of adequate models of skill acquisition in the teaching and coaching of sport. However, the relationship between motor learning and practitioners over the years has not been as effective as it could have been. One reason for this may have been that the adoption by many researchers of non-representative laboratory experiments at the expense of applied work has produced few practical, empirically verified recommendations for physical education teachers and coaches (see Hoffman, 1990 ; Locke, 1990). To alleviate this problem, Newell and Rovegno, (1990) proposed that the theory and practice of motor skill acquisition required an inter-disciplinary perspective and the emergence of a new theory (the ecological approach to perception and action) was seen as the catalyst for future applied research using natural multiple degrees of freedom tasks. Indeed, the last few years has seen an increase in motor control research adopting this theoretical approach to the extent that the ideas and pedagogical approaches of the constraints-led approach are creating considerable interest and are being taken up by practitioners at all levels (Chow, Button, Renshaw, Shuttleworth & Davids, 2008; Rink, 2008). In this paper, I will briefly introduce the constraints-led approach and then discuss some of my research that has used the interceptive actions of cricket to demonstrate how motor control research can provide theoretical and applied advances to meet the needs of academia and the practitioner. I will finish by highlighting how adopting the ideas of a constraint-led approach can enhance contemporary teaching and coaching practice to facilitate positive learning experiences for children and adults.

CONSTRAINTS ON PERFORMANCE

The principles and ideas of ecological psychology and dynamical systems can be used to provide a framework for teaching and coaching practice (Renshaw, Davids, Shuttleworth, & Chow, in press). The constraints-led perspective sits within ecological psychology and contrasts with the cognitive models that have traditionally driven practice design in sports settings. Constraints have been defined as boundaries which shape the emergence of behaviour from a movement system (e.g., learner) seeking a stable state of organization (Newell, 1986). Newell classified constraints into three distinct categories to provide a coherent framework for understanding how movement patterns emerge during task performance. The three categories of constraints are: performer, environment and task. *Performer constraints* refer to the unique structural and functional characteristics of learners and include factors related to their physical, physiological, cognitive and emotional make up. *Environmental constraints* refer to physical factors such as the surroundings of learners including gravity, altitude and the information available in learning contexts, such as amount of light or level of noise in a gymnasium or sports field. Other important physical environmental constraints include the parks, backyards, empty spaces and alleyways that provide the backdrop for early sport experiences of many active children. Finally, *task constraints* are perhaps the most important constraints for movement educators because of their significance in learning. They include the goal of the specific task, rules of the activity and the implements or equipment used during the learning experience.

Research from a constraints-led perspective

A key performer constraint in fast ball sports is perceptual discrimination ability which has been shown to be a function of previous experience and exposure to appropriate sport specific information (Abernethy, 1987; Renshaw & Fairweather, 2000; Williams, Davids, & Williams, 1999). Thus, players need to have extensive experience of the movement patterns of 'opponents' in order to be able to effectively perceive in sport. These person-related factors (Heft, 2003) act to delimit affordance possibilities and as such play a significant role in determining the style of play adopted by individuals. Given the importance of high levels of perceptual skill for expert performance, interest has centred on the most effective ways to develop these skills in sports such as cricket. In a recent study, (Renshaw, Fairweather, Oldham, & Rotheram, 2004) examined the efficacy of implicit and explicit learning approaches by attempting to train club level cricket batters to identify legspin and googly deliveries by 'batting' against national wristspin bowlers on a full-size video-screen. Batters were split into groups where one group received no 'coaching' information about the bowler and ball types, while a second group were given 'important' information about the position of the wrist and hand at release. A third group acted as a control group; they watched videotape of test matches. Results show that both implicit and explicit video-based training programmes improve the perceptual discrimination capability of the batters in a coupled perception-action task that replicates batting in cricket. Furthermore, over the long-term, trends suggest that implicit practice may be more robust. Indeed, this was the case for performance against a new spin bowler of similar standard to the other bowlers at short-term retention. A key message for coaches and players is that providing additional 'coaching' advice to players provides no advantage in development of perceptual discrimination ability and that simple 'exposure' leads to an improvement in performance. The process of differentiating legspinner from googly deliveries appears to occur by batters attuning to the invariant information provided within the bowlers' body actions.

Knowledge of the way that informational constraints affect movement means that the coach can deliberately change the constraints to enable the player to attain new movement solutions. Since perception is specific to environmental properties uniquely constraining each performance situation, changing the ecological constraints of practice can deeply influence the movement behaviours that emerge (Beek, Jacobs, Daffertshoffer, & Huys, 2003). An important role of practice is to educate learners to pick up constraining perceptual variables rather than non-constraining (less relevant) variables in specific and relevant practice contexts (Jacobs & Michaels, 2002). However, practice environments have traditionally been adapted to manage the information load on learners by decomposition of the movement model into micro-task components. For example, in cricket, bowling machines allow accurate and stable projection of balls to enable acquisition of batting skill in isolation from game contexts. The problem is that experienced performers use pre-ball flight information to constrain coordination modes, as revealed by studies of cricket batting (e.g., Renshaw and Fairweather, 2000). In a recent study, we required high intermediate level batters to face a medium pace bowler and a bowling machine set to the same speed. Analysis of the kinematics of the forward defensive shots played by the batters demonstrated that using bowling machines leads to batsmen developing different timing and co-ordination patterns for balls delivered at the same speed (Renshaw, Oldham, Golds, & Davids, 2007). These findings bring into question the efficacy of using ball projection machines in training and suggest that the most effective way to improve both perception and action is by batting against real bowlers.

Run-ups are not stereotyped

The task demands of sporting run-ups provide motor control researchers with ecologically valid opportunities to develop theoretical understanding of movement control. Additionally, research can enhance understanding for the practitioner. In cricket bowling, given that 'no-balls' are a major problem for many bowlers, providing a better understanding of the run-up is of vital importance in order to help us understand why they occur. Of particular interest is to see if the widely held notion (in cricket circles) that run-ups are stereotyped is in fact correct. However, previous research in long jumping has shown that these run-ups are controlled by visual regulation with jumpers making step adjustments at the end of their run-ups in order to successfully hit the take-off board (Montagne, Cornus, Glize, Quaine, & Laurent, 2000). As cricket run-ups have a different nested task constraint embedded at the end of the run-up (i.e. bowling rather than jumping), theoretically, we wished to see if the task constraint of cricket bowling led to bowlers demonstrating similar locomotor control strategies as long jumpers. On a practical level, a further aim was to provide useful applied information that would increase coaches and bowlers understanding of bowlers' run-ups in order to develop a greater understanding of why no-balls are bowled. In order to explore these issues (Renshaw & Davids, 2004, 2006) examined the run-ups of 6 professional cricket bowlers. Analysis using both inter-trial and intra-trial analyses revealed that bowlers did not produce stereotyped run-ups. In fact although bowlers demonstrated inconsistent start points (they did not 'hit' their marker) they still managed to achieve very low levels of variability in their footfall position at the bound step (mean=0.11 m, range=0.08-0.16 m). In general, most bowlers achieved this level of accuracy by making step adjustments early and late in their run-up. On an individual level, bowlers demonstrated individualised run-up patterns, adjusting their steps 'on-line' as and when they needed to do so, providing further support for the prospective view of locomotor control. At an applied level, the results show that stereotyped 'error' free run-ups are not possible and that run-up variability is not only to be expected but is a functional component of successful bowling performance. In a related case-study (Renshaw, Rotheram, Kemshall, Wilkinson, & Davids, 2003) we found differences in step patterns when the umpire and/or stumps were not present during the run-up. Once again, these findings have important implications for the design of practice for sports skills and highlights the importance of maintaining appropriate information sources to facilitate the development of appropriate information-movement couplings.

New approaches to coaching

The research findings of the studies discussed in this paper allow us to make some points with regards to coaching. The traditional approach in which perceptual and decision-making skills are only introduced after 'basic' technical skills have been 'grooved' is strongly rejected by the findings of these studies. Coaches should strive for holistic development of players; however, this does not mean that approaches have to be over complex. In fact, coaching should be based on task simplification as opposed to the traditional methods that use task decomposition. To use task simplification it is important that the coach understands the key control parameters for movement tasks within a task, as this means that he can manipulate the underlying constraints to guide learners to discover their own solutions. Essentially, if a more hands-off approach to coaching is adopted, performers are capable of exploring the task constraints and solving problems in unique ways that are appropriate to their individual capabilities. Pedagogical approaches that fit in with this nonlinear pedagogy (Chow et al., 2006) include guided discovery, discovery learning and differential learning (see Schöllhorn et al., 2006). Additionally, the constraint-led model provides a suitable theoretical framework for coaches to use the Teaching Games for Understanding (Bunker & Thorpe, 1982) approach in their coaching. Adopting these approaches provides opportunities for performers to be autonomous, to demonstrate competence and relate to others while solving problems. Not only will this lead to the development of more intelligent games players, it is also likely to impact on the player's intrinsic motivation and hence their adherence to the practice and performance programme.

To conclude this section, Newell's (1986) constraints model provides an excellent conceptualisation to guide the design of practice because it adequately captures the rich range of diverse constraints acting on learners during skills learning and games participation. It provides a framework which emphasises the important interactions of personal, environmental and task constraints in a balanced perspective. Knowledge and adoption of the ideas of this approach will provide teachers and coaches with an alternative viewpoint to motor learning and lead to radical changes in the design of pedagogical practice.

Summary

In this paper I have demonstrated how adopting a constraints-led approach can enhance the quality of coaching experiences for children and adults. The principles discussed predicate an approach to instruction that facilitates learning using natural self organising processes under constraints. Coaches should adopt a more hand-off approach to learning and minimise potential disruption to performance by unnatural explicit instruction. The research findings discussed highlight the importance of promoting natural implicit learning by creating environments that enable performers to engage in exploratory behaviour to learn many fundamental movement skills without recourse to verbal instruction. However, this is not to say that teachers and coaches should merely allow 'free play' and hope that learners complete a set task/ game situation in whatever way the learners deem appropriate! Coaching requires a sound understanding of the principles of ecological psychology and dynamical systems theory as well as the principles of play for specific sporting activities. With this knowledge, coaches can manipulate informational, task and environmental constraints to provide developmentally appropriate practice activities that facilitate learners to discover movement solutions in dynamic settings.

References

Abernethy, B. (1987) Selective attention in fast ball sports II: Expert-novice differences. *Australian Journal of Science and Medicine in Sport, December*, 7-16.

Beek, P. J., Jacobs, D. M., Daffertshoffer, A. and Huys, R. (2003) Expert performance in sport: Views from the joint perspectives of ecological psychology and dynamical systems theory. In J. L. Starkes and K. A. Ericsson (eds.), *Expert performance in sports: Advances in research on sport expertise* 321-344. Champaign: Human Kinetics.

Booth, M., Wake, M., Armstrong, T., Chey, T., Hesketh, K. and Mathur, S. (2001) The epidemiology of overweight and obesity among Australian children and adolescents, 1995-1997. *Australian & New Zealand Journal of Public Health,* **25:** 162-169.

Bunker, D. and Thorpe, R. (1982) A model for the teaching of games in the secondary schools. *The Bulletin of Physical Education*, 5-8.

Chow, J. W., Button, C., Renshaw, I., Shuttleworth, R. and Davids, K. (2008) *Non-linear pedagogy: Implications for TGfU.* Paper presented at the Teaching games for understanding: celebrations and cautions.

Chow, J. Y., Davids, K., Button, C., Shuttleworth, R., Renshaw, I. and Araújo, D. (2006) Nonlinear pedagogy: A constraints-led framework to understanding emergence of game play and skills. *Nonlinear Dynamics, Psychology and Life Sciences,* **10:** 71-103.

Heft, H. (2003) Affordances, dynamic experience, and the challenges of reification. *Ecological Psychology,* **15:** 149-180.

Hoffman, S. J. (1990) Relevance, application, and the development of an unlikely theory. *Quest, 42,* 143-160.

Jacobs, D. M. and Michaels, C. F. (2002) On the paradox of learning and realism. *Ecological Psychology,* **14:** 127-140.

Locke, L. F. (1990) Why motor learning is ignored: A case of ducks, naughty theories, and unrequited love. *Quest 42,* 134-142.

Martens, R. (2004) *Successful coaching.* Champaign: Human Kinetics.

Montagne, G., Cornus, S., Glize, D., Quaine, F. and Laurent, M. (2000) A perception-action coupling type of control in long jumping. *Journal of Motor Behavior,* **32:** 37-43.

Newell, K. M. (1986) Constraints on the development of co-ordination. In M. G. Wade and H. T. A. Whiting (eds.) *Motor development in children: Aspects of co-ordination and control.* Dodrecht: Martinus Nijhoff, 341-360.

Newell, K. M. and Rovegno, I. (1990) Commentary-Motor Learning: Theory and Practice, *Quest, 42,*184-192.

Renshaw, I. and Davids, K. (2004) Nested task constraints shape continuous perception-action coupling control during human locomotor pointing. *Neuroscience Letters,* **369:** 93-98.

Renshaw, I. and Davids, K. (2006) A comparison of locomotor pointing strategies in cricket bowling and long jumping. *International Journal of Sport Psychology,* **37:** 1-20.

Renshaw, I., Davids, K., Shuttleworth, R. and Chow, J. W. (in press). Insights from Ecological Psychology and Dynamical Systems Theory Can Underpin a Philosophy of Coaching. *International Journal of Sport Psychology.*

Renshaw, I. and Fairweather, M. M. (2000) Cricket bowling deliveries and the discrimination ability of professional and amateur batters. *Journal of Sports Sciences,* **18:** 951-957.

Renshaw, I., Fairweather, M. M., Oldham, T. O. and Rotheram, M. *Picking the wristspinner: Assessing batters perceptual skills using spatial occlusion techniques.* Paper presented at the 2nd World Congress of Sport Science & Cricket, Cape Town, 2003.

Renshaw, I., Oldham, T. O., Golds, T. and Davids, K. (2007) Changing ecological constraints of practice alters coordination of dynamic interceptive actions. *European Journal of Sport Science,* **7:** 157-167.

Renshaw, I., Rotheram, M., Kemshall, S., Wilkinson, R. and Davids, K. (2003) *Manipulating perceptual information and cricket bowling: A case study.* Paper presented at the Proceedings of the International Australasian Winter Conference on Brain Research, Queenstown, New Zealand.

Rink, J. (2008) *Teaching games for understanding: celebrations and cautions.* Paper presented at the Teaching games for understanding International Conference 2008.

Schöllhorn, W. I., Michelbrink, M., Beckmann, H., Trockel, M., Sechelmann, M. and Davids, K. (2006) Does noise provide a basis for the unification of motor learning theories? *International Journal of Sport Psychology,* **37:** 186-206.

Waters, E. B. and Baur, L. A. (2003) Childhood obesity: modernity's scourge. *Medical Journal of Australia,* **78:** 422-423.

Williams, A. M., Davids, K. and Williams, J. G. (1999) *Visual perception & action in sport.* London: E & F.N Spon.

THE USE OF IMAGES IN AN ITEM-PERSON MAP

Haynes, John

School of Education, University of New England, Armidale, NSW, Australia

Paper Submitted for Review: 31 March 2008
Revision submitted and accepted: 24 June 2008

Abstract. Results that employ Rasch (1960) analysis are often reported in the form of an item-person map, which show the relationship between two key variables: item difficulty estimates and person ability estimates. The map displays results numerically in the form of logits. This paper presents the results of an analysis, using the *Quest* (Adams & Khoo, 1993) statistical package, for individuals (N = 117) who performed the fundamental skill of the basic forward roll, in gymnastics. Initially the *Quest* package was employed to confirm the underlying construct, namely, the quality of the individual's performance. However, within the realm of the psychomotor mode of learning, the use of numbers, as they are traditionally employed in reporting statistical results, may not achieve a 'user friendly' status for either physical education teachers or sporting coaches. Because data were recorded using digital images, the results could be reported through the linking of photographs to in the step difficulties previously only presented in numerical format. These images depict what the different aspects of the roll actually 'look like', whilst still maintaining the integrity of the item-person map. As such, the degree of difficulty were visually linked to the quality of student performances for the forward roll. The individuals who took part in this study represent an a sample who are at both ends of the limits of ability to perform a roll. Their ages ranged from 4 years to 42 years, and comprised three cohorts, namely children, young adults and older adults. This approach brings together previously unconnected elements of data, analysis and interpretation of results.

Keywords: Rasch, images

INTRODUCTION

This article reports on a novel method of presenting statistical data, which demonstrates that sometimes a series of pictures can convey, for some individuals, more information than a thousand numbers! Using snapshot pictures, gleaned from a digitally recorded movie, information can be rendered in a "less complicated" fashion that may enhance the accessibility of the data for a wider audience, more specifically for those individuals who may be visual learners. For the presentation of the data in question, the meaning and significance of the pictures are no less meticulous than if they were presented in numerical format.

The individuals who performed the gymnastics skill, termed, the basic forward roll, were filmed using digital video. These data were then transferred to CD Rom, which enabled freeze frame and slow motion analysis, as well as the extraction and transfer of individual frames into different electronic documents. Filming was conducted from three perspectives, namely, the side (laterally), front (anterior) and behind (posterior). For the purpose of analysis the froward roll was divided into three hypothesised sequences, the *beginning, bridging* and *end*. Following close deductive analysis of the visual data, a number observational cues were identified. These cues were found to be useful for assisting in the determination of the quality of a movement performance. The term *indicator,* was applied to the cues, and refers to the position of selected anatomical structures of the performer. In addition, more fine-grained subdivisions within each indicator were identified; these were termed *descriptors*. The application of the descriptors to a performance was finally employed to determine the quality of movement for the forward roll. This measure of movement quality has not previously been catered for in forward roll assessment models, such as those proposed, for example by Roberton and Halverson (2006) or Gallahue and Ozmun (2006). The descriptors were coded and subjected to the Rasch (1960) statistical analysis partial credit method to verify the underlying construct, i.e. the varying levels of quality performance of the forward roll (Haynes, Miller, Callingham, & Pegg, 2005).

THE TAXONOMY

For each of the three hypothesised sequences of the forward roll, there is a number of indicators, and for each indicator there is an hierarchical set of descriptors, which provide a more precise account of the limb and body position. Both indicators and descriptors are used to assist with deeper analysis of the quality of movements exhibited within and between cohorts. Table 1 shows the descriptors, the indicators and their accompanying abbreviation for each sequence of the forward roll.

Table 1: Taxonomy for Determining the Quality of the Forward Roll

Sequence	Descriptor	Indicator and code
Beginning	Position and placement of the hands on the surface relative to shoulders	1. Shoulder width (sw) 2. Close to shoulder width (c) 3. Wide of shoulder width (w)
	Position of the arms and the elbows	1. Straight arms (st) 2. Elbows bent back (bb 3. Elbow bent laterally (sb)
	Part of the head making contact with the surface	1. No head contact (nc) 2. Back of head (ba) 3. Crown of head (cr)
	Number of body parts in contact with the surface at the commencement of the roll	1. Two 2. Four 3. Three 4. Five 5. Six or more
Bridging	Position of the hips and knees	1. Straight-bend-contact (sbc) 2. Remain bent (b) 3. Stay straight (ss) 4. Bend straighten (bs)
	Position of the arms during rotation	1. Arms arc 180° (as) 2. Arms bent straighten (af) 3. Arms 'V' shape (av) 4. Arm rotation with body (ar) 5. Rotate to elbow (ae)
End	Position of the feet at the end of the rotation	1.Close to buttocks (cb) 2. Away from buttocks (ab) 3. Inconsistent (in)
	Leg movements	1. Together (lt) 2. Knees/feet apart (kfa) 3. Legs separated (la)
	Rotational factors contributing to errors when attempting to attain a standing position	1. Rise to stand unaided (ru) 2. Balance is lost after rising (lb) 3. Roll pauses (sm) 4. Uses hands to assist rise (uh) 5. Roll stops (rs)

The descriptors, which are presented in attenuated format and accompanied acronym, shown in Table 1, are arranged in hierarchical order from the most to the least ideal, based upon the descriptions provided by George (1980). The first descriptor from each indicator determines a performance of highest quality. The last descriptor for each indicator delineates the lowest quality. However, for each individual descriptor the degree of quality of performance level can vary.

For the beginning sequence, for example, the hand position for the first descriptor points to a high quality performance, the second to a medium, whereas the third represents low quality. Likewise for the arm/elbow the first descriptor relates to high quality, the second and third descriptor medium and the fourth descriptor low quality. The first descriptor for the head position indicates the highest quality, the second descriptor indicates slightly less quality, and the third descriptor indicates low quality. To gain a complete picture of a performance this process is repeated for the bridging and end sequences.

Highest quality performances display all the first mentioned descriptors. Medium quality performances may show some descriptors from high quality and some from medium quality. Low quality performances will show mainly descriptors at the end of each descriptor list. Performers may not always exhibit descriptors that place them in discrete high/medium/low categories, which suggests the possibility of the existence of a performance continuum.

RASCH ANALYSIS

Rasch measurements are particularly suited to investigations in the wide range of human sciences (Bond & Fox, 2001:189), which according to these authors is the only technique generally available for constructing such measures. Rasch "provides a useful method for attaining approximate measures that assist with the understanding of the processes underlying the reason why people and specifically chosen items behave in a particular way" (Bond & Fox, 2001:19).

The hierarchically ordered descriptors, shown in Table 1, were allocated a numeral to represent their ranking, from high to low quality. For coding purposes the descriptors judged to be the lowest quality were allocated a zero, through to highest quality, which was given a sequential numeral depending on the number of descriptors. For example, for the beginning sequence the descriptors were allocated the following numerals: for the hand position of "shoulder width" '2', close to shoulder width '1' and wide of shoulder width '0.' This procedure was repeated for each descriptor for the other indicators. Each sequence was allocated an alphabetical code. In addition, the codes for the beginning sequence are all given the initial 'B', for bridging sequences (or middle) have the letter 'M' was used, and for the end sequences, 'E'. Each indicator therefore has a unique cryptogram. Table 2 provides a summary of the indicator codes, the acronym for each indicator and a numerical descriptor range.

Table 2: Indicator Codes and Descriptor Range for the Three Sequences

Sequence	Indicator	Code	Descriptor Range
Beginning (B)			
	Hand position	BAH	0-2
	Arm/elbow	BEA	0-2
	Head contact	BHT	0-2
	Contact points	BCP	0-4
Bridging (M)			
	Hip/knee	MHK	0-3
	Shoulder/arm	MSA	0-4
End (E)			
	Foot placement	EFT	0-2
	Final Leg movements	ELM	0-2
	Final rotational movements	ERM	0-4

The descriptor range as shown in the final column in Table 2, reflects the range of the movement quality. Note that the range may differ for each indicator.

Table 3 shows an example of the coding for a single person. Notwithstanding, all participants (N = 117) were allocated a code based upon how their own individual performance matched the Taxonomy.

Table 3: Example Of Data Code For A Single Participant

Indicator Code	BAH	BEA	BHT	BCP	MHK	MSA	ELM	EFT	ERM
Descriptor Code	**0**	**0**	**1**	**1**	**2**	**1**	**1**	**1**	**2**

Row one, in Table 3, shows the indicator codes used for all participants. The second row, labelled descriptor code ranking, provides an example of a range numbers that may be applicable to a single participant.

Following the coding of each participant the software package, *Quest*, developed by Adams and Koo's (1993) for the Australian Council of Educational Research, was used to calculate Rasch scores. *Quest* output includes estimates of item difficulty and subject ability, and as the data for the forward roll was scored polytomously the partial credit form of the Rasch model was employed. The partial credit model specifically incorporates the possibility of having differing numbers of steps for different items on the same test (Masters, 1982). An additional feature of this model is that "when a variable indicating a single particular construct has been identified, within a targeted population, the measurement of the subject's ability (in this case movement quality) is independent of the set of items that were administered, and the item difficulty is independent of the set of persons used to calibrate the item" (Snyder & Sheehan, 1992: 88).

Attempting to provide a more objective analysis can provide further information about the subjects' movement quality whereby, "meaning is added to the qualitative analysis" (Bond & Fox, 2001:14). To this end an estimate of the difficulty ranking is provided for the items by comparing them with the subjects' success rates, producing fit statistics that aid in the identification of the discriminatory nature of the items. This information is usually presented numerically, however, due to the nature of the data employed in the analysis of the forward roll, still images could be substituted for the numbers.

RESULTS

The output from *Quest* includes the item estimates and reliability statistics. The item reliability index provides an indication of the degree to which the range and distribution of item difficulty levels is sufficient to differentiate between subjects of near equal movement quality.

Item Estimates

In summary, the item estimates produced a mean of 0.00, SD= 0.91 and a reliability estimate = 0.71. The fit statistics were close to 1 (1.01 and 1.00), for the unstandardised fit estimates with both the infit and outfit mean squares showing little spread from the ideal. The *t* values of 0.10 and 0.13 indicated that the items were useful for the sample of subjects. No items were too easy or too difficult, which also verified this output information.

Case Estimates

The output for the case estimates and reliability statistics indicated that the reliability of the person movement quality was high at 0.87. The mean of the infit squares at 0.99 and the outfit means squares at 1.00 were close to or the same as the Rasch modelled expectations of 1.00. Consequently, the standardised fit *t* values were around zero (-0.01 and 0.15). The mean person estimate (i.e. group average) was close to 0 indicating a well-matched item list.

Item Fit

The *Quest* program generates an item infit mean square map that identify those items with infit mean square values that fall inside and/or outside the interval of 0.77 and 1.30. This is the interval, suggested by

Wright and Stone (1979) as being an acceptable fit, however, Wright and Masters (1982) consider values within the threshold of .70 are acceptable

Results of the analysis indicated that there were no infit mean square values less than 0.77 or greater than 1.30, i.e., the items were all elements of the same construct, namely, the quality of the forward roll.

Item Person Map: Pictorial Representation

Quest software includes an item-person map in which person movement quality and item difficulty relations can be seen in the form of an item fit map. Estimates of fit and error can be tabulated along with movement quality and difficulty estimates. Item difficulty is expressed in terms of logits: zero is average, negative easier, and positive becoming more difficult. Person movement quality is estimated in relation to item difficulty estimates the higher the positive values the better the quality of movement (Bond & Fox, 2001: 33). The numerical version of the whole item person map may be found at the ICHPER: SD Conference (Haynes, 2006)

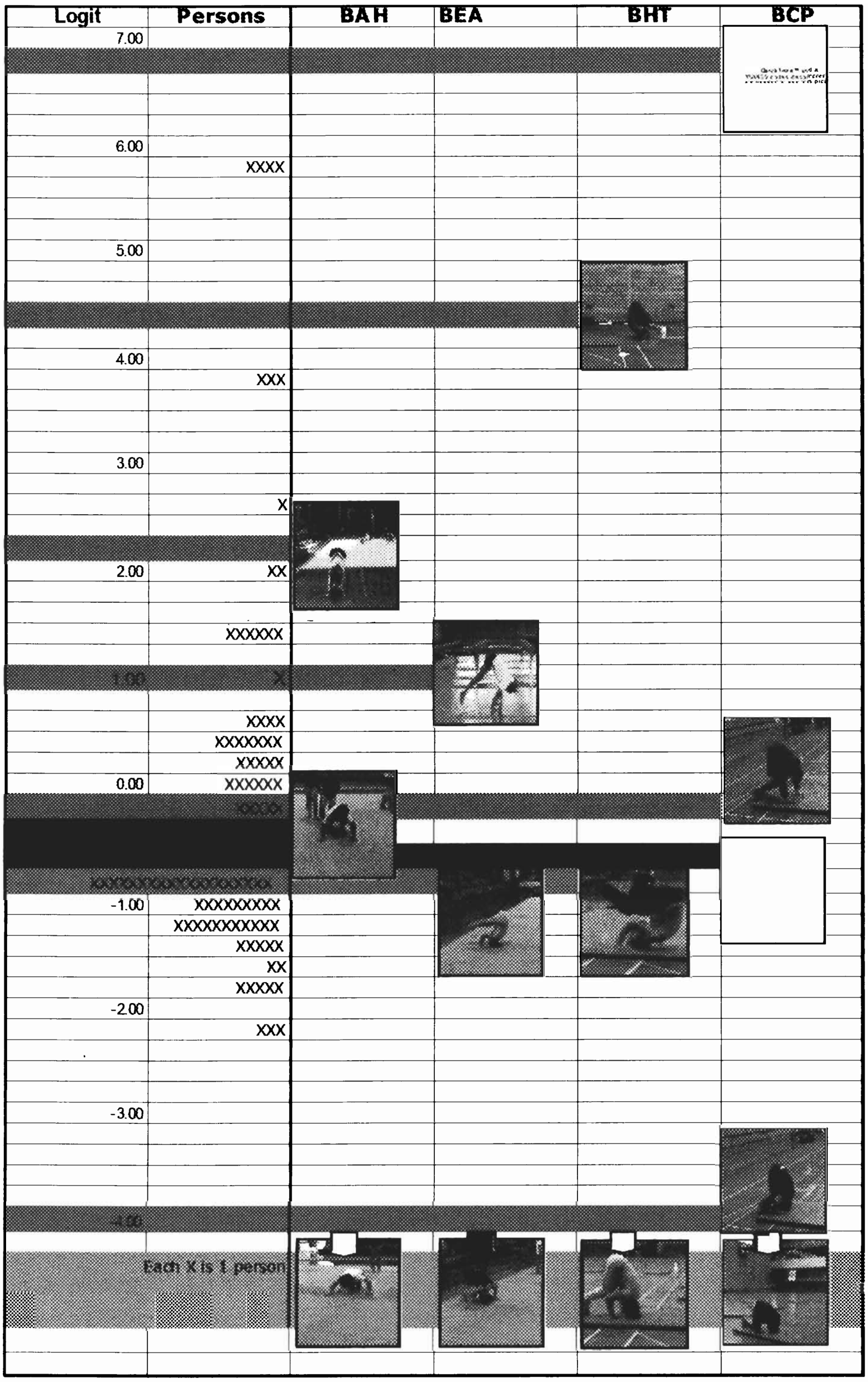
To reiterate, an item person map is usually expressed solely through the use of numbers, however, through the use of pictures, as well as numerals (which appear down the left hand side) a more user-friendly version of the map is produced. Figure 1 shows how pictures may be substituted for codes in the map of the data for the beginning sequence of the forward roll.

Figure 1 shows the item person map, using illustrations (pictures) instead of descriptor code acronyms. Each picture is placed on the map in the same position as its corresponding code, shown in the 'normal' item fit map format. The "coloured" horizontal line, which emanates from the picture, tracks back to the appropriate logit score.

The beginning sequence illustrations are presented under the indicator headings BAH, BEA BHT, and BCP. Each picture under these headings is arranged hierarchically from the highest quality at the top of Figure 1 to least quality at the bottom of Figure 1. The 'gap' between each descriptor picture represents the 'true' difference, in terms of the number of logits. That is, the distance between each picture within each column, vertically, is the amount of difference or change in performance required to move from one position to the next. This is because the measurement unit is a logit scale, i.e., an interval scale, which means the equal distances up and down that scale have equal value. This is unlike, for example, a 'Likert Scale' where the difference between measures, which use the terms 'agree' and 'strongly agree', are usually presented as ordinal measures. Persons and items are located on the pictorial map according to their movement quality and difficulty estimates, respectively.

In a similar fashion the bridging and end sequences may be presented as pictures. With some additional electronic "manipulation" the movements for all sequences could be illustrated using a 'video clip' (as can all the images), which may be presented using multimedia, e.g. via 'power point' presentation. The use of this mode of delivery is an additional instrument aiding clarification, whilst maintaining an accurate portrayal of the data.

Note that the pictures placed 'below the scale' do not indicate an accurate positioning, in terms of logits. They do, however, illustrate the performance of each descriptor at the lowest quality.



'X' indicates number of persons

Figure 1: Pictorial View of the Beginning Sequence of the Forward Roll

CONCLUSION

Many performances within the realm of human movement are examined, of necessity, using the visual sense. Examples of such performances include not only gymnastics, but could be useful for diving, dance and any other physical activity in which the quality of a movement is paramount.

Sport coaches and physical education teachers use the visual medium to learn about their athletes' performance, to teach and to correct errors. What this paper provided is a method of employing visuals to show, not only what each "level of performance" looks like, but also what the real gap is in terms of logits, between one level of performance and the next. To this end moving an individual from one "place" to another can be seen both in terms of how difficult the task may be as well as what the next level of performance looks like.

This paper presented the notion that pictures, gleaned from video recordings for the purpose of illustration, can still accurately demonstrate the step difficulty and level of quality of a particular movement. These images can under certain circumstances replace numbers. The main advantage is that sports people can see where they are going and have some idea how difficult it is going to be to achieve the next level.

References

Adams, R. and Khoo, S. (1993) *Quest - the interactive test analysis system*. Hawthorne, Victoria: ACER.

Bond, T. G. and Fox, C. M. (2001) *Applying the Rasch Model; Fundamental Measurement in the Human Sciences*. Mahwah New Jersey: Lawrence Erlbaum Associates.

Gallahue, D. L. and Ozmun, J. (2006) *Understanding Motor development: Infants, Children, Adolescents, Adults* (6th ed.). New York: McGraw-Hill.

George, G. (1980) *Biomechanics of Women's Gymnastics*. Englewood Cliffs NJ: Prentice Hall.

Haynes, J. (2006) *Objective measurement using pictorial data*. Paper presented at the ICHPER-SD: Fusion Down-under Recipes for Movement: Challenging perspectives and constructing alliances.

Haynes, J., Miller, J., Callingham, R. and Pegg, J. (2005) *Applying Item Response Modelling to confirm the underlying construct of a new process instrument in Gymnastics*. Paper presented at the Refereed Paper at the Australian Association for Research in Education. from http://www.aare.edu.au/.

Masters, G. N. (1982) A Rasch Model for partial credit scoring. *Psychometrica, 47*(2), 149-174.

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago MI: University of Chicago Press.

Snyder, S. and Sheehan, R. (1992) The Rasch Measurement model: An introduction. *Journal of Early Intervention,* **16(1):** 39-53.

Wright, B. D. and Masters, G. (1982) *Rating Scale Analysis: Rasch Measurement*. Chicago, IL: MESA Press.

Wright, B. D. and Stone, M. H. (1979) *Best test design*. Chicago: MESA Press.

ON THE ACCURACY OF A LOW-COST COMPUTERIZED FEEDBACK SYSTEM USED IN TABLE TENNIS TRAINING

Kornfeind, Philipp and Baca, Arnold

Department of Biomechanics, Kinesiology and Applied Computer Science, Faculty of Sport Science, University of Vienna, Austria

Paper Submitted for Review: 30 March 2008
Revision submitted and accepted: 12 June 2008

Abstract A low-cost system has been developed to determine the time intervals between a table tennis ball's impacts after a serve. A ball impact on the table causes a typical acoustic signal. Two microphones are used for recording this signal, both fixed in metallic boxes. The boxes are put near the net onto both halves of the table. The signals acquired from the microphones are pre-processed electronically and then fed to a microcontroller, which calculates the time intervals. The microcontroller is connected via a serial port to a PC, notebook or PDA, which then displays the results. The overall system is not bound to a specific table tennis table and can be transported easily to the environment (table, hall, etc.), where it is used. No user intervention is required between successive serves because of an automated system reset into a 'wait state' after a short period without an acoustic impact signal. Time intervals are presented to the player immediately after having served. One typical task in training, for example, is to play a short serve minimizing the time interval between the first and second impact on the opponent's side of the table. From physical considerations (speed of sound) a maximum error of not more than 4ms in the time intervals was expected. A table tennis robot and high-speed video recordings were used to evaluate the system. For the testing procedure 30 serves were performed. Three different impact scenarios were investigated. Durations were estimated simultaneously with the microphones and from the camera recordings. Time intervals of each trial were compared by calculating the absolute time difference for the results of both systems. The experimental results confirmed the assumptions and showed time differences less than 2ms for typical playing situations.

Keywords: feedback system, table tennis, low-cost

INTRODUCTION

Several authors have combined successfully computers and sensor technologies to build Computer-Human-Interaction (CHI) systems in table tennis. Some of them focused on giving ball position feedback in training situations (Baca & Kornfeind, 2004 & 2006; Hey & Carter, 2005) to assist athletes in their technique training. Others developed software- and hardware-interfaces for virtual reality scenarios in order to simulate real playing situations against virtual opponents (Rusdorf & Brunnett, 2005). Another interesting field of application results from the combination of physical activity and entertainment were the user's motion can act as input for computer games. For example, Ishii et al. (1999) designed a multimedia assisted table tennis game ("PingPongPlus") that incorporates sensing, sound and projection technologies. The number of network games is growing very fast, also in table tennis. It can be played virtually over the internet, even against two opponents (Mueller & Gibbs, 2007).

Computer-based systems may assist service training too. For a table tennis player the service is a key factor for winning a rally. Athletes spend considerable time in training to improve their serve. Serving speed plays an important role in addition to the spin and the positioning of the ball on the table. The latter is difficult to estimate from youth players. In order to practise table tennis serves independently, i.e. without the need of an assisting coach, an easy-to-use feedback system for the detection of the time intervals of services ("TimeCheck") has been constructed and manufactured (Figure 1). This system can determine the time interval between the ball's first and second impact on the table immediately after a serve (2-Ball mode). In the case of short serves, it also determines the time interval between the second and third impact (3-Ball mode).

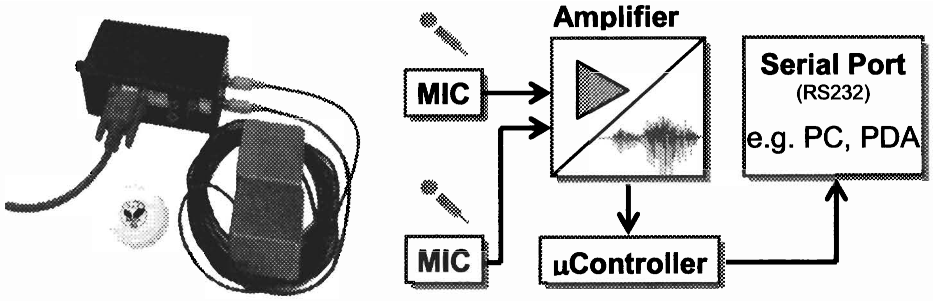


Figure 1: Components of TimeCheck and functional overview (block diagram).

During the development process the authors considered two different solutions for this task. The first one was based on accelerometers to measure the mechanical vibrations in the tabletop during the impact of the ball on the table surface. The alterative was an acoustic solution which measures the impact sounds with two microphones positioned on the table. Because of lower sensor costs and less sensitivity against disturbing vibrations the second concept was used. Typical values for time intervals of services were observed in the range between 250 and 500ms, depending on the skill level of players and the type of serve (e.g. top spin, back spin). In order to be applicable in practise, e.g. to recognize improvements, the time error of the system should therefore not exceed 5ms.

SYSTEM DESCRIPTION

The main components of the system are two microphones (sensors), an electronic box for signal processing and a computer running a simple program (Figure 2). The microphones acquire the acoustic information each time a ball hits the table and are positioned on the tabletop in front of the net (centre line, approx. 5cm distance to the net). After the signals have been pre-amplified (inside the microphones), input voltages above an adjustable threshold (→sensitivity) are detected in the comparator stage. Because of frequent changes in polarity of the input signal (microphone) the comparator produces a sequence of short pulses at TTL level (*TRUE: +5V, FALSE: 0V*). A timer IC (*NE555*) is then used to convert the pulse repetition into a trigger signal, which initiates the capturing procedure in the microcontroller (*PIC 16F628A*).

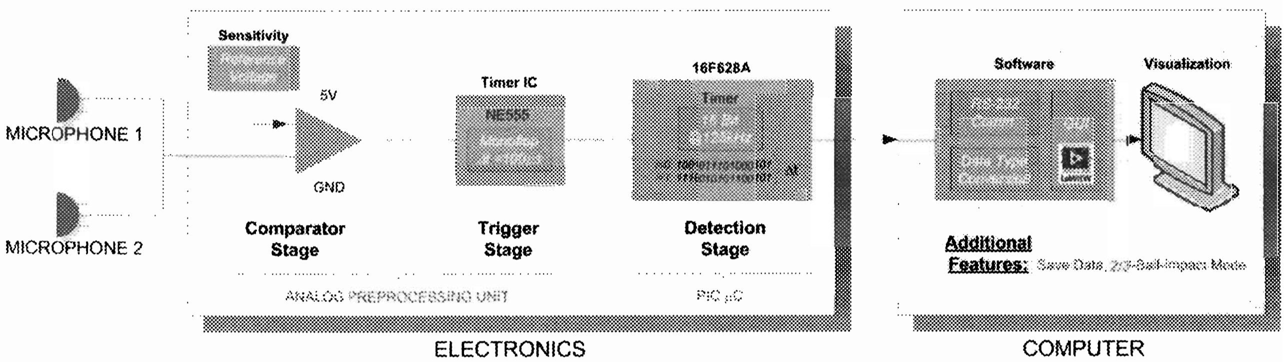


Figure 2: Detailed description of the components used and information flow.

Whenever a trigger occurs, a software routine inside the µC captures the actual value of an internal 16-Bit timer (resolution 8µs/cycle) and stores it into an array. Depending on the selected mode of operation (2 or 3 ball impacts), the differences between the stored timer entries are calculated and transmitted (*RS-232*) to the computer. Software written in LabVIEW™ (Figure 3) receives the data, converts them into standard data types and stores the time intervals in a file (optionally). Numerical and graphical display elements are used for visualization, whereby a graphical indicator can be used for fast comprehension without the need to read the numerical values (e.g. during playing).

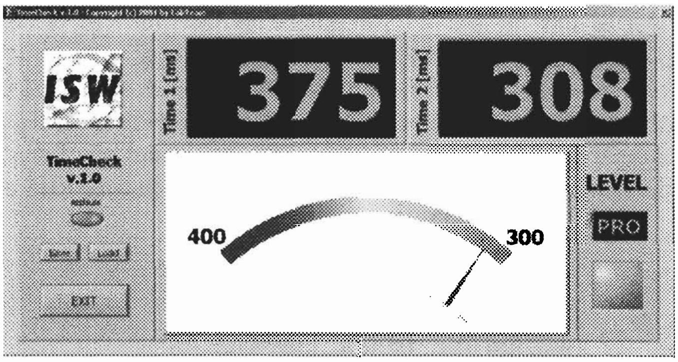


Figure 3: Screenshot of the GUI (LabVIEW™ program).

TIME DURATION ERRORS

In Figure 4, two different ball impact scenarios on the table are shown to describe the expected errors caused by the different latency periods of the acoustic sound signals. In case A (ball A, distances s_1, s_2) the ball first hits the tabletop near the end line so that the distance between impact point and sensor is comparatively long (s_1=1.00m). The second impact point is near the positioned microphone which results in a shorter distance (s_2=0.15m). Assuming a constant sound velocity in the air (c=343m/s @20°C) the signal propagation delays can be calculated as $t_1= s_1/c$=0.0029s and $t_2= s_2/c$=0.0004s. Calculating now the absolute time difference $\Delta t=t_1-t_2$, an error of about 2.5ms can be expected. In the second case (ball B, distance s_3), the ball hits the table in equal distances to the microphone positions. The delay can be calculated as $t_3= s_3/c$ for both table halves. Since the time difference will be zero, this case can be characterized as the best one. Considering the worst case, where the first impact position is directly at the corner (crossing point of side- and end line, $s \approx 1.5$m) and the second one exactly on the microphone (s=0m) a maximum time error of about 4.4ms can be expected.

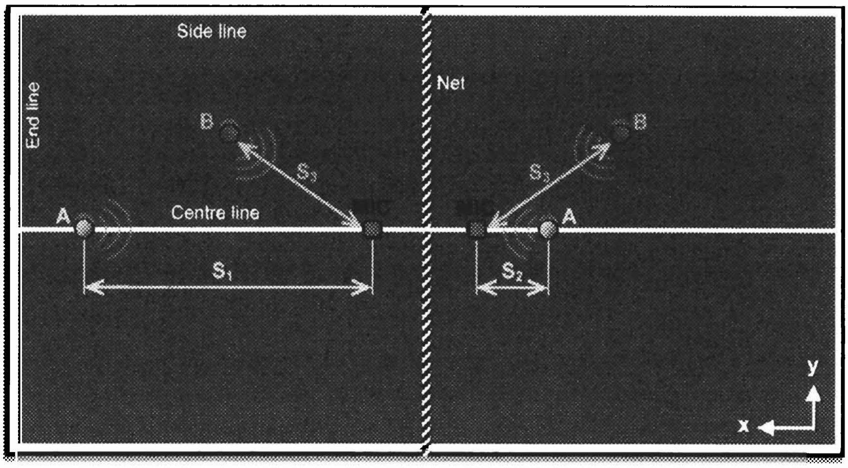


Figure 4: Two impact scenarios with different impact positions.

METHODS

To validate the results of the TimeCheck device, optical measurements with two synchronized high-speed cameras (MotionPro 2000, RedLake, USA) have been done simultaneously. The cameras were positioned beside the table and were adjusted in height shortly above the tabletop (Figure 5). Using additional lights it was possible to record serves with a frame rate of 2000Hz (Δt=0.0005s) and a shutter time of 1/4000s (to avoid motion blur). Three series of 10 serves each, categorized into different impact pair positions (see Table 1) were tested and recorded using a ball robot machine (Robo Pro Plus, TIBHAR, Germany). Each impact pair positions is defined by P1 (first impact point) and P2 (second impact point). After adjusting the ball robot to a predefined impact pair position, a ball was served on the table. Variations in ball positioning caused by the ball robot machine were less than ±2.5cm. The results of the TimeCheck device (in ms) were registered and the videos of the high-speed cameras were analyzed using the manufacturer's software (MiDAS). Time indices of each impact were identified by selecting manually the frame where the ball changes its direction. Time differences were calculated. Since a maximum resolution error of 0.5ms may be assumed for both measurements, the maximum resolution error for this difference should be 1ms. Mean and maximum values of the absolute values of the time differences were calculated.

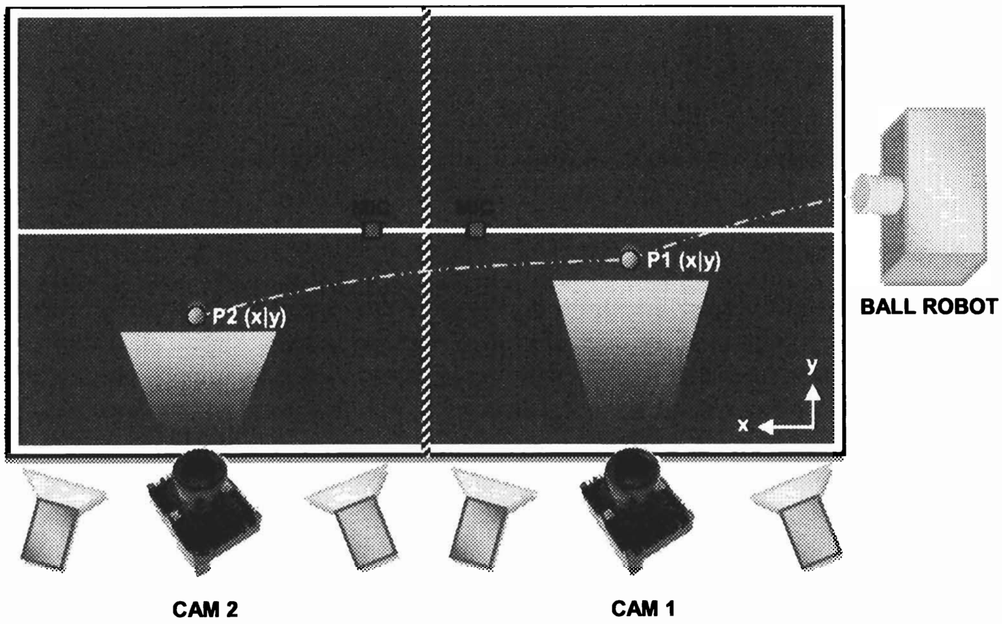


Figure 5: Measurement setup for the tests: two cameras + external lights, two microphones (connected to electronic box) and ball robot machine.

Table 1. Approximate ball impact coordinates of the different impact pair positions on the table.

Impact pair position	P1 coordinates (x | y) [m]	P2 coordinates (x | y) [m]	s_1 [m]	s_2 [m]	Δs [m]
1	0.65 | 0.55	2.49 | 0.35	0.53	0.98	0.45
2	0.40 | 0.76	2.19 | 0.76	0.74	0.59	0.15
3	0.15 | 0.76	2.39 | 0.76	0.99	0.79	0.20

RESULTS

Both test methods resulted in similar time intervals, which changed slightly over the different impact pair positions. Mean values and standard deviations for each method as well as for the time interval differences are listed in Table 2. The standard deviations of the time intervals depend on variations caused by the ball robot machine. They are, however, negligible when calculating the time differences between the two different measurement methods. None of the trials resulted in a differences greater than 1.5ms.

Table 2: Time interval results from TimeCheck versus the high-speed cameras.

Location	TimeCheck [ms]	HS-Camera [ms]	absolute differences [ms]
1	336.80 ± 10.10	336.35 ± 9.89	0.45 ± 0.42
2	355.40 ± 15.82	356.05 ± 15.72	0.65 ± 0.39
3	576.70 ± 13.15	577.85 ± 13.21	1.15 ± 0.39

DISCUSSION AND CONCLUSION

The validity of the device developed (TimeCheck) could be shown. In typical impact situations not even half of the maximum expected error was observed. Based on these findings it may be concluded that athletes can trust the feedback information given in training.

Acknowledgements

We would like to give our special thanks to our interns, Johannes Magerl and Bernhard Navratil, for providing us with excellent assistance during the time measurements.

References

Baca, A. and Kornfeind, P. (2004) Real time detection of impact positions in table tennis. In M. Hubbard, R. D. Metha, J. M. Pallis (eds.) *The Engineering of Sport 5 Vol. 1.* ISEA, Sheffield, UK, 508–514.

Baca, A. and Kornfeind, P. (2006) Rapid feedback systems for elite sports training. *IEEE Pervasive Computing.* **5(4):** 70-76.

Ishii, H., Wisneski, C., Orbanes, J., Chun, B. and Paradiso, J. (1999) PingPongPlus: Design of an athletic-tangible interface for computer-supported cooperative play. *Proceedings of the ACM SIGCHI conference on Human Factors in Computing Systems.* ACM, Pittsburgh, Pennsylvania, USA, 394-401.

Hey, J. and Carter, S. (2005) Pervasive Computing in sports training. *IEEE Pervasive Computing.* **4(3):** 54.

Mueller, F. and Gibbs, M. (2007) Building a table tennis game for three players. *Proceedings of the ACE conference.* ACM, Salzburg, Austria, 179-182.

Rusdorf, S. and Brunnett, G. (2005) Real time tracking of high speed movements in the context of a table tennis application. *Proceedings of the ACM symposium on Virtual reality software and technology.* ACM, Monterey, California, USA, 192-200.

ASSESSING GOODNESS OF FIT AND OPTIMAL DATA SIZE FOR A BROWNLOW PREDICTION MODEL

Bailey, Michael [1] and Clarke, Stephen [2]

1 Department of Epidemiology & Preventive Medicine, Monash University, Melbourne, Australia

[2] Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia

Paper Submitted for Review: 21 February 2008
Revision submitted and accepted: 16 May 2008

Abstract. The Brownlow medal is the highest individual honour that can be bestowed upon an AFL footballer. In each of the 176 home and away matches for a season, votes are assigned to the three best players (3 – first, 2 – second, 1- third) by the umpires that preside over the game. With the use of an ordinal logistic regression model retrospectively applied to past data, Bailey and Clarke (2002) constructed a 13 variable model that has successfully been used to identify the leading candidates for each Brownlow medal count. This paper seeks to build on this work by identifying an additional 12 variables relating to player and match statistics that are highly significant predictors of the number of votes received ($p<0.001$). We then use a range of various measures of goodness of fit, to explore the difference between statistical significance and practical significance by determining how much benefit each additional variable adds to the prediction process. By varying the size of the training data and holdout samples it is possible to determine the optimal size for training data, along with measuring the detrimental effects of over-fitting the data. Whilst it is possible to use mathematical models to aid in the prediction of the Brownlow medal, there is clearly a limit to the benefit achieved. This paper identifies this limit and determines how much data are required to achieve the optimal solution.

Keywords: Brownlow medal, Ordinal logistic regression, Goodness of fit

INTRODUCTION

During the 2000 AFL football season, discussion arose as to the best possible way to predict the winner of the Brownlow medal, both before and during the actual count. With a large number of match performance statistics readily available it was felt that a mathematical modelling process might well assist in the objective assignment of a player's probability of polling votes. Based on data collected from the 1997, 1998, and 1999 seasons, an ordinal logistic regression model was originally constructed and applied to the 2000 season. Predicted votes for each match were then tallied over the season to provide players predicted totals for the year. Following considerable success and media attention, the ordinal regression model was further enhanced for the 2001 season. Although 25 variables were now identified as being statistically significant predictors ($p<0.001$), the practical benefit of some variables came into question. Additionally, it was questioned just how much data were optimally required to accurately develop such a model. To answer these questions, a range of practical measures were developed, and with seven complete seasons of data, a comprehensive analysis was conducted exploring the practical benefit of each variable. Training datasets of differing sizes were further incorporated to ascertain the optimal amount of data required to build such models.

Database

A database was constructed that comprised of data collected from each regular season AFL match played between 1997 and 2003 (1232 games). For each game, an array of individual match statistics is readily available, both in the newspapers and via the internet. Whilst some predictors of votes are created from the past history of the players, most predictors are derived from match statistics.

Multivariate model

Using the number of votes polled as the outcome, a multiple ordinal logistic regression was constructed. Seven seasons with 176 matches per season and 44 players per match (7744 data points per season) were used to progressively construct a 25-variable multivariate model with all variables in the model statistically significant at a level of $p<0.001$. While backward elimination selection techniques were also used in model construction, to show the relative importance of each variable, results are presented and graphed in stepwise fashion. From Table 1 the 25 stages of development can be seen.

In addition to the 19 first order effects that were identified, there were six interactions between variables that were found to be statistically significant ($p<0.001$). Each interaction term was comprised of a first order effect, suggesting that the interaction term was acting as a fine tuning process for relationships that were not perfectly linear.

Table 1: List of variables (in order) found to be statistically significant predictors of votes polled ($p<0.001$)

Stage	Variable	Description
1	Disposals	Number of kicks & handballs
2	+ Result	Margin of victory
3	+ Win	Player's team won or lost
4	+ (Result*Win)	Interaction between variables Result & Win
5	+ Hit outs	Tap outs by ruckmen
6	+ Standout	Standout player on ground
7	+ Best players	Best players as given by AFL website
8	+ Good average	How often players previously polled when not expected
9	+ Bad average	How often players didn't polled when expected to
10	+ (Good*Bad)	Interaction between variables Good and Bad average
11	+ Goals	Number of goals kicked
12	+ Marks	Number of marks taken
13	+ (Marks*Goals)	Interaction between variables Marks & Goals
14	+ (Disposals*Goals)	Interaction between variables Disposals & Goals
15	+ Position	Named position of player on ground
16	+ (Disposal*Position)	Interaction between variables Disposals and Position
17	+ Inside 50	Number times player sends the ball inside 50 metre arc
18	+ Scoring shots	Number of team scoring shots
19	+ Distinct appearance	Unusual skin or hair colour
20	+ Captain	Player is the team captain
21	+ Average votes	Total number of votes polled in past counts
22	+ Tackles	Number of tackles laid for match
23	+ Rebounds	Number of times the player rebounded the ball from defence
24	+ Frees for	Number of frees issued to player
25	+ (Disposal*Best)	Interaction between variables Disposals and Best players

Goodness of fit

Goodness of fit can be determined at two different levels relating to the prediction of votes for each game, or the prediction of the votes that each player will poll for the season.

When fitting an ordinal logistic regression, a clear guide to the goodness of fit of the model can be gauged from the log likelihood estimate. Because a change in -2log likelihood can be well approximated by a chi-square distribution, the statistical significance of additional variables can readily be determined. By examining Figure 1 from left to right, we can see the corresponding reduction in -2log likelihood as each new variable is added to the model, with the greatest improvements occurring over the first seven stages of the model.

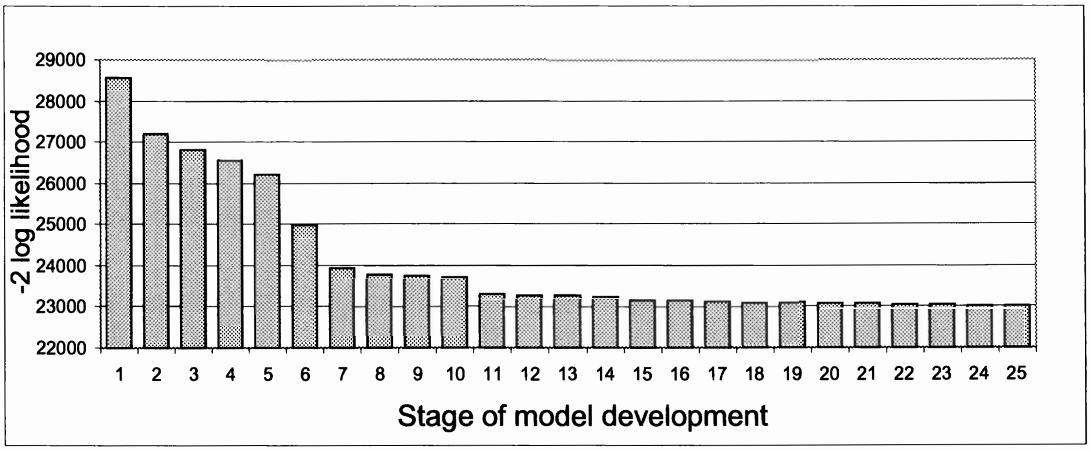


Figure 1: -2 log likelihood for each stage of model development

Average Rank

Using the 25-variable multivariate ordinal regression model, each player was assigned a probability of polling three votes, two votes or one vote for each match. A predicted value for the number of votes that each player was expected to get for each game was created by the following formula -

$$Predicted\ Votes = 3 * Pr(3\ votes) + 2 * Pr(2\ votes) + Pr(1\ vote) \quad (1)$$

Based on the predicted number of votes, the 44 players for each match could then be ranked in accordance with their predicted vote total. The average rank of players that poll votes provides a practical way to measure the goodness of fit for models, with a lower rank indicating a better fit to the data. For example, from Figure 2 it can be seen that using a model with only disposals as a predictor, the average rank of all players that polled three votes was about six. With the full 25-variable model the average rank of players that poll three votes is only 3.2.

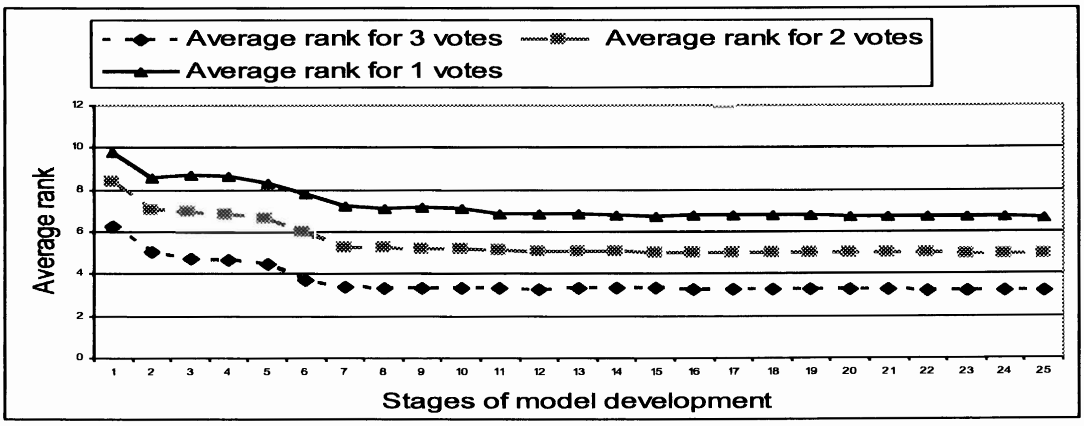


Figure 2: Average rank for vote getters as variables are added to the model

Top three for the game

Having ranked players for each match according to their predicted vote total, we can also measure how often the leading ranked player is awarded three votes. From Figure 3 we can see that the leading ranked player actually polls three votes, 43% of the time. Similarly, the leading ranked player has a 73% chance of polling any votes (three, two or one). Conversely, the player who polls three votes will be ranked within the top three players in 72% of all games. Once again, there is a clear indication from Figure 3 that the majority of improvement occurs over the first 10 stages of model development, with negligible improvement after that.

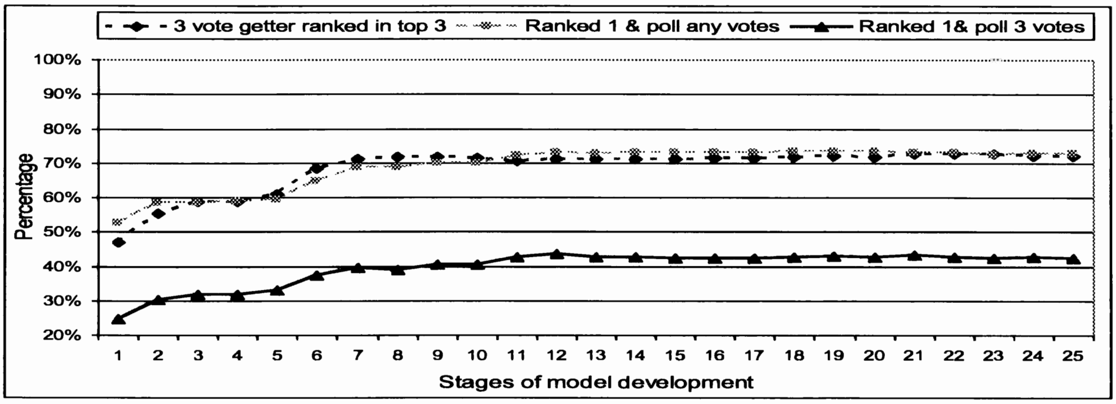


Figure 3: Within game goodness of fit measures for stages of model development

In addition to determining goodness of fit at a match level, players predicted vote totals can be aggregated to provide player predictions for the season.

Goodness of fit for season

By tallying each player's predicted number of votes for each match as given by equation 1, and aggregating for the 22 matches for the season, it was possible to derive a predicted Brownlow total for each player for the year. A simple measure of goodness of fit can then be derived by measuring the Absolute Average Error (AAE) between the predicted total and the actual total for each player for the season. When predicting the yearly total for players, of particular interest is the performance of the players most likely to win the Brownlow medal rather than all players, thus the AAE for the leading 20 predicted players is also considered.

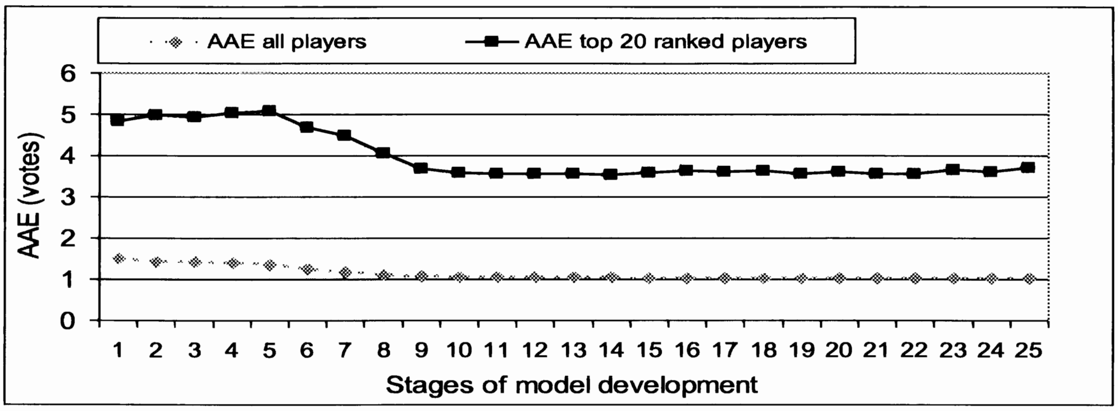


Figure 4: AAE for seasonal vote totals for 25 stages of model development

From Figure 4 we can see that most additional benefit can be achieved with the first 10 of the 25 statistically significant predictors. The annual AEE for all players can be reduced to approximately one vote, but there is higher variability in the leading players with an AEE of about three and a half votes. Based on the AAE, it is then possible to determine the proportion of players that the modelling approach can accurately predict to within one vote. From Figure 5, we can see that the modelling approach tops out after the addition of about 10 variables, successfully predicting 68% of all players to within one vote of the actual total. Similarly, Figure 5 also shows that the model can successfully predict 83% of players within two votes, and 90% of players within three votes.

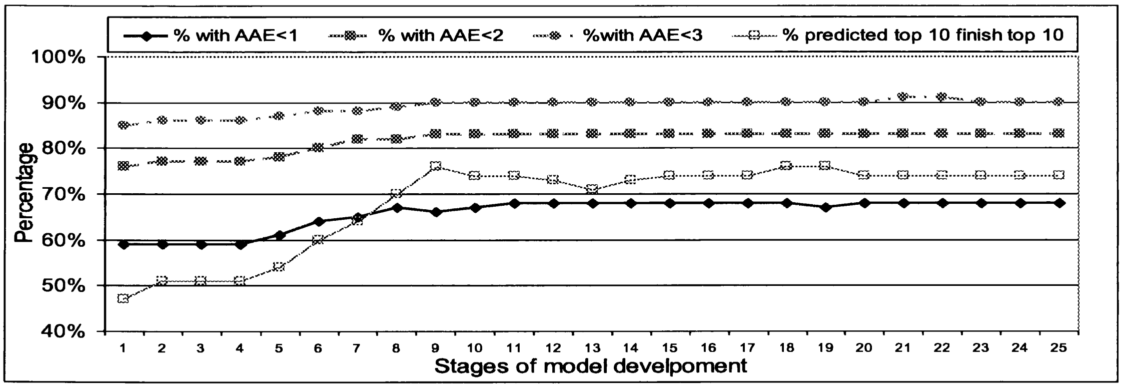


Figure 5: Predicted top 10 who finish in top 10 and AAE <1, 2 & 3

Top 10 for season

By measuring the proportion of players that were ranked by the models to finish in the top 10 and actually did finish in the top10, we can further gauge predictive capacity. From Figure 5 we can see that after the addition of the first 10 variables, little practical benefit could be achieved with regards to accurately identifying the leading 10 contenders to win the Brownlow medal, with the modelling approach having about 74% accuracy in predicting leading players.

Measuring the bias of over-fitting the data

All previously defined criteria for goodness of fit have been determined by using all seven years of data as both a training set to determine parameter estimates and a holdout sample to determine predictability, thus over-fitting the data. In order to measure the bias of over-fitting the data, the same 25 parameter model is applied to six additional training datasets, with lengths ranging from one to six years. Goodness of fit has been determined from holdout samples also ranging from one to six years, with the addition of training and holdout samples adding to seven years. Thus, when the training dataset was 1997, the corresponding prediction model was applied to the remainder of the data (1998-2003). When two years worth of training data were used (1997 & 1998), the holdout sample used was 1999 to 2003.

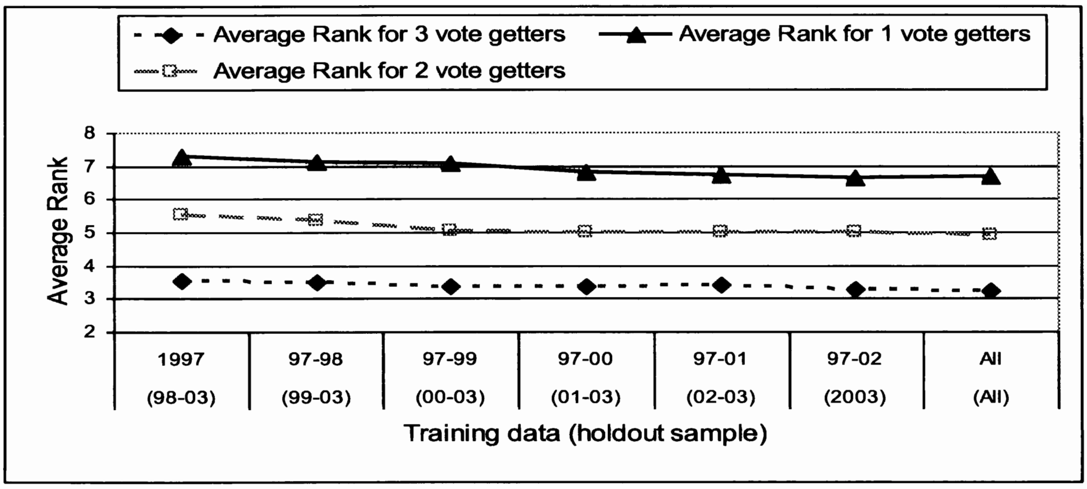


Figure 6: Average rank with differing training and holdout sample sizes

By considering the average rank for the players that polled votes, it is possible to gauge the relative accuracy of differing sets of training data. From Figure 6 it can be seen that when only the 1997 season was used as a training dataset, the average rank for players polling three votes was 3.5. When three seasons worth of data were used (97-99), the average rank for players polling three votes was reduced to 3.35. When the data were over-fitted, by developing on all seven seasons and then reapplying to the same data, the average rank for players polling three votes was reduced to 3.22.

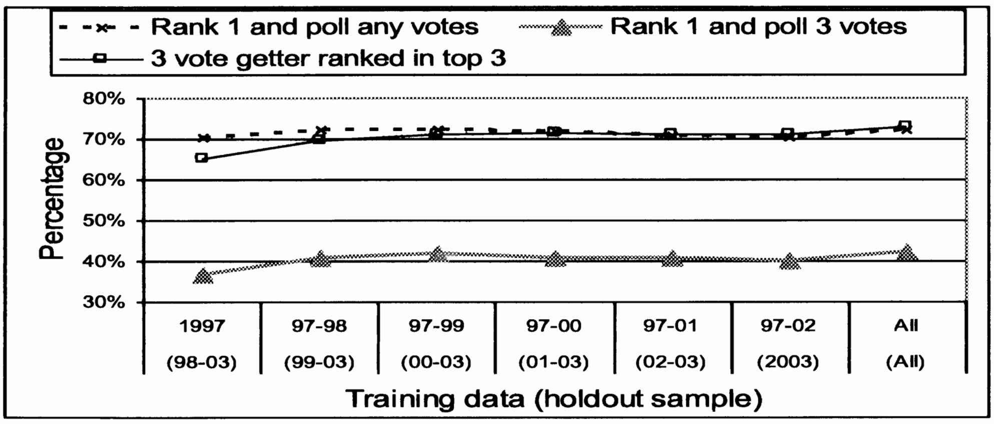


Figure 7: Within game measures of goodness of fit

When only the 1997 season was used as a training dataset, the leading ranked player according to the modelling process would poll three votes 36% of the time. By increasing the training data set to two years, the three vote getter could successfully be identified 40% of the time, but it would appear that little benefit could be gained by using more than two years worth of training data, as the percentage of best players successfully identified alters little between two and six years. When the data are over-fitted, the bias of over fitting can be approximated at 2% with the three vote getter identified 42% of the time.

When only the 1997 season is used as a holdout sample, the leading ranked player would poll votes 70% of the time. This figure improves slightly to 72% with either two, three of four years worth of holdout data, but drops again slightly when five of six years of training data are used; perhaps reflecting increased variability in the holdout sample. When the data are over-fitted, the leading ranked player polls votes 72% of the time.

The final measure considered in Figure 7 is how often the player who polls three votes is ranked in the top three positions by the models. When only the 1997 season is used as a training sample, the three vote getter was only ranked in the top three 65% of the time. This figure improves to 70% with two years worth of training data, but shows little improvement after that. When over-fitted, the three vote getter is ranked in the top three 72% of the time.

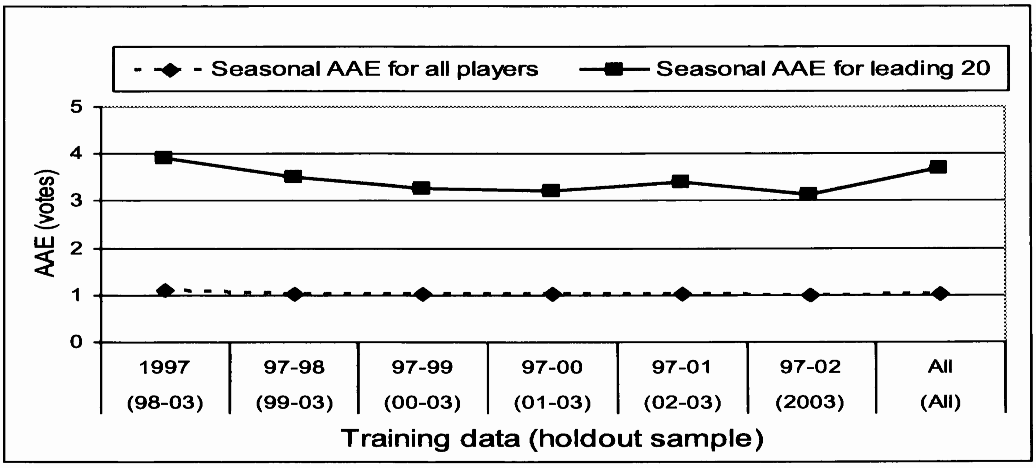


Figure 8: Seasonal AAE for all players and leading 20 players

When considering criteria for goodness of fit for the season, the bias associated with over-fitting the data is less pronounced. When only the 1997 season is used, the AAE between each player's predicted and actual number of votes polled for the season is 1.1, whilst for the leading 20 ranked players it is 3.9 (see Figure 8). When at least two seasons are used for the holdout sample, the AAE for all players is reduced to one, whilst for the leading 20 it is reduced to 3.5. Although there is no difference in AAE for all players when the data are over-fitted, the AAE for the leading 20 players is higher when the data are over-fitted.

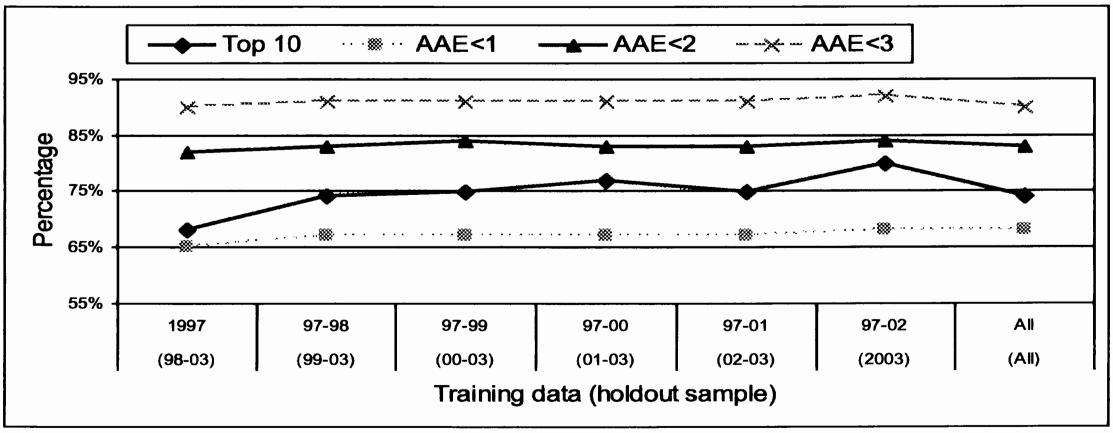


Figure 9: Season measures of goodness of fit for differing holdout samples

When only the 1997 season is used as a training dataset, the total votes polled by each player can be identified to within one vote for 65% of all players. It can be seen from Figure 9 that when the size of the training dataset is doubled, this figure improves to 67%. When six years of training data are used, this figure improves to 68%, but no additional benefit is apparent by over-fitting the data as the over-fitted data can still only identify 68% of all players to within one vote of their seasonal tally.

Similarly, the percentage of players that are accurately identified to within two or three votes, appears to improve when the training dataset is increased from one to two years, but little further improvement can then be achieved, and when over-fitted, results are slightly worse.

The percentage of players that are accurately identified as finishing the season in the top 10, shows slight improvement as the size of the training data increases, although as only 10 players are considered each year, there is higher variability is this measure of goodness of fit.

DISCUSSION

While 25 variables were found to be highly statistically significant predictors for the number of votes polled, the practical benefit of many of these variables is questionable, with a 10 variable model producing similar results.

This research suggests that goodness of fit of prediction models is dependent upon the quality and quantity of data used to construct the model. A minimum of about three seasons (528 games) is required from which to develop models, and although the quality of models does continue to improve with larger holdout samples, the improvement appears minimal. The similarities between the Brownlow medal and a horse race, reaffirms the finding of Benter (1994) in developing multivariate models for horse racing in Hong Kong. Benter states "… the minimum amount of data needed for adequate model development and testing samples is in the range of 500 to 1000 races. More is helpful, but out-of-sample predictive accuracy does not seem to improve with development samples greater than 1000 races".

CONCLUSION

Building on work previously developed by the authors, this paper uses ordinal logistic regression to identify match and player features that can be linked to the polling of Brownlow votes in AFL football. By applying this prediction model to matches played throughout the course of the AFL season, each individual player was assigned a probability of polling votes. By aggregating individual predictions, each player could then be assigned a probability of winning the Brownlow medal. A series of measures were developed to determine goodness of fit along with the minimal size of holdout samples required to accurately develop such models. Results suggest a minimum data of 500 matches and a maximum of 10 variables are required to build an effective model.

References

Bailey, M. and S. R. Clarke (2002) Predicting the Brownlow medal winner. In G. Cohen and T. Langtry (eds.) *Proceedings of the sixth Australian Conference on Mathematics and Computers in Sport*. Bond University, 56-62.

Benter, W. (1994) Computer based horse race handicapping and wagering systems: a report. *Efficiency of racetrack betting markets*. L. Hausch, Ziemba. San Diego, Academic Press Limited, 183-198.

AN ALGORITHM TO PLOT AN AFL TEAM'S PERFORMANCE IN REAL TIME USING INTERACTIVE PHASES OF PLAY

Ryall, Richard and Bedford, Anthony

School of Mathematical and Geospatial Sciences, RMIT University, Australia

Paper Submitted for Review: 29 January 2008
Revision submitted and accepted: 16 June 2008

Abstract. Presenting statistical predictions that are simultaneously representative of a team's likelihood of winning and graphically simple enough to be widely interpretable remains a constant challenge for the sport statistician. This study focuses on the process involved in transforming a mass of performance variables from "live-streaming" data into a single web-based phases of play plot. The algorithm utilised regresses a number of performance variables to a single binary variable (win/loss) for streamed AFL data, and evaluates each team's chances of winning for live and post-match analysis using logistic regression. Summary game data was gathered from 176 home and away games from the 2006 AFL season and used to forward predict the entire 2007 home and away AFL season. Using transaction data provided by ProWess Sports a series of algorithms were written for use in Excel. This algorithm extracts the required performance variables, runs the logistic regression model, and generates the relative phase plot in real time; where relative phase describes the overall interaction between the two teams. Statistically, the plot provides an effective representation of the state of the game, illustrating the calculated probability of the home team winning relative to the opposition at any point in time. Graphically the plot is enhanced by adding images of a players guernsey when a goal is scored. This combination provides the viewer with an objective probability assessment of the current state of the game. It is also useful as an instantaneous coaching tool. The final product delivers both a single statistical measure and graphical representation of the state of the game at any point in time for use by coaches and the general football public alike.

Keywords: AFL, algorithm, phases of play.

INTRODUCTION

Australian rules football is the countries most popular winter sport. The Australian Football League (AFL) oversees the administration of the game. At the elite level of the competition, known as AFL football, currently 16 teams play a regular season of 22 games, and potentially four weeks of finals. In a regular game of AFL there are two teams of 22 players, of which only 18 are permitted on the field at any one time, with the remaining four players "benched". Players are then rotated on and off the bench, at the coaches request, for many reasons, including rotating "fresh legs" off the bench, and players coming off the field due to coach instigated disciplinary actions (e.g. conceding a 50-metre penalty). Each team typically comprises three full-backs, three half-backs, three midfielders, three ruck/rovers, three half-forwards, three full-forwards and four players on the bench. Each player's position can change throughout a match, and there is no restriction on where a player can move. There is currently no limit to the numbers of interchanges allowed.

AFL football remains one of the hardest games to analyse in world sport due to the wide array of actions each player can take, for example, from kicking a goal, to picking up the 'loose ball', to shepherding (blocking) a player away from the ball. The average AFL game consists of approximately 2,500 unique transactions, with each transaction consisting of up to three actions, or unique statistics, (e.g. kick long; kicking to a contest; inside 50) attributed to one of the 44 players contesting a game. In this research, we consider breaking down these performance variables into a single probability describing the state of the game. Previous analysis has focused merely on the frequency of event occurrences, such as number of kicks in an AFL match, forehand winners in tennis, corners in soccer etc. Gréhaigne et al. (1997) state that such sports are made up of complex interactions between team/player performance variables and thus frequencies of these performance variables can't capture the full complexity of team and dual sport as an interactive system.

Kelso (1984), in a physical sense, was the first to introduce the idea of phase transitions using human bimanual coordination. The experiment involved asking participants to cycle their hands at the wrist in the horizontal plane in an asymmetrical mode. He found that initially phase relations were anti-phase stable, that is, synchronous motion of the wrists in the opposite direction. As the frequencies were increased through instructions the phase difference became more evident however, after this transition period the two motions were now in-phase stable, that is, synchronous motion of the wrists in the same direction. This is known as one-way transition (anti-phase → in-phase) since the jump from anti-phase stable to in-phase stable occurs at a critical point but the reverse is not true regardless of the cycling frequency. Relative phase is defined as phase-lag, that is, when the two motions are anti-phase stable this represents 0^{o} of relative phase, similarly in-phase represents 180^{o} of relative phase. The results of the study suggest that relative phase is an adept collective variable which can be used to address the change in phase relation in bimanual coordination.

The application of phases in a physical sense to a sporting contest is in its infancy. Several recent papers (McGarry et al., 2002; Palut and Zanone, 2005) proposed the idea that two teams/players interact together in a sporting contest in terms of dynamic system theory, that is, in an active-reactive nature. This idea is utilised and discussed later.

McGarry et al. (1999) set out to find whether championship squash could be described in terms of dynamic system theory. The experiment conducted involved data from previous studies by McGarry and Franks (1995, 1996) from the 1988 Men's Canadian Open Squash Championship. The aim was to determine whether system perturbations are subject to perceptual detection and whether the level of expertise of the observer was associated with the detection of perturbation 'onsets' and perturbation 'offsets'. A perturbation onset is the transition from stable → unstable. Similarly a perturbation offset is the transition from unstable → stable. Sixty rallies were analyzed at random by six expert and non-expert observers to identify which shot/s, if any, caused a perturbation onset or perturbation offset and the reason why the perturbation occurred. If three or more expert observers agreed upon a perturbation 'onset' within a single shot then it was considered valid. They concluded that not only were expert observers significantly more likely to agree upon when a perturbation onset occurred but also why the perturbation onset exists.

Franks and Miller (1986) found that coaches have the same level of difficulty in remembering critical events as eyewitnesses have in recalling criminal events. Franks and Miller (1991) showed that coaches can't accurately recall pertinent sequential information prior to a critical event occurring. This led them to develop a new method to train coaches to observe and remember. They proposed the idea to train the observational skills of coaches using a video training method. The results suggested that although coaches were incapable of remembering more than 40% of pertinent sequential information, coaches can be trained to observe and remember sequential information prior to a critical event occurring. This finding suggests to us that a simple reflective measure is needed to assist coaches in event recall.

Bedford and Baglin (2006) looked at describing ice hockey (NHL) from a probabilistic sense; their objective was to describe a team's game day performance using individual team and relative phase plots. They concluded that the results of the study gave an accurate measure of a team's performance at nearly all points in a game. Their phase plots were smoothed retrospectively; therefore the predictive power of the phases at specific intervals in the match was not investigated. They also suggested that these phases of play at a player based level should be investigated in order to assist the coach in fielding the best team for the best scenario at any time in the match.

This research addresses these limitations in current quantitative sports analysis by giving a graphical and statistical representation of the state of the game at any point in time. The subject of this research, phases of play, can be described as a complex dynamical system whereby teams/players fluctuate between invariant (stable) and variant (unstable) states (in phase and anti-phase respectively). The idea is simple, recorded behaviours/statistics would yield instabilities in the system - using this information coaches could objectively change future behaviours from past performances.

METHOD AND RESULTS

ProWess Sports provide comprehensive statistics to media outlets, AFL clubs, and subscribers via both their website (www.pro-stats.com.au) and dedicated software. Previously they had concentrated on detailed post-match statistics, however in mid 2007 they decided to develop a website dedicated to live-match statistics (www.realfooty.com.au/livestats). The website refreshes live statistics approximately every 30 seconds and displays the following 20 performance variables: kicks, handballs, marks, inside 50's, tackles, spoils, hitouts, 1st possession from an umpire control situation, clearances, goals, behinds, rushed behinds, frees for, marks inside 50, turnovers, goals from general play, goals from free kicks, goals from marks, goals from kick ins and goals from stoppages. An artificial variable, cumulative win was created and is simply defined as the percentage of games won prior to the commencement of a match. Summary game data was then gathered from ProEdge, software developed by ProWess Sports. Due to different definitions of data supplied, turnovers were excluded from the statistical analysis.

Firstly the contribution of each performance variable to a team winning a game needs to be considered. Stewart et al. (2007) set out to find which individual performance variables in AFL were important, and how much each variable contributed to a team winning a match. The objective was to identify inefficiencies in the market for recruiting professional AFL players. This was done by regressing 51 "primary variables" to a single variable margin using Ordinary Least Squares regression (OLS). Since margin was used as the dependant variable goals, behinds and rushed behinds had to be excluded from the model, as their inclusion is an exact predictor of margin. This meant that the final model would be biased against forwards, in particular full forwards.

Bedford and Baglin (2006) applied logistic regression to NHL summary game data for the season 2005-2006 for use in the forward prediction of season 2006-2007 based on 19 performance variables. Win/loss was used as the dependent variable, as the research focused on what contributes to a win (or loss) rather than to scoring a goal (or not). As such, score could not be ignored as an independent variable as it in itself is an outcome of a perturbation in the phases of play. Therefore logistic regression was applied to the previously mentioned 20 performance variables (excluding turnovers) on the 2006 AFL season to use in forward predicting the 2007 AFL season (see Table 1). Separate logistic regression models were applied to home and away games due to overwhelming evidence of home ground advantage (e.g. Clarke (1993)). By including cumulative win percentage into the model, the model initialises prior to the start of the match. These two models correctly classified 90.75% of wins.

Table 1: Logistic Regression Output

Variable	Coefficient		Significance		OR	
	Home	Away	Home	Away	Home	Away
Cumulative win % (CUM)	1.616	0.450	0.088	0.601	5.030	1.568
Kick (KCK)	0.013	0.055	0.545	0.016	1.013	1.057
Handball (HBL)	-0.025	-0.027	0.012	0.011	0.975	0.973
Mark (MRK)	0.009	-0.058	0.725	0.024	1.009	0.943
Inside 50 (I50)	-0.041	-0.035	0.366	0.510	0.960	0.966
Tackle (TKL)	-0.011	-0.007	0.520	0.708	0.989	0.993
Spoil (SPL)	-0.027	-0.022	0.505	0.598	0.973	0.979
Hitout (HIT)	-0.012	0.059	0.738	0.060	0.988	1.061
1st Possession (1ST)	-0.054	-0.232	0.450	0.002	0.947	0.793
Clearance (CLE)	-0.012	0.007	0.873	0.929	0.988	1.007
Goal (GLS)	0.610	0.482	0.001	0.026	1.841	1.619
Behind (BHS)	-0.195	0.133	0.045	0.144	0.823	1.142
Rushed Behind (RUS)	0.250	-0.122	0.050	0.337	1.284	0.885
Free Kick For (F/F)	-0.077	-0.139	0.121	0.014	0.926	0.870
Mark Inside 50 (M50)	0.167	0.135	0.047	0.155	1.182	1.145
Goal from General Play (GFG)	0.073	0.047	0.713	0.826	1.075	1.048
Goal from Free Kick (GFF)	-0.126	0.032	0.628	0.904	0.882	1.033
Goal from Mark (GFM)	-0.547	-0.172	0.009	0.456	0.579	0.842
Goal from Kick In (GFK)	0.095	-0.319	0.655	0.223	1.100	0.727
Goal From Stoppage (GFS)	0.115	0.225	0.396	0.124	1.122	1.252

Notably, running a separate logistic regression model on the binary variable win/loss using only cumulative win % correctly predicts 62.68% of all games. This suggests future performances are somewhat dependant on past performances. From Table 1, kicks have a positive effect on win, whereas handballs have a negative effect. Tackles and spoils are also negative which is probably due to the player that is in the act of tackling/spoiling not having possession of the ball (see (Stewart, Mitchell and Stavros 2007) for greater treatment) and therefore unable to score a goal and win. Goals from general play are worth the most for home and away sides compared to goals from free kicks and goals from marks (Note that GLS = GFG + GFF + GFM i.e. GFK/GFS must also be either GFG/GFF/GFM).

The logistic regression model is then utilised on the cumulative statistics of the home and away team after every each transaction. The probability of team *i* winning at time *t* regardless of opposition is given by:

$$P_i(x_t) = \frac{e^{\text{logit}_i(x_t)}}{1+e^{\text{logit}_i(x_t)}} \tag{1}$$

where x_t is the cumulative statistics of team *i* at time *t*; *i* = 0 (away) and *i* = 1 (home).

Hence the probability of the home team winning relative to the opposition is given by

$$\text{relative}(x_t) = \frac{P_1(x_t)}{P_0(x_t)+P_1(x_t)} \tag{2}$$

The main aim of this research was to construct live statistical predictions that are both representative of a team's likelihood of winning and graphically simple enough to be widely interpretable for coaches and the general football public alike. This was achieved by integrating interchange and transaction data, where interchange data comprises a time series of when players were rotated off the "bench" and transaction data contains detailed descriptions of event occurrences and when those events occurred. This gives the viewer a real time objective probability assessment of a team's performance, and which players contribute to high phase (more than 50% chance of winning) and low phase (less than 50% chance of winning).

The phases were investigated in terms of its reliability. Therefore the previously mentioned regression model was applied to the 2007 AFL season summary game data at quarter time, half time and three quarter time and compared with the % correct games classified by score, at the same time intervals, for all games (see Table 2). The results suggest that score is a better predictor at half time and three quarter time than the phases. However the phases at games end indicates that 9.25% of the time the team behind on the scoreboard should have won the game. This is an important finding as it indicates factors other than score tipped the result in favour of the losing side. To our mind, this begs further investigation – exactly why did the losing team not win?

Table 2: Percentage of games correctly classified by score and the phases at specific intervals

Quarter	Score	Phases
1	68.82%	71.68%
2	78.82%	74.57%
3	92.35%	82.08%
4	100%	90.75%

If the games in which the phases incorrectly predicted the winner are removed from the data set, then the phases are a better predictor at quarter time, half time and score is only a marginally better predictor at three quarter time as shown in Table 3. This suggests that there may be some variables (possibly kicks) which in significant quantities for the same team (which is known as "tempo" football) may override the fact that the margin is significantly different. If the model can be adjusted for this specific problem this will significantly increase the % correctly classified earlier.

Table 3: Percentage of games correctly classified by score and the phases at specific intervals in which the phases correctly classified the final result

Quarter	Score	Phases
1	69.67%	74.52%
2	77.42%	78.98%
3	90.97%	89.17%
4	100%	100%

From (2) the average relative phase of the team by each player can be calculated relative to the time that player spent on ground (TOG). Previous analysis in team sports with the trait of measurable player independence, such as baseball, bases a player's individual performance on individual quantifiable statistics. However, in AFL football, if we accept the idea that a players individual performance (specifically key position players such as full forwards) is dependent on their teammates then we arrive at an inadequate method for rating a players impact on the a game as a whole. However for players that are on the field for the entire game, such as midfielders, their measurable impact on the game relative to time on ground does not give a true representation on their impact to the game given they have not left the field. Further, some players receive an inordinately high phase probability due to the fact they happen to be on the field when the team is in high phase without contributing to the system. This is seen as a limitation and worth further investigation.

Table 4: 2007 AFL Home and Away Season Round 3 Carlton vs. Essendon.

Carlton				Essendon			
Player no.	Player name	TOG	Rank	Player no.	Player name	TOG	Rank
6	Simpson	105:32	1	19	Hille	66:23	1
2	Russell	108:06	2	30	Ryder	107:46	2
8	Whitnall	82:09	3	25	Lucas	117:20	3
4	Gibbs	99:11	4	4	Watson	67:47	4
28	Cloke	65:46	5	13	Lovett	111:36	5
44	Carrazzo	114:58	6	1	Johnson	126:31	6
34	Wiggins	91:16	7	10	McVeigh	126:31	7
17	O'hAilpin	113:41	8	18	Lloyd	126:31	8
24	Stevens	122:15	9	22	Michael	126:31	9
25	Fevola	126:31	10	29	Davey	126:31	10
29	Scotland	126:31	11	31	Fletcher	126:31	11
30	Waite	126:31	12	33	McPhee	126:31	12
32	Thornton	126:31	13	11	Peverill	121:22	13
33	Houlihan	126:31	14	5	Hird	104:05	14
7	Bentick	84:48	15	26	Heffernan	110:06	15
14	Fisher	99:22	16	7	Jetta	86:13	16
19	Betts	108:18	17	24	Stanton	107:44	17
3	Murphy	108:30	18	8	Winderlich	112:26	18
12	Lappin	97:45	19	2	Dyson	53:21	19
1	Walker	100:27	20	20	Slattery	111:30	20
5	Kennedy	95:46	21	6	Monfries	75:22	21
11	Ackland	73:18	22	27	Laycock	63:54	22

Table 4 ranks the players average team phase relative to TOG. Notably, the two ruckmen for Essendon appear at opposite ends of the table. Hille was the dominate ruckmen for the team, and clearly performed better than his teammate, who performed poorly given relatively the same opportunity.

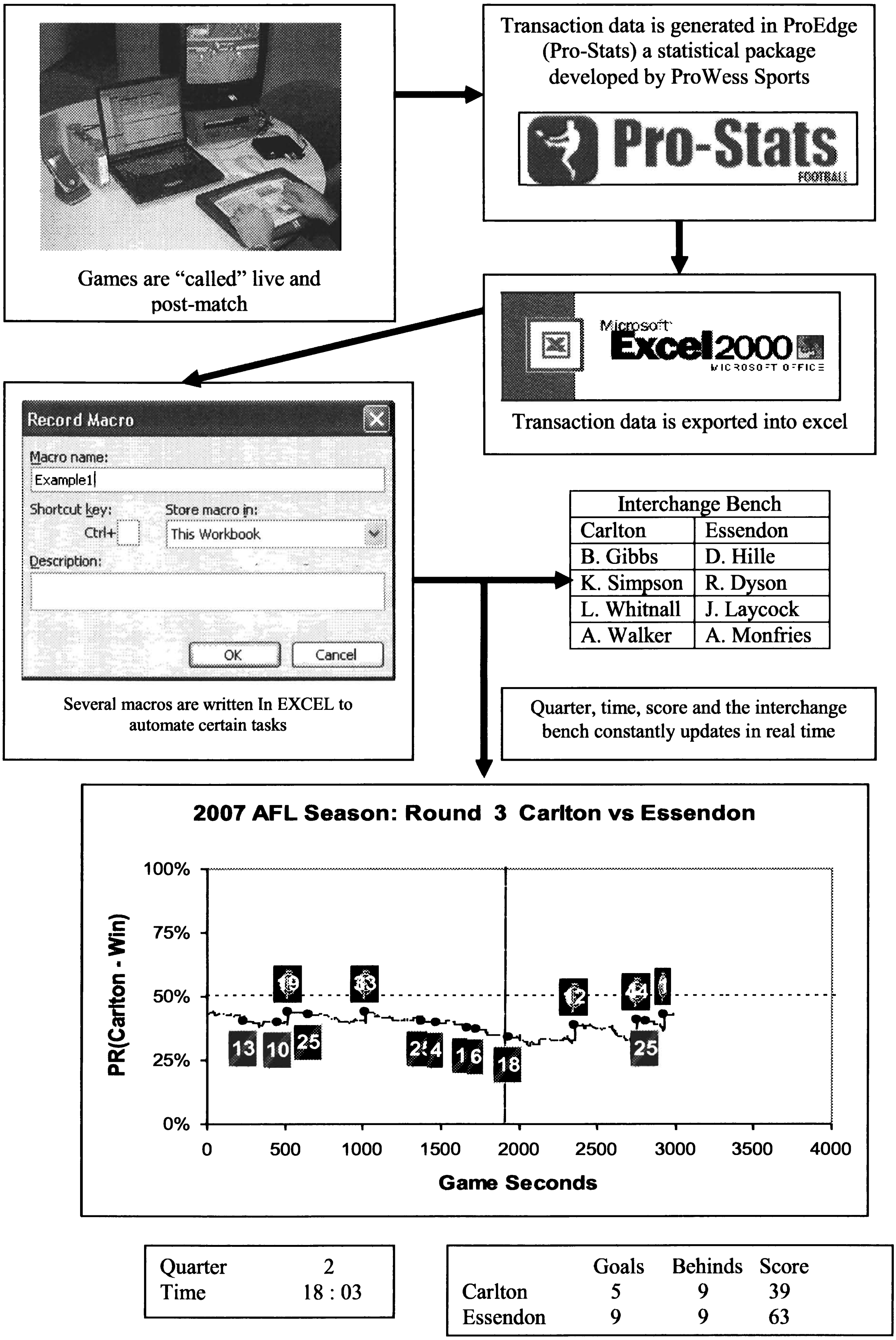


Figure 1: Algorithm of the procedures involved in generating the interactive relative phase plot

Figure 1 above represents the complicated algorithm required to generate the interactive relative phase plot. Although the transaction data is collected post-match the variables which are extracted are only those which match the real footy website live statistics. Therefore the model can theoretically be run live. This game contained many swings, with the eventual winner, Carlton, predicted to win even when they were behind.

CONCLUSION

The final product delivers both a single statistical measure and graphical representation of the state of the game at any point in time from a team perspective. By incorporating interchange data the impact of each player has on the system relative to time on ground is investigated. Applying these live measures during a game will assist coaches in being able in select the players whom have the biggest impact relative to time on ground. The statistical measures used are widely interpretable for use by coaches and the general football public alike.

Acknowledgements

The authors would like to thank ProWess Sports (http://www.prowess.com.au) for their outstanding support, including the use of their data on this project, helpful feedback on the process, and continued support of the RMIT Sports Statistics Research Group.

References

Bedford, A. and Baglin, J. (2008) Evaluating an ice hockey team's performance using interactive phases of play. To appear in the *IMA Journal of Management Mathematics.*

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

Franks, I. M. and Miller, G. (1991) Training coaches to observe and remember. *Journal of Sports Sciences,* **9:** 285-297.

Franks, I. M. and Miller, G. (1986) Eyewitness testimony in sport. *Journal of Sport Behaviour,* **9:** 38-45.

Kelso, J. A. S. (1984) Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology,* **15:** R1000-R1004.

Lames, M. (2006) Modelling the interaction in game sports - Relative phase and moving correlations. *Proceedings of the 8th Australasian Conference on Mathematics and Computers in Sport.* 29-34.

McGarry, T., Khan, M. A. and Franks, I. M. (1999) On the presence and absence of behavioural traits in sport: An example from championship squash match-play. *Journal of Sport Sciences,* **17:** 297-311.

McGarry, T., Anderson, D. I., Wallace, S. A., Hughes, M. D. and Franks, I. M. (2002) Sport competition as a dynamical self-organizing system. *Journal of Sport Sciences,* **20:** 771-781.

Palut, Y. and Zanone, P.-G. (2005) A dynamical analysis of tennis: Concepts and data. *Journal of Sport Sciences,* **23:** 1021-1032.

Stewart, F. M., Mitchell, H. and Stavros, C. (2007) 'Moneyball' applied: Econometrics and the identification and recruitment of elite Australian footballers. *Journal of Sports Economics,* **2:** 231-248.

MULTI-LEVEL MODELS FOR PLAYER PERFORMANCE IN AFL FOOTBALL

Meyer, Denny and Jackson, Karl

Swinburne University of Technology, Melbourne, VIC, Australia

Paper Submitted for Review: 25 March 2008
Accepted Without Revision

Abstract. It is difficult to measure the performance of individual players in the case of complex team sports such as AFL football. Methods based on the number and types of possessions and disposals fail to account for the effects of position and the performance of teammates and opposition players. In this paper the equity approach of O'Shaughnessy (2006) is used in an attempt to overcome this problem. In the context of spatial maps, O'Shaughnessy (2006) suggested that the equity of each possession could be measured by following each chain of play through to the next score. In this paper equity values similar to those developed by O'Shaughnessy will be used to measure player performance for each season. This is done by accumulating the total equity over all possessions and then standardising using time on field for each player. This produces a performance measure for each player in each season which can be used to assess the importance of draft pick, age and height for each position while controlling for the ability of each player's team. This is done by using a two-level multilevel model with player characteristics at level 1 and team characteristics at level 2. The results show that, for all but one of the positions considered, team performance has a positive relationship with player performance. Greater height was a disadvantage only for Key Defenders while greater age was an advantage for all positions except Key Defenders. Low draft picks performed particularly well in the Key Forward, Ruckmen and Midfielder positions. These results have important implications for player selection.

Keywords: draft pick, individual performance, team performance, multi-level models

INTRODUCTION

While it is team performance that is of most interest in AFL football, the performance of individual players is also of great importance. For instance Chu (2000) has reported that "social loafing", the term used to describe decreased expenditure on group tasks (Greenberg et al., 1997), is reduced when individual performance is measured. Bracewell (2003) regards ability as synonymous with performance, suggesting that performance can be quantified using a player's technical skill set. Other authors such as McKenna et al. (1987) and Deutsch et al. (1999) have suggested that performance of individuals be measured in terms of physical effort or work rate rather than skills. However, all of these approaches over-simplify player performance in the case of complex team sports such as AFL football where, according to Salmella and Regnier (1983), "mini-performances must be conceived of within a team context, considered against the strengths and weaknesses of other team mates and the demands of each position". This paper analyses player performance within this wider context.

O'Shaughnessy (2006) has suggested a tool which can be used for this purpose. This tool uses the equity of each possession which is measured by following each chain of play through to the next score. This approach appears to address the above problems, in that it measures the effect of location, pressure and skill for each possession. O'Shaughnessy used these equity values to develop equity contour maps for AFL football for various activities. In a parallel paper, Jackson (2008) uses equity values similar to those developed by O'Shaughnessy to produce a player rating system. In this paper, a variation to the rating system introduced in Jackson (2008), the equity per minute, will be used to measure player performance for each season. Equity per minute is calculated by accumulating the equity over all possessions and then standardising using time on field for each player.

The purpose of this paper is to determine the effects of player characteristics and team performance on this equity performance measure. Using four seasons of AFL data, supplied by Champion Data, regression and multi-level models have been used for this purpose. There are two levels of data. At the first level we have the player data for each season: the player's age, height, his position and, in the case of drafted players, draft pick number. At the second level there is data for each team's performance, including the average points scored per match for each season.

This paper provides useful information for teams because it will determine the importance of the draft pick for player success, while indicating the importance of player height and age for each position. Most importantly it will test whether these effects are moderated by team performance; for example is height more important in stronger teams. It is expected that the effect of height on player performance will depend on position. However, it is also expected that for all positions older players and lower draft picks will have better performance. In addition it is expected that players that belong to better teams will show better individual performance in any season.

METHODOLOGY

The AFL data was supplied by Champion data in three files; demographic player data for players including position, age, height, type of draft and draft number in the case of the National Draft; total equity for each player in each season as well as the total time on field; match results including scores for home and away teams in each season.

As mentioned above, the equity values used in this paper were calculated as described in Jackson (2008). Possessions were split into four phases (set, uncontested, loose & hard) and equity values were calculated as the average scoreboard value of each phase of possession at each location across the entire AFL ground. By combining these equity values with performance statistics recorded by Champion Data, an equity rating system was established. This equity rating is a measure of the scoreboard contribution of individual players within a game, which rewards players that consistently improve the position of their team rather than players that build up large numbers of cheap possessions.

After combining the player data for demographics and equity, descriptive analyses were conducted on the data, with an ANOVA analysis showing significant differences in equity levels depending on position. Linear regression analyses were then carried out separately for each position in order to test whether performance relationships were linear and whether error distributions were homogeneous with normal distributions. These analyses were carried out considering only players who were selected in the National Draft. Partial regression plots which control for the effect of other predictor variables were used to assess the linearity assumption, while residual plots and residual statistics were used to assess the distributional assumptions for the multi-level models.

However, the study obviously involves data for two levels, a player level (level 1 or individual level) and a team level (level 2 or group level). In past studies, the most common approach for the statistical analysis of multilevel data was to disaggregate the data to the individual level, thus treating individuals without reference to their group. Such conventional regression methods tend to focus too much on the individual and too little on the social or institutional contexts in which individuals are located. Multilevel models make it possible to analyse the levels of these structures simultaneously, so consideration about the appropriate level of analysis becomes redundant (Plewis, 1998). Multilevel modelling is a relatively new data analysis technique, in that it has been developed only over the past ten years. As mentioned in Heck and Thomas (2000), although there are numerous books to help in understanding univariate and multivariate data analytic methods using conventional methods of analysis, there are very few books that have included an understanding of multilevel analytic techniques. Similarly, these techniques are not generally available in commonly used statistical software packages.

In multilevel modelling we want to know how a number of level 1 and level 2 variables affect a particular outcome variable. The aim of the analysis is to determine the direct effect of the individual and group level explanatory variables, and to determine if the explanatory variables at level 2 serve as moderators of the level 1 relationships (Hox, 1995). Hence, by focusing attention on the levels of the hierarchy in the population, multilevel modelling enables a better understanding of where and how effects are occurring (Browne et al., 2001). Through examining the variation in outcomes that exists at different levels, more refined theories can be developed about how explanatory variables at each level contribute to outcomes.

An analysis that models the way in which players are grouped within teams in any season has a number of advantages. As Goldstein (1995) stated, it enables data analysis to obtain statistically efficient estimates of regression coefficients. By using the clustering information it provides correct standard errors, confidence intervals and significant tests which will generally be more conservative than the traditional significance tests, which are obtained by ignoring the presence of clustering and assuming that observations at level 1 are independent of other level 2 or contextual factors.

The 2-level models that will be investigated for each position are shown below with P_{ij} indicating the performance for the *i*th player in the *j*th team/season. The Greek letters indicate model parameters and it is assumed that the error terms, e_{ij} and u_j are random but normally distributed with constant variance. The variables in this model are described in Table 1.

$$P_{ij} = \beta_{0j} + \beta_{1j} DRAFT_i + \beta_{2j} AGE_i + \beta_{3j} HEIGHT_i + e_{ij}$$
$$\beta_{kj} = \gamma_{k0} + \gamma_{k1} SCORE_j + u_j$$

for k=0, 1, 2 and 3.

Table 1: Variables for Multi-Level Model

Construct	Variable	Measure
Player Performance	P	Total equity for season divided by minutes on field
Draft Pick	DRAFT	Draft Number
Age	AGE	Years
Height	HEIGHT	Centimetres
Average team score	SCORE	Average team score for season after subtracting the mean match score for all teams in all seasons
Position	POSITION	Key and General Forward, Ruckman, Key and General Defender, Midfielder

Draft number is only available for player selection based on the national draft and it is expected that the model will change depending on position. For this reason the analysis has been performed separately for each position considering only players who were selected through the national draft. Only in the case of Utility players was no analysis performed on account of the small number of players (24 Tall Utility players and 16 General Utility players).

RESULTS

Descriptive Statistics

The data was collected by Champion Data, the official provider of AFL statistics, for the 2004-2007 seasons. However, only players who took part in the 2007 season were included in the analysis. As a result players in the 2004 and 2005 seasons were under-represented in the data (19% and 23% respectively) while the 2006 and 2007 seasons were over-represented (27% and 31%).

Table 2 demonstrates the physical requirements of the various positions only for the 62.4% of players who were selected through the national draft. In the case of height there was a significant difference between the positions ($F(5,376) = 142.598$, $p<.001$). In addition there was a significant difference between positions for equity score per minute on field (F(5,780) = 25.309, p<.001), with Key Defenders performing significantly worse than other positions and Forwards and Midfielders performing significantly better than the other positions. This is consistent with the results presented in Jackson (2008), where it was found that players playing in the forward line were significantly advantaged by the equity ratings. The height measurements relate to levels at the end of the 2007 season but the age variable was corrected for each season.

Table 2: National Draft Player Statistics 2004-2007 for 2007 players only

	Number		**Mean Player Characteristics (SD)**			
Position	**Players**	**Seasons**	**Height (cm)**	**Age (yrs)**	**Draft Number**	**Equity by Season**
General Defence	49	107	187.04 (4.40)	22.16 (2.91)	34.00 (20.61)	.08 (.03)
General Forward	49	98	183.08 (4.31)	22.47 (3.59)	33.29 (18.91)	.11 (.03)
Key Defence	38	76	193.08 (2.38)	22.84 (3.93)	35.18 (20.04)	.07 (.02)
Key Forward	43	77	193.86 (3.14)	22.30 (3.64)	30.26 (22.74)	.10 (.04)
Midfielder	153	379	183.62 (4.23)	23.09 (23.62)	30.38 (21.42)	.10 (.03)
Ruckman	50	49	197.80 (4.19)	21.87 (2.80)	30.90 (22.56)	.08 (.03)

The data is most rich for midfielders. There were 153 Mid-fielders playing in an average of 2.54 seasons during the period 2004 to 2007. As Table 2 shows, the average height of these players was 183.62cm with an average age of 23.09 years and an average draft number of 30.38. For these players Figure 1 suggests that the distribution of equity per minute in any season was roughly normally distributed with a mean value of 0.10 equity points per minute and a standard distribution of 0.03 equity points per minute. Similarly shaped distributions were obtained for the other positions.

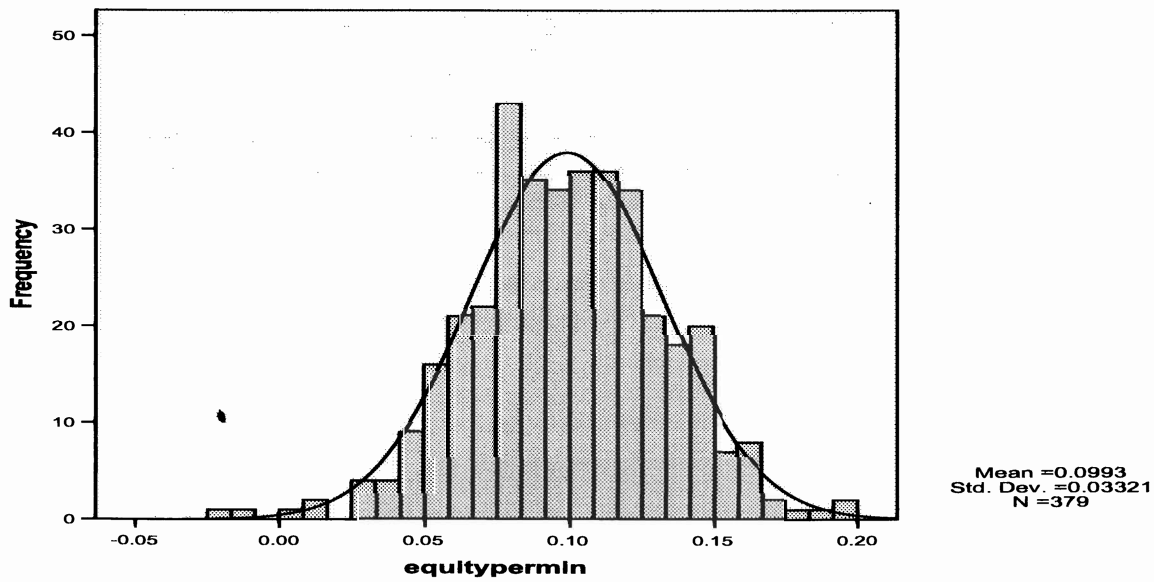


Figure 1: Distribution of Equity per Minute for Drafted Mid-Fielders

Regression Analysis

An initial regression analysis was performed for each position, identifying significant predictors in terms of height, age and Draft Number while ignoring the clustering of players within teams in any season. The purpose of this analysis was to identify outliers and to test the assumption of linearity and normality using appropriate plots.

Table 3: Summary of Regression Analyses for Individual Performance (Equity Per Minute)
Significant partial correlations (P.C.) (** p<.001, * p<.05)

Position	Midfielders	General Defence	General Forward	Key Defence	Key Forward	Ruckmen
Outliers Deleted	0	0	2	1	2	2
R-Square (%)	7.9	25.2	22.5	8.2	34.4	12.1
Height P.C.	-.014	-.09	-.113	**-.282***	.044	-.139
Age P.C.	**.259 ** **	**.469 ** **	**.466 ** **	-.085	**.575 ** **	.262
Draft Number P.C.	**-.152****	-.134	-.109	.105	**-.271***	**-.307***

These results in Table 3 suggest that shorter Key Defenders have an advantage while older Key Defenders do not. The age coefficient for Ruckmen was not quite significant (p = .082). Lower draft picks appear to perform particularly well in the position of Midfielder, Key Forward or Ruckman. However, the probable nesting of player performance within teams within any season made the assumption of residual independence untenable in the above regression analyses. This analysis has merely served to justify the linear and normality assumptions of the following multi-level analysis while identifying outliers for deletion purposes.

Multi-Level Models

In addition to the advantages described above, the Multi-Level Models allow a measure of team performance to be included as a predictor of player performance. The measure of team performance used in Table 4 is the average match score for a team in any season after subtracting the mean score for all teams in all seasons. A positive score means that the team has produced an above average performance in the season while a negative score means that the team has produced a below average performance in the season.

Table 4 shows the results when the Multi-Level Model described above is fitted separately for each position. It was found that although team performance, as measured by average match score, did have an impact on player performance for all but Ruckmen, the effects of height, age and draft number were not moderated by team performance. For none of these variables did the coefficient (β_{ij}) depend on the Score variable.

Table 4: Summary of Multi-Level Models for Individual Performance (Equity Per Minute)
(** p≤.001, * p<.06)

Position	**Midfielders**	**General Defence**	**General Forward**	**Key Defence**	**Key Forward**	**Ruckmen**
Coefficients*100						
Intercept	8.9952	9.3736	10.9754	66.7048*	-3.7924	34.3253
Score	.0773**	.0665**	.1190**	.0653*	.0778*	-.0085
Height	-.0269	-.0688	-.0529	-.3048*	.0124	-.1762
Age	.2771**	.5323**	.4307**	-.0655	.5329**	.4575*
Draft Number	-.0217*	-.0203	-.0076	.0135	-.0282*	-.0402*

The Multi-Level Model results were very similar to those obtained from the regression analyses. Age had a significant positive impact on player performance for all positions except Key Defenders. Also, as found in the regression analysis, height had a significant impact only in the case of Key Defenders. As before, it was found that draft number had a significant impact in the case of Midfielders, Key Forwards and Ruckmen.

DISCUSSION

Player performance differs significantly between positions, suggesting that the equity measure used in this study is not transferable between positions. In particular, players in the Key Defence position perform significantly worse than other positions according to this measure while Forwards and Midfielders perform significantly better than the other positions. Because of the differences between the distributions of the explanatory variables and the equity ratings between positions, analysis had to be performed separately for each position.

As expected it was found that player performance was better in the case of better teams. Only in the case of Ruckmen was this result found to be non-significant. Key defensive players are unique in that age is not an advantage for these players while being short is a decided advantage. It appears that low draft players are particularly important for the positions of Midfielder, Key Forward and Ruckmen. This is an expected result, since the better players that are nominated for a given draft are taken early, and hence have a low draft number.

CONCLUSION

The results suggest that teams should prioritise draft picks in the case of Midfielders, Key Forwards and Ruckmen. Having more talented players in these positions appears to be more important than in the case of general forwards, general defenders and key defenders. For key defensive players, it seems as though shorter players perform the best in equity per minute. It is expected that shorter players possess more speed than taller players, which may help to explain how shorter key defenders are outperforming taller players in the same position.

The results suggest interesting strategies in terms of the career development of players. Older players outperform younger players in all but one position. This is not unexpected, since it often takes several seasons for younger players to mature into dominant footballers. Seldom does a first or second year player dominate in his given position.

The under-representation of 2004 and 2005 players is a limitation of the above analysis. It means that there is probably an under-representation of players who withdrew from the AFL competition at an early age and it is difficult to predict what effect this would have had on the results.

Finally the equity measure used in this analysis is regarded as only a preliminary measure by Jackson (2008). The results of this paper confirm that the equity measure considered in this paper is not fair to all positions. This measure needs to be modified in order to take into account the opportunity offered by each position. Once this has been done it will be appropriate for the current multi-level analysis to be repeated, in order to confirm the variation in performance effects that was observed across positions in this analysis.

Acknowledgements

The authors wish to thank Champion Data for the use of their data.

References

Bracewell, P. J. (2003) Quantification of Individual Rugby Player Performance through Multivariate Analysis and Data Mining. PhD thesis. Massey University, New Zealand.

Browne, W. J. and Rasbash, J. (2001) Multilevel Modelling. To appear in Bryman, A. and Hardy, M. (eds.), Handbook of Data Analysis. URL: http://www.maths.nottingham.ac.uk/personal/pmzwjb/materials/mmsage.pdf

Chu, M. (2000) Team Dynamics. Advanced Coaching Course. Pukekohe: NZRFU.

Deustch, M., Kearney, G. A., and Rehrer, N. J. (1999) A comparison of competition work rates in elite club and Super 12 Rugby Union. *Journal of Sports Science*, **17:** 809-810.

Goldstein, H. (1995) Multilevel Statistical Models (2^{nd} Ed). London: Edward Arnold.

Greenberg, J. and Baron, R. A. (1997) Behaviour in Organizations. New Jersey: Prentice-Hall.

Heck, R. H. and Thomas, S. L. (2000) An Introduction to Multilevel modelling techniques. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hox, J. J. (1995) Applied Multilevel analysis. Amsterdam 1995 URL http://www.geocities.com/joophox/publist/amaboek.pdf

Jackson, K. (2008) A player rating system for Australian rules football using Field Equity measures. In J. Hammond (ed.) *Proceedings of the Ninth Australiasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM) Coolangatta, Queensland, Australia (in press).

McKenna, M. J., Patrick, J. D., Sanstom, D. R. and Chennels, M. H. D. (1987) Computer-Video analysis of activity patterns in Australian Rules Football. Science and Football. Eds. Reilly, T., Lees, A., Davids, K. and Murphy, W.J., 274-281.

O'Shaughnessy, D. (2006) Possession versus Position: Strategic Evaluation in AFL. *Journal of Sport Science and Medicine*, **5(4):** 533-540.

Plewis, I. (1998) Social Research Update Issue 23: Multilevel Models Winter 1998 online version: http://www.soc.surrey.ac.uk/sru/SRU23.html

MEASURING CONSISTENCY IN PERFORMANCE USING DATA TRANSFORMATIONS

Sargent, Jonathan and Bedford, Anthony

School of Mathematical & Geospatial Sciences, RMIT University, Melbourne, VIC, Australia

Paper Submitted for Review: 29 January 2008
Revision submitted and accepted: 16 June 2008

Abstract. This study outlines univariate data transformation methods designed to force a near-normal distribution on the 2007 Australian Football League so as to identify players who perform consistently well. Analysing 7,700 individual player cases in the 2007 AFL season, and a selection of performance variables (kicks, handballs etc), a more robust application than simple player 'averages' is developed. The consistency statistic, a variation on the Coefficient of Variation, reveals the players whose round-by-round performances vary the least from a personal and league standard while favouring players who have played the majority of games in the chosen measurement period. In realising each player's standard, a "time-on-ground" variable is introduced to estimate a player's final match performance should they be absent from the match at any stage. Incorporating a log transformation, this variable rewards strong performance in partial matches whilst leaving poor performance relatively unchanged. An initial finding is the performance variables (supplied by ProWess data) commonly possess a right-skewed distribution implying any performance much less than average is unacceptable. Box-Cox power transformations are separately applied to the time-on-ground transformed performance data to create an approximately normal distribution, an assumption for unbiased use of the consistency statistic. The intimate relationship between these two transformations will be detailed in this paper with the accompanying results. Player performance consistency becomes highly regarded from a coaching perspective, particularly for team selection and player "match-ups". Player consistency also aids in player selection for participants in fantasy leagues. It is perceived that this model will translate simply to other team sports.

Keywords: Box-Cox Power Transformations, Time-on-Ground Transformation, Consistency Measure

INTRODUCTION

Much like an investor expects consistent returns from his/her portfolio, so too does a sporting coach expect consistent returns from his/her "investments". Why is the issue of performance consistency so important? Of fund management, Marquardt (2008) remarks the more regularly managers beat their peers and their benchmarks, the greater the likelihood that skill, rather than luck, is driving their performance. Furthermore, consistent performers give you more certainty over your money. It means the returns on your investments are less volatile. Sporting coaches highly value consistency in their players' performances in the same way: "You can't let poor performance go unnoticed - even from a superstar. The same goes for good performance - performance is all-important; that's what you need to respond to on a consistent basis" (Shula and Blanchard, 1995).

This paper seeks to isolate consistently good performers in the 2007 AFL season (Australian rules football's premier competition) using equations derived from the coefficient of variation. Initially, univariate data transformations designed to force a near-normal distribution are trialled. A major objective is to achieve at least a symmetrical distribution from which to calculate the consistency measurements.

The first data transformation in the research recognises the inability of a team competitor to achieve at least what is expected of him/her due to injury; a constant distortion in performance analysis. Using "time-on-ground" data, it will be shown that a competitor who plays partial matches can have a final quantified performance estimate based on the actual performance achieved to the moment of his/her removal from the contest (James et al., 2005). The paper will establish the contribution to near-normality a "time-on-ground" transformation contributes to the distribution of performances in the 2007 AFL season.

The consistency equation incorporates a performance average across the league. A proven right-skewed distribution of AFL performance variables forces the analysis to focus on methods of achieving symmetry in the data if the mean and variability (known, not being inferred) are to be used in non-biased measurement. With this knowledge, Box-Cox (1964) transformations are applied to the time-on-ground transformed AFL performance distribution to remedy skewness and kurtosis. As a result, it will be proven that competitors located around the mean of the Box-Cox transformed data set have a greater probability of appearing in higher order consistency than if the data were left untransformed. The pre and post-transformation performance data distributions will be evaluated by observing histograms.

Much interest surrounds selecting consistent performers in fantasy sport leagues, or "dream teams". Waldman (2005) believes a fantasy team of players performing consistently at a desired level is more valuable than higher scoring, but more erratic players. This paper considers two types of consistency measure – a personal and league consistency, based on equations derived from the coefficient of variation. It is important to establish that the consistency measure will only return logical results when applied to an appropriate position-based performance measure in the notational analysis. For example, it is worthless measuring consistency in goal kicking for a defender, given that it is a defender's role to prevent goals. Another consideration is that the consistency measure is not necessarily identifying the "best" players in a league, but rather those that display the least variability in their game-to-game performances.

The paper will conclude with a comparison of the most consistent players based on different orderings of data transformations applied to the different consistency measures. The benefits to coaching, as well as correlation between the consistency results and club champion awards will be discussed to ascertain the true worth of this analysis.

METHOD AND RESULTS

Hughes and Bartlett (2002) identify performance as any combination of quantifiable performance variables within a match that constitute effective team play. Moreover, they discuss the concept of notational analysis, or the performance of individual team members in team sports based on "open skills" (kicks, goals etc). Some fantasy sport competitions or "dream teams" employ simple algebraic equations (see (1)) to determine player j's performance, Y_j after each match, with the sum of each performance value X_i recorded for each player in a match, assigned an arbitrary coefficient.

$$Y_j = 3X_k + 2X_h + 3X_m + 6X_g + X_b + 4X_t + X_{ho} + X_f - 3X_{fa} \qquad (1)$$

where: k = kick, h = handball, m = mark, g = goal, b = behind, t = tackle, ho = hit-out, f = free-for, fa = free-against

Although this formula is proven to bias certain positional plays (Sargent & Bedford, 2007), in the interests of simplicity the same performance measure is to be employed to demonstrate the benefits of data transformations, before calculating consistency in the latter stages of this work. See James et al. (2005) for research on performance indicators for specific positions in rugby union.

Figure 2A displays the untransformed, standardised performance for each player case with μ=0 and σ=1, by (1). The initial distribution is right-skewed (0.502 by (9)) with negligible kurtosis (-.012 by (10)). Before calculating consistency statistics for AFL players, this paper will discuss in detail the necessary transformations found to force a symmetrical and hence nearer-normal distribution on the player performance data in Figure 2A. All calculations have been performed using SPSS, except when optimising log likelihood functions for Box-Cox transformations, where a script exports the transformed data to EXCEL (Solver) then back to SPSS for consistency measurement.

Time-on-ground transformation

In many team sports, a competitor faces the reality of not realising a full game due to injury, substitution or send-off, hence depriving him or her the opportunity to achieve at least an expected quantifiable performance. For the purposes of this research, it is deemed unfair that a player's consistency should be jeopardized for playing a partial match due to injury sustained in that match. James et al. (2005) describe data transformations to account for this shortcoming in their performance measurements in rugby union.

The exponential nature of the transformation (see Figure 1) rewards team competitors who have performed well in a partial match, while leaving poor performances in partial matches relatively unchanged. Assuming 125 minutes in a game of AFL (four 30 minute quarters plus time on), each player's performance measurement could be expressed as a function of time spent on the ground.

$$\Delta_1 = x_i\left(\frac{125}{y}\right) \tag{2}$$

where: x_i = frequency of performance variable x in game i; y = time on ground (minutes)

This provides an acceptable transformation, however the exponential nature of the function (see Figure 1) dictates that one performance point in one minute of playing time will be extended to 125 for the match. This is an overly generous forecast and is improved by the use of (3).

$$\Delta_2 = x_i\left(\sqrt{\frac{125}{y}}\right)\left[\left(\log_{10}\frac{125}{y}\right)+1\right] \tag{3}$$

Given that taking the square root and/or logarithm of cases in an asymmetric data set are basic but (sometimes) effective near-normal transformation methods (Chinn, 1996), it seems logical that this method be preferred to (2). In addition, (3) yields a far more realistic estimation of final performance in a partial match – although a one minute-one performance point match is rare, a forecast score of 35 by (3) is prudent.

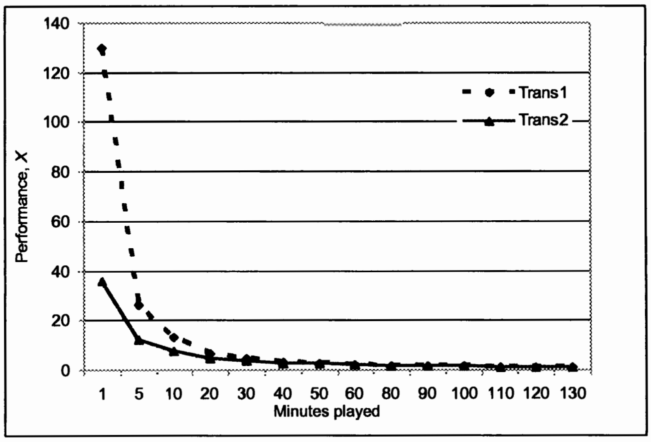


Figure 1: Time-on-ground transformation

It will now be shown that when (3) transformed data is transformed again using Box-Cox methods, symmetry is achieved in the performance data distribution.

Transformations to approximate normality

Right-skewed distributions are certainly not uncommon among sporting performance variables. Poisson techniques (commonly applied to other right-skewed data such as waiting times) have been utilised in modeling the distribution of goals scored by competing teams (Karlis, 2003). Commonly employed when desiring a normally distributed sample on which to perform parametric tests, non-linear data transformations, such as power transformations are extremely effective in removing asymmetry from data sets (Smith, 1998).

In this research, data transformations are to be applied to the population (league) data to force an approximately normal distribution, from which to take player samples.

What are the parametric benefits of a symmetrical distribution? A successful transformation of a skewed dataset to a reasonably symmetrical one implies a central measure, such as a mean or median becomes more suitable for parametric estimation (Smith, 1998). Further, we note from classic normal distribution theory that a sample from a normally distributed population is approximately normal. Given the consistency measures employed later in this paper are developed from the ratio of the second (σ) to the first (μ) moments of the normal distribution, it is appropriate that each statistic be derived from at least an approximately normal data set.

The first transformation discussed in this research was developed by Tukey (1957) who introduced a power transformation of the form:

$$x(\lambda)=\begin{cases} x_i^{\lambda} & (\lambda \neq 0) \\ \log x_i & (\lambda = 0) \end{cases} \tag{4}$$

Box-Cox (1964) developed maximum-likelihood methods in data transformations. The foremost Box-Cox power transformation is of the form:

$$x(\lambda)=\begin{cases} \dfrac{x_i^{\lambda}-1}{\lambda} & (\lambda \neq 0) \\ \log x_i & (\lambda = 0) \end{cases} \tag{5}$$

The Box-Cox shifted power transformation becomes:

$$x(\lambda)=\begin{cases} \dfrac{(x_i+\lambda_2)^{\lambda}-1}{\lambda_1} & (\lambda \neq 0) \\ \log(x_i+\lambda_2) & (\lambda = 0) \end{cases} \tag{6}$$

Transformation (4) and (5) requires that $x > 0$ and (6) that $x > -\lambda_2$. Given the vector of performance data observations, $x_i = x_1,..,x_n$, the power λ in (5) can be selected so that it maximizes the logarithm of the likelihood function (7) (Chinn, 1996).

$$f(x,\lambda)=-\frac{n}{2}\log\left[\sum_{i=1}^{n}\frac{(x_i(\lambda)-\bar{x}(\lambda))^2}{n}\right]+(\lambda-1)\sum_{i=1}^{n}\log(x_i) \tag{7}$$

where $\bar{x}(\lambda)=\frac{1}{n}\sum_{i=1}^{n}x_i(\lambda)$ is the arithmetic mean of the transformed data.

When considering the shifted transformation (6), the second term in (7) becomes:

$$(\lambda-1)\sum_{i=1}^{n}\log(x_i+\lambda_2) \tag{8}$$

Rigby and Stasinopoulos (2004) suggest the shifting parameter, λ_2 can be optimised so as to minimise kurtosis, however its effect on this data was negligible, and the resulting distribution too similar to (5) for further consideration in consistency measurement. By way of (5) the resulting distribution (see Figure 2d) is symmetrical but the density around the mean suggests kurtosis remains an issue. Applying (10) to the TB data, a kurtosis of -0.242 supports this.

$$Sk = \frac{\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^3}{n} \tag{9}$$

$$Ku = \frac{\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^4}{n} - 3 \tag{10}$$

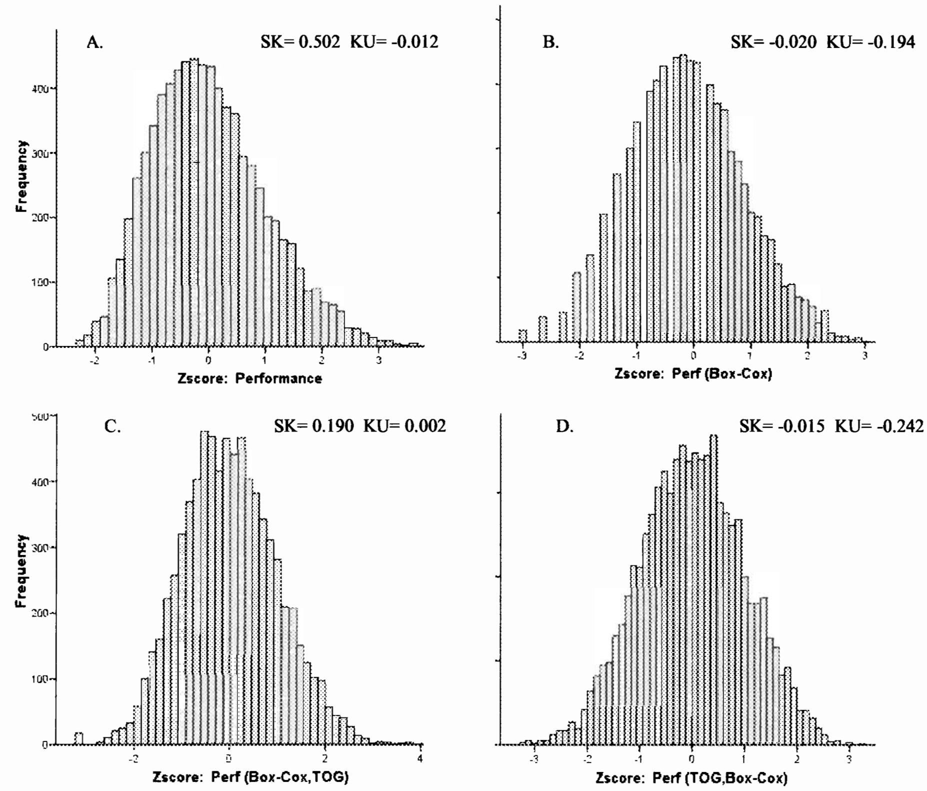


Figure 2: Histograms of transformations with skewness and kurtosis coefficients

From Figure 2, the untransformed standardised performance data (A) possesses a right skew without kurtosis. Running a Box-Cox power transformation (5) treats skewness but introduces kurtosis (B); the right half of the distribution is denser than the left. Applying (3) to the (5) transformed data treats kurtosis but reintroduces skewness (C); (D) is the standard normal distribution of performance data transformed by (3) then (5). Skewness is virtually non-existent, yet kurtosis is still present, however the Shapiro-Wilk test proves that (D) is the closest to normality ($p = 0.046$).

Consistency measurement

Having performed important transformations on the asymmetric performance data, it is now possible to take a series of non-random samples (each player's season performances) from the population (league) to arrive at consistency measures for each player. Three consistency measures are considered.

Coefficient of Variation: $$CV_k = \frac{\sigma}{\bar{x}} \quad (11)$$

CV has been incorporated in the past to measure performance consistency in sport such as basketball (based on points scored) (Manley, 1988). The limitation associated with CV is the reality of players who have played two to three games appearing in the top ranks of consistency. Based on average performance, two to three games is not sufficient proof a competitor is a consistently good performer.

Personal consistency measure: $$PCM_K = \frac{\sigma}{n\bar{x}} \quad (12)$$

The PCM formula, given by (12), discounts players with limited matches in a season by dividing performance standard deviation by the number of games played in the season, n. Concurrently, the method attracts higher averaging players into the top rankings.

Observing variability in week-to-week performance, coaches and fantasy league participants are keen to know how consistently their players are performing at least at a desired level (Waldman 2005). The desired level is dependent on several factors governing each player's ability to perform, but for the purposes of this research, an expected level will be defined as the average quantifiable performance across the 2007 AFL season, as given by μ. LCM, as given in (13), overcomes a problem defined by Waldman (2005) of simple standard deviations penalizing players who consistently and erratically score over the established norm (see Table 1) by only measuring consistency in performance above the league average.

League consistency measure:

$$LCM_k = \frac{\sigma_{(x_i-\mu)}}{\sum_{i=1}^{n}(x_i-\mu)} \quad (x_i>\mu) \quad (13)$$

where: σ = standard deviation of player k's performances, $\bar{x}$ = mean of player k's performances, μ = average league performance, x_i = player k performance in round i, n = number of games played by player k

Bivariate correlation calculated between all three consistency measures (transformed and untransformed) and club champion awards (1 to 10 voting system where 1 is best) are designed to give an insight into the predictive power of consistency. It should immediately be noted that transformed data consistency measures (denoted by subscript t in Table 1) each have a higher correlation than their untransformed equivalent. The top half of Table 1 shows top ranked players by the LCM_t method. Significant correlation (at α=0.10) between these consistencies and club champion votes show this method to be the best predictor of player awards out of all methods trialed. This is expected given the consistency is only calculated on player's games which exceed the league average. Note its impact for Harvey (club champion for Kangaroos): CV_t ranks him at 236, while LCM_t ranks him at 6.

The bottom half of the table ranks players by PCM_t. With a lower consistency-award correlation and average score, it is isolating players, who although may not all be recognized as the best performers, are showing the least variability around their expected personal performance from week to week. This information is potentially of great interest to coaching staff who want to monitor their "up-and-coming" players. Players who achieve top ten rankings in PCM_t and LCM_t (Bartel, Tuck and Foley) should be of great interest to coaches and fantasy league participants, given the chosen performance variable (1). These players are consistently achieving their personal and league expectations, benefiting from skill rather than luck.

Table 1: Ranked consistent players by LCM (top half) and PCM (bottom half) where *t* denotes transformed data

Player	Games	Score	CV*t*	PCM*t*	LCM*t*	CV	PCM	LCM
Corey	21	109.29	96	41	1	28	10	1
O'Bree	21	82.52	94	38	2	84	43	62
Thompson (Ad)	22	103.14	86	24	3	33	4	3
Bartel	20	120.3	35	10	4	66	45	2
McDonald (Me)	21	98.38	71	25	5	39	15	16
Harvey (Ka)	22	99.18	236	97	6	143	57	8
Tuck	22	90.59	44	4	7	40	8	35
Lewis	22	99.55	150	54	8	63	22	14
Foley	22	91.45	33	2	9	41	9	30
Burns	22	87.05	92	27	10	71	26	24
Johnson (WB)	21	103.81	19	1	27	16	1	6
Foley	22	91.45	33	2	9	41	9	30
Van Berlo	22	85.91	37	3	18	57	19	61
Tuck	22	90.59	44	4	7	40	8	35
Akermanis	19	76.16	31	5	81	18	5	121
Milburn	21	87.14	39	6	88	19	2	46
Watson	19	88.47	32	7	11	44	34	106
Stiller	22	87.5	48	8	73	70	25	93
Crawford	21	94.19	42	9	13	23	3	18
Bartel	20	- 120.3	35	10	4	66	45	2
Correlation			0.134	0.219	0.226	0.065	0.196	0.220

CONCLUSION

Box-Cox transformation methods are successful in forcing approximate normality on univariate team sport performance data. It was shown that when following a logarithmic "time-on-ground" transformation (James et al., 2005), a Box-Cox power transformation (Box & Cox 1964) is most effective in forcing an approximately normal distribution on league performance data with known parameters in preparation for consistency measurement. When calculated on the approximately normally distributed league data, the consistency measurements gave a broader insight into the players most consistently reaching a desired level from week-to-week. The coefficient of variation method too easily accounted for players who had played limited games, hence does not provide enough evidence of consistency. The PCM method gave a good indication of players who weren't necessarily the highest averaging players in the league, but were consistently achieving their own expected performance on a consistent basis. The LCM method also proved to have the highest correlation with player "best and fairest" awards in 2007, hence promoting the predictive capabilities of the model. These findings provide both coaches and fans a far more accurate rating of ongoing consistency that those reported, be it through simple averages, or even more advanced time adjusted data. With these findings, we aim to further our analysis on a player ratings model by incorporating the transformations outlined in this work.

Acknowledgements

The authors wish to thank ProWess Sports (www.prowess.com.au) for use of their data.

References

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society.* **26**: 211-252.

Chinn, S. (1996) Choosing a transformation. *Journal of Applied Statistics*. **23**: 395-404.

Hughes, M. D. and Bartlett, R. M. (2002) The use of performance indicators in performance analysis. *Journal of Sports Sciences*. **20**: 739-754.

James, N., Mellalieu, S. D. and Jones, N. M. P. (2005) The development of position-specific performance indicators in professional rugby union. *Journal of Sports Sciences*. **23**: 63-72.

Karlis, D. (2003) Analysis of sports data using bivariate Poisson models. *The Statistician.* **52**: 381-393.

Manley, M. (1988) *Martin Manley's Basketball Heaven.* Doubleday, New York.

Marquardt, K. (2008) Consistency counts. *Kiplinger's Personal Finance: 2008 Mutual Funds.* 50-53.

Rigby, R. A. and Stasinopoulos, D. M. (2004) Smooth centile curves for skew and kurtotic data modeled using the Box-Cox power exponential distribution. *Statistics in Medicine.* **23**: 3053-3076.

Sargent, J. and Bedford, A. (2007) Calculating Australian Football League player ratings using a T4253H Elo model. *Proceedings of the first IMA International Conference on Mathematics in Sport.* Manchester, UK. 186-191.

Shula, D. and Blanchard, K. (1995) Winning strategies revealed for scoring big-in any game. *Manager's Magazine*. **70**: 3.

Smith, P. (1998) *Into Statistics*. Springer, Singapore.

Tukey, J. W. (1957) The comparative anatomy of transformations. *Annals of Mathematical Statistics.* **28**: 602-632.

Waldman, M. (2nd August 2005) http://fftoday.com/articles/waldman/05_gc_consistency_1.htm. Fantasy Players Network. Accessed: 1st March 2008.

MODELLING OUTCOMES IN VOLLEYBALL

Barnett, Tristan [1], Brown, Alan [2] and Jackson, Karl [1]

[1] Faculty of Life and Social Sciences, Swinburne University of Technology, Australia

[2] Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Australia

Paper Submitted for Review: 16 March 2008
Revision submitted and accepted: 17 June 2008

Abstract. A Markov Chain model is applied to volleyball to calculate win probabilities and mean lengths with the associated variances, conditional on both the scoreboard and the server. A feature of this model is that it predicts outcomes conditional on both the scoreboard and the serving team. The inclusion of the serving team in the event space is an essential requirement of this model, and arises from the rule in volleyball that the winner of each point in a set must serve on the following point. The average probability of a team winning a point on serve is less than 0.5, and so rotation of serve is commonplace. The key to the analysis of an evenly contested set is the observation that, from the situation where the scores are level (after at least 46 points have been played), the team that wins the set must eventually win two successive points. If the two points are shared then the score is level once again, although a rotation of server has occurred. This scoring structure, combined with the method of rotating the serve, distinguishes volleyball from other racket sports such as tennis, squash, badminton and table tennis. Results from the model indicate that it is advantageous to be the receiver on the opening point of a set and the team that wins the toss at the start of the fifth set (if the set score reaches 2-all), has an advantage for the remainder of the match. However, due to the rotation of serve after each set, there is no advantage for either side in being server or receiver at the start of the match.

Keywords: volleyball, Markov Chain model, scoring systems

INTRODUCTION

Markov Chain models are widely used in modelling sporting outcomes. Kemeny and Snell (1960) recognized that tennis could be modelled by Markov Chains. Tennis has four levels (point, game, set, match) and the time to play the match is not fixed, but rather depends on a player winning 2 or 3 sets. Other racket sports contain a similar structure to tennis and Markov Chain models could be developed to model these sports. For example, Clarke and Norman (1979) use Markov Chain models to compare North American and International squash scoring systems.

A Markov chain model can be applied to volleyball in a similar approach to racket sports. In volleyball, scoring consists of three levels: point, sets and match. A coin is tossed to determine the first serve of a match. At the start of each set, the team that was receiving first in the previous set becomes the server for the next set. If the set score reaches 2-all, then the toss of a coin decides the server for the start of the final set. Each team can win a point while either serving or receiving. The first team to win three sets wins the match. Each set is played as a 25-point set with the exception that after the set score reaches 2-all, the final set is played as a 15-point set. If the score reaches 24-all in a 25-point set, then play continues indefinitely until one team has obtained a two point lead. Similarly, if the score reaches 14-all in a 15-point set then the play continues indefinitely until one team has obtained a two point lead.

In this paper, a Markov Chain model is applied to volleyball to predict outcomes conditional on both the scoreboard and the server. The inclusion of the server in the event space is an essential feature of the model for volleyball because of the rule on serving in this sport, and distinguishes it from models for other racket sports such as tennis, squash, badminton and table tennis.

PROBABILITY OF WINNING

Point

Volleyball has the added complication of having 6 players that make up a team, rather than just the one player each side as occurs in racket sports. To simplify the analysis we will assume throughout that the probabilities of winning a point by each player in a team on their respective serves are identical and constant, irrespective of the score. Therefore, the model consists of two parameters, the probabilities of team A and team B winning a point on their respective serves. The probabilities of winning a point on serve are represented as follows:

	A winning point	B winning point
A serving	p_A	$q_A = 1 - p_A$
B serving	$q_B = 1 - p_B$	p_B

Set

The conditional probabilities of a team winning a set from point score (a, b) in a 25-point set are represented as follows:

	A winning set	B winning set
A serving next	$P(A \mid A,a,b)$	$P(B \mid A,a,b) = 1 - P(A \mid A,a,b)$
B serving next	$P(A \mid B,a,b)$	$P(B \mid B,a,b) = 1 - P(A \mid B,a,b)$

The conditional probabilities of a team winning a set from point score (a, b) in a 15-point set are represented as follows:

	A winning set	B winning set
A serving next	$P^*(A \mid A,a,b)$	$P^*(B \mid A,a,b) = 1 - P^*(A \mid A,a,b)$
B serving next	$P^*(A \mid B,a,b)$	$P^*(B \mid B,a,b) = 1 - P^*(A \mid B,a,b)$

We will adopt a similar notation for all other occasions where we need to distinguish between a 25-point set and a 15-pont set without further comment.

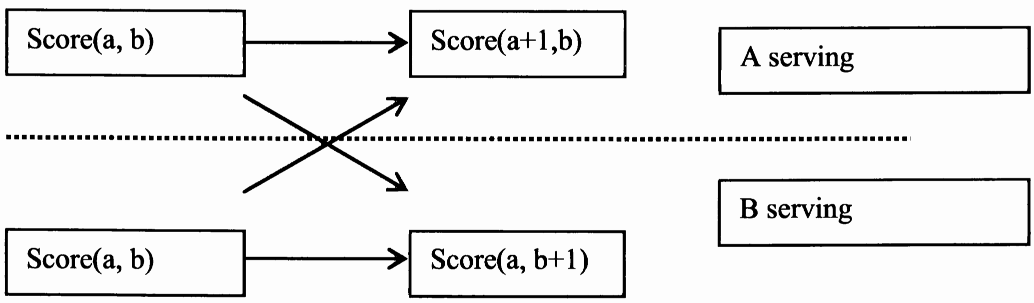


Figure 1. One step transitions between states of play.

Figure 1 illustrates the one step transitions between the various states of play. We set up a Markov chain model for the case of team A winning a 25-point set using backwards recurrence formulas, as follows:

$$P(A \mid A,a,b) = p_A P(A \mid A,a+1,b) + q_A P(A \mid B,a,b+1)$$
$$P(A \mid B,a,b) = p_B P(A \mid B,a,b+1) + q_B P(A \mid A,a+1,b)$$

The boundary values are:

$P(A \mid A,a,b) = 1$ if $a = 25, 0 \leq b \leq 23$
$P(A \mid B,a,b) = 0$ if $b = 25, 0 \leq a \leq 23$
$P(A \mid A,24,24) = p_A^2 / D$
$P(A \mid B,24,24) = p_A q_B (1+ p_A p_B - q_A q_B) / D$

where $D = (1 - q_A q_B)^2 - p_A q_A p_B q_B$

A method for determining the final boundary values, P(A|A,24,24) and P(A|B,24,24), is explained below. It is however straight forward to check using the recurrence formulas and boundary conditions that

$P(A \mid A,23,23) = P(A \mid A,24,24)$, and
$P(A \mid B,23,23) = P(A \mid B,24,24)$.

Similar formulas can be developed for a 15-point set. In the case where A wins a 15-point set the backwards recurrence formulas are

$P^*(A \mid A,a,b) = p_A P^*(A \mid A,a,+1,b)+ q_A P^*(A \mid B,a,b+1)$
$P^*(A \mid B,a,b) = p_B P^*(A \mid B,a,b+1) + q_B P^*(A \mid A,a+1,b)$

The boundary values are:

$P^*(A \mid A,a,b) = 1$ if $a = 15, 0 \leq b \leq 13$
$P^*(A \mid B,a,b) = 0$ if $b = 15, 0 \leq a \leq 13$
$P^*(A \mid A,24,24) = p_A^2 / D$
$P^*(A \mid B,24,24) = p_A q_B (1+ p_A p_B - q_A q_B) / D$

The formulas to cover the cases where B wins the set are obvious.

Table 1 represents the probability of team A winning a 25-point and 15-point set for different values of p_A and p_B from the start of the set. The average probability of winning points on serve in men's volleyball is about 0.25. Therefore the values of p_A and p_B were chosen to reflect this value. The results indicate that the team receiving first has an advantage in winning the set. This is not an unexpected result, since the receiving team has the first opportunity at an attack.

Table 1: The probability of team A winning a 25-point and 15-point set for different values of p_A and p_B from the start of the set.

	25-point set		15-point set	
p_A, p_B	A serving	B serving	A serving	B serving
0.30, 0.30	0.48	0.52	0.47	0.53
0.30, 0.29	0.51	0.56	0.49	0.56
0.30, 0.25	0.63	0.69	0.59	0.66
0.25, 0.25	0.47	0.53	0.46	0.54
0.25, 0.24	0.50	0.57	0.48	0.57
0.20, 0.20	0.46	0.54	0.45	0.55

MATCH

The probabilities of a team winning a 25-point set from its beginning are represented as follows:

	A winning set	B winning set
A serving first point of set	$g_{AA} = P(A \mid A,0,0)$	$g_{AB} = P(B \mid A,0,0) = 1 - g_{AA}$
B serving first point of set	$g_{BA} = P(A \mid B,0,0)$	$g_{BB} = P(B \mid B,0,0) = 1 - g_{BA}$

The conditional probabilities of a team winning a match from set score (c,d) are represented as follows:

	A winning match	B winning match
A to serve first point of set	$G(A \mid A,c,d)$	$G(B \mid A,c,d) = 1 - G(A \mid A,c,d)$
B to serve first point of set	$G(A \mid B,c,d)$	$G(B \mid B,c,d) = 1 - G(A \mid B,c,d)$
Toss to serve first point of set	$G(A \mid *,c,d)$	$G(B \mid *,c,d) = 1 - G(A \mid *,c,d)$

The boundary values for team A winning a match from set score (c, d) are given by

$$G(A \mid A,c,d) = 1 \text{ if } c = 3, 0 \le d \le 2$$
$$G(A \mid B,c,d) = 0 \text{ if } d = 3, 0 \le c \le 2$$

A toss for serve is required at the start of the match, and the serve rotates at the start of each subsequent set unless the set score reaches 2-all, when another toss for serve is required. Thus when (c, d) = (0, 0) or (c, d) = (2, 2) the formula for the toss is

$$G(A \mid *,c,d) = 0.5 * [\, G(A \mid A,c,d) + G(A \mid B,c,d) \,]$$

The recurrence formulas after the first toss, and before the fifth set are

$$G(A \mid A,c,d) = g_{AA}\, G(A \mid B,c+1,d) + g_{AB}\, G(A \mid B,c,d+1)$$

$$G(A \mid B,c,d) = g_{BB}\, G(A \mid A,c,d+1) + g_{BA}\, G(A \mid A,c+1,d)$$

The recurrence formulas after the toss for the 15-point fifth set are

$$G(A \mid A,c,d) = g^*_{AA}\, G(A \mid B,c+1,d) + g^*_{AB}\, G(A \mid B,c,d+1)$$

$$G(A \mid B,c,d) = g^*_{BB}\, G(A \mid A,c,d+1) + g^*_{BA}\, G(A \mid A,c+1,d)$$

However when (c,d) = (2, 1) or (c,d) = (1, 2) one of the possible outcomes at the end of the set is a level score of 2 sets all, and recurrence formulas require modifications to allow for the toss.

$$G(A \mid A,2,1) = g_{AA}\, G(A \mid B,3,1) + g_{AB}\, G(A \mid *,2,2)$$

$$G(A \mid B,2,1) = g_{BB}\, G(A \mid *,2,2) + g_{BA}\, G(A \mid A,3,1)$$

$$G(A \mid A,1,2) = g_{AA}\, G(A \mid *,2,2) + g_{AB}\, G(B \mid B,1,3)$$

$$G(A \mid B,1,2) = g_{BB}\, G(A \mid A,1,3) + g_{BA}\, G(A \mid *,2,2)$$

When the boundary conditions are applied we obtain the simplification:

$$G(A \mid *,2,2) = 0.5 * [g^*_{AA} + g^*_{BA}]$$

The formulas for the probabilities of team B winning a match are obvious. Table 2 represents the probability of team A winning a match for different values of p_A and p_B from the start of the match. The results indicate that there is no advantage in serving or receiving at the start of the match. At the start of the fifth set, another coin toss is used to determine the serving team, and as given in Table 1, it is an advantage for the remainder of the match to be receiving first in this final set.

Table 2: The probability of team A winning a match for different values of p_A and p_B from the start of the match.

	match	
p_A, p_B	A serving	B serving
0.30, 0.30	0.50	0.50
0.30, 0.29	0.56	0.56
0.30, 0.25	0.77	0.77
0.25, 0.25	0.50	0.50
0.25, 0.24	0.56	0.56
0.20, 0.20	0.50	0.50

Finding the boundary conditions at the end of a set

The key to determining boundary values such as P(A | A,24,24) and P(A | B,24,24) , is to observe that, from the situation the scores are level when at least 46 points have been played, the team that wins the set must eventually win two successive points. If the two successive points are shared then the score is level once again, although a rotation of server may have taken place. Figure 2 illustrates the one step transitions between the various states of play after the scores are level.

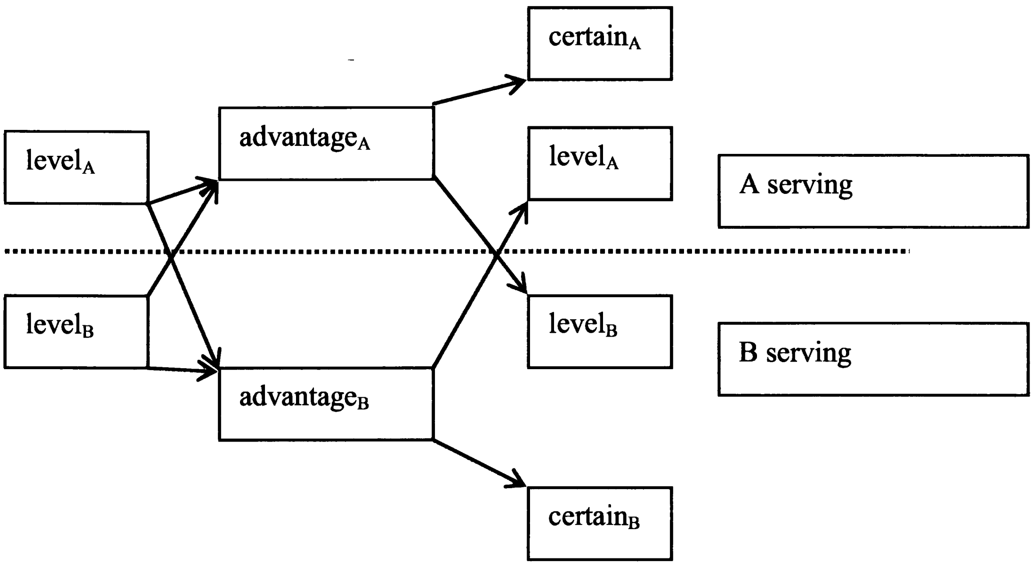


Figure 2: One step transitions between states of play, in a set after scores are level and when at least 46 points already played.

From the independence of the outcome of successive points we can calculate the two step transitions, and use this to eliminate the advantage states from the diagram, as shown in Figure 3.

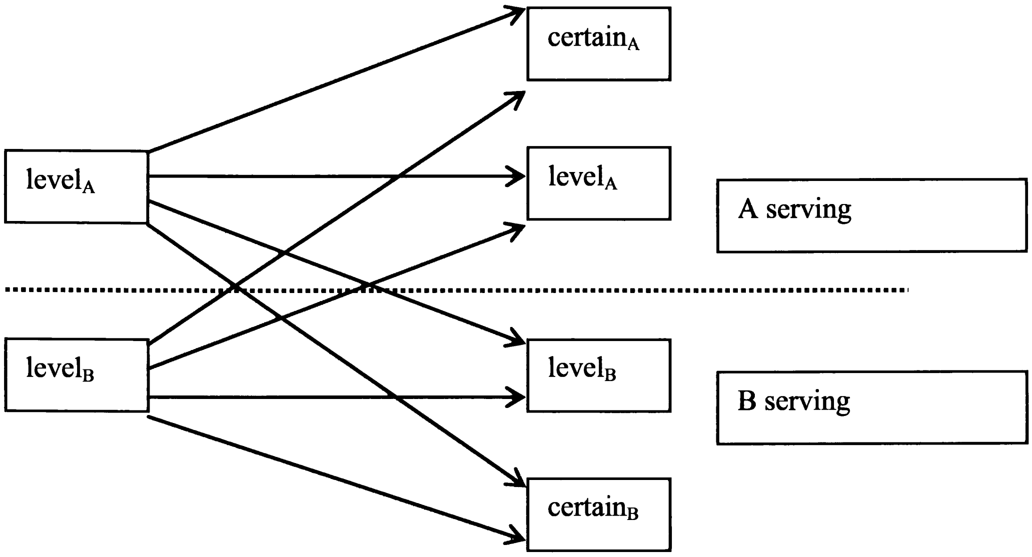


Figure 3: Two step transitions between states of play, in a set after scores are level and when at least 46 points already played.

We consider the case of A winning the set after reaching various states. To do this we must set the appropriate boundary conditions:

$c_A = P(A \mid A,25,23) = P(A \mid B,25,23) = 1,$

and

$c_B = P(A \mid A,23,25) = P(A \mid B,23,25) = 0.$

To simplify the notation let

$w_A = P(A \mid A,24,24)$, and $w_B = P(A \mid B,24,24)$.

Since

$w_A = P(A \mid A,23,23) = P(A \mid A,24,24) \quad \{= P(A \mid A,25,25) = \ldots\ldots\}$

and

$w_B = P(A \mid B,23,23) = P(A \mid B,24,24) \quad \{= P(A \mid B,25,25) = \ldots\ldots\}$

we can use the backwards recurrence relations already given to obtain, after two steps

$$w_A = p_A^2 c_A + q_A q_B w_A + p_A q_A w_B + q_A p_B c_B$$
$$w_B = q_B p_A c_A + p_B q_B w_A + q_A q_B w_B + p_B^2 c_B$$

When the boundary conditions are taken into account these equations simplify to

$$w_A (1 - q_A q_B) - p_A q_A w_B = p_A^2$$
$$w_B (1 - q_A q_B) - p_B q_B w_A = p_A q_B$$

Solving this pair of simultaneous equations leads to

$$w_A = p_A^2 / [(1 - q_A q_B)^2 - p_A q_A p_B q_B]$$
$$w_B = p_A q_B (1 + p_A p_B - q_A q_B) / [(1 - q_A q_B)^2 - p_A q_A p_B q_B]$$

The same argument can be used to develop the boundary conditions for the case where team A wins a 15-point set, leading to identical results. Results for the cases where team B wins a set can be obtained by symmetry, using similar arguments.

NUMBER OF POINTS PLAYED IN A SET

Mean number of points in a set

The mean number of points remaining in a 25-point set from point score (a, b) are represented as follows:

A serving next	$M(A\|a,b)$
B serving next	$M(B\|a,b)$

The backwards recurrence formulas are as follows:

$$M(A|a,b) = 1 + p_A M(A|a+1,b) + q_A M(B|a,b+1)$$
$$M(B|a,b) = 1 + p_B M(B|a,b+1) + q_B M(A|a+1,b)$$

The boundary values are:

$$M(A|a,b) = 0 \quad \text{if } a = 25,\ 0 \leq b \leq 23$$
$$M(B|a,b) = 0 \quad \text{if } b = 25,\ 0 \leq a \leq 23$$
$$M(A|24,24) = 2[1 + p_A q_A - q_A q_B]/D$$
$$M(B|24,24) = 2[1 + p_B q_B - q_A q_B]/D$$

where $D = (1 - q_A q_B)^2 - p_A p_B q_A q_B$

Variance of the number of points in a set

The variance of the number of points remaining in a 25-point set from point score (a, b) are represented as follows:

A serving next	$V(A\|a,b)$
B serving next	$V(B\|a,b)$

The backwards recurrence formulas are as follows:

$$V(A|a,b) = p_A V(A|a+1,b) + q_A V(B|a,b+1) + p_A q_A [(M(A|a+1,b) + M(B|a,b+1)]^2$$
$$V(B|a,b) = p_B V(B|a,b+1) + q_B V(A|a+1,b) + p_B q_B [M(B|a,b+1) + M(A|a+1,b)]^2$$

The boundary values are:

$$V(A|a,b) = 0 \quad \text{if } a = 25,\ 0 \leq b \leq 23$$
$$V(B|a,b) = 0 \quad \text{if } b = 25,\ 0 \leq a \leq 23$$
$$V(A|24,24) = 4q_A[p_A + q_B + 3p_A p_B q_B - 2q_A q_B^2 - p_A^2 q_A - p_A p_B q_A q_B^2 - p_A q_A^2 q_B^2 + p_A^2 p_B q_A q_B + q_A^2 q_B^3]/D^2$$
$$V(B|24,24) = 4q_B[p_B + q_A + 3p_A p_B q_A - 2q_A^2 q_B - p_B^2 q_B - p_A p_B q_A^2 q_B - p_B q_A^2 q_B^2 + p_A p_B^2 q_A q_B + q_A^3 q_B^2]/D^2$$

Similar recurrence formulas can be devised for the mean and variance of the number of points remaining in a 15-point set and the mean and variance of the number of sets remaining in a match. Table 3 gives numerical results of the number of points in a 25-point set for different values of p_A and p_B.

Table 3: The mean and standard deviation of the number of points in a 25-point set for different values of p_A and p_B.

	Mean points in 25-point set		Standard deviation of the number of points in a 25-point set	
p_A, p_B	A serving	B serving	A serving	B serving
0.30, 0.30	47.0	47.0	4.7	4.7
0.30, 0.29	47.1	47.1	4.8	4.8
0.30, 0.25	47.2	47.0	5.0	5.0
0.25, 0.25	47.8	47.8	5.4	5.4
0.25, 0.24	47.9	47.8	5.5	5.5
0.20, 0.20	48.9	48.9	6.7	6.7

CONCLUSIONS

This paper has demonstrated that the use of Markov chains can be used to model outcomes in volleyball conditional on both the scoreboard and the server. Results from the model indicate that it is advantageous to be the receiver on the opening point of a set and the team that wins the toss at the start of the fifth set (if the set score reaches 2-all), has an advantage for the remainder of the match. However, due to the rotation of serve after each set, there is no advantage for either side in being server or receiver at the start of the match. Similar models could also be applied to beach volleyball, where the rotation of serve in beach volleyball is the same as standard volleyball.

References

Clarke, S. R. and Norman, J. M. (1979) Comparison of North American and international squash scoring systems - analytical results. *Research Quarterly* **50(4)**: 723–728.

Kemeny, J. G. and Snell, J. L. (1960) *Finite Markov chains*. Princeton, New Jersey, USA.

ARTIFICIAL INTELLIGENCE MODELLING OF THE RELATIONSHIP BETWEEN TRAINING AND PERFORMANCE IN ATHLETES

Churchill, Tania, Sharma, D. and Balachandran, B.

Faculty of Information Sciences and Engineering, The University of Canberra, ACT, Australia

Paper Submitted for Review: 25 March 2008
Revision submitted and accepted: 25 June 2008

Abstract. How an athlete trains is the most influential factor in the myriad of variables which determine performance in endurance sports. The questions that face all coaches and athletes is how should I train to: a) produce peak performance; and b) produce a peak when desired? This paper reviews attempts to use mathematical models to quantify the relationship between training and performance. The theoretical underpinnings of these models are identified, touching on exercise physiology and load quantification techniques. The paper then identifies limitations to existing systems modelling approaches. The four key limitations discussed are: that the input for the model is training load (none of the training load quantification methods used in the literature accurately describe physiological load for the different types of sessions a cyclist commonly undertakes); simplification of the model input (training load) into a single variable limits the applicability of the model in the real world; that to estimate model parameters, regular performance tests are required (this is not practical for professional cyclists); and the model parameters need to be recalculated regularly, as an athlete's fitness changes. Possible techniques to overcome the identified limitations are then discussed, focussing on the domain of professional cyclists. The paper then discusses a number of different modelling techniques which may be applicable in this domain. Consideration is given to the use of Artificial Neural Networks (ANNs) which have yielded positive results in some studies, and to stochastic optimisation. The paper concludes with a suggested approach to overcoming existing limitations with systems modelling techniques.

Keywords: Modelling, training load quantification, artificial neural network.

INTRODUCTION

Performance in endurance sports is determined by a myriad of factors, from physiological and psychological parameters to technological and environmental factors. Extensive research effort has gone into attempting to understand the factors that influence performance, and the relationships between them. The consensus is that by understanding the factors that influence performance, these factors can then be manipulated to produce peak performances when desired.

From the 1970's onwards, numerous studies have focused on modelling physiological responses to training input using linear mathematical concepts (Banister, Calvert, Savage & Bach, 1975; Banister, Carter & Zarkadas, 1999; Busso, 2003; Busso, Carasso & Lacour, 1991).

There are currently a number of software tools available on the market (CyclingPeaks WKO+, RaceDay Performance Predictor™), which calculate an arbitrary measure of training load to model training and performance. These approaches follow on from the TRIMP (training impulse) method proposed by Bannister and Calvert (1980).

Linear modelling approaches have limitations in describing the relationship between training inputs and response, as the response to a given input will change over time. Some more recent researchers have looked at using non-linear Artificial Intelligence (AI) techniques - namely neural networks - to model training response relationships (Hohmann, Edelmann-Nusses & Hennerberg, 2001). The results from this early work is promising, with neural network models outperforming conventional linear models. However, these techniques have not been used to model training response relationships in cycling.

Performance modelling for cycling presents a number of unique challenges. These include: difficulty in objectively assessing performance, as performance in road racing involves numerous facets, including a strategic component (i.e. the rider with the highest average power, for example, is not necessarily the winner of the race (Martin et al., 2001); and difficulty in objectively quantifying training load as cyclists undergo many different types of training sessions, with varying intensities and durations. Interval sessions have varying work to rest ratios, and varying numbers of repetitions. These factors affect how fatiguing the training is, but do not indicate how training load, fatigue and fitness should be modelled).

QUANTIFICATION OF TRAINING LOAD

In order to model training input and performance output, training load must be quantified so it can be used as a system input. There has been great difficulty in finding a way to effectively quantify training load using a single term (Foster et al., 2001) .Taha and Thomas (2003) suggest that the parameters of intensity, frequency and duration need to be taken into account.

Busso and Thomas (2006) conclude that the specificity of the activity also needs to be considered, possibly by multifactorial models that do not simplify training loads into a single variable. This approach would, however, create a very complicated model.

Pattern of load is another consideration which impacts on physiological cost. Accelerating to a given speed requires greater power application and hence comes at a greater physiological cost than maintaining a given speed. Similarly, maintaining the same power at the end of an interval comes at greater physiological cost than maintaining that power at the beginning of an interval. No quantification models currently look at the pattern of loading and the consequent physiological cost of different load patterns.

TRIMP Research

Banister proposed a quantifying training load using a unit of training termed a TRIMP (TRaining IMPulse) (Banister, 1991). A TRIMP is derived from multiplying the duration of training by the relative training intensity (measured by heart rate). A multiplying factor is applied in order to weight higher heart rate (HR) ratios proportionally higher.

$$TRIMP = T(\text{min}) \times \frac{HRex - HRrest}{HRmax - HRrest} \times y$$

Taha and Thomas's paper (2003) describe calculating a TRIMP using currently accepted methods of measuring heart rate intensity. An absolute measure of percentage of maximum heart rate is used, while Banister used the Karvonen method of heart rate reserve.

TRIMPs have been successfully used to model the relationship between training and performance in many studies (e.g. Banister, Carter & Zarkadas, 1999). TRIMPs are a relatively easy measure to calculate, and require only heart rate monitors, which are readily available and moderately priced. TRIMPs also have the advantage of being able to be calculated across most sports and training activities, for example running and cycling.

Heart rate is a poor method of evaluating very high intensity exercise (Foster et al., 2001). Although TRIMPs can be calculated for a range of sports, they are not an appropriate load quantification method for strength training, as heart rate only measures the cardiovascular load of exercise.

Heart rate is affected by many factors, which impacts on its reliability as a measure of load. Factors which can affect heart rate include temperature, hydration, sleep, overtraining and caffeine (Jeukendrup, 2002). There are claims that the variability of heart rate impacts on the validity of quantification methods based on heart rate (Skiba, 2008).

Rating of Perceived Exertion

Foster et al. (2001) proposed a method of quantitating training using rating of perceived exertion (RPE). The RPE method involves getting athletes to rate their exertion for a session on Borg's RPE scale - a scale of

1 to 10 (Borg, 1985, cited in Foster et al., 2001). The 2001 study suggested that a TRIMP obtained from RPE (duration * RPE) was highly correlated with a modified TRIMP calculated from heart rate zones.

RPE has been used as the input of a systems model to model training and performance in an elite sprinter. The authors concluded that the resultant model was a potentially powerful tool for assessing the effects of training on athletic performance (Suzuki et al., 2006).

Advantages of RPE include: it can be used across many different exercise modalities, including strength training; it requires no equipment; and it is simple to calculate. The disadvantages of this approach include: it is a subjective measure; it relies on individual's memory of the session; differences in rating exist between individuals; and some intra-subject variation exists.

Using Power to Quantify Training Load

Power is the amount of energy generated per unit of time. In cycling, it is the amount of energy transferred to the pedals (in watts) per second. SRM power monitors are commonly used to measure power in cycling. The SRM system provides reliable measurements of power (Gardner et al., 2004), as well as measurements of cadence, speed and distance.

Power is a direct reflection of exercise intensity- whereas heart rate is a response to the exercise intensity (Jeukendrup, 2002). Heart rate is affected by external factors, and also experiences a delay in responding to changes in exercise factors. As a direct reflection of work output, power is not affected by these limitations.

Power has potential as the basis for quantifying training load. How can power data be analysed though in order to provide us with useful information? Taking the mean power of a session is a common summary statistic. Power output is highly variable however, so the mean does not necessarily accurately reflect the physiological demands of a session.

Training Stress Score (TSS) was proposed by Coggan in 2002 as a load quantification method using power. An intensity factor (IF) is used to normalise the TSS to functional threshold power (or the power that can be maintained in a 60 minute time trial). This allows for easier comparison off TSS between athletes.

TSS has a number of advantages: it is based on power, which is a direct measurement of training stimulus (work rate), rather than a response to the stimulus (such as heart rate); it is relatively simple to calculate (Coggan, 2007a); and by using normalised power as the basis for the calculation, TSS applies increased loading for high intensities, which is in keeping with physiological principles.

It can only be used, however, in sports where power can be easily measured (i.e. cycling). Training load from cross training or strength work cannot be measured. Power monitors are also still not in widespread use, due mainly to their relatively high cost.

The model has not been validated by any scientific studies (Coggan, 2007a). A similar algorithm (BikeScore), however, has received some initial support from a study which demonstrated the usefulness of the algorithm as an input function for systems-based performance modelling. Using the load quantification technique it was possible to accurately model performance (Skiba, 2007).

Summary of Load Quantification Methods

None of the currently accepted load quantification methods consider all of the factors identified as important in developing an accurate measure of physiological load.

The power output of a cyclist is a direct reflection of exercise intensity. Heart rate responds to the exercise intensity. The load quantification algorithms using power weight a certain power output the same whether it occurs at the start of a session when an athlete is fresh or at the end when they are fatigued. The occurrence of cardiac drift (the continuous increase in heart rate that usually occurs during prolonged moderate-intensity exercise (Jeukendrup, 2002) is cited as a criticism of using HR to measure training load (e.g. Skiba, 2008). Maintaining the same power at the end of an effort comes at greater physiological cost than maintaining that power at the beginning of an effort. It is proposed that HR more accurately models the physiological cost in this situation than power.

It is suggested that a combination of TSS, TRIMP and RPE would be powerful, giving a three dimensional picture of 1) the work the athlete is doing 2) how they are physiologically responding to the work they are doing and 3) how they feel while performing the work.

SYSTEMS MODELLING

Systems theory creates a mathematical model as an abstraction of a dynamic process. The system has at least one input, and one output, which are related by a mathematical representation called a transfer function (Busso & Thomas, 2006).

Existing Models – Impulse-Response Model

Banister (1991) proposed that the relationship between training and performance could be conceptualised by the idea that repeated training bouts contribute to two factors- fitness and fatigue. At any point in time, this relationship can be expressed as the formula Fitness – Fatigue = Performance. This is commonly referred to as the impulse-response model.

In order to model this relationship, Banister (1991) considered that a proportion of the training impulse (TRIMP) defined the fitness impulse (p(t)) and the fatigue impulse (f(t)). The TRIMP is therefore weighted by a multiplier (initially $K_1 = 1$ for fitness and $K_2 = 2$ for fatigue). Between training bouts, fitness and fatigue decline, modelled by an exponential decay equation. Fitness and fatigue decay at different rates- this is modelled by relative decay time constants. Banister (1991) defines the decay constant for fitness (r_1) initially as 45 days and fatigue (r_2) as 15 days.

Busso (2003) proposed a nonlinear model of the effects of training on performance, based on the assumption that the fatigue induced by a training session varies according to the preceding training load. The gain term for the fatigue impulse is mathematically related to training dose using a first-order filter. This refinement of Banister's model improved the fit of the model for previously untrained subjects trained on a cycle ergometer. It is important to note, however, that model parameters estimated in untrained subjects may not be representative of athletes in real situations (T. Busso, 2003; T. Busso & L. Thomas, 2006). In fact the fit of the same model in a study using highly trained participants in real training conditions was poorer (Thomas, 2008).

There are a number of limitations of the impulse-response model. There are claims it is overly simplified (Busso & Thomas, 2006); it requires frequent performance measures and the models assume that a greater amount of training leads to better performance. Previous studies have reported that the impact of training loads on performance has an upper threshold, beyond which training doesn't elicit further adaptations (Fry et al., 1992; Morton 1997 cited in Hellard et al., 2006). The model has not been linked to underlying physiological processes (Taha & Thomas, 2003); and the inability of the model to accurately predict future performance (Taha & Thomas, 2003; Busso & Thomas, 2006).

Performance Manager

Dr Andy Coggan (2007b) developed the Performance Manager concept, an attempt to overcome some of the limitations with the impulse-response model. His work is based on the recognition that performance is generally greatest when training is first progressively increased to a high level, to build fitness, and then tapered to eliminate residual fatigue.

This model eliminates the gain factors k_1 and k_2 (the gain factors are used to convert the impulse into separate fitness and fatigue impulses). This makes the model simpler by removing the difficulty of establishing appropriate values for the gain factors. It does, however, make the resultant fitness and fatigue components relative indicators of changes in performance ability, rather than absolute predictors. It also allows the substitution of simpler exponentially-weighted moving averages for the fitness and fatigue components (Coggan, 2007b).

The Performance Manager model has not been validated by any scientific studies. As with the impulse – response models, it does not consider the specificity of the training.

Summary of Systems Modelling

A workable model needs to be based on simplified abstractions of the underlying complex structures. The question is how much of the underlying structure should be incorporated into models of the relationship between training and performance (Taha & Thomas, 2003). A balance needs to be struck between complexity and ease of use to arrive at the greatest accuracy for the least computation cost.

Busso and Thomas (2006) concluded that in order for systems modelling techniques to be of practical use for coaches developing training programs, new modelling strategies should be considered in accordance with the specificity of the activity.

Existing models have four key limitations:

1. The input for the model is training load. None of the training load quantification methods used in the literature accurately describe physiological load for the different types of sessions a cyclist commonly undertakes.
2. Simplification of the model input (training load) into a single variable limits the applicability of the model in the real world
3. To estimate model parameters regular performance tests are required, however this is not practical for professional cyclists.
4. The model parameters need to be recalculated regularly, as an athlete's fitness changes.

Strategies to address these limitations are discussed in the following sections. It is proposed that by using artificial intelligence techniques in the modelling process, a more robust model, which will overcome some of the identified limitations in the existing models, can be developed.

POTENTIAL MODELLING APPROACHES

Artificial Neural Networks

Neural networks are flexible, adaptive learning systems, which can find patterns in observed data enabling the development of nonlinear systems models that can make reliable predictions (Samarasinghe, 2006). The non-linear modelling techniques of neural networks have been suggested as potentially appropriate for use in modelling the relationship between training and performance (Hellard et al., 2006).

Silva et al. (2007) used neural network technology to create models of swimming performance. They concluded that this approach was valid for resolving complex problems such as performance modelling.

Neural networks have been used to model the competitive performances of an elite swimmer, on the basis of training data (Hohmann et al., 2000) . The study concluded that neural networks are excellent tools to model and even predict competitive performances on the basis of training data. The precision of the neural network prediction was much higher than that achieved by a conventional regression analysis.

Hohmann et al. (2001) concluded that there were a number of advantages of using neural networks, including: the approach is more robust in handling issues with input data- such as noise or limited data; and the neural network can model nonlinear transformation of the relationship between training and performance, that is as the athlete's fitness changes over time. Banister (1991) found that the constants for the model will only serve for a period from 60 to 90 days before the process of iterative modelling and resetting the model constants needs to be repeated.

Stochastic Optimisation

Physiological data, such as heart rate, have been successfully modelled as stochastic time series. Stochasatic optimisation incorporates randomness in the optimisation process. It provides a good fit for models with incomplete parameterisation. A promising study used a stochastic optimisation method, the Alopex algorithm, to obtain optimal parameter values for a model of heart rate response to exercise (Zakynthinaki & Stirling, 2007). This approach may be appropriate for the conceptually similar problem of finding the optimal parameter values for a systems model of training and performance.

SUGGESTED APPROACH

The ability of systems models to provide an accurate enough fit to allow for prediction of performance in the real world has been questioned (Busso & Thomas, 2006; Thomas et al., 2008) . Research into modelling techniques such as ANN have shown encouraging results (Hohmann et al., 2001) and it is suggested that further studies are required to investigate the potential of this and other AI techniques,

It is proposed that the fit of three models, using the potential modelling approaches discussed in this paper, be compared with the fit of the traditional impulse-response model proposed by Busso (2003), the Control Model. The Control Model will use a single value load quantification input, such as TSS.

It is proposed that a combination of inputs be used in creating the ANN, looking at three dimensions of training, such as the actual work performed (using power, and quantified using the TSS algorithm); physiological response to the work performed (using HR and the TRIMP load quantification method) and how the athlete feels completing the work using RPE (Model 1).

The second model proposed is a hybrid model, using ANN to optimise the parameters for Busso's impulse-response model (Model 2).

The third model proposed will use stochastic optimisation methods to optimise the parameters for Busso's impulse-response model (Model 3).

Comparing the fit of these three models will provide some insight into whether the use of AI techniques has the potential to overcome the limitations of existing systems models and produce a model that has the potential of being a useful predictor of performance.

Reference List

Banister, E. W. and Calvert, T. W. (1980) Planning for future performance: Implications for long term training. *Canadian Journal of Applied Sport Sciences. Journal Canadian des Sciences Appliquées au Sport*, **5(3):** 170-6.

Banister, E. W. (1991) Modeling Elite Athletic Performance. In *Physiological Testing of the High Performance Athlete* (Second) (pp. 403-424). Champaign: Human Kinetics Books.

Banister, E. W., Calvert, T. W., Savage, M. V. and Bach, T. (1975) A systems model of training for athletic performance. *Australian Journal of Sports Medicine* **7:** 57-61.

Banister, E. W., Carter, J. B. and Zarkadas, P. C. (1999) Training theory and taper: validation in triathlon athletes. *European Journal of Applied Physiology and Occupational Physiology*, **79(2):** 182-191.

Busso, T. (2003). Variable dose-response relationship between exercise training and performance. *Medicine & Science in Sports & Exercise*, **35(7):** 1188–1195.

Busso, T., Carasso, C. and Lacour, J. R. (1991) Adequacy of a systems structure in the modeling of training effects on performance. *Journal of Applied Physiology*, **71(5):** 2044-2049.

Busso, T. and Thomas, L. (2006). Using mathematical modeling in training planning. *International Journal of Sports Physiology and Performance*, **1:** 400-405.

Coggan, A. (2007a) *Quantifying Training Load.* Retrieved June 17, 2007, from http://www.cyclecoach.com/andycoggantrainingload/.

Coggan, A. (2007b) *The Scientific Inspiration for the Performance Manager*. Retrieved May 30, 2007, from http://www.cyclingpeakssoftware.com/power411/performancemanagerscience.asp.

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S. (2001) A new approach to monitoring exercise training. *Journal of Strength & Conditioning Research*, **15(1):** 109-115.

Fry, R. W., Morton, A. R. and Keast, D. (1992) Periodisation of training stress - a review. *Canadian Journal of Sport Sciences*, **17(3):** 234-40.

Gardner, A. S., Stephens, S., Martin, D. T., Lawton, E., Lee, H. and Jenkins, D. (2004) Accuracy of SRM and power tap power monitoring systems for bicycling. *Medicine & Science in Sports & Exercise*, **36(7):** 1252-8.

Hellard, P., Avalos, M., Lacoste, L., Barale, F., Chatard, J. C. and Millet, G. P. (2006) Assessing the limitations of the Banister model in monitoring training. *Journal of Sports Sciences*, **24(5):** 509-520.

Hohmann, A., Edelmann-Nusses, J. and Hennerberg, B. (2001) Modeling and prognosis of competitive performance in elite swimming. *XIX International Symposium on Biomechanics in Sports, June 20-27, San Francisco, USA*, 54-57.

Hohmann, A., Edelmann-Nusser, Juergen and Hennegerg, B. (2000) A nonlinear approach to the analysis & modeling of training & adaptation in swimming. Retrieved June 26, 2007, from http://coachesinfo.com/article/index.php?id=152&style=printable.

Jeukendrup, A. (2002) *High-Performance Cycling*. Champaign: Human Kinetics.

Martin, D. T., McLean, B., Trewin, C., Lee, H., Victor, J. and Hahn, A. G. (2001) Physiological characteristics of nationally competitive female road cyclists and demands of competition. *Sports Medicine*, **31(7):** 469-477.

Samarasinghe, S. (2006) *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*. Florida: CRC Press.

Silva, A. J., Costa, A. M., Oliveira, P. M., Reis, V. M., Saavedra, J. and Perl, J. (2007) The use of neural network technology to model swimming performance. *Journal of Sports Science and Medicine*, **6:** 117-125.

Skiba, P. (2007) *Evaluation of a Novel Training Metric in Trained Cyclists*. Retrieved February 21, 2008, from http://www.acsm-msse.org/pt/re/msse/fulltext.00005768-200705001-02596.htm;jsessionid=H8nJGfF8nxCnrJGwNG3KmMzQvWqdfnJ2QvwSm7wRJLQSN2pvTTzQ!-1428189930!181195629!8091!-1?index=1&database=ppvovft&results=1&count=10&searchid=2&nav=search.

Skiba, P. (2008) *Analysis of Power Output and Training Stress in Cyclists*. Retrieved from http://www.physfarm.com/Analysis%20of%20Power%20Output%20and%20Training%20Stress%20in%20Cyclists-%20BikeScore.pdf.

Suzuki, S., Sato, T., Maeda, A. and Takahashi, Y. (2006) Program design based on a mathematical model using rating of perceived exertion for an elite Japanese sprinter: a case study. *Journal of Strength & Conditioning Research*, **20(1):** 36-42.

Taha, T. and Thomas, S. G. (2003) Systems modelling of the relationship between training and performance. *Sports Medicine*, **33(14):** 1061-1073.

The Performance Manager. (2006) Retrieved May 29, 2007, from http://www.cyclingpeakssoftware.com/pmc_summit.pdf.

Thomas, L., Mujika, I. and Busso, T. (2008) A model study of optimal training reduction during pre-event taper in elite swimmers. Journal of Sports Sciences, **26(6):** 643.

Zakynthinaki, M. and Stirling, J. (2007) Stochastic optimization for modeling physiological time series: Application to the heart rate response to exercise. Computer Physics Communications, **176(2):** 98-108.

AN INVESTIGATION INTO THE APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNIQUES TO THE PLAYER SELECTION PROCESS AT THE AFL NATIONAL DRAFT

McCullagh, John and Whitfort, Tim

Department of Computer Science and Computer Engineering, La Trobe University, Australia.

Paper Submitted for Review: 20 March 2008
Revision submitted and accepted: 10 June 2008

Abstract. The aim of this study is to investigate the potential of Artificial Intelligence techniques to assist recruiting managers in the Australian Football League (AFL) National Draft selection process. Currently players are selected based largely on the subjective judgement of recruiting managers, coaches and scouts. The major source of information for making decisions in the AFL National Draft will always be recruiting staff observing games. However there is a large amount of untapped data available which has the potential to improve the decision making process. This includes body composition, flexibility, anaerobic and aerobic power, visual tests, TAIS (Test of Attentional and Interpersonal Style) tests, psycho-motor tests, medical and psychological reports. Full player profiles are also provided on each player which include assessments on numerous skills and personal attributes. Extensive subjective evaluations on each player describing strengths, weaknesses, skills etc, have been written by previous coaches. This study uses Multi-layer Perceptrons (MLPs) to investigate the relationship between the player's data and their rating. A player rating system has been devised, and a rating determined for each player who has been through the AFL National Draft and has been in the AFL system for at least three years. Preliminary results from this study suggest that the extra data available to recruiting staff may have the potential to assist in improving the success of selecting players in the AFL National Draft.

Keywords: AFL National Draft, Artificial Intelligence, Multi-layer Perceptrons

INTRODUCTION

Numerous studies have reported a positive relationship between certain characteristics for junior elite players and future success in their respective sports. Pyne et al. (2005) investigated the relationships between anthropometric and fitness tests from the AFL Draft camp and the career progression of these players in AFL Football. The results demonstrated that the 20 m sprint, jump, agility and shuttle run tests are a factor with the career progression of AFL footballers. Tschopp et al. (2003) conducted a study to evaluate the predictive value of physiological, medical, psychological, anthropometric, social and personal characteristics for medium term success in junior elite soccer players. Height, isokinetic strength of the knee flexors and age at entry to the football club were statistically significant predictors. The study also concluded that the implementation of a multidisciplinary assessment of elite junior players would be most effective at age 15 due to greater homogeneity with increasing age. A study involving players competing in the Australian under 16 basketball championships demonstrated that the elite players could be distinguished from others on a number of anthropometric (height, sitting height, arm span) and physiological variables (speed, agility, vertical jump, basketball throw, aerobic endurance). This research also demonstrated that the relationship between the anthropometric and physiological variables and the player status was position dependent (Hoare, 2000).

The US National Football League Combine is the NFL's equivalent to the AFL National Draft. Players are tested on a variety of measures and the results are presented to all clubs to assist in the drafting process. McGee and Burkett (2003) conducted a study to assess the relationships between the player test results and draft status and demonstrated that the Combine results can be used to predict accurately the draft status of certain positions (running backs, wide receivers and defensive backs). Other positions were not accurately predicted. The research findings could be used to determine which position an athlete is most suited to and in which position an athlete is most likely to be successful. The study also found that the first and second round drafted athletes were taller, heavier and faster over 10, 20 and 40 yards as well as scored higher in agility runs, vertical and broad jumps. McDavid (1977) devised a test battery for American football players using both football skills and motor ability test items. The results demonstrated that the tests had discriminatory power and indicated that football potential may be predicted. The tests could be used as a screening device for football potential as well as a technique to identify relative strengths and weaknesses in players.

The inclusion of a skill assessment component to complement the other forms of player data is seen as important to gain an overall view of potential players in the AFL National Draft. This is supported by research in other sports which indicate the importance of skill assessment in the talent identification process. Elferink-Gemser et al. (2004) found that the most discriminating variables between elite and sub-elite junior hockey players were dribbling skills, tactical knowledge and motivation. Williams and Reilly (2000) suggest that identifying talent for young soccer players is best achieved on skill and ability rather than physical size. They state that the prediction of future elite players from anthropometric measurements alone may be unrealistic in younger age groups because performance could be affected by the players' rate of physical growth and maturation. NFL coaches and assistants conducted skill drills which were filmed at the NFL Combine. These drills were specific to the individual playing positions and clubs received a video of each player (Sheehan, 2007). Skill assessment has been achieved in the AFL by subjective analysis of players by scouts and coaches during match play. The use of tests to evaluate skills for talent identification in the AFL has not been favoured (Turnbull, 2007).

The aim of this study is to investigate the applicability of MLPs to predict the future playing ability of players in the AFL National Draft from all forms of data available including anthropometric, psychological and skill assessment.

ARTIFICIAL INTELLIGENCE

Artificial Neural Networks (ANNs) are very loosely modelled on the human brain. Multi-layer Perceptrons (MLPs) are a type of ANN that have been used because of their pattern matching ability in complex problem spaces. An MLP can be conceptualized as a black-box, non-linear model, where a collection of inputs (attributes) are presented to the MLP and one or more results (outputs) are produced. MLPs' use a supervised learning approach, learning from training examples, adjusting weights to reduce the error between the correct result and the result produced by the network. MLPs endeavour to determine a general relationship between the inputs and outputs provided (Smith, 1993). Once trained MLPs can be used to predict outputs based on input data alone. An example of a Multi-layer Perceptron is shown in Figure 1. Player attributes such as 3km and shuttle run performance are supplied to the MLP as inputs, and an output produced which is a prediction of the player's rating.

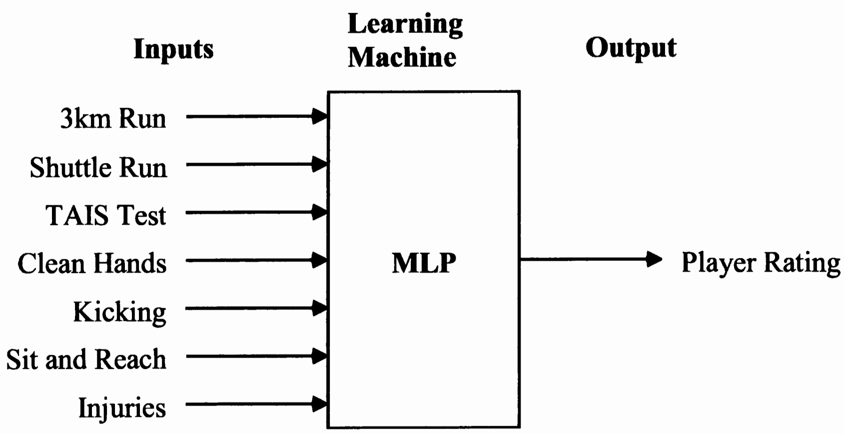


Figure 1: Multi-layer Perceptron as a black-box approach.

EXPERIMENTAL DESIGN

Player Data and Ratings

The player data set includes body composition, flexibility, anaerobic and aerobic power, visual tests, TAIS (Test of Attentional and Interpersonal style) tests, psycho-motor tests, skill assessments and subjective assessments on strengths, weaknesses and personal attributes. In total 58 input attributes and 310 player examples were used in this study. The 58 inputs were selected as they were the attributes tested across all years between 1999 to 2004. Skill assessments such as marking, kicking and clean hands were assessed numerically by coaches. Other attributes such as big game ability, coachability and aggressiveness were quantified by the use of a set of criteria to analyse the coaches written reports for each player.

Players in the draft from 1999 to 2004 were given a rating out of 10, where the rating indicates their value in the game today. The ratings used in this study are the average of two subjective assessments from people experienced in the AFL drafting process. Table 1 was used as a guide to assist in this assessment. Players were required to be in the AFL system for at least 3 years for an assessment to be made of their ability. As a result players from the 2005, 2006 and 2007 National Draft were excluded from the current study.

Table 1: Player ratings and descriptions

Rating	Description
≥8	Elite AFL player
7	Very good AFL player
6	Good AFL player
5	Plays a majority of games in the seniors (≥ 80%) and is regarded in the top 22 in the team
4	Just outside of the top 22 in the team. Plays < 80% of games in the seniors
3	Plays a majority of games in the reserves. Not thought of as a regular senior AFL player at this stage.
2	Unlikely to become a regular AFL player. Minimal or no AFL games
1	Drafted but no impact
0	Not drafted

The numbers of players in each ratings group are shown in Table 2 below.

Table 2: Frequency of players for each ratings group

Rating	Number of players
≥ 1	310
≥ 5	121
≥ 6	74
≥ 7	33
≥ 8	11

A second set of ratings were used for the classification experiments. The classes are outlined in Table 3.

Table 3: Classification classes used for experimentation

Class	Description	Rating
GOOD	Range from: plays the majority of games in the seniors to elite	≥ 5
AVERAGE	Range from: not a regular senior player to no impact at all	< 5

The correlation between the rating and player attributes was investigated. The top 5 correlations were: vision (-0.191), big game ability (-0.170), shuttle (-0.164), durability (-0.152) and clean hands (-0.147). Shuttle was the only attribute in the top five correlations that wasn't in the skill category.

The authors are currently developing a ratings system which will combine the subjective ratings of three experts together with club best and fairest votes and games played to produce a more robust player rating system. However it is acknowledged that there is no perfect ratings system.

MLP Parameters

A training set and a testing set were used in all experiments. The training set consisted of player data from the 1999 to 2002 National Drafts (214 examples), and the testing set consisted of data from the 2003 and 2004 National Drafts (96 examples). The MLPs were trained using the parameters specified in Table 4 below. The parameters were derived by conducting a series of trials involving varying parameters and assessing the effect on the MLP's output. Once trained the MLP's performance was assessed on the testing set to provide a measure of performance on unseen data.

Table 4: MLP parameters

Parameter	Value
Architecture	58 - 5 - 1
Learning rate	0.3
Momentum	0.2
Epochs	2000

RESULTS AND DISCUSSION

Two sets of experiments were conducted in this study. The first set of experiments, called the Regression Experiments, use a numeric rating between 1 and 10 as the MLP's output. The second set of experiments referred to as the Classification Experiments uses two rating classes, GOOD and AVERAGE as the output.

The results for the MLPs were compared with the performance of the recruiting managers (RMs) for the draft selection process. For this to be achieved the following assumption was made. The draft order for each year can be viewed as a measure of the accumulation of knowledge from recruiting managers across the 16 AFL clubs, even though each club works independently. The RMs' predictions used in this study are based on the draft order for each year.

Regression Experiments

The results in Table 5 show the performance of recruiting managers and the performance of the MLP on the testing data. The correlations are calculated as an average across the two years of the testing data (2003 & 2004).

Table 5: Correlations for RM versus rating and MLP versus rating for the testing set.

	Correlation
RM v Rating	-0.43
MLP v Rating	-0.23

The result presented for the recruiting managers is the correlation between the actual draft order for a particular year and the current rating for a player. It would be expected that a higher draft order (i.e.: 1, 2, 3 etc) would result in a higher rating and vice versa, resulting in a negative correlation. The MLP output a rating for each player in the testing set and this was converted to a draft order for both the 2003 and 2004 draft. The result presented for the MLP is the correlation between the MLP's predicted draft order and the current rating for each player. Both the RM and MLP correlations were negative as expected, however the recruiting managers clearly outperformed the MLPs' predictions.

Classification Experiments

The classification experiments used two player classes based on their rating. The classes are outlined in Table 3. To allow a comparison to be made between the MLPs' classification performance and the recruiting managers, the following reasoning has been used. An analysis of the draft data indicates that on average approximately 20 players each year achieved a rating of 5 or above. This information was used as the basis to develop an RM classification for each player in the draft. As previously stated, the draft order each year is taken as an estimation of the accumulation of knowledge from the recruiting managers across the 16 clubs. Therefore if each player selected in the top 20 is assigned a GOOD rating, and players selected below 20 an AVERAGE rating, this may be used as an estimation of the RM's classification.

Table 6 shows the performance of the recruiting managers and the MLP across the entire testing set, while Table 7 shows the respective percentage correct for GOOD predictions.

Table 6: RM and MLP predictions on the testing set.

Scenario	Percentage Correct
RM predictions	68.1%
MLP predictions	60.6%

Table 7: Performance of RMs and MLPs GOOD predictions on the testing set.

Scenario	Percentage Correct
RM predicts GOOD	45.0%
MLP predicts GOOD	37.2%

The results in both cases indicate that the recruiting managers clearly outperform the MLP. The RMs' predictions are 12.4% higher than the MLP on the entire testing set, and 20.1% higher for GOOD predictions. This would be expected due to the extensive networks that clubs have developed in the recruiting areas, and as a result, the vast amount of knowledge held on players in the draft that is not available to the MLP.

Further analysis of the GOOD predictions shows that while there was a certain amount of disagreement between the RMs' and the MLPs' classification, there were also numerous times when they agreed. Table 8 shows the percentage correct for GOOD predictions where the recruiting managers predicted correctly, the MLP predicted correctly and both predicted correctly.

Table 8: GOOD predictions on the testing set: 3 scenarios.

Scenario	Percentage Correct
RM predicts GOOD, MLP disagrees	33.0%
MLP predicts GOOD, RM disagrees	19.1%
RM and MLP predict GOOD	54.6%

The results show that the RM clearly outperforms the MLP on the percentage correct of GOOD predictions. In the cases where the RM and the MLP both agree on their GOOD prediction, the percentage correct was 54.6%. This compares favourably with the results in Table 7 where the RM predictions are 45% correct and the MLP 37.2% correct for GOOD predictions. While the low numbers used in these experiments make it difficult to draw conclusions, it is interesting to note that on their own MLPs do not approach the success achieved by the recruiting managers, however, they may have potential to be used as a secondary source of information to confirm or improve a decision.

CONCLUSION

The experiments conducted in this research are preliminary in nature and it is difficult to derive any definite conclusions from the results. The draft selection process is a very difficult task as evidenced by the success rate of the draft predictions. The recruiting managers largely base their decisions on a subjective analysis of players by their large network of recruiting staff and scouts, with the use of other forms of data of secondary importance.

The results from this research indicate that the recruiting managers clearly outperform the MLP in the regression and classification experiments. This would be expected as the recruiting managers commonly have information on players over a number of years including numerous subjective opinions from games on skills and fitness, performance in key games, family background, improvement over time etc, including all of the data available to the MLP. In comparison the MLP is limited to data on body composition, flexibility, anaerobic and aerobic power, visual tests, TAIS tests, psycho-motor tests, skill assessments and subjective assessments on strengths, weaknesses and personal attributes. While all of the tests used including the fitness, visual and psycho-motor tests provide an excellent snapshot of a player's profile, the skill assessment is a difficult component to accurately capture.

As each team is restricted to a few draft picks per year it is essential that RMs make the best choices. The results presented suggest that MLPs may have potential in supporting the decision making process for RMs. This is supported by the results showing that the recruiting managers alone were 45.0% correct with their GOOD predictions, whereas when the recruiting managers and the MLP both agreed on a GOOD prediction the percentage correct increased to 54.6%. Further research is being conducted in this area.

Acknowledgements

The authors would like to thank Stephen Wells, Kevin Sheehan and Col Hutchinson for their cooperation and assistance in this study.

References

Elferink-Gemser, M.T., Visscher, C., Lemmink, K. and Mulder, T. (2004) Relation between multidimensional performance characteristics and level of performance in talented youth field hockey players. *Journal of Sports Sciences*. **22:** 1053-1063.

Hoare, D.G. (2000) Predicting success in junior elite basketball players – the contribution of anthropometric and physiological attributes. *Journal of Science and Medicine in Sport.* **3(4):** 391-405.

McDavid R.F. (1977) Predicting potential in football players. *Research Quarterly*. **48(1):** 98-104.

McGee K.J. and Burkett L.N. (2003) The national football league combine: a reliable predictor of draft status. *Journal of Strength and Conditioning Research.* **17:** 6-11.

Pyne, D., Gardner, K., Sheehan, K. and Hopkins, W. (2005) Fitness testing and career progression in AFL football. *Journal of Science and Medicine in Sport.* **8(3):** 321-332.

Sheehan, K. (2007) Review of NFL combine. *Report conducted for the Australian Football League.*

Smith, M. (1993) *Neural networks for statistical modelling*. Van Nostrand Reinhold, New York, USA.

Tschopp M., Biedert R., Seiler R. (2003) Predicting success in Swiss junior elite soccer players: a multidisciplinary 4-year prospective study. *Journal of Sports Sciences*. **22:** 563 (Abstract).

Turnball, J. (2007) Analysis of the research and literature into the methods of successfully identifying and developing talent in sport from a global perspective. *Report to the Australian Football League*.

Williams, A.M. and Reilly, T. (2000) Talent identification and development in soccer. *Journal of Sports Sciences*. **18:** 657-667.

A WAVE-CATCHING MODEL FOR BODYSURFING

de Mestre, Neville

Bond University, Gold Coast, Australia

Paper Submitted for Review: 20 March 2008
Revision submitted and accepted: 7 June 2008

Abstract. Bodysurfing has been a popular summer pastime in Australia, New Zealand and other parts of the world for more than one hundred years. The most difficult aspect when learning the associated skills of bodysurfing is to launch oneself onto the wave just as it is about to break and become a moving surf front. The skill is easier to acquire when standing on a sandbank, but is more difficult when trying to launch oneself in deep water onto the wave just as it is about to break. A mathematical model of catching a breaking wave, by swimming onto it, is presented here. It will be validated by showing that when the bodysurfer does not swim for the wave it will pass through the bodysurfer's floating position and will not be caught. On the other hand if the bodysurfer is swimming for the wave and it reaches the bodysurfer just as it is about to break, the person may be accelerated up to the wave speed because of the forward drag induced by the water particle velocity in the crest of the wave. This paper does not consider the dynamics of surfing down the sloping front of a spilling wave (roller) nor the technique used to ride a plunging wave (dumper), but concentrates on modelling the act of catching either type of wave.

Keywords: bodysurfing, wave-catching

INTRODUCTION

Bodysurfing is the art of catching and riding a broken surf front without the aid of any craft. The most difficult part of the technique is to launch onto the wave at the correct moment, so that the bodysurfer can reach the speed of the wave and then be carried along in the turbulence of the surf front. It should be noted that once a wave has broken it is extremely difficult for even the strongest swimmer to catch the wave in deep water, but it is easy for anybody to catch the wave when standing on a shallow sandbank. When a shallow sandbank is present, bodysurfers can achieve a sufficient launching speed by pushing off the sandbank into a horizontal position using a strong thrust of their legs. Now the wave speed near ocean shorelines is regularly faster than the swimming speed of even the fastest human swimmer and so, in deeper water beyond the wave-breaking region, no swimmer can catch a wave. The relevant wave speed necessary to catch a wave seems only to be achieved by a swimmer just as the wave is about to break. The mechanism for this has not been completely understood, and no physical model has been put forward to explain why a human can actually catch and ride a wave unassisted by any floating device. It is the aim of this paper to provide a mathematical and physical model for catching a wave. A model for riding a wave will be addressed later. Consequently the mechanism for assisting bodysurfers to swim onto and catch a wave must lie in the forces available as the wave is about to break, and not before nor afterwards.

THE PHYSICAL MODEL

As ocean waves approach a sloping beach they grow in height, lose speed and eventually the front face of the wave becomes so steep that it breaks to become a turbulent surf front of tumbling water and entrapped air (Peregrine 1983; Battjes 1988). For traveling ocean waves it is known that the water particles within each wave move forwards at the crest and backwards in the trough. Research by Peregrine et al. (1980) has shown that the water particles at the crest of a breaking wave are travelling at a much faster speed W than the wave speed U. Computations show that the value of W could be as much as 2U.

To catch a wave in deep water, bodysurfers have to wait in the wave-breaking region and, with suitable timing, accelerate themselves from rest to try to match the wave speed U at the position where the wave starts to break. Therefore the mathematical condition that the bodysurfer will catch the wave will be that his or her forward speed at the breaking crest will equal U.

The propulsive forward swimming force generated by the bodysurfer in still water is denoted by P newtons. If his or her maximum swimming speed is S m/s then

$$P - kS^2 = 0$$

where k is the bodysurfer's resistance coefficient in still water when swimming. The air drag on the small part of the bodysurfer's anatomy above the water surface can be considered to be relatively insignificant compared with the water drag on and below the surface.

Most competent swimmers cover 50 m in times ranging from 25 s (very fast) to 60 s (slow and casual). Thus S varies from 0.8 to 2 m/s, and a typical value for the purpose of this paper will be taken as 1.5 m/s. For a swimmer of mass 75 kg, a typical value for k is 30 kg/m (Bellemans, 1981). This produces an average propulsive force P equal to 62.5 newtons. As the front of any wave passes through a surfer's position, the water particle velocity w on the front will change from a negative value at the foot of the wave, through zero part-way up the face, to a maximum W at the crest. At some stage the water particle velocity w will match the swimmer's forward velocity S, and there will be no forward or backward drag on the swimmer at all. Let this instant denote the initial time t = 0. The position of the wave crest at that instant is chosen as the origin for a 2-dimensional co-ordinate system with Ox in the direction of the wave's velocity and Oy vertically upwards (Figure 1). Let the position of the bodysurfer initially be (x_0, y_0) with a typical value of x_0 being 2m ahead of the crest. The time of interest for this investigation of the dynamics of catching a wave is therefore $0 \leq t \leq t_1$, where $t_1 = x_0 / (U\text{-}S)$ is the time for the crest to reach the bodysurfer with U assumed constant for this short interval of time.

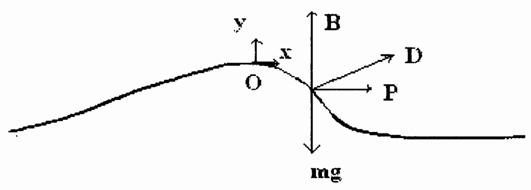


Figure 1

The forces on the swimmer at any time t between starting the swim and when the crest overtakes him or her are

(i) weight (mg) vertically down,
(ii) buoyancy (B) vertically upward,
(iii) propulsive force (P) towards the shoreline,

(iv) drag (D) assisting the swimmer because the water is moving faster than the swimmer. This drag is at an angle to the horizontal but near the crest the main thrust of the water particles will be towards the shoreline.

Resolving these forces in the horizontal and vertical directions for Newton's laws of motion yield

$$m\, d^2x/dt^2 = P + D\cos\psi \qquad (1)$$

$$m\, d^2y/dt^2 = B + D\sin\psi - mg \qquad (2)$$

where ψ denotes the angle that the drag makes with the horizontal.

For the time interval under consideration, it can be assumed the swimmer is near the crest and that the horizontal component of the swimmer's velocity is much greater than the vertical component. Thus time derivatives with respect to y are much smaller than time derivatives with respect to x, and the angle ψ will be very small. Hence equation (2) can be approximated by

$$0 \cong B - mg$$

showing that the surfer's weight is approximately balanced by the buoyancy force. Using the same reasoning, an approximate version of equation (1) is

$$m\, d^2x/dt^2 = P + k\,(w - dx/dt)^2 \qquad (3)$$

showing that the bodysurfer is thrust forward by the water particle drag in the region of the wave crest. Equation (3) can be used to predict whether or not the bodysurfer catches the wave.

ANALYSIS

For the period of time from when the swimmer's speed matches the water particle speed to when he or she is overtaken by the crest, the value of w ranges from dx_0/dt to a maximum W (>U). This is modelled linearly as

$$w = dx_0/dt + (W - dx_0/dt)\, t/t_1$$

where t_1 is the time when the crest reaches the swimmer. Thus equation (3) is now

$$m\, d^2x/dt^2 = P + k\,[(W - dx_0/dt)\, t/t_1 + dx_0/dt - dx/dt]^2$$

With $z = (W - dx_0/dt)\, t/t_1 + dx_0/dt - dx/dt$, the equation of interest becomes

$$m\,[(W - dx_0/dt)/t_1 - dz/dt] = P + k z^2$$

or

$$(m/k)\, dz/dt = A^2 - z^2$$

where $A^2 = (mW - m\, dx_0/dt - P t_1)/(k t_1)$. Integration of the reciprocal of this equation yields

$$k t/m = 1/(2A) \ln\left|\frac{A+z}{A-z}\right| + C$$

When t=0, z=0 and so C=0. Thus

$$z = A \tanh(A k t / m)$$

However the expression of interest is the speed of the bodysurfer at the crest, namely dx_c/dt, which occurs when $t = t_1$. Thus

$$dx_c/dt = W - A\tanh(A k t_1 / m) \qquad (4)$$

CALCULATIONS

To see if the bodysurfer can catch the wave, it remains to determine from equation (4) whether dx_c/dt reaches the value of U for typical values of the parameters used.

Consider first of all the case P = 0 when the bodysurfer does not swim for the wave at all. Then $A = \sqrt{(m W/(k\, t_1))}$ and $dx_0/dt = 0$. Typical values of the parameters are m = 75, k = 30, $t_1 = 1.33$ and U = 3.0 (6 knots approximately). Note that U is faster than any human swimmer can achieve unassisted.

Calculations based on equation (4) in the present model are shown in Table 1. They show that even when the water particle speed (W) at the crest is double the wave speed (U), the swimmer cannot possibly catch the wave if he or she remains at rest (P=0) as the wave approaches. Practical bodysurfing verifies this.

However for P = 62.5, when the bodysurfer attempts to swim for the wave, it is seen also from Table 1 that he or she will catch it in this case if W = 4.4 or more.

Table 1: (U = 3.0, $t_1 = 1.33$)

W	dx_c/dt (P = 0)	A	dx_c/dt (P = 62.5)
3.2	0.91	1.05	2.67
3.4	1.01	1.22	2.71
3.6	1.11	1.36	2.75
3.8	1.22	1.49	2.81
4.0	1.34	1.61	2.88
4.2	1.46	1.73	2.95
4.4	1.58	1.83	3.02
4.6	1.71	1.93	3.11
4.8	1.84	2.03	3.19
5.0	1.97	2.12	3.28
5.2	2.10	2.20	3.38
5.4	2.24	2.29	3.48
5.6	2.38	2.37	3.58
5.8	2.52	2.45	3.69
6.0	2.66	2.52	3.80

The calculations were repeated with the swimmer reaching water particle speed when only one metre ahead of the position of the wave crest, instead of two metres. For this situation the swimmer has probably started too late and the calculations show that the bodysurfer cannot catch this wave for any W<2U as t_1 is now halved in value and dx_c/dt only reaches U for the extreme case W = 6.0.

Finally the assistance offered by swimfins should be referred to. These can increase the swimmer's forward propulsive force P by up to 10%. This decreases the value of the parameter A, and allows the swimmer to catch the wave even when it is not about to break and W is much less than 2U. In some cases a bodysurfer with swimfins can even swim onto a broken surf front.

CONCLUSION

The model shows that the main wave characteristic in determining whether or not a bodysurfer can catch a wave is the water particle velocity near the wave crest. Once a wave has been caught, then a different model is needed which will involve a body sliding down a moving inclined plane of water. This paper does not address that situation.

The model predicts that a swimmer can only catch a wave when he or she propels themselves forward just before the wave reaches them. The timing of the bodysurfer's forward propulsion is critical, and is achieved by lots of practice.

References

Battjes, J. A. (1988) Surf-zone dynamics. *Annual Review of Fluid Mechanics* **20:** 257-293.

Bellemans, A. (1981) Drag force exerted by water on the human body. *American Journal of Physics* **49:** 367-368.

Peregrine, D. H. (1983) Breaking waves on beaches. *Annual Review of Fluid Mechanics* **15:** 149-178.

Peregrine, D. H., Cokelet, E. D. and McIver, P. (1980) The fluid mechanics of waves approaching breaking. *Proceedings of the Conference on Coastal Engineering 17th*, 512-528.

THE ACCURACY OF INTRA-STROKE VELOCITY USING A GPS-BASED ACCELEROMETER IN KAYAKING

Janssen, Ina [1], Sachlikidis, Alexi [1] and Hunter, Adam [1,2]

[1]Department of Biomechanics and Performance Analysis, Australian Institute of Sport, Canberra, ACT, Australia

[2] University of Canberra, Canberra, ACT, Australia

Paper Submitted for Review: 28 March 2008
Accepted Without Revision

Abstract. Advances in micro-technology have resulted in the development of devices such as the minimaxX unit (Catapult Innovations, Victoria, Australia), which combine triaxial accelerometers with a Global Positioning System (GPS). These units are designed to provide athletes, coaches and scientists with detailed information on athletic performance and workload, potentially providing an objective basis for technique enhancement, exercise prescription and injury management. The purpose of this study was to assess the accuracy of the algorithm which is used to determine the instantaneous velocity (100 Hz) of a sprint kayak based on the GPS and accelerometers within the minimaxX. Four minimaxX units of varied accelerometer sensitivities (2g or 6g) and GPS frequencies (1Hz or 5Hz) were positioned on a kayak paddled by a skilled male participant. The paddler performed 15 trials ranging in stroke rates between 57-107 strokes/min. The velocity of the minimaxX was compared with the velocity derived from high speed footage (100 Hz) (Phantom, Vision Research). The pattern of the velocity waveform across minimaxX and video is highly related. The minimaxX with a setting of 2g and 1Hz had the greatest error and consistently under-reported the kayak velocity by on average 0.32 m/s (R^2 = 0.9139). The minimaxX setting of 6g and 1Hz had the smallest average error under-reporting 0.14 m/s (R^2 = 0.9752). For a performance measure this is a substantial discrepancy in magnitude, potentially resulting in a race difference of 18.5 seconds over 1000 m (4.00 m/s vs 4.32 m/s). The results of this study indicate the velocity waveforms were confidently portrayed; however, the magnitude of the intra-stroke velocity was consistently being under-reported for the all of the minimaxX units trialled. The findings of this study have led to refinements of the algorithms used to calculate velocity specifically for kayaking with these upgrades subject to future validation.

Keywords: GPS-based accelerometer, velocity, canoe

INTRODUCTION

Advances in micro-technology have resulted in the development of devices, such as the minimaxX unit, with Global Positioning System (GPS) functionalities which aim to provide athletes, coaches and scientists with detailed information on athletic performance and workload. GPS is a proven effective positioning technology that is capable of determining positions at centimetre level precision when used in kinematic differential mode (Zhang et al., 2004). A GPS receiver calculates position by measuring the distance to three or more satellites. Nondifferential GPS units provide a functional tool for researchers due to their low cost and lightweight design compared with differential GPS (Townshend et al., 2008) and has been shown to provide an accurate estimation of speed during human locomotion and constant cycling speeds along straight paths (Witte & Wilson, 2004; Townshend et al., 2008). In sprint kayaking, the oscillations in boat movement and velocity challenge the GPS to provide accurate measures as they do no travel at consistent speeds.

Accurate measurement of intra-stroke velocity in a sporting environment such as sprint kayaking could assist coaches and athletes by providing an objective basis for technique enhancement and may also assist in developing optimal race profiles and establishing combinations for multiple person boats (Pelham et al., 1993). Previous researchers have relied on high speed video and impellers to measure velocity (Pelham et al., 1993), but this has been an indirect, impractical and time-consuming method. To ensure ecological validity of providing feedback to athletes and coaches kayak performance, on-water accelerometer could be used (MacFarlane et al., 1997; Peach et al., 1996). Various on-water triaxial accelerometers have been used and validated for this purpose for sprint kayaking (Peach et al., 1996; Robinson et al., 2002; Pelham et al., 1993).

One such accelerometer with GPS functionalities is the minimaxX unit (Catapult Innovations, Victoria, Australia). The minimaxX is a unit which contains a triaxial accelerometer (KXM52-1050) and three gyroscopes (ADXRS300) which sample at 100Hz, three magnetometers which sample at 40Hz (HM55B) and a non-differential GPS unit (Fastrax) which samples at 1Hz or 5Hz. The minimaxX are small, light (65-70g), unobtrusive and currently there are approximately 500 minimaxX units in elite sports use around Australia (Catapult Innovations, 2007) being used in a range of sports including cricket, hockey, tennis, rowing and kayaking (Reid et al., 2008). However there is a paucity of information regarding the accuracy of the intra-stroke velocity produced by the minimaxX units in sprint kayaking. Before these units could be used effectively they need to be assessed to determine the accuracy of the algorithms which combine the data from the accelerometers and the GPS.

The purpose of this study was to assess the accuracy of the algorithms which are used to determine the instantaneous intra-stroke velocity of a kayak based on the GPS and accelerometers in the minimaxX. Additionally, this study aimed to establish if the different GPS settings that are possible for the minimaxX produced comparable velocity measures. The quantification of the variation in the accelerometer would allow for more accurate interpretation of velocity data and determine if the minimaxX is an appropriate on-water measurement tool of intra-stroke velocity for sprint kayaking.

METHODS

Approval for the study was obtained from the Australian Institute of Sport Ethics Committee and written informed consent was obtained from the participant prior to the study. One participant was considered acceptable as the physical and physiological characteristics of the subject have no effect on the accuracy of GPS measurements (Larsson & Henriksson-Larsen, 2001). The participant (a skilled male paddler - age 30 years, body mass 74 kg, height 1.75 m) conducted his own warm up prior to commencing the data collection trials. Data collection involved the participant performing 15 trials ranging in stroke rate between 57 and 107 strokes per minute. Each trial involved the participant accelerating the boat (a flatwater racing kayak (K1) 'Blade' (Van Dusen copy) manufactured by 'Competition kayaks') to a constant velocity along a marked still water course before entering the capture area which was 15m in length.

The boat was setup with four minimaxX units (Catapult Innovations, Victoria, Australia) with accelerometers of varied sensitivities (2g or 6g) and GPS sampling frequencies (1Hz or 5Hz) (Table 1) and a video-minimaxX synchronisation unit attached 30cm behind the cockpit (2.5m from the nose) (Figure 1). The trigger for the synchronisation unit was attached so the participant could activate it during each trial with his knee. Reference points on the boat were measured for calibration and analysis of the video footage, including: boat length (5.190m) and distance from bow to minimaxX position (3.390m).

High speed video cameras (Phantom, Vision Research) with a sampling rate of 100 Hz were used to collect each trial to coincide with the sampling rate of the minimaxX. The camera was positioned on the bank of a lake with the lens axis perpendicular to the kayak line of travel at a distance of approximately 30m.

In order to synchronise the video and minimaxX, a light-emitting diode (LED) was encased with a magnetic coil and positioned on the kayak in between the minimaxX units with an activation trigger positioned inside the cockpit which was activated by the paddler in the kayak (Figure 1). When the trigger was activated, a magnetic pulse was sent to the minimaxX and recorded by its magnetometers while the LED lit up and was visible by the high speed video cameras. Trials where the LED did not activate in view were eliminated from analysis resulting in a total of eleven trials suitable for analysis.

The total boat length was measured to enable a scale factor and calibration to be calculated in the subsequent analysis (Sanders & Kendal, 1992). The known distance of the kayak allowed for the boat area to be calibrated over the water and throughout the entire trial, eliminating and accounting for potential issues if the kayak was not travelling exactly perpendicular to the camera position. The following reference points in the lateral view were digitised in the Phantom software (Vision Research): bow, front of the cockpit, position of the minimaxX and LED units and the stern. To minimise error associated with the digitising process, the position data were filtered through a low-pass 4th order Butterworth digital filter with a cut-off frequency of 2Hz. A Fast Fournier transformation was used to select the cut-off (MATLAB version 7.2; The MathWorks, Inc.). Instantaneous velocities of the kayak were derived from the filtered bow position data. The first frame where the LED was visible was identified for synchronisation with the minimaxX. Pilot testing demonstrated this method of data collection and synchronisation was a valid and reliable method to measure velocity.

Table 1: Individual settings for the minimaxX units used in the testing. Sensitivity of the unit indicates the range of the accelerometer and the GPS sampling reports the number of times the velocity is adjusted.

minimaxX unit	Sensitivity	GPS Sampling
A	6g	1Hz
B	6g	5Hz
C	2g	1Hz
D	2g	5Hz

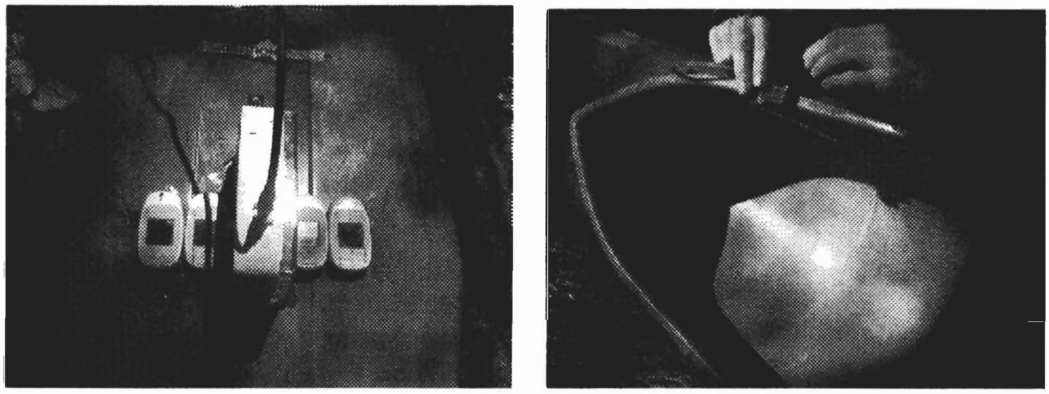


Figure 1: The position of four triaxial accelerometer minimaxX units mounted on a kayak with an encased LED in the middle. Two minimaxX units are on either side of the LED; the activation trigger is positioned inside the cockpit of a kayak and is attached to the encased LED placed behind the cockpit.

Data from the minimaxX units were downloaded, GPS and accelerometer data combined and exported as a CSV file, using Logan software (Version V21.9). The sample at which the magnetometer increased by 100 units was identified as the synchronisation point with the video. Acceleration signals in the x-direction from the minimaxX were integrated between each GPS reading to determine instantaneous velocity. The minimaxX data was filtered with a 4th order Butterworth digital filter and a cut-off of 3Hz to reduce artefacts of the GPS integrated with velocity.

A digitising reliability analysis was performed for one trial with the average and maximum residual error calculated. A complete double stroke cycle was defined from the catch on the right side of the kayak (in view of the camera) until the next catch on right side. A total of 13 double strokes were identified due to the elevated stroke rate of the last two trials enabling two full double strokes within the camera view. Video-derived velocities were used as the reference measure. To graphically display these relationships, Bland-Altman (Bland & Altman, 1986) scatter plots with bias and limits of agreement were calculated. The coefficient of determination (R^2) was calculated for each minimaxX to determine to proportion of variability in the data set that was accounted for by a statistical model. Paired *t*-tests were used to compare the accuracy of the video-derived velocity and the GPS-derived velocity with significance set at $p<0.05$. Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS Version 15.0, SPSS inc, Chicago, IL). To compare the closeness in amplitude of the video-derived velocity and the GPS-derived velocity, the root of the mean of the squared differences (RMSe) was also calculated (Mayagoitia et al., 2002).

RESULTS

The results of the reliability analysis for the average and maximum residual error range were 0.0071m and 0.011m, respectively. MinimaxX unit D malfunctioned and thus no data was available from this unit. At the first right-side paddle entry, the greatest difference of 0.27 m/s was observed between the video and minimaxX unit C with the minimaxX under-reporting the kayak velocity (Figure 2). At the second right-side paddle entry (time 4.2 sec), the difference was observed for minimaxX C as it under-reported velocity by 0.43 m/s.

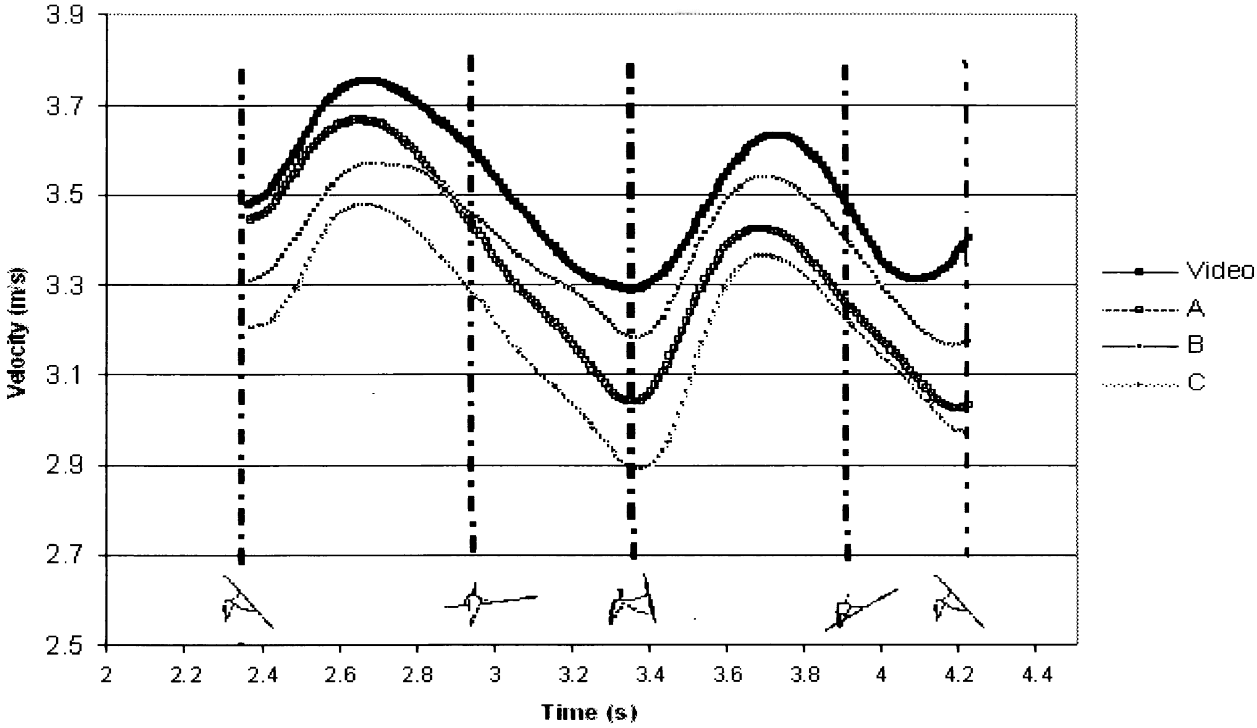


Figure 2: A typical comparison of kayak instantaneous velocity for one double stroke from right entry to right entry shown for all functioning minimaxX units (A, B, C) and the video for a typical trial at 64 strokes/minute.

In order to determine the total variation in the velocity reported between the video and minimaxX units, the velocities of all 13 trials were combined for a total of 2117 data points. The RMSe for the amplitude of all the trials for the functioning minimaxX units are reported in Table 2. MinimaxX units A and C under-reported kayak velocity by on average 0.32 m/s, which was the highest average error out of the minimaxX units tested. MinimaxX unit B had the smallest average error of 0.14 m/s. The R^2 of the GPS-derived velocity produced by each minimaxX unit with the video-derived velocity demonstrated that minimaxX B had the smallest variability whilst minimaxX A had the highest (Table 2).

Table 2 – Coefficient of determination (R^2) between the minimaxX and video-derived velocity measurements. The root mean square error (RMSe) indicates the velocity differences of the minimaxX from the video.

minimaxX unit	R^2	RMSe
A	0.9043	0.32 m/s
B	0.9752	0.14 m/s
C	0.9139	0.32 m/s

The following Bland-Altman scatter plots (Figure 3) represent the plot of error for all 13 double strokes of the specific minimaxX and the video-derived velocity. MinimaxX A under-reported velocity by 0.32m/s for one trial, when the RMSe was recalculated without the flawed trial the velocity still under-reported by 0.31m/s.

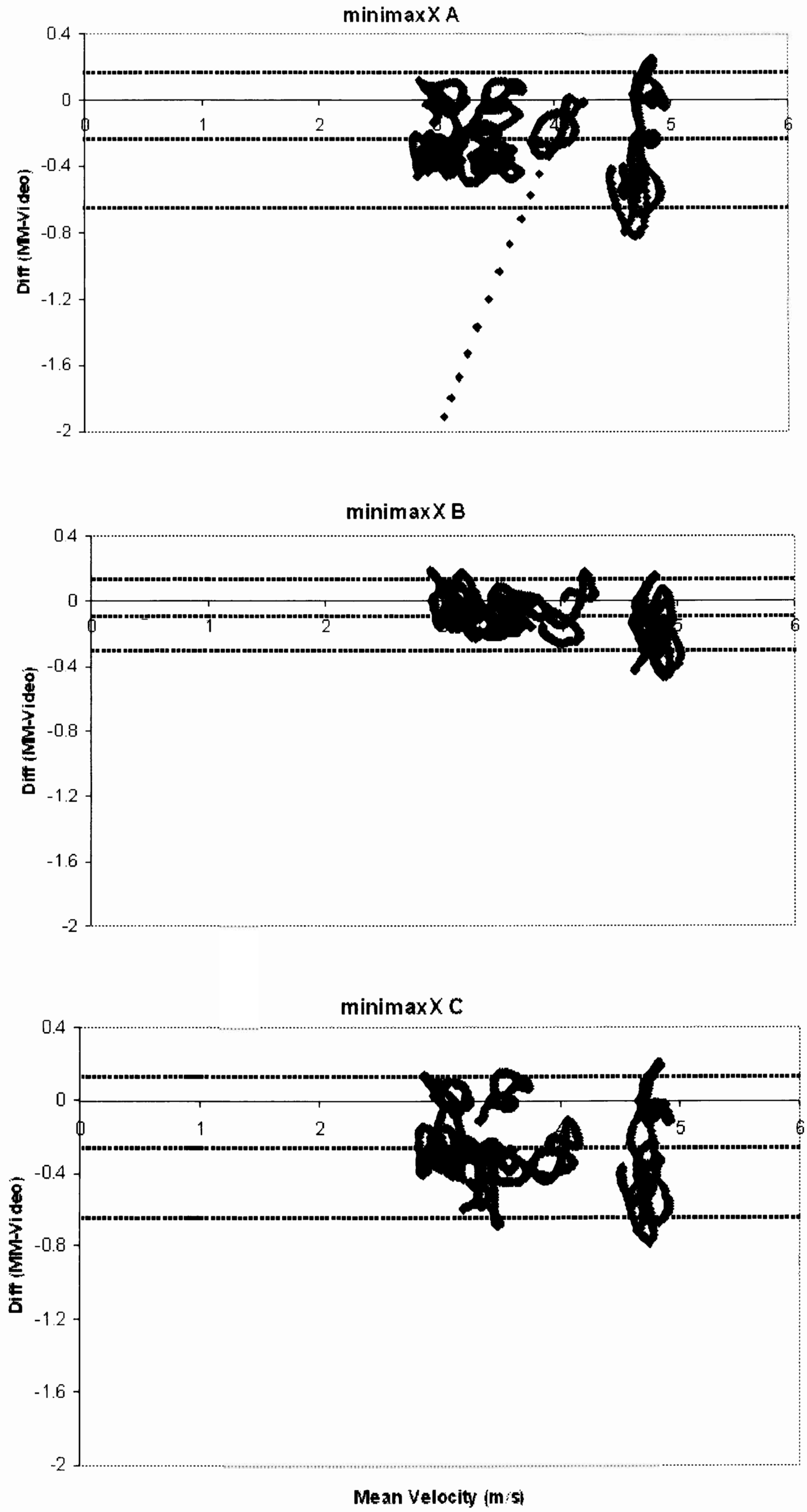


Figure 3: Bland-Altman plot of the difference in overall velocity between GPS-derived minimaxX and video-derived measurements of unit A, unit B, and unit C. The bias and 95% limits of agreement (dashed lines) are also shown.

Analysis of the differences between each of the minimaxX units revealed that minimaxX units A and C reported similar velocities (Table 3). The greatest variation occurred between minimaxX units A and B. The average velocities the minimaxX produced were statically significant when compared with the video-derived velocity for all settings trialled (Figure 4).

Table 3 – The relationship between minimaxX units with different sensitivity and GPS sampling settings.

minimaxX unit	R^2
A vs. B	0.9378
B vs. C	0.9559
A vs. C	0.9674

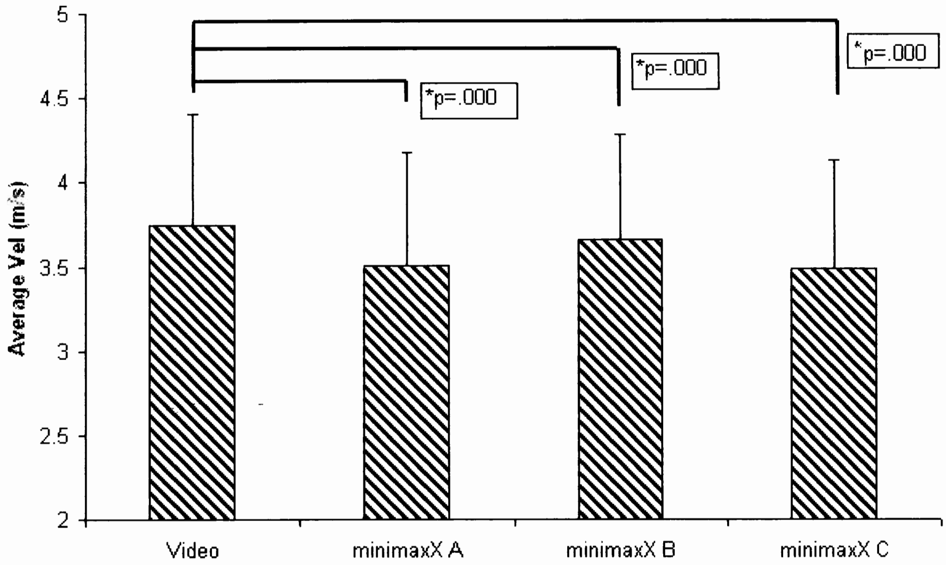


Figure 4 – Average velocity (m/s) and standard deviation of all trials for video-derived and minimaxX-derived measurements. *significantly different compared with video-derived measurements at $p<0.05$.

DISCUSSION

The purpose of this study was to assess the accuracy of intra-stroke velocity using the minimaxX and to our knowledge this is the first investigation to assess the accuracy for sprint kayaking. The intra-stroke velocity waveform that was produced from the minimaxX was highly related to that of the video-derived velocity. This suggests that the minimaxX intra-stroke velocity pattern can be used as a coaching tool to portray ascending and descending kayak velocities and compare the symmetry of the right and left stroke. Additionally, each of the different minimaxX settings produced similar patterns and therefore can be used interchangeably to observe velocity waveforms.

The magnitude of the intra-stroke velocity was under-reported by all of the minimaxX units when compared with the video. These results demonstrated that currently the use of the minimaxX is not an accurate representation of the magnitude of intra-stroke velocity for sprint kayaking. MinimaxX A and C with settings of 1Hz had the greatest error and consistently under-reported the kayak velocity by on average 0.32 m/s (R^2 = 0.9139). For a performance measure in sprint kayaking, this is a substantial discrepancy in magnitude, potentially resulting in a race difference of 18.5 seconds over 1000 m (4.00 m/s vs. 4.32 m/s). Conversely, previous researchers found nondifferential GPS to be an accurate estimation of velocity during locomotion velocities from 1.06 to 9.62 m/s, with mean errors of 0.01 m/s (Townshend et al., 2008). Multiple factors may have led to increased variability during the current testing which may include error in the calculation of velocity, measurement instrument (accelerometer, GPS) error, the orientation of the minimaxX on the kayak, digitizing error or velocity synchronization errors. The most variability was present in minimaxX A with 9.66% of the variability accounted for by other variables.

The second aim of this study was to determine if minimaxX units with different settings were interchangeable. This data indicated that minimaxX units with GPS sampling of 1Hz (A and C) produced similar errors when compared to video. Additionally, these two units had the lowest variability between minimaxX velocity measures. As a result, this data suggested that units with settings of 1Hz are interchangeably though further reliability testing should be conducted in order to confirm this. Conversely, this data indicated that minimaxX units of different GPS settings (1Hz vs 5Hz) are not inter-changeable and comparable in regards to magnitude of velocity. Higher GPS sampling indicates that the minimaxX is adjusting its velocity based on GPS position more often. A minimaxX with a sensitivity of 2g indicates that the range of the accelerometer is two times gravity while a 6g accelerometer has a range of six times gravity. Therefore, a 2g accelerometer is more sensitive to kayak movements compared to a 6g unit and thus could potentially provide more detailed information of the kayak. Based on these results, the minimaxX in its present form on average consistently under-reported the intra-stroke velocity for kayaking. Based on the variability and error analysis, out of the three possible minimaxX settings used in this study the minimaxX unit B with a setting of 6g at 5Hz provided the most accurate intra-stroke velocity measurements compared with high speed video. Unfortunately, minimaxX unit D with settings of 2g and 5Hz malfunctioned and therefore no conclusions regarding those settings can be made.

The outcomes of the current study supports similar results that GPS slightly under estimated true speed at cycling velocities of 2.1 to 10.8 m/s (Witte and Wilson, 2004). Similar to our results, Edgecomb and Norton (2006) found a significant difference between the distance measured by a different GPS tracking system and a trundle wheel pedometer in Australia identifying systematic errors associated with the GPS tracking system which is theorized to be related to the resolution capability of the satellite tracking technology. Conversely, Macleod and Sunderland (2007) found no systematic bias present in the GPS tracking system manufactured by the same company as those used by Edgecomb and Norton (2006), however it was conducted in England. GPS information is dependent on satellite position; therefore the location of the testing is critical when comparing previous studies as more accurate GPS data is available in Europe due to the receiver positions.

In minimaxX unit A, the velocity for one trial was substantially lower than the video. The GPS signal appeared to have dropped out briefly and the velocity was unable to retain its true velocity. Interestingly, this was the only minimaxX to have dropped out of the three that were simultaneously recording. The cause for the GPS drop out is unidentified and it is unknown why only one of the minimaxX units experienced this problem. A loss of GPS signal on average of 1.6 times per subject during a 30 minute field test has been found in past studies (Larsson and Henriksson-Larsen, 2001). The accuracy of GPS is influenced by the number of satellites available to the receiver and therefore satellite lock is important to consider.

The findings of this study have led to refinements of the algorithms in Logan that are used to calculate velocity specifically for sprint kayaking with these upgrades subject to future validation. Further studies are essential to determine the variability within a given accelerometer and between units with the same settings. A static calibration trial is also necessary to determine the validity of positional measurements.

CONCLUSION

The results of this study indicated that velocity waveforms from the minimaxX were confidently portrayed; however, the magnitude of the intra-stroke velocity was consistently being under-reported by all of the minimaxX units trialled. Based on these results, in its current form the minimaxX units are functional as a coaching tool to measure stroke pattern and symmetry; however, is not an accurate measure of intra-stroke velocity for biomechanical purposes. There is potential for the minimaxX to be an appropriate measurement instrument however further adjustments to reduce present errors need to be addressed.

References

Bland, M. and Altman, D. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. **1**: 307-310.

Catapult Innovations (2007) *Minimax Newsletter*, December. Victoria, Australia

Edgecomb, S. and Norton, K. (2006) Comparison of global positioning and computer-based tracking systems for measuring player movement distance during Australian Football. *Journal of Science and Medicine in Sport*. **9:** 25-32.

Larsson P. and Henriksson-Larsen K. (2001) The use of dGPS and simultaneous metabolic measurements during orienteering. *Medicine and Science in Sports and Exercise*. **33:** 1919-1924.

MacFarlane, D., Edmond, I. and Walmsley, A. (1997) Instrumentation of an ergometer to monitor the reliability of rowing performance. *Journal of Sports Science*. **15:** 167-173.

Macleod, H. and Sunderland, C. (2007) Reliability and validity of a global positioning system for measuring movement patterns during field hockey. *Medicine and Science in Sports and Exercise*. **S39(5):** 209-210.

Mayagoitia, R., Nene, A. and Veltink, P. (2002) Accelerometer and rate gyroscope measurement of kinematics: an inexpensive alternative to optical motion analysis systems. *Journal of Biomechanics*. **35:** 537-542.

Peach, J., Pelham, T., Carter, A., Holt, L. and Burke, D. (1996) A method of motion analysis for self propelled aquatic craft. In T. Bauer (ed.) XIII International Symposium for Biomechanics in Sport: Proceedings. ISBS, Thunder Bay, Ontario, 75-78.

Pelham, T., Holtz, L., Burke, D. and Carter, A. (1993) Accelerometry for paddling and rowing. In J.Hamill, T. Derrick and E. Elliott (eds.) Biomechanics in Sport XI : proceedings of the XIth Symposium of the International Society of Biomechanics in Sports. ISBS, Amherst, Massachusetts, 270-273.

Reid, M., Duffield, R., Dawson, B., Baker, J. and Crespo, M. (2008) *British Journal of Sports Medicine*. **42(2):** 146-151.

Robinson, M., Holt, L. and Pelham, T. (2002) The technology of sprint racing canoe and kayak hull and paddle designs. *International Sports Journal*. **6(2):** 68-85.

Sanders, R. and Kendal, S. (1992) A description of Olympic flatwater kayak stroke technique. *Australian Journal of Science and Medicine in Sports*. **24(1):** 25-30.

Townshend, A., Worringham, C. and Stewart, I. (2008) Assessment of speed and position during human locomotion using nondifferential GPS. *Medicine and Science in Sports and Exercise*. **40(1)**: 124-132.

Witte T. and Wilson A. (2004) Accuracy of non-differential GPS for the determination of speed over ground. *Journal of Biomechanics*. **37**: 1891-1898.

Zhang, K., Deakin, R., Grenfell, R., Li, Y., Zhang, J., Cameron, W., and Silcock, D. (2004) GNSS for sports – sailing and rowing perspectives. *Journal of Global Positioning Systems*. **3(1):** 280-289.

THE INFLUENCE OF CREW WEIGHT ON SAILING PERFORMANCE IN TAIPAN CATAMARANS

Tonkes, Elliot

Energy Edge Pty Ltd, Australia

Paper Submitted for Review: 30 March 2008
Revision submitted and accepted: 28 May 2008

Abstract. This article investigates the impact of crew mass on the racing performance of Taipan catamarans. A regression on data from recent national titles derives the sensitivity of finishing times and establishes the existence of a moderate relationship. A higher powered test is applied using the matched nature of the data to confirm that lighter crews perform better in lighter winds, while the advantage diminishes for stronger conditions. A physical model is introduced calculating perturbation in boatspeed resulting from marginal increase in crew mass. The model takes into account additional sail forces which can be extracted from more righting moments, at the cost of higher hull drag. The results are compared with an existing yardstick formula (Texel, 2008) which provides time penalties for lower weight.

Keywords: sailing, yardstick

INTRODUCTION

In the sport of competitive sailing, the objective is to sail around a defined course faster than one's competitors. Successful racing involves strategy and tactics along with good yachtsmanship. Of course, the ability to win races is improved if boatspeed can be increased, and a clear, but nontrivial, relationship exists between a boat's weight and the speeds achievable. In the contemporary racing environment, an overriding concern is the reduction of weight on the boat. In the Taipan catamaran class, this concern has also led to a preference for lightweight crews. The Taipan class is a modern one-design high performance two-person 4.9 metre catamaran, sporting a wing mast and high aspect-ratio sloop rig. The minimum mass of the boat is 102 kilograms, and no boats deviate substantially from this. The objective of this paper is to determine the extent to which relative performance deviations can be explained by crew weight. Bethwaite (1995) has undertaken similar studies for the Tasar class of dinghy to suggest limitations on Tasar crew weights.

Yachts use a combination of aerodynamic lift on sails and hydrodynamic resistance on the hull, keel and rudder to generate propulsion (Townend, 1984). When a yacht sails roughly with the direction of the wind, the forces generated by the sails are reasonably aligned with the direction of travel. While a yacht cannot sail directly into the wind, it is capable of sailing at approximately 45 degrees towards the wind. Under such circumstances, the sails generate a substantial lateral force, which is resisted by the keel in the water, while a relatively small forward component of force drives the boat forward. The large lateral sail forces are generated at altitude and act to tip the boat sideways. Large yachts possess a heavy keel for righting moments but the ballast for small dinghies and catamarans is supplied by the crew themselves by either leaning out or by suspending their bodyweight outboard on a trapeze wire. In stronger winds, larger crew weight can be beneficial to offer more righting moments to extract additional force from the sails which translates to additional driving force and faster speeds. In lighter winds, any additional ballast is not needed and a higher weight leads to greater hull drag and performance deterioration. The theory underlying hull drag in smooth waters is well established (for example Tuck et al., 1996) but in choppy water is not so well understood (Harris et al., 1999).

Intuition dictates that light crews should perform better in lighter wind conditions, while heavier crews should achieve superior results in stronger winds. In a real-world racing environment, the effects are more complex because increased crew weight is beneficial on the upwind legs of the course, but additional righting moments are not needed on the downwind legs. Bethwaite (1995) drew conclusions that for a *planing* sailboat, lighter crews will display superior performance in all windstrengths, but most apparently in 8 knots or more. The results were related to the threshold speed where a hull transitions from a displacement mode (speed is limited by the wave drag) to a planning mode (hydrodynamic forces lift the hull).

The primary approach of this study is to analyse the observed finishing times for a fleet of Taipan catamarans over a sample of racing heats and to relate to crew weights. A further analysis is performed on the physics of sailing by taking linear approximations to calculate the effect of perturbing crew weight. Results are compared with the Texel Rating yardstick formula. Finishing times and crew weights have been taken from three Australian national championships. A uniform course was sailed for each heat, being close to an upwind/downwind course. All heats were sailed in flat, protected waters without significant currents and most races enjoyed relatively constant windstrength in accordance with the Taipan racing rules. The relative performance of each competitor in each heat is determined by comparing finishing times against the fleet's performance. Analysis of the data enables us to establish if crew weight provides a significant explanation for performance, if crew weight explains the difference between strong and light wind performance, if the statistics concur with predictions based on physics and if the Texel Rating yardstick is consistent with observations.

DATA

Data Sources

The statistical results in this paper are based on data from the 2001, 2002 and 2003 Taipan Catamaran Australian Championships held at various locations (Table 1). All events provided heats that were conducted in a range of windstrengths, and in general, the windstrength remained constant throughout the heat apart from several heats under mixed conditions. Finishing times and windstrengths were provided by the hosting race committees. Crew weights were collected by the author. The minimum crew mass over all heats was 112 kg and the maximum 175 kg.

					Number of heats by windstrength:				
Year	**Number of sloops**	**Yacht mass mean/stdev**	**Crew mass mean/stdev**	**Location**	**Light (≤10 kn)**	**Medium (10-14 kn)**	**Strong (≥14 kn)**	**Mixed Conditions**	**Total**
2001	35	104 (2) kg	137 (14) kg	Waterloo Bay (Qld)	4	1	5	0	10
2002	47	104 (2) kg	133 (13) kg	Pt Stephens (NSW)	4	2	5	0	11
2003	35	N/A	135 (9) kg	Gippsland Lakes (Vic)	4	2	2	2	10

Table 1: Locations of national titles, yacht and crew weights, numbers of competitors and number of heats

Data Filtering

Heat 4 (2003) was eliminated from all analysis as the windstrength built considerably midway through the race. Heat 7 (2003) was eliminated from all analysis as the wind was shifty and many competitors protested abandonment was suitable. The distribution of finishing times for each heat displays consistent left skewness, which is explained by capsized or crippled yachts continuing to finish well behind the fleet. To attenuate this effect, a recursive filter was instituted (Clewlow, 2000) to remove anomalous outcomes outside of three standard deviations from the mean.

Taipan regattas are open to all competitors and a filter was applied to remove several novice sailors from the analysis. Yachts with finishing times consistently worse than two standard deviations behind the mean were eliminated from the data set (three yachts from the 2002 regatta and one from 2003). To perform *relative* statistical performance analysis, yachts that could not provide finishing time data for at least one strong-wind and one light-wind heat were eliminated from the analysis (an additional yacht from the 2001 regatta and two from the 2003 regatta).

MODELS

Definition of Statistical Performance

We define a measure of performance P (relative to the fleet) for a given boat in a given heat. Let

$$P = \frac{\overline{T} - T}{\overline{T}}$$

where T is the yacht's finishing time and $\overline{T}$ is the fleet's mean finishing time. Negative P implies below average performance while positive P means above average performance. Based on the objective of yacht racing, there is merit in a measure of performance based upon finishing rankings or normalised performance, but the present definition facilitates a connection between a yacht's performance and its mean boat speed.

Absolute Statistical Performance Model

The figure below illustrates for a typical heat and over all heats how performance corresponded with crew mass. For Heat 1 2001, the gradient of the regression line yields the *sensitivity of performance to crew mass* is -0.0025 per kilogram (in this instance, an additional 0.172 minutes for every additional kilogram over a 68 minute race). The gradient over all heats was -0.0014.

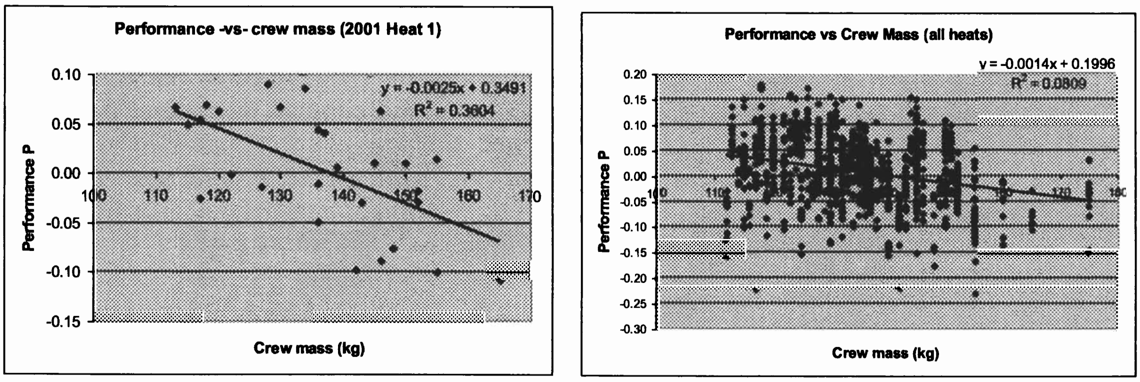


Figure 1: Scatterplot of performance against crew weight (a) Heat 1 2001 and (b) all heats. Linear regressions in solid.

Figure 2 plots the sensitivity of performance to crew mass for each heat against the windstrength for all championship heats. The result demonstrates that the sensitivity is generally negative over all wind strengths, that is, an increase in crew weight will correspond to degradation in performance. Visual inspection of Figure 2 suggests that in light wind, a smaller crew mass yields better performance (more negative sensitivity) while in moderate and heavy wind the sensitivity appears less marked. Using ANOVA, the hypothesis that the sensitivity is negative for light winds ($\leq$ 10 knots) is supported at the 99% level of confidence. The sensitivity is not significantly negative for strong winds ($\geq$ 14 knots) with p-value 0.25. The hypothesis that the sensitivity of performance to crew mass is stronger at light winds compared with strong winds (that is, seeking a positive regression coefficient in Figure 2 by a one-sided t-test) is supported at the 90% level of confidence, but not at 95%.

Relative Statistical Performance Model

Anecdotally, it appears that the most serious competitors appreciate the influence of crew mass and compete with lightweight crews. The fact that consistent negative sensitivity coefficients are found might be fully explained by the possibility that the better sailors intentionally sail with light crews rather than any causal relationship between performance and crew mass. A superior statistical analysis utilises the extra information that the same boat competed in heats with various windstrengths. The matched nature of the data permits a higher-powered test and can be used to generate *relative* performance comparisons, rather than the *absolute* results above. All races have been categorised as light, medium or strong windstrengths. Although these distinctions are somewhat arbitrary, a Taipan catamaran with average crew weight is fully powered in around 12 knots of wind. Beyond 12 knots, the crew will shed power when sailing to windward. Analogous to the Portsmouth Yardstick system (Portsmouth, 2008) we have taken winds up to, and including 10 knots to be classified as *light* conditions, winds 10-14 knots as *medium* and windstrengths beyond 14 knots as *strong*.

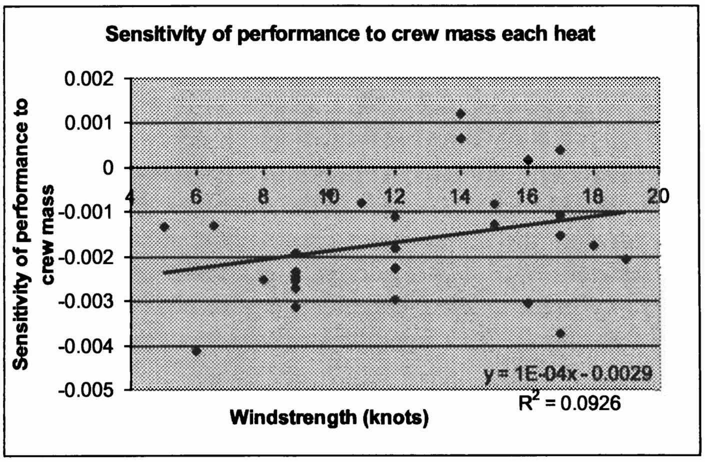


Figure 2: Sensitivity of performance to crew weight against windstrength over all heats and linear regression.

For each competitor, we determined the performance index P for each heat. For each competitor over all heats in each regatta, the means of the performance indices over each of the light wind, medium wind and strong wind heats were calculated and denoted P_l, P_m and P_s respectively.

We construct the quantity $P_s - P_l$ for each yacht to represent the difference in performance between strong and light windstrengths. A positive value of $P_s - P_l$ indicates that a yacht performs relatively better in strong wind than in light wind. To test the hypothesis that strong winds benefit heavy crews, while light winds benefit light crews, we seek statistical significance that $P_s - P_l$ is increasing in crew mass.

Figure 3 plots $P_s - P_l$ against crew mass for the combined data set (all heats), where each data point represents a single yacht that competed in both light and strong winds. The figure demonstrates that there is a moderate relationship in the expected direction: lighter crews have a better relative performance in light winds.

The R-squared value gives an indication that around 10% of the discrepancy in performance between light and strong conditions can be explained solely by the crew mass. The coefficient to the regression has a one sided p-value of 5×10^{-4} indicating that at the 99% level of confidence a significant relationship exists between the relative performance of a yacht across windstrengths based on the mass of the crew.

Based on this data, the crew mass which will suffer least variability in performance between light and strong windstrengths is given by the intercept in Figure 3 at 140 kg. However the shallow gradient of the linear best fit yields a large standard deviation and a 90% confidence interval for the intercept is an unhelpful range of 88 to 335 kg.

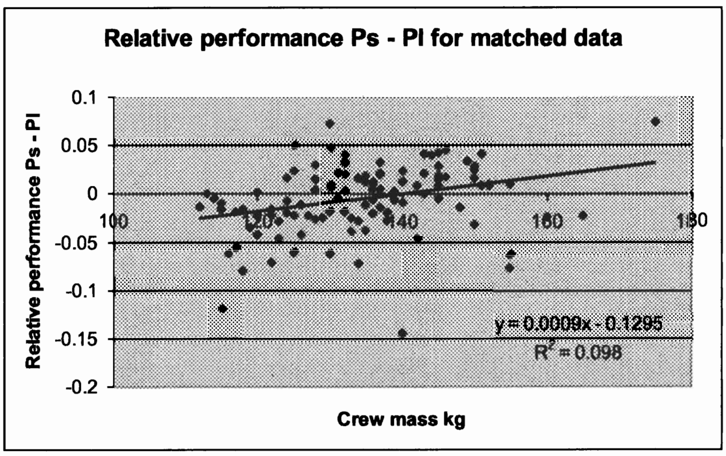


Figure 3: Relative performance $P_s - P_l$ based on crew weight over all heats and regression.

Physical Model

A hull moving through water is subject to drag forces from skin friction, form drag and wave drag (Tuck et al., 1996). The hydrodynamic resistance of catamaran hulls is reviewed by Sahoo et al. (2007).

Taipan hulls are slender and operate in the displacement or forced modes (Bethwaite, 1996) with a relatively small wake. Similar to the models of Philpott et al. (1993), we model hull drag D by a quadratic dependency upon boatspeed and linearly on wetted surface area according to $D = ½ \rho (C_f + C_w) A v^2$. Here C_f is a skin friction coefficient and C_w is a wave drag coefficient which depends upon hull geometry and Froude number, but for our purposes remains approximately constant in a neighbourhood of given hull speed. While this relationship is very suitable for skin friction which is the dominant drag component at low speeds, the model becomes less accurate at high speeds as wave drag begins to dominate. A Taipan hull closely resembles a rowing shell and Brearley et al. (1998) derive a quadratic relationship in a neighbourhood of rower's racing speed. For a semi-cylindrical form with either one or both hulls in the water, Dudhia (2008) justifies that the wetted surface area closely fits a square-root dependency on total mass (boat and crew).

A small reduction in mass leads to less displacement and wetted surface area and therefore reduced hull drag and supports a higher hull speed. In turn, an increase in boatspeed leads to a perturbation of the relative wind vector towards the direction of the yacht's bow and an *increase* in the relative windspeed when sailing upwind and a *decrease* when sailing downwind. The perturbation to sail forces becomes relatively complex at this stage, as a subtle playoff occurs between the changed relative wind vector and the resultant force vector from sail pressures, however the change in force $F' - F$ is small in comparison to the reduction in hull drag. The concept is illustrated in the figure below. The yacht's speed is v, the true wind vector is u and the relative wind in u_{rel} while the sail produces a lift L which is resolved into force F in the direction of travel (which must be precisely counteracted by the hull drag). Upon a reduction in hull drag, the yacht experiences faster speed v', the relative wind changes to u_{rel}' and so does the sail force L' and its resolved component F'.

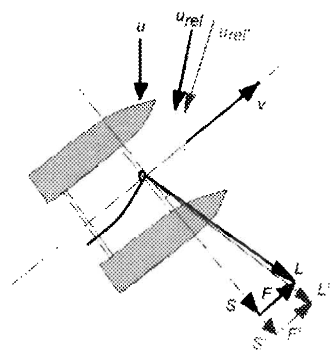


Figure 4: Changes to relative wind and sail force for upwind sailing based on increase in boatspeed.

Light weather and downwind sailing

In light weather sailing (all angles) and in downwind sailing (all wind conditions) additional weight on the boat is purely a burden which creates additional hull drag.

The drag force on the boat must be balanced by the forward driving force from the sails and so $D = F = c\, m^{1/2}\, v^2$: for some constant c. Since driving force F remains approximately constant, applying the calculus to a small perturbation dm in mass yields

$$dv = \frac{\partial v}{\partial m} dm \Rightarrow \frac{dv}{v} = -\frac{1}{4}\frac{dm}{m} \Rightarrow \frac{dT}{T} = \frac{1}{4}\frac{dm}{m}.$$

In the context of the performance P, we have $dP = -1/4\; dm/\overline{m}$. Based on $\overline{m}$ = 102 + 135 kg this yields dP = -0.00105 dm. In other words, performance P will *deteriorate* by an additional 0.00105 for every additional kilogram on the yacht.

Upwind sailing in strong wind conditions

In strong wind conditions, Taipans sail with only one hull in the water for the entire course, but with the windward hull only just out of the water, resulting in the mast remaining near vertical. Let L be the total force from the sail, resolved into lateral force S and force F in the direction of travel. Force S acts at altitude h to tip the boat sideways and is countered by a moment offered by the boat's weight B at a moment arm b and the crew's weight W at a moment arm r. By increasing crew weight by dW, the total force S that can be sustained is increased by dS. The driving force increases by the same proportion, $dF / F = dS / S$.

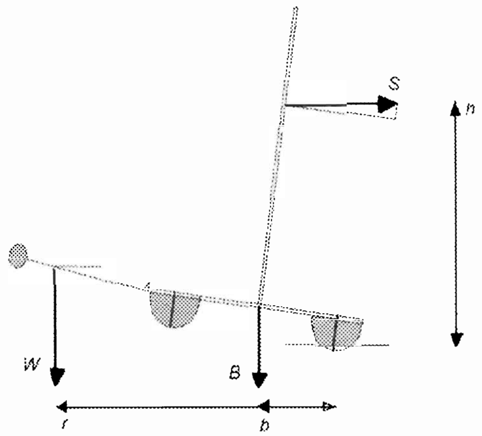


Figure 5: Front view of forces acting on upwind sailing Taipan, producing moments around the (leeward) hull in the water.

Balancing moments yields $rW + bB = hS$. Taking differentials yields

$$rdW = hdS \Rightarrow r\frac{dW}{S} = h\frac{dS}{S} \Rightarrow \frac{dS}{S} = \frac{r}{h}\frac{dW}{S} = \frac{rdW}{(rW + bB)}.$$

Noting that driving force F remains proportional to the lateral force S yields

$$\frac{dF}{F} = \frac{dW}{\left(W + b/r\,B\right)} \qquad (*)$$

Taking a first order perturbation of the force balance $F = c\,m^{1/2}\,v^2$ yields

$$\frac{dv}{v} = \frac{1}{v}\left(\frac{\partial v}{\partial F}dF + \frac{\partial v}{\partial m}dm\right) = \frac{1}{v}\left(\frac{1}{2}\frac{v}{F}dF - \frac{1}{4}\frac{v}{m}dm\right) = \frac{1}{2}\frac{dF}{F} - \frac{1}{4}\frac{dm}{m}.$$

Substituting from expression (*) and noting that the increment in total mass dm is attributable entirely to the increment in crew mass, $dW = g\,dm$ then

$$\frac{dv}{v} = \frac{1}{2}\frac{gdm}{\left(W + b/r\,B\right)} - \frac{1}{4}\frac{dm}{m}.$$

The beam of a Taipan catamaran is 2.43m, so b = 1.22. The centre of mass of an average person is approximately 1m, and so r = 3.43. We use W = 135g as a mean crew weight. Inserting numerical values,

$$dv/v = (1/2 \times g/(135g + 1.22/3.43 \times 102g) - 1/(4 \times (135 + 102)))\,dm = 0.00186\,dm.$$

In the context of a yachting race, this implies that a heavier crew will achieve a superior upwind performance (0.00186 for every additional kg) but worse downwind (-0.00105 for every additional kg). Taipans take around 1.2 times longer to complete an upwind leg as opposed to the same distance downwind (independent of wind strength). Consequently, the theoretical performance sensitivity over an entire course is

$$\frac{dv}{v} = dP = (1.2 \times 0.00186 - 1 \times 0.00105) / (1.2 + 1)\,dm = 0.00054\,dm.$$

Comparison of observations with physical model

Regressions of performances against crew weight for the collection of light, and strong wind races let us test the sensitivities predicted from the physical models. Results indicate that lighter crews achieve better performance improvements than predicted by the physical model. As mentioned earlier this may be explained by better sailors competing with lightweight crews than a causal relationship.

Table 2: Sensitivities of performance to crew weight (regression coefficient): observed and physical predictions. The statistically derived sensitivities in light, medium and strong winds are all negative with 95% confidence.

Conditions	Regression coefficient (90% confidence)	Physical model
Light winds	-0.00196 (-0.00242, -0.00150)	-0.00105
Medium winds	-0.00152 (-0.00219, -0.00084)	N/A
Strong winds	-0.00090 (-0.00139, -0.00039)	0.00054

YARDSTICK SYSTEMS

Yardstick systems are common in yachting to calculate corrected times in order that different classes of yacht can compete fairly. There are two main types of yardsticks: (i) formula based yardsticks based on key properties of the boat such as sail area, weight and length, and (ii) statistically calibrated yardsticks which use historical mixed-fleet race results. Examples of (i) are the Texel Rating system (Texel, 2008) and ISAF Small Catamaran Handicap Rating System (ISAF, 2008) while examples of (ii) are the Portsmouth Yardstick (Portsmouth 2008) and Victorian Yardstick systems (VYC, 2008). The Texel rating system incorporates crew weight in its formulation and recent enhancements in 2007 also make provisions for windstrength. The widespread use of the Texel system in catamarans motivates a comparison of this particular yardstick against the Taipan results.

The current Texel formula (Texel 2008) is:

$$TR = \frac{100 \times RW^{0.3}}{0.99 \times RL^{0.3} \times RSA^{0.4}}$$

where *TR* is the Texel rating, *RW* is the combined crew and boat mass, *RL* is the boat's rated length and *RSA* is the rated sail area. The 2007 Texel rating for a sloop Taipan is 107 over all windstrengths. To make race comparison with other yachts, a corrected time is calculated as 100/*TR* times the elapsed time, and corrected times are used to establish placings.

Since all Taipans satisfy the same length and sail measurements, the only variable is the combined crew and boat mass. If the Texel system is fair, then an increase in the total mass of a yacht should degrade performance by precisely the proportion accounted for in the corrected time. Let *CT* and *T* and *TR* (respectively *CT'* and *T* + *dT* and *TR* + *dTR*) be the corrected time and elapsed time and Texel rating for a yacht with total mass *RW* (respectively *RW* + *dRW*). If the corrected times are fair, then *CT* = *CT'*, leading to

$$CT = \frac{T}{TR} \times 100 = \frac{T + dT}{TR + dTR} \times 100 \Rightarrow \frac{dT}{T} = \frac{dTR}{TR}.$$

Linearising the equation and assuming a small change in crew weight, the increment in the yardstick is

$$dTR \approx \frac{\partial TR}{\partial RW} \times dRW = TR \times \frac{0.3}{RW} \times dRW.$$

Inserting numerical values (boat mass 102 and crew mass 135 kg) yields $dTR/TR = 0.00127\, dRW$. For an increase in elapsed time, the performance *P* will be affected $dP = -\, dT/\overline{T} \approx -dT/T = -dTR/TR = -0.00127\, dRW$. In other words, the Texel rating predicts that the sensitivity of performance to crew mass, dP/dRW is –0.00127. Table 2 indicates that historical Taipan results are consistent with a sensitivity of around –0.00196 in light conditions and and –0.0009 in strong conditions. We conclude that in light winds, the Texel correction is not sufficient (that is, light crews hold an unfair advantage).

CONCLUSIONS

We have found that the crew mass has a statistically significant impact on performance across all windstrengths. As intuitively expected, in stronger winds a heavier crew can expect a better position in the fleet than in light wind races. A higher powered test was applied using the matched nature of the data to confirm that crew weight is a significant predictor to explain the difference in performance for a given boat between light wind and strong wind races. The regression suggests that around 140 kg is the crew weight which yields the most consistent performance across light and strong windstrengths. A simple physical model was unable to predict the degree of variation observed in historical race data, which may be explained by more complex drag phenomenon or by a propensity for better sailors to compete with lighter than average crews.

The Texel Rating system provides light Taipan crews with an unfair advantage in light weather. If a handicapping system was introduced based on crew weight and windstrength, the historical observations suggest that for light wind races, for every kilogram the crew weight falls below (resp. above) 140 kg, the finishing time is should be adjusted by a reduction (resp. an extension) of 0.196 minutes for every 100 minutes of racing. In medium winds the extension is 0.152 minutes and in strong winds is 0.09 minutes per 100 minutes. Although the data provides a statistically significant relationship, it holds a wide relative error. The physical phenomena driving the dependences are not at all trivial. In agreement with Bethwaite (1995), "performance equalization for crews of different weights is turning out to be a complicated business".

References

Bethwaite, F. (1995) Performance equalization for smaller stature crews, World Tasar Class Association website at http://www.tasar.org/racing/equal.html, accessed 1 February 2008.

Bethwaite, F. (1996) *High Performance Sailing*, Waterline, Great Britain.

Brearley, M. N., de Mestre, N. J. and Watson, D. R. (1998) Modelling the rowing stroke in racing shells, *Mathematics Gazette*, **82:** 1-16.

Clewlow, L. and Strickland, C. (2000) *Energy Derivatives,* Lacima Group, Sydney.

Dudhia, A. (2008) *Physics of Rowing*, web site at http://www.atm.ox.ac.uk/rowing/physics, accessed 25 May 2008,

Texel ratings (2007 version), Texel rating webpage http://www.texelrating.org, Catamaran and Trimaran Club of the Netherlands, accessed 1 February 2008.

Harris, D., Thomas, G. and Renilson, M. (1999) Downwind Performance of Yachts in Waves, in *Proceedings of the 2nd Australian Sailing Science Conference*, February 1999, Hobart, Tasmania

International Small Catamaran Handicap Rating System (SCHRS) (2007 version) web site at http://www.schrs.com/, International Sailing Federation, accessed 1 February 2008.

Philpott, A.B., Sullivan, R.M. and Jackson, P.S. (1993) Yacht velocity prediction using mathematical programming, *European Journal of Operational Research*, **67:** 13-24.

Portsmouth yardstick for multihulls (2004 version) at http://www.ussailing.org/portsmouth/tables04/tables04mh.htm, US Sailing Foundation, accessed 1 February 2008

Sahoo, P.K., Salas, M., Schwetz, A. (2007) Practical evaluation of resistance of high speed catamaran hull forms – Part I, Ships and Offshore Structures, **2:4**, 307-324.

Tuck, E.O. and Lazauskas, L. (1996) Low drag rowing shells. In N. de Mestre (ed.) *Third Conference on Mathematics and Computers in Sport*, Bond University, Australia.

Townend, S. (1984) *Mathematics in Sport*, Halsted Press, NY.

VYC Victorian Yardstick website http://www.vic.yachting.org.au/site/yachting/vic/downloads/Yardsticks05_06.pdf, Yachting Victoria, accessed 1 February 2008.

CONCURRENT VALIDATION OF THE FACTOR STRUCTURE OF THE DECATHLON

Heazlewood, Timothy

School of Exercise Science, Australian Catholic University National, Sydney, Australia

Paper Submitted for Review: 7 April 2008
Revision submitted and accepted: 17 June 2008

Abstract. The Men's Decathlon in the sport of athletics consists of ten events, which are held on two consecutive days. Coaches and sport scientists have suggested training programs based on three models. An exploratory factor analysis (Heazlewood, 2006) based on the top IAAF performances for the Decathlon in 2005 was conducted, and revealed five factors that provided the simplest and most interpretable solution and a fourth model in designing training for the Decathlon. The five factors derived reflected throwing, sprint-hurdle, long sprint-endurance, jumping and pole vault abilities. The major focus of this investigation was to assess if the factor solution derived from the 2005 data would be replicated utilising IAAF Decathlon performances of 2006 (n=134) and 2007 (n=149). Athletes who had scores for all events and nonsignificant wind assistance were used in the analysis. The analysis replicated the 2006 model based on principle component analysis, Cattell's scree test for determining the number of significant factors and varimax rotation with Kaiser normalisation. Factor loadings greater than .5 were considered significant. The 2006 and 2007 data explained 74.5 % and 74.5% of the explained variance respectively, identical to 2005 data. Factor comparisons to compare the factor structure of the different samples were based on the pooled data method, factor invariance test via AMOS, and comparison measures method. Results based on 2006 and 2007 data did not display simple factor structure for the events as in the 2005 data. The javelin did not load with the shot put and discus, the 2005 and 2006 data loaded the 1500m across two factors, whereas on the 2007 data the 1500m loading on one factor and with the 2006 data, the jumps did not load on one factor as in 2005. However, the pole vault was identified as a unique factor using the 2005, 2006 and 2007 data. These findings suggest that greater factor complexity might exist in terms of the interrelationships among events and further revisions in training approaches.

Keywords: multigroup factor analysis, AMOS, structural equation modelling.

INTRODUCTION

The Men's Decathlon consists of ten events, which are held on two consecutive days. The events are contested in the following order; first day consists of 100m, long jump, shot put, high jump and 400m; and second day of the 110m hurdles, discus, pole vault, javelin and 1500m (International Association of Athletics Federations, 2005). Points are awarded for performance, based on times and distances in each event using a set of outdoor scoring tables (IAAF, 2004). The sum of the ten event scores produces the total points for the Decathlon and final awarding of places. Many coaches and athletes hypothesize that decathletes should possess speed, power and strength for the sprint/hurdles, jump and throw events. Endurance and aerobic ability are also thought to be required for the 1500m. Based on this model of factors or hypothesized constructs underpinning the Decathlon, a four factor model for the ten events was suggested (Mackenzie, 2006).

Some coaches have suggested that up to nine factors are responsible for the Decathlon, identifying factors such as aerobic endurance, gross strength, skill, relative strength, running speed, mobility, explosive strength, speed endurance and strength endurance. Three conceptual models have been presented as to the motor fitness constructs that are important for the Decathlon. However, it must be emphasized that these models have been developed without substantive multivariate empirical support. The three models are:

1. To group the individual events into training sessions based on the human energy systems that are thought to underpin each event (Kyprianou, 2007). That is, ATP-PC dependent sprint-power events such as 100m, 110m hurdles, throws and jumps; the anaerobic glycolytic event the 400m; and the aerobic event the 1500m.
2. The conceptual overlap of motor fitness components for each event such as sprints, throws and jumps (Lawler, 1987; Marra, 1985; Myers, 1988; Mackenzie, 2006; Kyprianou, 2007). The conceptual model using this approach is identified in table 1.
3. That each event in the Decathlon represents a unique motor skill and must be trained as a unique event (Brockburger, 2000; Siris, 1987).

Table 1: The motor fitness elements of the Decathlon events as conceptualized by Mackenzie (2006).

Event	Aerobic Endurance	Gross Strength	Skill	Relative Strength	Running Speed	Mobility	Explosive Strength	Speed Endurance	Strength Endurance
100m	Low	Med	Med	High	High	High	High	Med	-
Long Jump	-	Low	High	High	High	High	High	-	-
Shot Putt	-	High	High	Med	Low	Med	High	-	-
High Jump	-	Low	High	High	High	High	High	-	-
400m	Med	Low	Low	High	High	Med	Low	Med	High
110m Hurdles	-	Med	High	High	High	High	High	Med	-
Discus	-	High	High	Med	Low	High	High	-	-
Pole Vault	-	Med	High	High	High	Med	Med	-	-
Javelin	-	Med	High	High	Low	High	High	-	-
1500m	High	-	-	Low	Low	-	-	-	Med

Heazlewood (2006) conducted an exploratory factor analysis where the data were the performances from the ten events using the IAAF 2005 international rankings for the Men's Decathlon. He hypothesized that once the factor structure of the different events was identified; coaches, athletes and sports scientists could modify training recommendation to maximise training transfer between events as well as to competition. A fourth model based on this evidence for designing training for the Decathlon was developed. This model was derived from a five factor model that provided the simplest and most interpretable solution. The statistical solution applied principle component analysis, Cattell's scree test for determining the number of significant factors and varimax rotation with Kaiser normalization (Hair et al., 2006). The five factors derived (Heazlewood, 2006) were defined as throwing, sprint-hurdle, long sprint-endurance, jumping and pole vault abilities.

The major focus of the current research was to evaluate if the factor structure derived from the IAAF 2005 Decathlon data would be replicated or concurrently validated with the Men's Decathlon international scores and rankings using IAAF 2006 and 2007 data. The research objective was to understand in a more empirical manner the interrelationships between the ten events in the Decathlon by applying multivariate confirmatory factor analysis using the multigroup approach.

Research Questions

A number of significant research questions were generated by the review of literature. Specifically,

1. Will multigroup confirmatory factor analysis, derived from competition performance, confirm the five factor model generated by Heazlewood (2006) or will a completely new model be derived based on each year indicating the relationship of the ten events to the factors is evolving over time?
2. Will the derived factor structure enable further modification of Decathlon training, where events that display significant factor loadings with specific factors can be grouped together for training in an attempt to maximize training efficiency and transfer between events?

METHODS

The primary objectives of confirmatory factor analysis (Kim & Mueller, 1978; Costello & Osborne, 2005; Hair et al., 2006) is to determine if the number of factors and the loadings of measured variables on them conform or replicate what is expected on the basis of pre-established theory. In addition,

1. To identify underlying constructs or factors that explains the correlations among a set of variables.
2. To test hypotheses about the structure of the variables.
3. To summarise a large number of variables with a smaller number of derived variables (factors).
4. To determine the number of dimensions to represent a number of variables.
5. To achieve the simplest and pragmatically more meaningful factor solution.
6. The simultaneously analysis of multiple groups provides a test of significance of any differences found between the groups.
7. It can be inferred that if no differences exist between the groups, or if the group differences affect only a few model parameters, the multigroup approach provides for more efficient parameter estimates than single group models and approaches utilising factor analysis.

The major focus of this investigation was to assess if the factor solution derived from the IAAF 2005 data (n=122) would be replicated utilising IAAF Decathlon performances of 2006 (n=134) and 2007 (n=149). Many of the athletes appear on all three data sets. The data sets represent actual competition performances recognised by the IAAF and of sufficiently high performance to receive an official IAAF ranking for the event. Athletes who had scores for all events and nonsignificant wind assistance were used in the analysis. The data for the track and field events were recorded to two decimal places by the IAAF using the 2005, 2006, and 2007 data sets and were recorded in seconds for all track events and in metres for all field events. The 1500m times were converted to seconds to facilitate the statistical calculations.

The analysis replicated the Heazlewood (2006) model by applying principle component analysis, Cattell's scree test for determining the number of significant factors and varimax rotation with Kaiser normalization. Factor loadings greater than .5 were considered significant based on sample sizes approximating 120 cases (Hair et al., 2006). Factor comparisons to compare the factor structure of the different samples were based on the pooled data method using dummy variable coding for group association, factor invariance test via AMOS, and the comparison measures method. The pooled data method is where the data for two samples is pooled, adding a dummy variable where the coding represents group membership. The factor loadings of this dummy variable indicate the factors for which the groups' mean factor scores would be most different. Factor invariance test via AMOS software is a structural equation modeling technique, which tests for deterioration in model fit when factor loadings are constrained to be equal across sample groups (Arbuckle & Wothke, 1999).

The comparison measures method requires computation of various measures which compare factor attributes of the two samples. Specifically, 1. Root mean square, which is the root mean square of the average squared difference of the loadings of the variables on each of two factors; 2. The coefficient of congruence is the sum of the products of the paired loadings divided by the square root of the product of the two sums of squared loadings and; 3. The salient variable similarity index is based on classifying factor loadings into positive salient ones (over +.1), hyperplane ones (from -.1 to +.1), and negative salient ones (below -.1). Hyperplane loadings, which approach 0, indicate variables having only a near-chance relationship to the factor. The solutions were derived using SPSS version 16 (2007) and AMOS 16 (2007) statistical software.

RESULTS

The 2006 and 2007 data when analyzed individually, explained 74.5 % and 74.5% of the explained variance respectively based on the initial factor solution to identify the number of significant factors, identical to 2005 data. As indicated previously, factor comparisons to compare the factor structure of the different samples were based on the pooled data method, factor invariance test via AMOS, and comparison measures method.

The pooled data method where the factor loadings of this dummy variable indicate the factors for which the groups' mean factor scores would be most different, indicated the major source of difference was for factor 5, when the 2005 data was compared to the 2006 data, although the loadings on most of the events do not occur on factor 5 in this solution.. This solution is based on the rotated component matrix illustrated in table 2.

Table 2: Pooled data method based on comparisons of the 2005 with 2006 data.

Rotated Component Matrix

Event	Component				
	1	2	3	4	5
100m	.029	.824	-.207	-.144	.041
Long jump	-.222	-.207	.059	.789	.021
Shot put	.795	-.066	.020	.095	-.041
High jump	.333	.052	-.159	.557	-.310
400m	.343	.655	.425	.227	.213
110m Hurdles	-.250	.678	.080	-.139	-.123
Discus	.735	-.218	.141	-.095	-.007
Pole vault	.129	-.079	-.735	.208	.283
Javelin	.675	.167	-.048	-.034	-.044
1500m	.378	-.195	.632	.286	.230
Group	-.075	.016	-.092	-.094	**.871**

When the 2005 data was compared to the 2007 data the major source of difference between the two samples is on factor 3, which are the 400m and 1500m. This solution is based on the rotated component matrix illustrated in table 3. When the 2006 data was compared to the 2007 data the major source of difference between the two samples is on factor 5, which is pole vault. This solution is based on the rotated component matrix illustrated in table 4.

When the 2006 data was compared to the 2007 data the major source of difference between the two samples is on factor 5, which is pole vault. This solution is based on the rotated component matrix illustrated in table 4. In the 2005 data the shot, discus and javelin loaded on factor 1 (throwing ability) whereas in the 2006 data the javelin was loaded on factor 1 and factor 2 indicating factor complexity, and for the 2007 data it was loaded significantly on factor 3. It is important to note that the identical factor solution was utilised with all three Decathlon samples.

Table 3: Pooled data method based on comparisons of the 2005 with 2007 data.

Rotated Component Matrix

Event	Component				
	1	2	3	4	5
100m	-.009	.804	-.105	-.061	.185
Long jump	-.052	-.287	-.071	.672	-.320
Shot put	.816	-.063	.100	.081	.002
High jump	.161	.021	.137	.692	.223
400m	.239	.703	.442	-.109	-.108
110m Hurdles	-.193	.719	-.110	-.021	-.157
Discus	.774	-.072	.125	-.136	.042
Pole vault	.051	-.043	-.020	.028	.902
Javelin	.632	.101	-.092	.362	.057
1500m	.277	-.205	.732	-.212	-.145
Group	.076	-.076	**-.704**	-.262	-.106

Table 4: Pooled data method based on comparisons of the 2006 with 2007 data.

Rotated Component Matrix

	Component				
	1	2	3	4	5
100m	.813	-.108	.201	.031	-.046
Long jump	-.645	-.230	.300	.265	.158
Shot put	-.091	.866	.196	.110	.008
High jump	-.128	.079	.715	.031	-.060
400m	.648	.020	.192	.602	-.054
110m Hurdles	.620	-.247	-.295	-.020	.201
Discus	-.028	.867	.111	.119	-.043
Pole vault	.024	.002	.380	-.273	-.688
Javelin	.103	.225	.597	.090	.049
1500m	-.097	.222	.064	.822	-.028
Group	-.014	-.035	.275	-.342	**.723**

The information in table 5 represents the rotated component matrix for all samples with group included as a dummy variable. The loading of the group variable on factor 5 indicates that this factor appears to be the main source of difference between the 2005, 2006 and 2007 samples. However, in this solution it is interesting to note that most other events, except the small loading for the 1500m, load significantly on the other four factors. Table 5 is also very similar to the original solution utilising only the 2005 sample, where the different factor constructs defined as throwing (factor 1), sprint-hurdle (factor 2), jumping (factor 3) and pole vault (factor 5) abilities loaded predominantly with one factor. It is interesting to note the long sprint-endurance factor displayed factor complexity and loaded with factor 1, 4 and 5).

Table 5: Pooled data method based on comparisons of the 2005, 2006 and 2007 data with the group association dummy variable included.

Rotated Component Matrix

	Component				
	1	2	3	4	5
100m	-.030	.824	-.044	.200	.087
Long jump	-.157	-.324	.722	-.235	.039
Shot put	.811	-.074	.077	.001	-.029
High jump	.227	.053	.711	.189	-.086
400m	.288	.732	.046	-.302	-.240
110m Hurdles	-.307	.642	-.159	-.085	.038
Discus	.793	-.098	-.098	-.025	-.053
Pole vault	.119	-.058	.035	.773	-.136
Javelin	.551	.185	.306	.124	.076
1500m	.433	-.087	.021	-.527	-.412
Group	.023	-.027	-.034	-.090	**.910**

The omission of the group association dummy variable and pooling the 2005, 2006 and 2007, essentially replicated the original 2005 factor structure. This pooled model using all the data sets is represented in table 6. Specifically, throwing ability with factor 1 with shot (.812), discus (.767) and javelin (.650) displaying high factor loadings; sprint-hurdle ability with factor 2 with high loadings for 100m (.831), 400m (.709) and 110m hurdles (.646); jumping ability with factor 4 displaying high loadings with long jump (.734) and high jump (.702); and pole vault ability with factor 5, where the only significant loading was for pole vault (.959). It is interesting to note the long sprint-endurance ability displayed some factor complexity as the 400m loaded across factor 2 (.709) and factor 3 (.562), however it is important to recognise that the 1500m loaded significantly on factor 3 (.922).

The five factor solution using the AMOS approach for the factor invariance test based on a structural equation modeling techniques, which tests for deterioration in model fit when factor loadings are constrained to be equal across sample groups (Arbuckle, 1997; Arbuckle & Wothke, 1999) did not solve for two sample method or three sample method. This was an outcome of the model as not sufficiently identified to solve the structural equations and derive the factor loadings or coefficients, especially with pole vault as a unique factor not linked to the other four factors. When the pole vault was linked to the fourth latent factor of jump ability (observed variables of high jump and long jump), in an attempt to identify the model sufficiently, the AMOS solution was still not identified using least squares and maximum likelihood factor approaches.

Table 6: Pooled data method based on comparisons of the 2005, 2006 and 2007 data. The group association dummy variable has been omitted.

Rotated Component Matrix

	Component				
	1	2	3	4	5
100m	.003	.831	-.109	-.057	.139
Long jump	-.174	-.322	.083	.734	-.152
Shot put	.812	-.104	.160	.044	.019
High jump	.271	.065	-.016	.702	.177
400m	.154	.709	.562	.041	-.035
110m Hurdles	-.224	.646	-.134	-.157	-.229
Discus	.767	-.130	.211	-.127	.024
Pole vault	-2.931E-5	-.023	-.023	.030	.959
Javelin	.650	.170	-.112	.270	-.010
1500m	.154	-.122	.922	.033	-.015

DISCUSSION AND CONCLUSION

Results based on 2006 and 2007 data did not display simple factor structure for the events as in the 2005 data. The javelin did not load with the shot put and discus; the 2005 and 2006 data had the 1500m across two factors whereas on the 2007 data the 1500m loading on one factor, the jumps did not load on one factor as in 2005, which suggests the jumping events may require slightly different abilities based on these findings. However, the pole vault was identified as a unique factor using the 2005, 2006 and 2007 data. Although, the pole vault it loaded on different factors for each sampled year. It is important to note that the sample of international athletes were almost complete based on the IAAF data and represented the high performance athletes who compete in the sport of athletics. The lower ability athletes were not evaluated in this study as these athletes do not appear on the IAAF data due to the fact that the lower performing athletes have not met IAAF performance criteria. If a more representative sample of all decathlete abilities were available it would be interesting to assess if they displayed similar variability in the factor structure when they competed over a number of years.

The pooled data resulting in a factor solution very similar to the 2005 data and the factor structure for the ten Decathlon events derived by Heazlewood (2006), where the throwing events, sprint-hurdle events, jump events (except pole vault) and long sprint-1500m tended to load with a unique factor does suggest that clustering the events in this manner for training sessions may result in more optimal and efficient training programs. However, emphasizing that the pole vault tended to load as a unique construct with factor 5, which can be defined as the unique construct of pole vault ability, this event can probably be trained for in isolation from other events. The following conclusions were an outcome of the research. Specifically:

1. The factor structure of the Decathlon appears to vary from year to year as the factor loadings and factor structure changed with different data sets. Although the amount of initial explained variance utilising the five factors model of throwing, sprint-hurdle, long sprint-endurance, jumping and pole vault abilities. A model, which initially provided a clustering of training activities based on a simple and interpretable factor structure.
2. The 2006 sample suggested a more complex factor structure, especially for the 100m, 110m hurdles, 1500m, javelin and high jump as these events loaded across a number of factors, whereas the 2007 was more like the 2005 sample in terms of a simple factor structure.

3. The results may indicate performance variability resulting in inconsistent factor loadings and factor structure. However, it would be expected that underpinning performance constructs should be reflected in consistent performances in related events at this high performance level, such as reflecting the throw ability construct. This suggests the throwing events form a useful training cluster.
4. Developing training programs based on events clustered under specific factors such as throws, jumps, sprints, might be the logical event clusters to follow based on the general trends of the 2005 and 2007 samples. Both 2005 and 2007 are World Athletic Championship years, whereas the 2006 sample is not the World Championship year and may reflect different training and event performance interactions and a possible artifact within long term Decathlon training programs.
5. Finally, pooling all the Decathlon data sets (2005, 2006 and 2007) reproduced a simple and interpretable factor solution for all events, except for the 400m, which loaded across the two factors of sprint ability and long sprint 400m - 1500m ability. The interpretable factor loadings related to specific events and provided an empirical basis to develop training programs to enhance training transfer between the clustered events. Events, which are thought to represent underpinning motor fitness abilities.

References

Arbuckkle, J. and Wothke, W. (1999) *Amos 4.0 Users' Guide*. SmallWaters Corporation, Chicago, USA.

Blockburger, S. (2000) Development and training for the Decathlon. *Decathlon 2000*. Retrieved November 23, 2005 from Web site: http://www.decathlon2000.ee/eng/10athlon.php?id=4

Costello, A. and Osborne, J. (2005) Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. [Electronic version]. *Practical Assessment, Research & Evaluation*. **10 (7):** 1-9.

Hair, J. E., Block, W., Babin, B.J., Anderson, R. E. and Tatham, R. L. (2006) *Multivariate Data Analysis 6th Ed.* Pearson - Prentice Hall, Upper Saddle River, USA.

Heazlewood, T. (2006) Factor structure of the Decathlon: implications for training. Conference presentation at *CISC 2006 – 13th Commonwealth International Sport Conference: People, Participation and Performance- Scientific Program.* 9-12 March 2006. Melbourne, Australia.

International Association of Athletics Federations. (2005). *Competition Rules 2004 – 2005*. IAAF, Monaco.

IAAF Global Athletics (2004) *IAAF Scoring Tables for Combined Events*. IAAF, Monaco.

Kim, J. O., and Mueller, C. W. (1978). *Factor Analysis: Statistical Methods and Practical Issues*. Sage, London, UK.

Kyprianou (2007) Developing the Decathlon technical model. *Decathlon 2000*. Retrieved December 11, 2007 from Web site: http://www.decathlon2000.ee/eng/training.php?art=1158

Lawler, P. (1987) Chasing points in the Decathlon. *Modern Athlete and Coach*. **24(4):** 3-6.

Mackenzie, B. (2006) Decathlon. *Sports Coach*. Retrieved January 5, 2006, from Web site: http://www.brianmac.demon.co.uk/decath/index.htm

Marra, H. (1985) Decathlon hints. *Modern Athlete and Coach*. **23(4):** 15-18.

Myers, B. (1988) Training for the jumps and multi-events. *Modern Athlete and Coach*. **26(3):** 7-10.

Powers, S. K. and Howley, E. T. (2004) *Exercise Physiology: Theory and Applications to Fitness and Performance 5th ed*. McGraw Hill, Sydney, Australia.

Siris, P. (1987) Evaluation procedures for the selection of potential Decathletes. *Modern Athlete and Coach*. **25(1):** 18-20.

SPSS Inc. Software (2007) *SPSS Version 16.0* [Computer Program]. SPSS Inc, Chicago, USA.

A MEASURE OF COACH'S INFLUENCE ON GAMES IN THE RUN: WHO IS MOST ACCOUNTABLE?

Baglin, James, Bedford, Anthony, Ladds, Monique, and Makhoul, Pamela

School of Mathematical and Geospatial Sciences, RMIT University, Australia

Paper Submitted for Review: 29 February 2008
Revision submitted and accepted: 16 June 2008

Abstract. In many sports, the measure of a team's success is often attributed to the coach. Notably, when a team performs poorly, it is the coach that is sacrificed as a ritual scapegoat. Extensive research has evaluated the validity of this mentality, yielding conflicting results. However no research, to our knowledge, has attempted to measure exactly what a coach can influence during a game and how different sports compare. In this research, we consider a number of professional sports on the level of influence coach's have on their player's/team's outcome during a match. By considering level of communication, strategic ability, substitutions available and officiating influence which are all determined by the rules of a sport, we develop a coaching influence rating model. This model yields a single measure which is used to rank the degree of influence a coach has during a match and enables comparison of coaching influence across the rated sports. Then, assuming that coaching turnover is related to coach accountability, we validate the coaching influence model using coaching turnover data for the ranked sports for the previous ten seasons. We find our ratings model and coach turnover are strongly correlated, indicating that the more influence a coach has in the run, the higher the turnover of coach's in that sport. The results of the analysis not only make for interesting analytical discussion but also provide a framework for assessing and researching the exact impact that a coach has on the day of a sporting event.

Keywords: Coaching, turnover, firings, coach influence

INTRODUCTION

With the advances in technology, billion dollar budgets, wealth of support staff, technical teams, and advanced training methods, the pressure on coaches to get their players and support team to perform is critical. Wood (2008) states the "...[role of the] coach will be many and varied, from instructor, assessor, friend, mentor, facilitator, chauffeur, demonstrator, adviser, supporter, fact finder, motivator, counselor, organizer, planner and the Fountain of all Knowledge..." no less, all this on top of the technical and analytical roles. So the question is how much of a teams performance is actually attributable to the coach? Previous research conducted on coach influence has typically looked at the effect that coaching change has on a team's performance. These studies ask the question of whether firing a coach is effective in improving a team's performance. Early research suggested that firing baseball managers increased team performance in the short-term but this improvement soon regressed to the mean (Fabianic, 1994).Further studies have been inconsistent. Audas, Dobson and Goddard (1997) found that short-term performance did increase after firing a coach in soccer whereas Balduck and Buelens (2007) concluded that firings had no short-term impact on a team's performance.

Other studies on the four most popular North American sports, basketball, baseball, ice hockey, and American football have established that coaching change does seem to increase short-term performance although in the following season performance did not change (McTeer, White, & Persad, 1995). Koning (2003) argued that apparent team performance increases after a coaching change disappears after controlling for opposition quality and Bruinshoofd and ter Weel (2003) contended that this apparent improvement was due to regression to the mean and not due to the coaching change itself. More surprisingly, Audas, Dobson and Goddard (2002) have even found that managerial change in football (soccer) can have an adverse effect on performance. According to this research it would appear that sacking a coach is a highly controversial strategy to improve a team's performance. However, many teams in many different sports still fire coaches or put coaches in positions where they are forced to resign in a ritualistic scapegoat manner (Gamson & Scotch, 1964).

Our research aims to assess the possibility that the amount of influence a coach can have on a game in-the-run is related to the rate of coaching turnovers and firings. We develop a new model used to determine the level of influence in major worldwide team sports, and assess the correlation between the influence level and coach turnover.

METHOD AND RESULTS

To effectively evaluate the influence of a coach in-the-run we developed a Coach's Influence Rating Model (CIRM). This level of influence is directly related to what the rules of a sport enable a coach to do. Given the high variability between sporting rules, this model was developed keeping in mind that any sport should be able to be rated using the CIRM. This rating is best interpreted only as a way of comparing coaching influence among different sports and should not be generalised beyond that purpose. To construct the CIRM we had to determine what a coach can do to influence the results of a game keeping in mind sports of all types. Three sources of information gathering where used including the literature, sport rule books, and a group Delphi. Literature on coaching competency identified motivation and game strategy (Myers, Feltz, Maier, Wolfe, & Reckase, 2006) as two important but broad *on-the-day* coaching influences. Analysis of sport rule books enabled a further break-down of the different areas of influence. Because motivation would be troublesome to quantify or rank, the group Delphi decided that motivation should be related to the ability to communicate with players. Thus *communication* became the first factor of CIRM. We later added a second component called field size to this factor as the Delphi agreed that that larger the field the harder it is to communicate with players. The second factor, *game strategy*, was defined by on-field strategic changes and strategic substitution. Both of these factors were further determined by how often and how many changes could be made. The last factor identified was influence over game *officiating*. For example, NFL (American football) allows coaches to make two challenges to a game official's decision using video replay. Therefore, we proposed that coaching influence on-the-day was a function of their ability to communicate with players, make strategic changes both on the field and through substitution, and influence officiating.

The next step in the development of CIRM was to determine how to score each sport on each of the three coaching influence factors. While many methods were proposed in the Delphi, difficulty with resulting large variation in coaching influence indicators informed us that a simple approach was the best. Therefore, an ordinal system of scoring each factor was developed. This scoring system along with the CIRM is shown in Table 1. The first factor, communication, was determined by the sub-factors *contact* and *field size*. Contact enables a sport to be ranked on the coach's ability to communicate with players. This ranges from two extremes; no contact (e.g. Tennis), to full contact (e.g. Basketball). We also added the concept of indirectness if the coach had to communicate through another person (e.g. runners in AFL). Field size was included to take into account the coach's ability to communicate effectively especially in stadiums full of spectators. There was debate over how to break this factor into level but the agreed upon conventions were fields (e.g. Football, Rugby and Cricket) and courts or rinks (e.g. Basketball and Ice Hockey).

The second factor, game strategy, was divided into *on-field changes* and *substitution*. On-field changes refer to the changes that coach's can make between different positions on the field and substitution refers to the changes that the coach can make off a bench if one exists. Both of these sub-factors were measure by two determinants: the *timing of changes* and the *quantity of changes*. Timing of changes rates a sport in terms of when changes can be made. Sports can range from no changes at all (e.g. tennis) to changes anytime during the play of a match (e.g. ice hockey). Quantity of changes scores a sport in terms of how many possible combinations or changes a coach can make during a match. Some sports can only make very subtle changes (e.g. switching forwards in netball) whereas others can make substantial changes (e.g. substitution of players virtually at will in basketball).

The final factor that was included was *influence over officiating*. This aimed to score sports on a coach's ability to influence official decisions during a game. With most sports there are clear rules that prevent coaches and player from questioning official rulings, however, some sports allow either the captain, or even the coach, to voice their concerns. We proposed that sports which allow a coach to have an impact on game officiating give coaches the potential ability to have a greater effect on the outcome of the game compared to sports that do not, even if it may be detrimental to a teams chances of winning. A measure for indirect influence was also included in sports where coaches had to question officiating decisions through team captains (e.g. Ice Hockey). The broad equation for CIRM is given by

$$CIRM = Communication + Game Strategy + Officiating\ Influence \quad (1)$$

where values of variables are taken from Table 1.

Table 1: The Coach's Influence Rating Model with Scoring System

Coaching Influence Factor	Coaching Influence Factor determinants	Scoring	Example Sport – AFL
Communication	1.1 Contact	0 = Zero contact at match/event 1 = Contact between periods/events 2 = Contact between plays 3 = Contact any time (-1 if contact is indirect)	2*
	1.2 Field Size	0 = Large field 1 = Court/Rink	0
Game Strategy	2.1 Timing of Changes	0 = No changes 1 = Between periods or injuries 2 = Between plays 3 = On-the-fly (anytime during play)	3
	2.2 Quantity of changes	1 = Less than a ¼ of a team 2 = Less than or equal to ½ 3 = Greater than ½ but less than a full team 4 = Whole team	4
	2.3 Timing of Substitutions	0 = No changes 1 = Between periods or injuries 2 = Between plays 3 = On-the-fly (anytime during play)	3
	2.4 Quantity of Substitutions	1 = Less than a ¼ of a team 2 = Less than or equal to ½ 3 = Greater than ½ but less than a full team 4 = Whole team	1
Officiating Influence	3. Officiating Influence	0 = No Influence 1 = Indirect influence 2 = Direct influence	0
CIRM score			13

*AFL football is subtracted a point as the coach must communicate to the team via runners (indirect communication).

Data on coaching turnover was sourced from a number of different websites for each sport. While many websites had information about coaches and the years of their coaching tenure, many sites did not specify the nature of the turnover. This required extensive research using a number of different internet resources, and where information was not available, email and telephone calls were made to respective team officials. We coded coaching turnover as simply a change in coaches for any reason such as resignations, contract conclusions, and terminations. We then coded a coaching turnover as a coach firing if the turnover was involuntary on the coach's behalf, e.g. termination or non-renewal of coaching contract, or forced resignation due to the hiring of another coach. Contract conclusions could go either way. For example, a coach could choose not to continue with a team at the end of a contract even though they were offered a contract extension or whether the contract ended and no extension was offered. The later was coded as a firing. We also had to determine if any of the listed coaches were designated as interim coach. Changes of interim coaches were coded as turnovers but never as a firing. In the end we had two measures of coaching changes; an all in measure of coaching turnover and a more specific measure of involuntary firings. Both were deemed to provide related but unique indications of the nature of coaching turnover in a sport. It is also important to note here that our information about turnovers and firings are only as good as the validity of the sites that reported the coaching change.

Sporting websites were consulted for data pertaining to coaching turnovers over a period of ten seasons for each sport. Table 2 lists the sports included in the analysis and the websites used to gather the required data. These sports were chosen because of the ease of access to information about coaches and coaching turnover, and were restricted to team sports.

Table 2: Coaching Turnover Information Sources

Sport	Source
MLB	www.sportecyclopedia.com, www.baseball-almanac.com, www.wikipedia.org, www.nytimes.com
NFL	www.sportecyclopedia.com, www.nytimes.com
NHL	www.sportecyclopedia.com, www.nhl.com, www.nytimes.com
NBA	www.sportecyclopedia.com, www.basketball-reference.com, www.wikipedia.org, www.nytimes.com
EPL	http://www.soccerbase.com/, http://en.wikipedia.org, http://news.bbc.co.uk/sport/
NRL	http://www.nrlstats.com/, http://en.wikipedia.org, http://www.highbeam.com
AFL	http://en.wikipedia.org, http://afl.allthestats.com/, http://news.google.com/archivesearch
S14	http://en.wikipedia.org, http://news.google.com/archivesearch
IC	http://www.cricinfo.com/homepage/index99.html, http://news.google.com/archivesearch

Note. NBA = National Basketball Association, NHL = National Hockey League, NFL = National Football League (American rules), MLB = Major League Baseball, EPL = English Premier League (Soccer), NRL = National Rugby League, AFL = Australian Football League (Australian rules), S14 = Super 14 (Rugby Union), IC = International Cricket.

RESULTS

Rule books and the Delphi procedure were used to classify each team sport on each of the factors. Where there was uncertainty or disagreements, respective sporting experts were contacted and queried to make a final decision. The results of the CIRM analysis for the sports analysed ranked and listed in descending order are shown in Table 3. The sport with the most in-the-run coaching influence is basketball whereas cricket is scored quite low.

Table 3: Coaching Influence Ratings as Scored by CIRM

Sport	CIRM Factors							
	1.1	1.2	2.1	2.2	2.3	2.4	3	Total
Basketball	3	1	3	4	2	4	2	19
Ice Hockey	3	1	3	3	3	4	1	18
American Football	3	0	2	4	2	4	2	17
Baseball	2	0	2	3	2	3	2	14
Australian Football	2	0	3	4	3	1	0	13
Football (Soccer)	3	0	3	3	2	1	1	13
Rugby League	2	1	3	4	2	1	0	13
Rugby Union	1	0	3	1	2	2	0	9
Cricket	1	0	2	3	0	0	0	6

The coaching turnover data that was gathered for each sport was converted into a turnover measure used to compare all sports. This required a measure that would control for the number of teams in a sporting league and the differing length of seasons. The length of a season was an important variable to control for as it varies substantially between sports (e.g. NHL 28 weeks vs. NFL 17 weeks). We believed that we had to control for season length as sports that had longer season would have more opportunities to turnover coaches as opposed to a sport with a shorter season. The season length was measured in weeks from the week of the first game to the last game of the regular season. If this period varied from season to season we recorded the season length as the median number of weeks per season over the ten seasons gathered. We did not use games played as a measure of season length because it created massive variability (e.g. MLB 162 games per season compared to NFL 16 games per season). This ended up considerably skewing the turnover measures for different sports when compared to others. Weeks of season was found to offer the best compromise of controlling for season length but at the same time not creating excessive variability. Thus, coaching turnover for each season of the sport in question was measured by

$$\text{Season Turnover} = \left(\frac{x_i}{y} \Big/ z_i \right) \quad (2)$$

where x_i = coach turnovers for a season, y = the number of teams in the competition, and z_i = the number of weeks of the respective sport's regular season. Once this was calculated for each sport for each of the ten seasons, the average coaching turnover was obtained by

$$\text{Coaching Turnover} = \frac{\sum \left(\frac{x_i}{y} \Big/ z_i \right)}{n} \quad (3)$$

where n = the number of seasons that (2) is calculated for. This gives a measure of *coaching turnover* defined as the average coaching turnover per team per season. We also calculated *coaching firings* by substituting coach firings in x_i of (2) and (3). This gave a measure of the average firings per team per season. The correlation between these two measures was high at $r = .79$ ($p = .012$ [Spearman's $\rho = .98$, $p < .001$]). Coaching Turnover and Firings for each sport are shown in Table 4.

Table 4: Coaching Turnover and Firings for the Last 10 Seasons

Sport	Turnovers	Firings	Teams*	Season Weeks	Coaching Turnover (2)	Coaching Firings (2)	CIRM Score
NBA	109	67	30	24	0.0155	0.0095	19
NHL	107	73	30	28	0.0147	0.0100	18
NFL	73	48	32	17	0.0137	0.0090	17
MLB	87	58	30	26	0.0114	0.0076	14
EPL	82	45	20	39	0.0107	0.0059	13
AFL	36	18	16	22	0.0102	0.0051	13
NRL	33	18	16	26	0.0086	0.0046	13
S14**	41	16	14	12	0.0277	0.0110	9
IC	42	9	10	52	0.0083	0.0018	6

* = Team numbers vary for each season because of defunct teams or relegation.
** = Outlier

Eight out of the nine sports seem to conform relatively well to the CIRM except for the Super 14 competition. The relationship between coaching turnover and CIRM score with the Super 14 outlier included was not significant ($r = -.017$, $p = .97$ [Spearman's $\rho = .51$, $p = .16$], Figure 1). However, with Super 14 removed there was a significant positive trend between coaching turnover and CIRM scores ($r = .9$, $p = .003$ [Spearman's $\rho = .98$, $p < .001$], Figure 2). For coaching firings the Super 14 also acted as an apparent outlier. With the Super 14 included in the data, the relationship between coaching firings and CIRM score was positive but not significant $r = .58$ ($p = .11$[Spearman's $\rho = .49$, $p = .18$], Figure 3). Again, with the influence of the Super 14 removed, the correlation returned a strong positive linear trend between coaching firings and CIRM score ($r = .96$, $p < .001$[Spearman's $\rho = .95$, $p < .001$, Figure 4). Overall, it appears that coaching turnover and firings seem to co-vary with on-the-day coaching influence as measured by the CIRM. However, the Super 14 outlier complicates this conclusion.

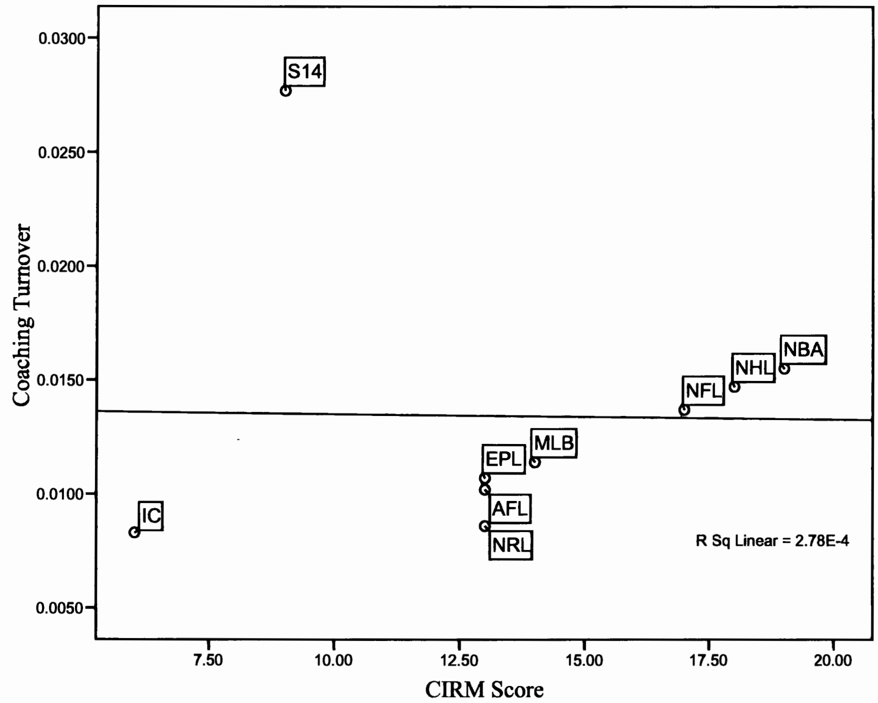


Figure 1: The relationship between coaching turnover and the CIRM (Super 14 included)

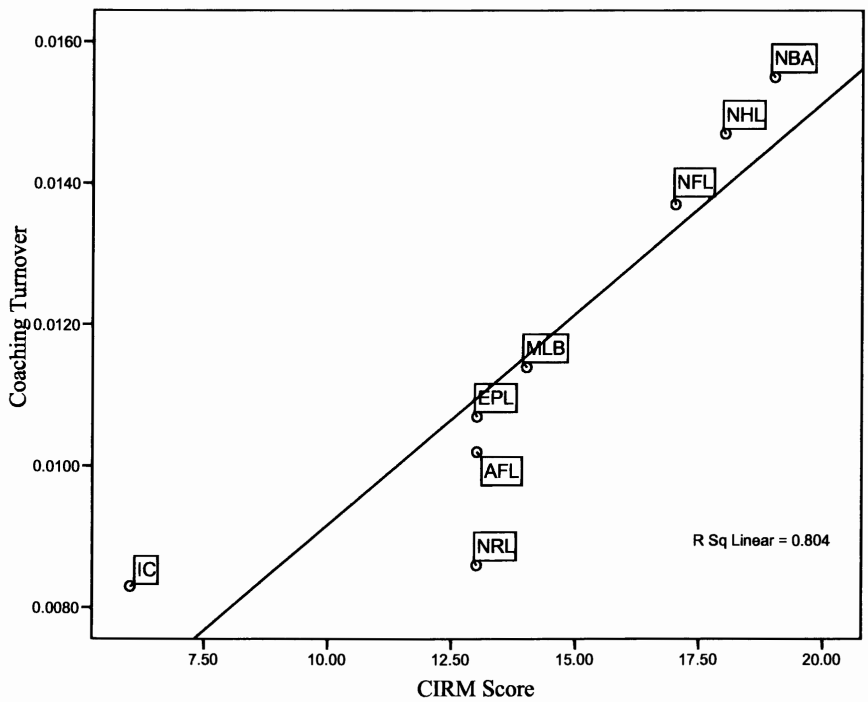


Figure 2: The relationship between coaching turnover and the CIRM (Super 14 excluded).

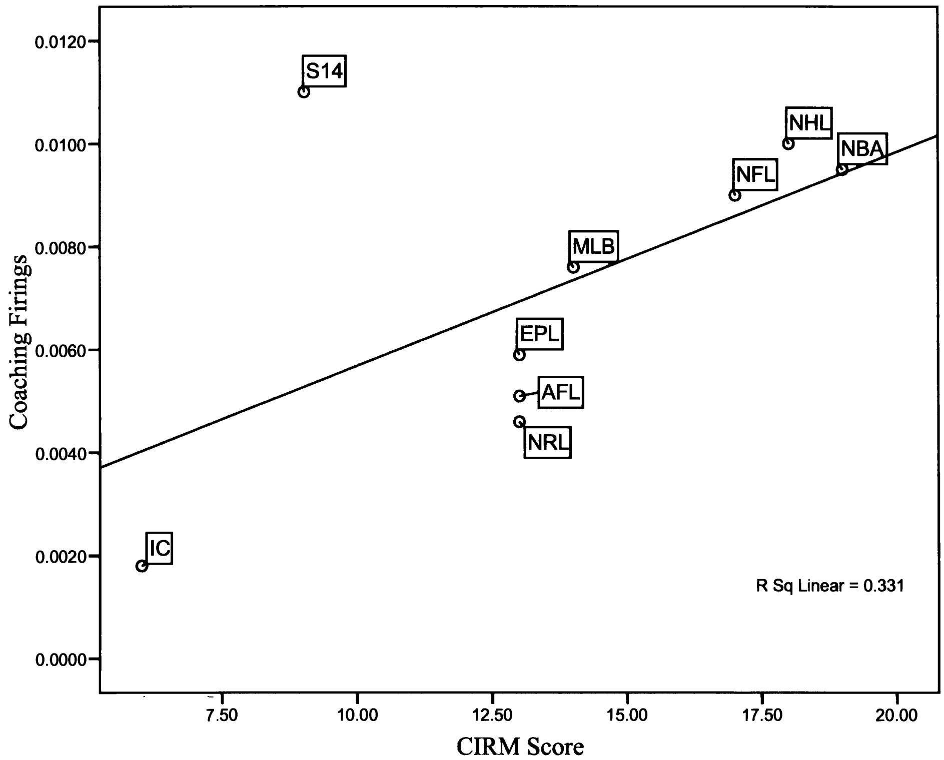


Figure 3: The relationship between coaching firings and the CIRM (Super 14 included).

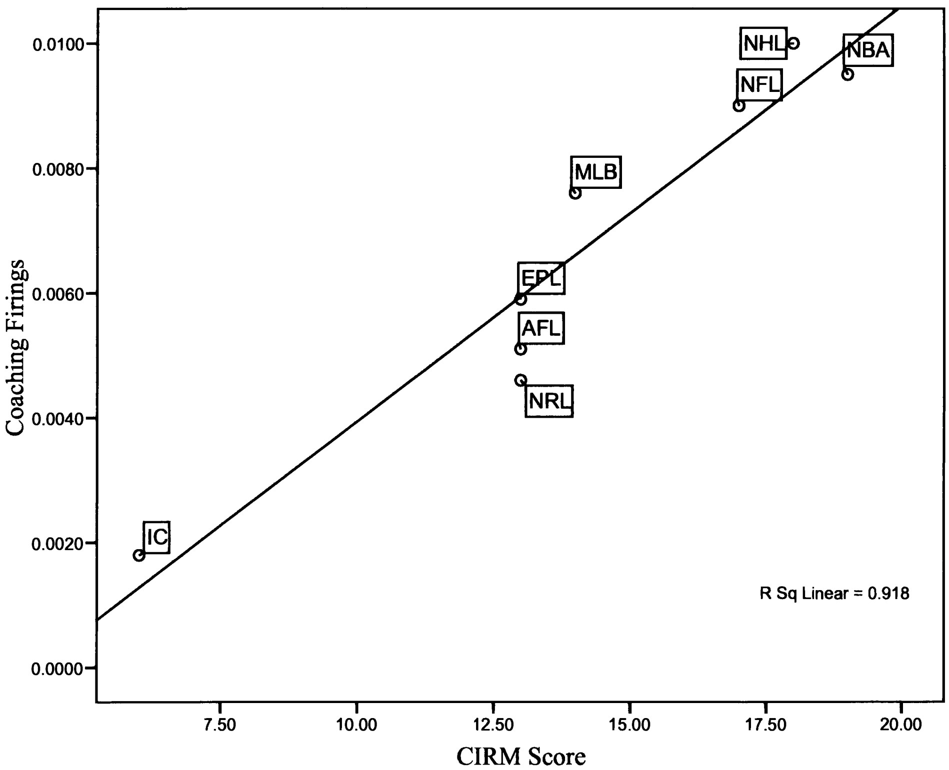


Figure 4: The relationship between coaching firings and the CIRM (Super 14 excluded)

CONCLUSIONS

Coaches are under increasing pressure to get the most out of players, and to obtain success for their club with some consistency. By determining the influence coaches have in-the-run we find that a strong positive correlation exists between coaching influence and coaching firings, with the exception of Super 14s. The rugby union super 14s level of firing is dramatically higher than the fit line, suggesting firings are too frequent given the level of influence coaches have in the code. The super 14s result is somewhat suspect given that the code is only a few years old in its current format. Removing the rugby union code from consideration, we find that the Australian rugby league competition fires somewhat lower number of coaches than the fit line, and MLB/NHL somewhat higher than the fit line. Overall we determine that coach turnover and firings is linearly related to the level of influence coaches have on a game. Further investigation is needed into other potential factors, including the influence of performance, out-of-run communication and training, and financial positions of each club.

Acknowledgements

We wish to thank Marli van Schalkwan (Zimbabwe Cricket); English Cricket Board; Gus Logie (West Indian cricket great); and Alex de la Mar (Holland Cricket Board) for their assistance in data retrieval.

References

Audas, R., Dobson, S. and Goddard, J. (1997) Team performance and managerial change in the English Football League. *Economic Affairs,* **17:** 30-36.

Audas, R., Dobson, S. and Goddard, J. (2002) The impact of managerial change on team performance in professional sports. *Journal of Economics and Business,* **54:** 633-650.

Balduck, A. and Buelens, M. (2007) Does sacking the coach help or hinder the team in the short term? Evidence from Belgian soccer. Ghent University.

Bruinshoofd, A. and ter Weel, B. (2003) Managers to go? Performance dips reconsidered with evidence from Dutch football. *European Journal of Operational Research,* **148:** 233-246.

Fabianic, D. (1994) Managerial change and organizational effectiveness in Major League Baseball: Findings for eighties. *Journal of Sport Behavior,* **70:** 69-72.

Gamson, W. A. and Scotch, N. A. (1964) Scapegoating in baseball. *The American Journal of Sociology,* **70:** 69-72.

Koning, R. H. (2003) An econometric evaluation of the effect of firing a coach on team performance. *Applied Economics,* **35:** 555-564.

McTeer, W., White, P. G. and Persad, S. (1995) Manager/coach mid-season replacement and team performance in professional team sport. *Journal of Sport Behavior,* **18:** 58-68.

Myers, N., Feltz, D. L., Maier, K. S., Wolfe, E. W. and Reckase, M. D. (2006) Athletes' evaluations of their head coach's coaching competency. *Research Quarterly for Exercise and Sport,* **77:** 111-121.

Wood, R. (2008) The Role of the Coach in Sport. *Rob's Home of Fitness Testing*, www.topendsports.com.

GETTING TO THE RIGHT PLACE AT THE RIGHT TIME: TESTING THEORIES OF FLY BALL CATCHING IN VIRTUAL REALITY

(ABSTRACT ONLY)

Fink, Philip [1], Foo, Patrick [2] and Warren, William [3]

[1] Institute of Food, Nutrition, and Human Health, Massey University, NZ

[2] Department of Psychology, University of North Carolina at Asheville, USA

[3] Department of Cognitive and Linguistic Sciences, Brown University, USA

Abstract Submitted: 25 July 2008
Presentation accepted: 28 July 2008

How does an outfielder know where to run to catch a fly ball? The "outfielder problem" is a particularly challenging variant of the object interception problem, and carries implications for theories of perceptual control of action. It may seem obvious that the fielder must predict the landing point based on an internal model of the ball's trajectory, but two prevailing theories demonstrate that the fielder may simply be led to the right place at the right time by coupling visual information to movement without invoking an internal representation of the trajectory. In the Optical Acceleration Cancellation (OAC) theory, the fielder runs so that the tangent of the ball's elevation angle (tan α) increases at a constant rate, while keeping a lateral angle to the ball (ψ) constant. In the Linear Optical Trajectory (LOT) theory, the fielder runs so that the optical trajectory of the ball appears straight (i.e. in a constant direction γ). Whereas OAC predicts that the fielder's radial and tangential movements are independent, LOT predicts that they are linked. We tested these theories by using virtual reality to perturb the vertical motion of the ball in mid-trajectory. The results confirm the predictions of OAC that elevation angle velocity is kept constant, and that the fielder's radial and tangential movements are unlinked, casting doubt upon LOT. We present a simple model of fielder movement with independent control of radial and tangential velocity that reproduces these findings, and shows how fielders can catch fly balls without an internal model of the trajectory.

Keywords: Perception-Action, Virtual Reality, Modeling

MEASURING FUNDAMENTAL MOVEMENT ABILITIES IN CHILDREN AS THEY LEARN A GOLF PUTTING TASK

(ABSTRACT ONLY)

Maxwell, Jon [1], **Hammond, John** [2] and Masters, Rich [1]

[1] Institute of Human Performance, University of Hong Kong, HK

[2] Department of Sport, Coaching and Exercise Science, University of Lincoln, UK

Abstract Submitted: 29 January 2008
Presentation accepted: 28 March 2008

This experiment investigated relationships between children's fundamental movement ability and implicit/explicit learning conditions Explicit motor learning refers to increases in performance that are accompanied by an awareness of and ability to communicate details of the dynamics underlying performance. Explicit motor learning places a heavy load on working memory and it can be argued that the learner becomes reliant on the availability of memory due to habitual use. Implicit motor learning is the acquisition of a movement skill without associated explicit or declarative knowledge. Implicit learners are less aware of the underlying processes that characterise skilled performance. Importantly, motor performance in implicit learners places a lower load on memory and is less susceptible to disruption. Because there may be links between movement ability and cognitive ability, it is possible that reducing the load on memory, whilst acquiring skill, may help to enhance performance. It is also possible that implicit (errorless) learning environments may be beneficial for low ability children, compared with explicit (errorful) learning environments. Based on previous research, it was predicted that errorful learners would learn less when performing a primary motor task concurrently with a secondary task. The performance of errorless learners was expected to show no such negative effects. Additionally, high ability learners were expected to outperform low ability learners in the errorful condition only. Performance of high and low ability learners in the errorless condition was predicted to be equivalent. Finally, errorful learners were expected to report more technical rules about how they performed the motor task than errorless learners. The experiment was conducted in two parts; the first consisted of measuring a large group of children using a battery of movement tests. In the second part of the experiment, children whose fundamental movement ability measures were in the upper or lower quartiles were invited to attend a laboratory session to learn a golf putting skill. Low and high ability children were assigned to either an implicit or explicit learning condition. Learning consisted of 300 shots following errorless or errorful procedures. Following learning, all participants performed a 50 trial retention test, 50 trial secondary task transfer test, and a second 50 trial retention test. It was predicted that concurrent performance of a secondary tone-counting task would adversely affect the performance of explicit learners but not implicit learners. Additionally, the negative effects of a secondary task was expected to be more pronounced in low ability children following errorful learning conditions. The results supported these predictions; implicit learners were unaffected by the secondary task, as were high ability explicit learners. Low ability explicit learners suffered a drop in performance when required to perform the putting task and secondary task concurrently. Low ability errorful learners performed significantly poorer than the corresponding high ability group but only when required to perform the secondary task. The results suggest that implicit learning techniques can be efficacious, particularly for children with lower movement abilities.

Keywords: implicit learning, errorless learning, measurement of skill

TRENDS IN OLYMPIC AND COMMONWEALTH GAMES RECORDS FOR THROWING EVENTS

Hammond, John and Bishop, Dan

Department of Sport, Coaching and Exercise Science, University of Lincoln, UK

Paper Submitted for Review: 25 March 2008
Revision submitted and accepted: 18 June 2008

Abstract. Throwing events have been an integral part of the track and field program in both Olympic and Commonwealth Games since their inception. Most scientific studies of these events have concentrated on biomechanical analysis or the physical capacity requirements of the athletes. This paper examines and compares the trends over time of the gold medal results of three throwing events in the Olympics and the Commonwealth Games for male competitors. Data was collected from the 'Athletic's Almanac' web-site, for the period since inception of these games until the present day. The data was examined to identify the linear trends that exist for all three events. Similarities between results in the shot put and hammer of a steady increase in distances thrown over time were evident. Whereas, the discus records showed a steeper positive relationship over time. In addition, critical world-wide incidents, trends in social expectations and increase in sports science and technical knowledge were concluded to have an effect on the results of these athletic pursuits in terms of acceleration periods and plateaus of results. In summary, trends in general for records in these throwing events showed a steady rise from the outset of competition until the late 1960s to early 1970s. Since this period, there has been considerable tapering off of improvements in distances achieved in both the Olympic and the Commonwealth Games.

Keywords: shot put, discus, hammer throw

INTRODUCTION

Throwing events have been included in athletic tournaments dating back to the ancient Olympiads and more recently the highland games in Scotland which began in the early 19th century. With that background in the culture of sporting contests it is little wonder that throwing events would be included in the first modern Olympic Games and also then in the Commonwealth Games (beginning as the Empire Games). The weightier throwing implements of the shot put and hammer (7.26 kg) have direct derivations from the highland games and the more aerodynamic implements of javelin and discus can be tracked back to the Ancient Greek Olympics. Other distinctions have occurred within the four events, such as the hammer throw being the domain of males only until female competition was included in both games in more recent times. Another important distinction across the four events would be that, in general terms, the javelin can be seen to require distinctive physical capacities at the elite level that set it apart from the other three 'power' events. The former requiring a relatively less bulky and more athletic physique and the latter requiring enormous body dimensions and strength that nowadays matches weightlifters more closely than in the past (Dyson, 1977; Bloomfield et al., 1994; Norton et al., 1996).

Both technological change and globalisation have contributed to improving track and field records over the period (Munasinghe et al., 2001). Scientific studies of throwing events have been focussed mainly on technique or biomechanical and physical capacity requirements of the athletes. Linthorne (2001) investigated optimum release angle in the shot put and developed simple models of shot putting to explain the relations between release speed, height and angle in terms of the anthropometric and strength characteristics of the athlete. While Dapena et al. (2003) attempted to predict the distance a hammer would travel by tracking its centre of mass in conjunction with aerodynamic considerations.

There are a variety of reasons why the historical trends in records set might change and there have been some studies which have attempted to track and/or predict these. Dickwach, H. and Scheibe (1993) tracked performance developments in the throwing events and noted a stagnation and decline in performance levels in more modern days, that they attributed to more advanced doping controls during competitions, the introduction of out of competition dope testing and socio-political changes. This reasoning for the slowing of progress in throwing events was supported by Ueya (1992) in a paper taking a retrospective view of results in world championships. Other reasons put forward for change in trends in Olympic results include home advantage (Balmer et al., 2001), although this does not seem to be borne out by throwing events results. However, a more obscure and as yet unsubstantiated reason, is that of any cultural bias in the trends, ie. it is noticeable that there are a more than expected number of Irish-American (as indicated by the 'Irishness' of their surnames) champions and/or record holders in the hammer throw. In addition, there seems to be country groupings of athlete success in some throwing events that may well be over and above the norm for Olympic or World record success. This has yet to be investigated in any empirical studies but worthy of note.

APPROACH

This paper examines and compares the trends over time of the gold medal results of three throwing events in the Olympics and the Commonwealth Games, namely shot put, discus and javelin. Selection of these three events was due to anticipated similarities in results in these 'power' events for reasons outlined previously in this paper and that the javelin implement for men had changed substantially in 1986 and, therefore, tracking of those results for direct comparison was deemed inappropriate (Dickwach and Scheibe 1993). In addition, records for female competitors in the hammer throw are incomplete due to considerably delayed inclusion of this event in both games, so only male results were sought and tracked. Data was collected from the 'Athletic's Almanac' web-site, for the period since inception of these games until the present day. The accuracy of these data was corroborated using the Olympic.org (2004) and the CGF (2007) web-sites. The data was examined to identify the linear trends that exist for all three events. Given the nature of the data sets being retrospective rather than predictive, not all explanations for changes in trends were anticipated to be mathematical. Therefore, principles adapted from Kreighbaum and Barthels (1996) for interpretation of the data both by descriptive statistical and qualitative methods were adopted.

RESULTS AND DISCUSSION

Comparison of the trends between the Olympic and Commonwealth games within all three throwing events provides a clear indication as to the increases in athletic performance since the inception of both games. Figures 1, 2 and 3 show the trends in discus, hammer and shot put respectively. Both the discus throw and the shot put have shown constant increases in performance up to the late 1970's, the hammer throw has equally shown large increase in performance up to the late 1970's however the performance improvements over the first 40 years until the early 1970's was much more gradual.

After the late 1970's the rate of increase within athletic performance within the discus and the hammer throw tapers considerably. However, the shot put fails to show any marked increase within performance since this period, with the exception of the 1988 Olympics which showed a large increase from the 1984 games, since this point performance has shown a steady decline replicating similar performances to that of the late 70s and early 80s that has not been matched since. The performances within the shot put for the Commonwealth Games showed a marked decline in performance since the 1974 games up until the 1986 games where subsequent performances have increased back to the level of the 70s.

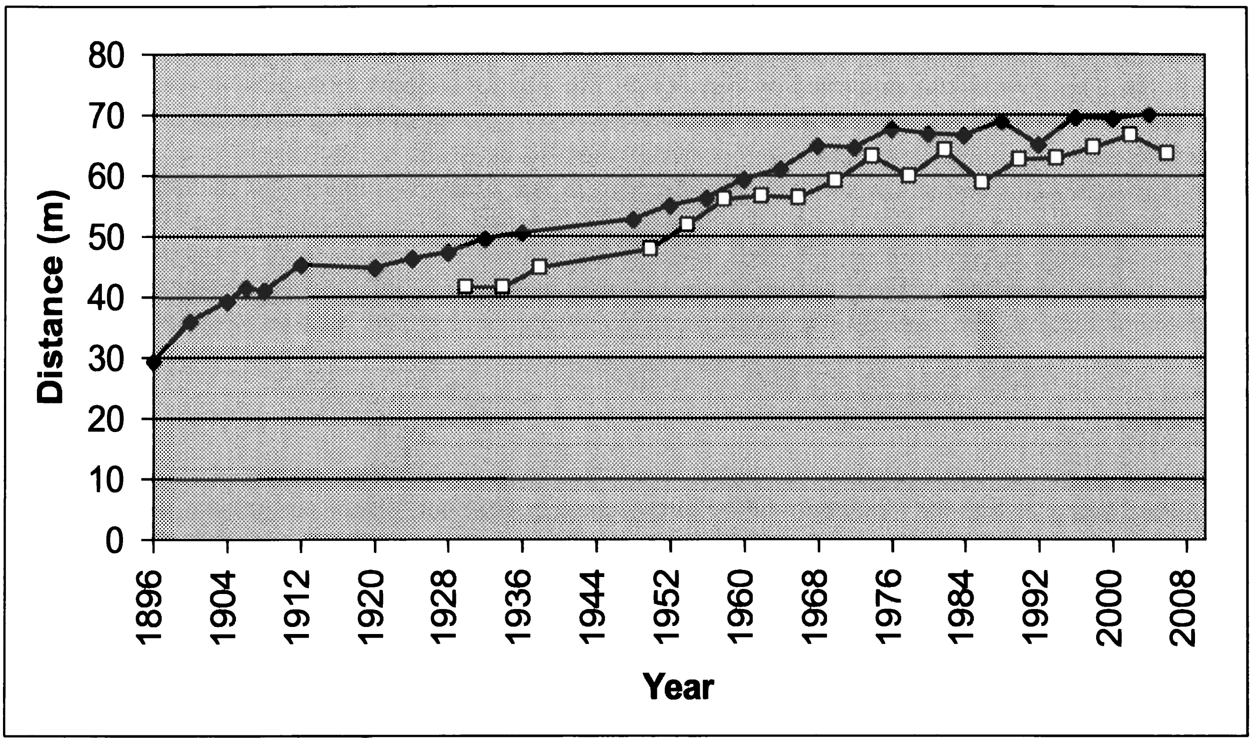


Figure 1: Trends in discus gold medal performances in Olympic (♦) and Commonwealth games (□).

The discipline that has shown the most improvement over the 70 year period was the hammer throw. The discus throw has shown similar levels of improvement and both disciplines have continued to show a steady rate of improvement to the present day. However, the shot put has shown a more restrained rate of improvement over the same time period, with approximately 15 – 20 % less increases demonstrated.

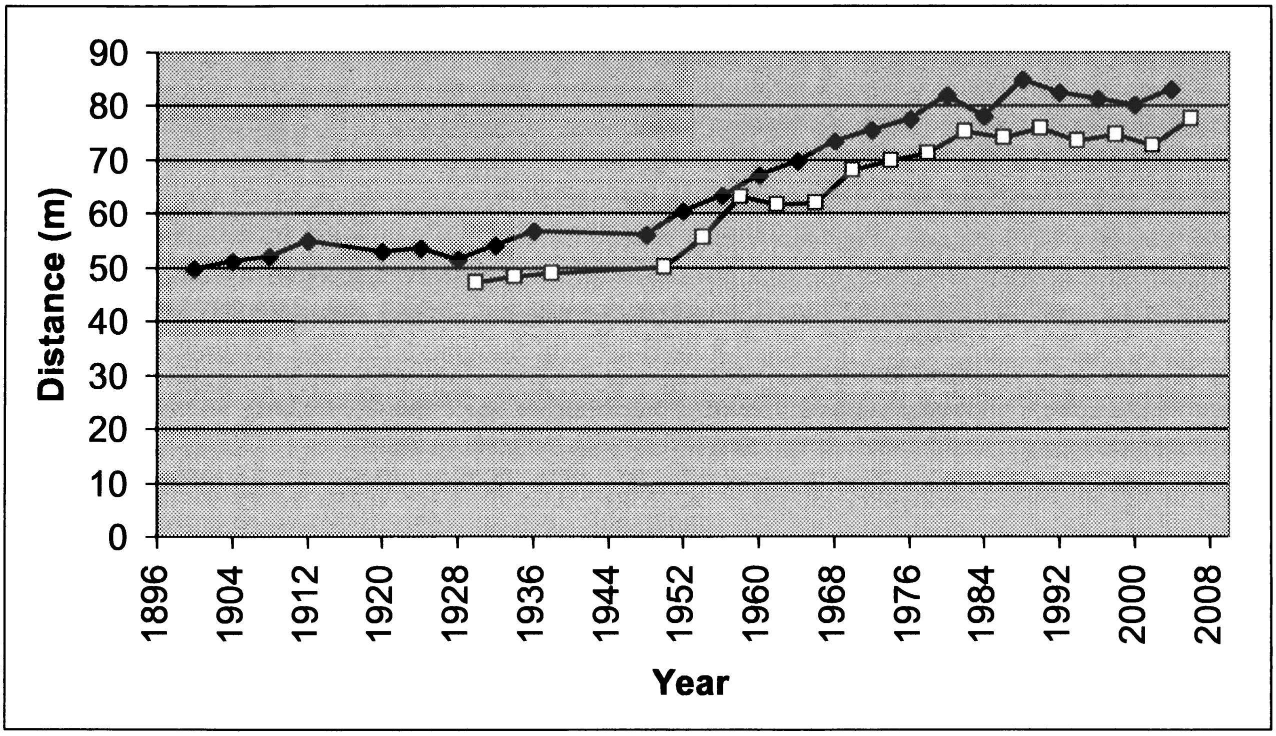


Figure 2: Trends in hammer gold medal performances in Olympic (♦) and Commonwealth games (□).

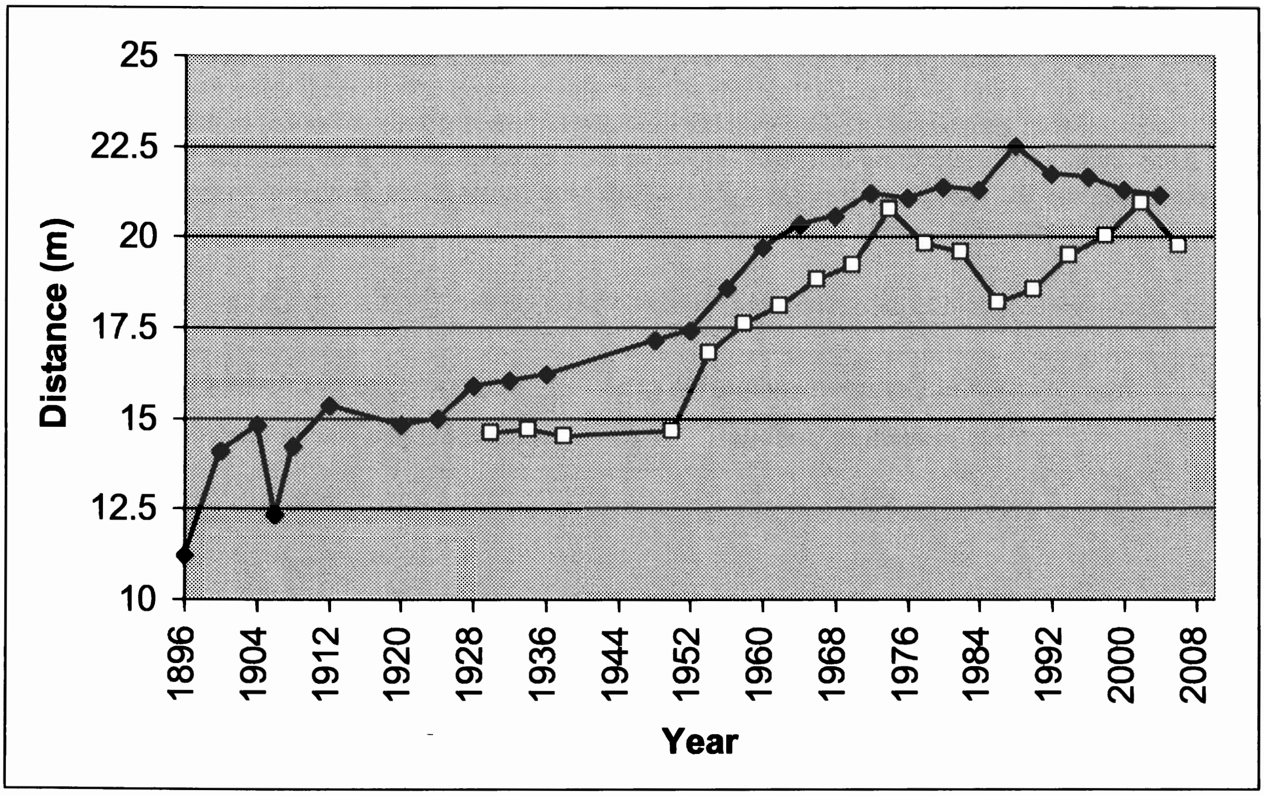


Figure 3: Trends in shot put gold medal performances in Olympic (♦) and Commonwealth games (□).

A line of best fit can be surmised from Figure 4 for each event linking both Commonwealth and Olympic distances and this approximates a flattened sigmoidal curve. With a slight gradient at the start of the recordings (Olympics), showing a plateau just before and at World War II, proceeding to a steeper rise after WW2 until a final plateau in more recent times. These 'fits and starts' are due to: world events, such as WW2; tighter monitoring of drug-assisted performance; and, improvements in technique and scientific support for elite athletes. Significant contributions made by changes to technique coincide with a steep increase or at least steady rise in performance in two events. Shot put distances increased markedly as a consequence of the introduction of the reverse glide ('Parry' technique) in the 1950s and less of an increase came from the rotational technique, introduced in 1976 and used alongside the glide (Dennis, 2008). Improvements in hammer throwing have been attributed to both a change in technique from two rotations in the circle to four (The Association of British Hammer Throwers, 2008) and to changes in the composition of the hammer head itself (Dennis, 2008). The Discus throw has not gone through such radical changes in technique and as a consequence has the flattest of the lines of best fit.

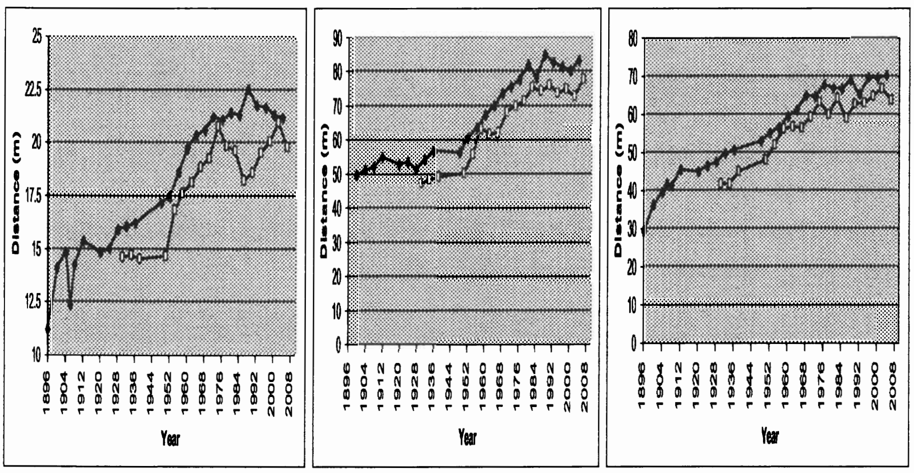


Figure 4: The three throwing events in comparison: left to right -shot, hammer, discus

Table 1 shows percentage improvement of the gold medal performances for the three events for the 1928–1938 (pre-war) period compared with 1996–2006 (modern) period. Within all three disciplines, since the early period of the Commonwealth Games there has been an increased improvement in distances thrown ranging from 2.9 to 5.8% more those distances thrown in the Olympics. Worthy of particular note is that the shot put increase was the greatest of these at 5.8% just after the considerable decline in performance in this event between 1976 and 1996.

Table 1. Improvement of gold medal distances for Hammer, Discus and Shot (1928 – 1938 against 1996 – 2006)[1].

Event	Improvement at Olympics comparing pre-war to modern periods	Improvement at Commonwealths comparing pre-war to modern periods
Hammer	52%	56%
Discus	49.1%	52%
Shot	33%	38.8%

From these improvements it is reasonable to suggest that performances within Commonwealth Games countries are becoming more comparable to the world class standard of the Olympics. This in part at least can be attributed to the standard and status of athletics increasing in Commonwealth countries hitherto impoverished in the field of sports performance and scientific support.

CONCLUSION

World-wide incidents, changes in socio-cultural expectation and increasing technical knowledge coupled with sports science support have had an effect on the trends in results of the throwing events reported in this paper. Records in these events showed a steady rise from the outset of competition, with a little plateauing at the period around WW2. The period after the World War showed the greatest range of change across the three events with tapering off in more recent times in terms of improvements in distances achieved. These trends were seen in both the Olympic Games and the Commonwealth Games.

References

Athletics Almanac (2007) *Welcome to Del's Athletics Almanac.* http://www.athletics.hitsites.de/index.php, (accessed Jan. 2008).

Balmer, N., Nevill, A. and Williams, M. (2001) Home advantage in the winter Olympics (1908-1998). *Journal of Sports Sciences.* **19:** 129-139.

Bloomfield, J., Ackland, T. and Elliott, B. (1994) *Applied Anatomy and Biomechanics in Sport.* Blackwell, Victoria.

CGF - Commonwealth Games Federation (2007) CG Records for athletics. *Commonwealth Games Records.* http://www.commonwealthgames.com/, (accessed Feb. 2008).

Dapena, J., Gutierrez-Davila, M., Soto, V. and Rojas, F. (2003) Prediction of distance in hammer throwing. *Journal of Sports Sciences.* **21:** 21-28.

[1] *To calculate the percentage improvement within the three disciplines the three games within the period of 1928 – 1938 were averaged against the three most recent games within the period of 1996 – 2006.*

Dennis, J. (2008) *The Throwing Events of Track and Field.* http://geocities.com/throws2000/. (accessed Feb. 2008).

Dickwach, H. and Scheibe (1993) Performance developments in the throwing events. *New Studies in Athletics.* **8:** 51-59.

Dyson, G. (1977) *The Mechanics of Athletics.* Hodder and Stoughton, London.

Kreighbaum, E. and Barthels, K. (1996) *Biomechanics: A Qualitative Approach for Studying Human Movement.* Allyn and Bacon, Boston.

Linthorne, N. (2001) Optimum release angle in the shot put. *Journal of Sports Sciences.* **19:** 359-372.

Munasinghe, L., O'Flaherty, B. and Danninger, S. (2001) Globalization and the rate of technological progress: what track and field records show. *Journal of Political Economy.* **109:** 1132-1149.

Norton and Olds (1996)

Norton, K., Olds, T., Olive, S. and Craig, N. (1996) Anthropometry and sports performance. In K. Norton and T. Olds (eds.) Anthropometrica. UNSW Press, Sydney, 287-352.

Terpstra, J. and Schauer, N. (2007) A simple random walk model for predicting track and field world records. *Journal of Quantitative Analysis in Sport.* **3:** 1-16.

Olympic.org (2004) Olympic record progression. *Official web-site of the Olympic Movement.* http://www.olympic.org/uk/utilities/reports/level2_uk.asp?HEAD2=10&HEAD1=5 (accessed Feb. 2008).

The Association of British Hammer Throwers (2008) *Help and Coaching Advice for Hammer Throwing.* http://www.olympic.org/uk/utilities/reports/level2_uk.asp?HEAD2=10&HEAD1=5 (accessed Feb. 2008).

Ueya, K. (1992) The men's throwing events. *New Studies in Athletics.* **7:** 57-65.

A RECURSION METHOD FOR EVALUATING THE MOMENTS OF A NESTED SCORING SYSTEM

Brown, Alan [1], Barnett, Tristan [2] and Pollard, Graham [3]

[1] Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Australia

[2] Faculty of Life and Social Sciences, Swinburne University of Technology, Australia

[3] Faculty of Information Sciences and Engineering, University of Canberra, Australia

Paper Submitted for Review: 16 March 2008
Revision submitted and accepted: 17 June 2008

Abstract. Nested scoring systems exist in several sports, e.g. tennis, badminton, table tennis, squash and volleyball. Recursive models are developed to study the influence that the scoring system has on the length of a match. Our examples are taken from the sport of tennis, which has rules covering points in a game, games in a set, and sets in a match. The assumption is made in the models that the transition probabilities from each state of the scoreboard depend only upon the score that has been reached, and is independent of the path by which that score was attained. This assumption enables us to show that, even in situations where the probability of a player winning a point on serve is not constant, there exists a simple calculus to determine higher moments of the number of points played at each level of the scoring system. A method is described for estimating the distribution of the total number of points in a match from its moments. This distribution is often multi-modal, and due care must be taken with its tails, as it is well known that a distribution is not uniquely determined by its moments.

Keywords: nested scoring systems, moment generating functions

INTRODUCTION

Nested scoring systems exist in several sports, e.g. tennis, badminton, table tennis, squash and volleyball. We are particularly interested in modelling situations where the probability of a player winning a point on serve is not constant.

Recursive models to evaluate the probability of winning a match under the various scoring systems are easy to develop provided sufficiently strong assumptions are made. Calculation with these models is straightforward when the assumption is made that the probability of the server winning a point is constant, and point outcomes are independent. However when attention is turned to the influence that a scoring system has on the length of a match, the analysis using recursive models becomes more difficult, especially when higher moments of the distributions are required. Monte Carlo methods have been used in the past, e.g. Pollard and Noble (2002, 2003, 2004) to consider cases in which the probability of winning a point is constant, and cases in which it is not constant. Our aim here is to show that methods can be developed to analyse these models using formulas that are both exact and fast to compute, and thus provide an alternative to simulation methods.

STOCHASTIC CALCULUS FOR MARKOV PROCESSES ON A LATTICE

Modelling processes on a graph

In tennis the match progresses as a player (or team) either wins or loses a point at each serve. The scoreboard accumulates the points won by each player, and at certain stages, depending on the rules, the scores accumulate at a higher level, and scores at the lower level are reset to zero. The rules may also impose various conditions on the rotation of serve between the players. Tennis has three higher levels known as game, set and match. We are interested as to how these levels nest within each other, and the construction the graph(s) for each level.

Modelling processes on a chain

Suppose the various states of the model form a chain. Denote the j^{th} state by E_j, and suppose the probability of being in the j^{th} state is $p\{E_j\} = p_j$.

Figure 1: Links in a chain

Consider the link in the chain from E_j to E_k, and denote the transition probability from state j to state k by

$$p\{(E_j, E_k)\} = q_{j,k}$$

The conditional probability of reaching state k after being in state j is then

$$p\{E_k \mid E_j\} = p_j\, q_{j,k}$$

Let $X_{j,k}$ be a random variable that is associated with the link in the chain from E_j to E_k. Denote its moment generating function by $M_{Xj,k}(t)$, where $M_{Xj,k}(t) = E[\exp(t\, X_{j,k})]$

We assume that all its moments exist. Let $S_k = S_j + X_{j,k}$ and assume S_j and $X_{j,k}$ are independent. Then using a standard lemma on the sum of independent random variables it is easy to obtain

$$M_{Sk}(t) = M_{Sj}(t)\, M_{Xj,k}(t)$$

We can combine probabilities and moment generating functions to define the conditional moment generating function (Cmgf) on a chain.

Cmgf for the transition step.

$$T_{Xj,k}(t) = q_{j,k}\ M_{Xj,k}(t)$$

This is obtained by combining the multiplicative factors for the step of the chain.

Cmgf for the conditional outcome $\{E_k \mid E_j\}$ on a chain.

$$C_{Sk}(t) = T_{Xj,k}(t)\, C_{Sj}(t)$$

This relation is a multiplicative form, and can be applied on the chain at each successive link.

$$C_{Sk}(t) = C_{S0}(t)\, T_{X0,1}(t)\, T_{X1,2}(t) \ldots\ldots\ldots\, T_{Xj,k}(t)$$

To initialise this relation we consider the initial state of the chain, E_0. If we set

$p_0 = 1$ and $S_0 = 0$, so that $M_{S0}(t) = 1$, which leads to $C_{S0}(t) = 1$.

MODELLING PROCESSES ON A LATTICE

A lattice is a generalization of a chain where there may be joins as well as branches at various nodes as in Figure 2.

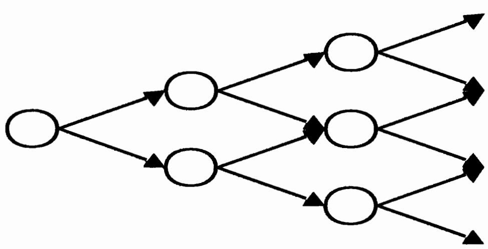


Figure 2: A lattice with 2-way branching and 2-way joining

It is straightforward to treat a branch exactly the same as a link on a chain. However the treatment at a join requires further consideration. We assume throughout that models of processes on a lattice have the following

A 1. Markov property: The transition probabilities from each state depend only upon that state, and not on the path by which the state was attained.

Suppose that there are n links that join at the k^{th} node. Since the links are disjoint, the probability of reaching state k, when the Markov property exists, is the sum of the transition probabilities weighted by the probability of being in the prior states.

$$p_k = p\{E_k\} = \sum_{j=1}^{n} p_j\, p\{E_k \mid E_j\} = \sum_{j=1}^{n} p_j\, q_{j,k}$$

Now considering the link from the j^{th} node, where

$M_{Sk}(t) = M_{Xj,k}(t)\, M_{Sj}(t)$ with probability $w_{j,k} = p_j\, q_{j,k} / p_k$, the probability weights satisfy

$$\sum_{j=1}^{n} w_{j,k} = \sum_{j=1}^{n} p_j\, q_{j,k}/ p_k = 1,$$

so it is quite easy to calculate the moment generating function $M_{Sk}(t)$ as an expectation

$$M_{Sk}(t) = \sum_{j=1}^{n} w_{j,k}\, M_{Xj,k}(t)\, M_{Sj}(t) = \sum_{j=1}^{n} p_j\, q_{j,k}\, M_{Xj,k}(t)\, M_{Sj}(t) / p_k,$$

which leads to

$$C_{Sk}(t) = \sum_{j=1}^{n} T_{Xj,k}(t)\, C_{Sj}(t)$$

These calculations for lattice models with the Markov property can be summarised by the following two rules:

(a) for each branch, multiply by the Cmgf of the transition to progress on each link,

(b) for each join, multiply by the Cmgf of the transition to progress on each link and add.

The flexibility of these lattice models and the simplicity of their calculus enable us to study a wide range of interesting cases. However the rules of tennis allow a countably infinite number of points to be played in a game. Hence we require additional assumptions before the lattice model is tractable. A sufficient assumption at the game level is the following:

A2. The probability of the server winning a point depends only upon the margin in points between the two players, and does not depend on the total number of points played.

A similar problem arises in an advantage set. A sufficient assumption at the set level is:

A3. The probability of the server winning a game depends only upon the margin in games between the two players, and does not depend on the total number of games played.

The two assumptions A2 and A3 are sufficient to restrict to a finite number the distinct points that must be modelled.

Often an additional assumption is added at the match level:

A4. The probability of the server winning a set depends only upon the margin in sets between the two players, and does not depend on the total number of sets played.

Assumption A4 is appended to the other assumptions purely for computational convenience.

APPLICATIONS OF THE STOCHASTIC CALCULUS TO TENNIS

A tennis match

The example given here is for a tennis match consisting of best-of-3 advantage sets. Many other variations are possible; nevertheless this example is sufficient to illustrate the ease with which nested levels of scoring can be handled in tennis, even when the probability of the players winning a point on serve is not constant.

Denote the two players by A and B. Let p_A and q_A be the probability of a server A winning and losing a point on serve respectively, where $p_A + q_A = 1$.

Let p_B and q_B be the corresponding symbols for server B. We will ignore stating this correspondence for the many other symbols that we introduce. To ensure that we have a Markov process, we assume throughout that both p_A and p_B depend only upon the scoreboard at the time, and are independent of the path by which that score was attained. We do not assume that p_A and p_B are constant, except where explicitly stated. We note however that our numerical tables are restricted to cases where p_A and p_B are constant, for brevity.

Modelling a point on serve

Assume player A is the server for a point. The scoreboard takes a two-way branch when a point has been played. The moment generating function for a single point is e^t. The Cmgf for a winning step in the game when the point scores are (a,b) is

$T(w,t|g,A,a,b) = p_A e^t$ and the Cmgf for a losing step is $T(l,t|g,A,a,b) = q_A e^t$

where p_A and q_A may themselves depend upon the scoreboard. In this notation, g indicates we are working at the game level, A indicates the server, whilst a and b indicate the respective point scores before the current point is played.

A classical or advantage game

Assume player A is the server in a game. Denote the probability that the point score in a game for server A reaches (a, b) by P(a,b|g,A). Denote the conditional moment generating function for the number of points played when reaching this score by C(a,b,t|g,A). Note that 0,1,2,3 correspond to love, 15,30, 40,.

The initial value is $C(0,0,t|g,A) = 1$,

and the forward recursions are given by

$C(a,b,t|g,A) = T(w,t|g,A,a-1,b)\, C(a-1,b,t|g,A) + T(l, t|g,A,a,b-1)\, C(a,b-1,t|g,A)$ for $0 < a \le 3, 0 < b \le 3$,
$C(a,b,t|g,A) = T(w,t|g,A,a-1,b)\, C(a-1,b,t|g,A)$ for $b = 0, 0 < a \le 3$ and for $a = 4, 0 \le b \le 2$,
$C(a,b,t|g,A) = T(l,t|g,A,a,b-1)\, C(a,b-1,t|g,A)$ for $a = 0, 0 < b \le 3$ and for $b = 4, 0 \le a \le 2$.

Player A wins the game when the point score reaches $a = 4, 0 \le b \le 2$, and player B wins the game when the point score reaches $b = 4, 0 \le a \le 2$. However when the score reaches (3, 3), also known as deuce, the game continues until one player is two points ahead, and this player wins the game. This rule permits the game to go on for an indefinite number of points. The forward recursions after deuce are given, for $n \ge 3$ by

$$C(n+1,n,t|g,A) = T(w,t|g,A,n,n)\, C(n,n,t|g,A)$$
$$C(n,n+1,t|g,A) = T(l,t|g,A,n,n)\, C(n,n,t|g,A)$$
$$C(n+2,n,t|g,A) = T(w,t|g,A,n+1,n)\, C(n+1,n,t|g,A)$$
$$C(n,n+2,t|g,A) = T(l,t|g,A,n,n+1)\, C(n,n+1,t|g,A)$$
$$C(n+1,n+1,t|g,A) = T(w,t|g,A,n,n+1)\, C(n,n+1,t|g,A) + T(l,t|g,A,n+1,n)\, C(n+1,n,t|g,A)$$

Let c and d denote the respective scores in games at the start of a game. The join of the instances where player A wins the game whilst serving is just one of four possible cases; the other cases are A losing whilst serving, B winning whilst serving and B losing whilst serving. The Cmgf for the number of points played when player A wins the game whilst serving is

$$T(w,t|s,A,c,d) = C(4,0,t|g,A) + C(4,1,t|g,A) + C(4,2,t|g,A) + C(5,3,t|g,A) \dots .$$

Two points are played between consecutive deuces. It follows that, for $n \geq 3$,

$$C(n+1,n+1,t|g,A) = R(t|g,A)\, C(n,n,t|g,A)$$

where $R(t|g,A) = T(w,t|g,A,n,n+1)\, T(l,t|g,A,n,n) + T(l,t|g,A,n+1,n)\, T(w,t|g,A,n,n)$

Assumptions A1 and A2 are sufficient to ensure that R(t|g,A) is does not depend on n. Likewise 2 points are played to win after a deuce, so for $n \geq 3$: $C(n+2,n,t|g,A) = T(w,t|g,A,n+1,n)\, T(w,t|g,A,n,n)\, C(n,n,t|g,A)$.

These relations are used to obtain a closed form of the Cmgf for the transition step in a set as

$$T(w,t|s,A,c,d)$$

$$= C(4,0,t|g,A) + C(4,1,t|g,A) + C(4,2,t|g,A) + T(w,t|g,A,4,3)\, T(w,t|g,A,3,3)\, C(3,3,t|g,A) /(1 - R(t|g,A))$$

This closed form deals with convergence, and greatly assists the speed of computation. In the special case where p_A is constant it is easy to recover standard results such as

$$T(w,t|s,A,c,d) = p_A^4 e^{4t} + 4\, p_A^4 q_A e^{5t} + 10\, p_A^4 q_A^2 e^{6t} + 20\, p_A^5 q_A^3 e^{8t} / (1 - 2p_A q_A e^{2t})$$

when $0 < p_A < 1$. In general, where the closed form is not so tractable, numerical calculations using the recursive relations are usually carried through up to the fourth moment.

Table 1: Cmgfs for a game, showing first three coefficients, when $p_A = 0.65$, $p_B = 0.60$.

A serve win	A serve lose	B serve lose	B serve win	all games
0.8296	0.1704	0.2643	0.7357	1.0000
5.0409	1.1473	1.8002	4.6840	6.3362
17.5823	4.3858	7.0574	17.3191	23.1723

The advantage set

In an advantage set the serve alternates each successive game. Denote the Cmgf for the number of points played in a winning game for player A by T(w,t|s,A,c,d) when A is serving, and by T(l,t|s,B,c,d) when B is serving. Denote the probability that the game score in a set with A serving first reaches (c, d) by P(c,d|s,A). Denote the conditional moment generating function for the number of points played when reaching this score by C(c,d,t|s,A).

Assume player A is the server *in the first game* of the set. The initial value is $C(0,0,t|s,A) = 1$.

When $c+d = 1 \bmod 2$, player A has just served, so the forward recursions are given by

$C(c,d,t|s,A) = T(w,t|s,A,c-1,d)\, C(c-1,d,t|s,A) + T(l,t|s,A,c,d-1)\, C(c,d-1,t|s,A)$ for $0 < c \leq 5$, $0 < d \leq 5$

$C(c,d,t|s,A) = T(w,t|s,A,c-1,d)\, C(c-1,d,t|s,A)$ for $d = 0$, $0 < c \leq 5$ and for $c = 6$, $0 \leq d \leq 3$, and

$C(c,d,t|s,A) = T(l,t|s,A,c,d-1)\, C(c,d-1,t|s,A)$ for $c = 0$, $0 < d \leq 5$ and for $d = 6$, $0 \leq c \leq 3$.

When $c+d = 0 \bmod 2$, player B has just served, so the forward recursions are now given by

$C(c,d,t|s,A) = T(l,t|s,B,c-1,d)\, C(c-1,d,t|s,A) + T(w,t|s,B,c,d-1)\, C(c,d-1,t|s,A)$ for $0 < c \leq 5$, $0 < d \leq 5$

$C(c,d,t|s,A) = T(l,t|s,B,c-1,d)\, C(c-1,d,t|s,A)$ for $d = 0$, $0 < c \leq 4$ and for $c = 6$, $0 \leq d \leq 4$, and

$C(c,d,t|s,A) = T(w,t|s,B,c,d-1)\, C(c,d-1,t|s,A)$ for $c = 0$, $0 < d \leq 4$ and for $d = 6$, $0 \leq c \leq 4$.

A player wins the set when his game score is two ahead after at least six games have been played. This rule permits the advantage set to go on for an indefinite number of games. When the score is level at (5, 5), the set continues until one player is two games ahead, and this player wins the set. If player A serves first, the forward recursions after the scores are level are given, for $n \geq 5$, by

$$C(n+1,n,t|s,A) = T(w,t|s,A,n,n)\ C(n,n,t|s,A),$$

$$C(n,n+1,t|s,A) = T(l,t|s,A,n,n)\ C(n,n,t|s,A),$$

$$C(n+2,n,t|s,A) = T(l,t|s,B,n+1,n)\ C(n+1,n,t|s,A),$$

$$C(n,n+2,t|s,A) = T(w,t|s,B,n,n+1)\ C(n,n+1,t|s,A), \text{ and}$$

$$C(n+1,n+1,t|s,A) = T(l,t|s,B,n,n+1)\ C(n,n+1,t|s,A) + T(w,t|s,B,n+1,n)\ C(n+1,n,t|s,A).$$

Let e and f denote the respective scores in sets. At the match level there are eight cases to consider. Let v and o denote whether an even and odd number of games is played in the set respectively. The Cmgf for the number of points in a set when A is serving first, and A wins in an odd number of games, is given by

$$T(w,o,t|m,A,e,f) = C(6,1,t|s,A) + C(6,3,t|s,A)$$

The Cmgf for the number of points in a set when A is serving first, and A wins in an even number of games, is given by $T(w,v,t|m,A,e,f) = C(6,0,t|s,A) + C(6,2,t|s,A) + C(6,4,t|s,A) + C(7,5,t|s,A) + \ldots$.

Since two games are played between consecutive level scores in a set it follows that

$$C(n+1,n+1, t|s,A) = R(t|s,A)\ C(n,n, t|s,A) \quad \text{for } n \geq 5,$$

where $R(t|s,A) = T(w,t|s,A,n,n)\ T(w,t|s,B,n+1,n) + T(l,t|s,A,n,n)\ T(l,t|s,B,n,n+1)$.

It is easy to check that $R(0|s,A) \neq 1$ provided $0 < p_A < 1$ and $0 < p_B < 1$. Using

$$C(n+2,n, t|s,A) = T(w,t|s,A,n,n)\ T(l,t|s,B,n+1,n)\ C(n,n, t|s,A) \quad \text{for } n \geq 5,$$

we obtain a closed form for a transition step in a match as

$$T(w,v,t|m,A,e,f) = C(6,0,t|s,A) + C(6,2,t|s,A) + C(6,4,t|s,A) + C(5,5,t|s,A)\ T(w,t|s,A,5,5)\ T(l,t|s,B,6,5)/(1 - R(t|s,A)).$$

The two cases where A serves first in a set but loses the set follow a similar pattern. There are another four cases in a set where B serves first, but we omit the details.

A sv fst win/ A next	A sv fst win/ B next	A sv fst lose/ A next	A sv fst lose/ B next
0.3513	0.3226	0.2922	0.0339
27.8935	17.3168	21.8576	1.8884
1287.6826	477.6359	929.6579	53.8005

Table 2: Some Cmgfs for a set, showing first three coefficients, when $p_A = 0.65$, $p_B = 0.60$.

The best-of-3 sets match

A rotation of serve occurs between the two players for successive games throughout the match. Denote the probability that the set score in a match reaches (e, f) with A to serve in the next set by P(e,f,A|m). Denote the conditional moment generating function for the number of points played when reaching this score with A to serve by C(e,f,A,t|m).

The two initial values are C(0,0,A,t|m) = 1, and C(0,0,B,t|m) = 0.

and the forward recursions are given by

$$C(e,f,A,t|m) = T(w,v,t|s,A,e-1,f)\ C(e-1,f,A,t|m) + T(l,o,t|s,B,e-1,f)\ C(e-1,f,B,t|m)$$
for e = 1 or 2, f = 0 and e = 2, f = 1,

$$C(e,f,A,t|m) = T(l,v,t|s,A,e,f-1)\ C(e,f-1,A,t|m) + T(w,o,t|s,B,e,f-1)\ C(e,f-1,B,t|m)$$
for e = 0, f = 1 and e = 1, f = 2,

$$C(e,f,B,t|m) = T(w,o,t|s,A,e-1,f)\ C(e-1,f,A,t|m) + T(l,v,t|s,B,e-1,f)\ C(e-1,f,B,t|m)$$
for e=1, f = 0 and e = 2, f = 1, and

$$C(e,f,B,t|m) = T(l,o,t|s,A,e,f-1)\ C(e,f-1,A,t|m) + T(w,v,t|s,B,e,f-1)\ C(e,f-1,B,t|m)$$
for e = 0, f = 1 or 2 and e = 1, f = 2.

Otherwise for e = 1, f = 1 we have the pair of recursions which encompass 4-way joins.

$$C(e,f,A,t|m) = T(w,v,t|s,A,e-1,f)\ C(e-1,f,A,t|m) + T(l,o,t|s,B,e-1,f)\ C(e-1,f,B,t|m) + T(l,v,t|s,A,e,f-1)\ C(e,f-1,A,t|m) + T(w,o,t|s,B,e,f-1)\ C(e,f-1,B,t|m),$$

and

$$C(e,f,B,t|m) = T(w,o,t|s,A,e-1,f)\ C(e-1,f,A,t|m) + T(l,v,t|s,B,e-1,f)\ C(e-1,f,B,t|m) + T(l,o,t|s,A,e,f-1)\ C(e,f-1,A,t|m) + T(w,v,t|s,B,e,f-1)\ C(e,f-1,B,t|m).$$

We can study the case where player B serves in the first game of the match by reversing the pair of initial values, or the effect of randomizing the toss by setting both initial values to 0.5.

2 set match A sv fst win	2 set match A sv fst lose	3 set match A sv fst win	3 set match A sv fst lose
0.4402	0.1132	0.3009	0.1456
55.4046	15.3568	59.6526	29.3918
3595.7443	1067.3448	6096.1103	3057.9805

Table 3. Some Cmgfs for a match, showing first three coefficients, when $p_A= 0.65$, $p_B= 0.60$.

RECOVERING DISTRIBUTIONS FROM MOMENTS

We have shown above that using lattice models with the Markov property and a few other modest assumptions it is feasible to calculate moments of the number of points in a match.

We can write $C(w,t|m,A,0,0) = P(w|m,A,0,0)\ M_{Sw|m,A}(t)$ which enables us to recover both the probability of a player A winning, and the moment generating function for the total number of points played in such a match. This may be put to practical use (Barnett and Pollard,2006). The estimation of the distribution of the total number of points in a best-of-3 sets match requires some care, since it may be bi-modal. It is well known that a distribution is not uniquely determined by its moments. To overcome these difficulties, the distributions for matches requiring 2 sets and those requiring 3 sets to complete were estimated separately. This can be done using the first four moments of each distribution and the Normal Power approximation (Pesonen, 1975). It is inappropriate to use this approximation in conjunction with the statistics for "All matches" as reported in the Table 4. Further examples of estimating the distribution can be found in Pollard, Barnett, Brown and Pollard, (2007).

Table 4: Some statistics for a match, when $p_A= 0.65$, $p_B= 0.60$.

Type of match	2 set	3 set	All matches
Probability	55.35%	44.65%	100.00%
Mean no of points	128.54	199.43	160.20
St dev no of points	22.13	35.16	45.44

APPLICATIONS

The rules for scoring systems differ both within and between various sports. We have taken examples from tennis, which is a sport with a rich set of rules covering points in a game, games in a set, and sets in a match. This variety should be sufficient to demonstrate the generality of our methodology. This stochastic calculus using conditional moment generating functions has a wide range of applications to other sports e.g. tennis, badminton, table tennis, squash and volleyball. The common feature is that the event space includes the scoreboard for each player (or team) and the server. The rules for incrementing the scoreboard and rotating the server vary both within and between sports.

References

Barnett, T and Pollard, G. (2006) Reducing injuries by substantially decreasing the likelihood of long tennis matches. *Medicine and science in tennis*. **11(2)**: 10-11.

Pesonen, E. (1975) NP-technique as a tool in decision making. *ASTIN Bulletin*. **8(3)**: 359-363.

Pollard, G and Noble, K. (2002) The characteristics of some new scoring systems in tennis. In G. Cohen and T. Langtry (eds.) *Proceedings of the sixth Australian Conference on Mathematics and Computers in Sport*. University of Technology, Sydney, 221-226.

Pollard, G. and Noble, K. (2003) The effect of changing the assumption that the probability of winning a point in tennis is constant. In S. Miller (ed.) *Proceedings of Tennis Science and Technology 2*. London, 341-346.

Pollard, G. and Noble, K. (2004) The effect of having correlated point outcomes in tennis. In R. Morton and S. Ganesalingam (eds.) *Proceedings of the seventh Australasian Conference on Mathematics and Computers in Sport*. Massey University, New Zealand, 266-268.

Pollard, G.H., Barnett, T., Brown, A. and Pollard, G.N. (2007) Some alternative men's doubles scoring systems. In S. Miller and J. Capel-Davies (eds.) *Proceedings of Tennis Science and Technology 3*. Roehampton, London, 301-309.

MOMENT GENERATING FUNCTION FOR A TENNIS MATCH

Pollard, Geoff [1] and **Pollard, Graham** [2]

[1] Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia

[2] Faculty of Information Sciences and Engineering, University of Canberra, Canberra, Australia

Paper Submitted for Review: 27 March 2008
Revision submitted and accepted: 11 June 2008

Abstract. The moment generating function of the number of points played using a particular scoring system can be used to estimate the actual distribution of the number of points played. Estimates can also be made of the tails of the distribution, so that an estimate can be made of the proportion of time very long matches will occur. In this paper an explicit expression for the moment generating function of the number of points in a game of tennis conditional on the server winning (and conditional on the server losing) is derived. These expressions are used to determine an explicit expression for the moment generating function of the number of points in a set of tennis conditional on player A winning (and conditional on player A losing) in an odd (and in an even) number of games. Finally, an explicit expression for the moment generating function of the number of points in a best of three advantage sets match of tennis is determined. The Normal Power Approximation and the first four moments from the moment generating function can be used to estimate the distribution of the number of points, which can then be used to estimate some useful statistics for a tennis match. Importantly, the methods in this paper can be used to determine the moment generating function of the number of points for any tennis scoring system, or indeed to determine the moment generating function for nested scoring sytems for many other sports.

Keywords: moment generating function of a nested scoring system, distribution of the number of points in tennis

INTRODUCTION

The moment generating function (mgf) of the number of points in (or the duration of) a tennis match is a very useful function, as it can be used to obtain a range of statistics or characteristics for that scoring system. For example, the mgf can be used to determine the mean, the variance, measures of skewness, and the higher moments of the distribution. A method for estimating the distribution itself using the Normal Power Approximation and the first four moments of a distribution, has been demonstrated (Brown et al., 2008; Pollard et al., 2007). Characteristics such as the tails of the distribution, for example the 2% and 98% points in the cumulative distribution, can then be estimated.

In this paper the mgf of the duration of a single game of tennis conditional on the server winning the game is determined, as is the mgf of a single game conditional on the server losing. These conditional mgfs are then combined in order to determine the mgf of the duration of a set conditional on player A (the better player) winning in an odd (and in an even) number of games. Explicit expressions for the mgfs of the duration of a set conditional on player A losing in an odd (and in an even) number of games are also given.

These conditional mgfs for a set of tennis are then combined to determine an explicit expression for the mgf of a best of three advantage sets match of tennis.

The analytical methods used in this paper can be used to determine the mgf for the duration of any tennis scoring system. They can also be used to determine the mgf for the duration of the scoring system of a wide range of nested scoring systems for other sports.

METHODS

The tennis scoring system is a triple-nested one. Points are played to determine the winner of a game, games are played to determine the winner of a set, and sets are played to determine the winner of a match.

The moment generating function for the duration of a game of tennis

Suppose player A has a probability p_a ($q_a = 1 - p_a$) of winning (losing) a point on his service. Then, assuming points are independent, the probability player A wins his service game is given by

$$ {}_aP_a = \sum_{i=0}^{2} {}^{3+i}C_i p_a^{\,4} q_a^{\,i} + 20 p_a^{\,3} q_a^{\,3} \sum_{i=0}^{\infty} p_a^{\,2} (2 p_a q_a)^i , $$

and the probability player A loses one of his service games is given by ${}_aP_b = 1 - {}_aP_a$. The mgf for the the number of points of one of his/her service games conditional on player A winning the game is given by

$$ {}_aG_a(t) = (p_a^{\,4} e^{4t} \sum_{i=0}^{2} {}^{3+i}C_i (q_a e^t)^i + 20 p_a^{\,5} q_a^{\,3} e^{8t} (1 - 2 p_a q_a e^{2t})^{-1}) / {}_aP_a . $$

The mgf for the duration of one of his/her service games conditional on player B winning the game, denoted by ${}_aG_b(t)$, can be written down similarly.

Now suppose player B has a probability p_b ($q_b = 1 - p_b$) of winning (losing) a point on his/her service. Then, again assuming points are independent, the probability player B wins his/her service game, ${}_bP_b$, and the mgf for the duration of one of player B's service games conditional on player B winning the game, denoted by ${}_bG_b(t)$, can also be written down, as can the mgf for the duration of one of player B's service games conditional on player A winning the game, denoted by ${}_bG_a(t)$.

The moment generating function of the duration of an advantage set of tennis

An *advantage* set of tennis is considered in this section. However, the same methodology can be used to consider a tiebreak set. It is assumed that player A serves in the first game of the set. Using $P(k,i)$ to denote the probability player A wins a set by 6 games to k games and wins exactly i of his/her own service games in the process, and assuming games are independent, the probability player A wins a set by 6 games to k games (for k = 0, 2, or 4), is given by

$$ {}_aP_{6/k} = \sum_{i=(6-k)/2}^{(6+k)/2} P(k,i) \text{ ,where} $$

$$ P(k,i) = {}_bP_a b((6+k)/2, i, {}_aP_a) b((4+k)/2, 5-i, {}_bP_a) \text{ and } b(n,r,p) = (n!/(r!(n-r)!)) p^r (1-p)^{n-r} . $$

The relevant equations can also be written down for the cases when k = 1 or 3. Further, the cases when player A *loses* the set k games to 6 can be written down for k = 0, 2, or 4 and for k = 1 or 3.

Noting that the probability player A wins a set in an odd number of games, ${}_aP_{a,o}$, is equal to ${}_aP_{6/1} + {}_aP_{6/3}$, the mgf for the duration of a set in which player A served first and won in an odd number of games, ${}_aS_{a,o}(t)$, is given by

$$ {}_aS_{a,o}(t) = \sum_{k=1,3} \sum_{i=(7-k)/2}^{(7+k)/2} P(k,i) {}_aG_a(t)^i {}_aG_b(t)^{((7+k)/2-i)} {}_bG_b(t)^{((k-7)/2+i)} {}_bG_a(t)^{6-i} / {}_aP_{a,o} . $$

A corresponding expression can be written down for ${}_aS_{b,o}(t)$ for the case in which player A serves first and loses the set in an odd number of games.

The probability that the games' score reaches 5 games each, $P_{5/5}$, is given by

$$P_{5/5} = \sum_{i=0}^{5} P_{5/5}(i)\text{, where } P_{5/5}(i) = b(5,i,{}_aP_a)b(5,5-i,{}_bP_a)\text{,}$$

and $P_{5/5}(i)$ denotes the probability that the games' score reaches 5 games each with player A winning exactly i of his own service games. Now supposing $C(t)$ represents the mgf of the duration of a set up to the score 5/5, given that 5/5 was reached, we have

$$C(t) = (\sum_{i=0}^{5} P_{5/5}(i)\,{}_aG_a(t)^i\,{}_aG_b(t)^{5-i}\,{}_bG_b(t)^i\,{}_bG_a(t)^{5-i})/P_{5/5}\,.$$

The mgf for the duration of two consecutive games shared by Players A and B is given by

$$D(t) = ({}_aP_{ab}P_{ba}G_a(t)_bG_b(t) + {}_aP_{bb}P_{aa}G_b(t)_bG_a(t))/({}_aP_{ab}P_b + {}_aP_{bb}P_a)\,.$$

The probability that player A wins the set given that the games' score reached 5 games each, P_{ad}, is given by

$$P_{ad} = \sum_{j=0}^{\infty} P_j \text{ where } P_j = {}_aP_{ab}P_a({}_aP_{ab}P_b + {}_aP_{bb}P_a)^j\,.$$

The mgf for the duration of the set beyond the score 5/5 given that 5/5 was reached and player A won the set, is given by

$$E(t) = (\sum_{j=0}^{\infty} P_jD(t)^j\,{}_aG_a(t)_bG_a(t))/P_{ad}$$

Thus, the mgf for the duration of a set given player A served first and won the set in an even number of games is given by

$${}_aS_{a,e}(t) = (\sum_{k=0,2,4}\sum_{i=(6-k)/2}^{(6+k)/2} P(k,i)_aG_a(t)^i\,{}_aG_b(t)^{((6+k)/2)-i}\,{}_bG_b(t)^{((k-6)/2)+i}\,{}_bG_a(t)^{6-i} + P_{5/5}P_{ad}C(t)E(t))/{}_aP_{a,e}$$

where ${}_aP_{a,e}$ is the probability that player A wins a set in an even number of games, given that he/she served first in the set, and is given by ${}_aP_{6/0} + {}_aP_{6/2} + {}_aP_{6/4} + P_{5/5}\,P_{ad}$.

The mgf for the duration of a set in which player A served first and won, ${}_aS_a(t)$, is given by

$${}_aS_a(t) = {}_aP_{a,o\,a}S_{a,o}(t) + {}_aP_{a,e\,a}S_{a,e}(t)\,.$$

In the same way, the corresponding mgfs for the duration of a set in which player A served first and lost, ${}_aS_{b,o}(t)$, ${}_aS_{b,e}(t)$, and ${}_aS_b(t)$ can be written down (using an obvious notation).

Further, the mgf for the duration of a set in which player A served first, is given by

$${}_aS(t) = {}_aP_{as\,a}S_a(t) + {}_aP_{bs\,a}S_b(t)$$

where ${}_aP_{as}$ (equal to ${}_aP_{a,e}$ + ${}_aP_{a,o}$) and ${}_aP_{bs}$ (equal to ${}_aP_{b,e}$ + ${}_aP_{b,o}$) are the probabilities players A and B respectively win a set in which player A served first.

Thus far the formulae for seven mgfs for a set with player A serving first in the set have been noted. Correspondingly, the seven mgfs for the case in which player B serves first in the set, can be written down.

These fourteen mgfs are all that is needed to write down the mgf for the duration of a best of three advantage sets match of tennis.

The moment generating function for a best of three advantage sets match of tennis

The mgf for the duration of a best of three advantage sets of tennis can be written down by making use of the above fourteen mgfs. It is noted that in the third set of the match it is not necessary to keep track of who wins the set and whether an even or odd number of games is involved. Correspondingly, if the same person wins the second set as wins the first set, it is not necessary to keep track of whether the number of games in the second set is odd or even. In total, given that player A serves in the first game of the first set, the match can be described in one of twelve ways, giving the following mgf for the match.

$$
\begin{aligned}
{}_aM(t) = {} & {}_aP_{a,e\,a}P_{as\,a}S_{a,e}(t)_aS_a(t) + {}_aP_{a,e\,a}P_{b,e\,a}S_{a,e}(t)_aS_{b,e}(t)_aS(t) + {}_aP_{a,e\,a}P_{b,o\,a}S_{a,e}(t)_aS_{b,o}(t)_bS(t) \\
& + {}_aP_{a,o\,b}P_{as\,a}S_{a,o}(t)_bS_a(t) + {}_aP_{a,o\,b}P_{b,e\,a}S_{a,o}(t)_bS_{b,e}(t)_bS(t) + {}_aP_{a,o\,b}P_{b,o\,a}S_{a,o}(t)_bS_{b,o}(t)_aS(t) \\
& + {}_aP_{b,e\,a}P_{bs\,a}S_{b,e}(t)_aS_b(t) + {}_aP_{b,e\,a}P_{a,e\,a}S_{b,e}(t)_aS_{a,e}(t)_aS(t) + {}_aP_{b,e\,a}P_{a,o\,a}S_{b,e}(t)_aS_{a,o}(t)_bS(t) \\
& + {}_aP_{b,o\,b}P_{bs\,a}S_{b,o}(t)_bS_b(t) + {}_aP_{b,o\,b}P_{a,e\,a}S_{b,o}(t)_bS_{a,e}(t)_bS(t) + {}_aP_{b,o\,b}P_{a,o\,a}S_{b,o}(t)_bS_{a,o}(t)_aS(t).
\end{aligned}
$$

The mgf for the duration of a match given that player B serves first in the first set, ${}_bM(t)$, can be written down correspondingly. Thus, assuming the players toss a coin to determine who serves first, the mgf for the match, M(t), is given by

$$M(t) = 0.5\,{}_aM(t) + 0.5\,{}_bM(t).$$

The mgf for a best of five set match of tennis can be written down similarly, although the expression corresponding to that for ${}_aM(t)$ above, is considerably more complicated.

CONCLUSIONS

Explicit expressions for the moment generating functions of the duration of a game, a set and a match, conditional on one player winning, and conditional on that player losing have been derived. The unconditional moment generating functions can also be written down. These functions can be used to derive several useful characteristics of the distribution of the duration of a match of tennis. For example, the mean, the standard deviation, measures of skewness, tail-point values of the distribution can be evaluated.

Importantly, the methods used in this paper can be applied to a wide range of other nested sports scoring systems, and indeed, to other sequential processes that are directly comparable to sports scoring systems.

References

Brown, A., Barnett, T. and Pollard, G. (2008) A recursion method for evaluating the moments of a nested scoring system. In J. Hammond (ed.) *Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM) Coolangatta, Queensland, Australia (in press).

Brown, A. , Barnett, T. Pollard, G. Lisle, I. and Pollard, G. (2008) The characteristics of various men's doubles scoring systems. In J. Hammond (ed.) *Proceedings of the Ninth Australiasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM) Coolangatta, Queensland, Australia (in press).

Pollard, G. H., Barnett, T., Brown, A. and Pollard, G. N. (2007) Some alternative men's doubles scoring systems. In S. Miller and J. Capel-Davies (eds.) Tennis Science and Technology 3, International Tennis Federation Licensing (UK) Ltd, Roehampton, London SW15 5XZ.

FOUR BALL BEST BALL 1

Pollard, Geoff [1] and Pollard, Graham [2]

[1] Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia

[2] Faculty of Information Sciences and Engineering, University of Canberra, Australia

Paper Submitted for Review: 27 March 2008
Revision submitted and accepted: 11 June 2008

Abstract. In this paper a four-ball-best-ball (4BBB) model for pairs of golf players is set up. The 4BBB match-play scoring system is seen to satisfy a basic requirement of fairness. It is shown that it is not strictly possible to rate individual players as 4BBB players. However, a (reasonably broad) class of individual players is identified such that it is possible to rate them individually as 4BBB players. The capacity of an individual to play birdies is seen to be a very important determinant in being a successful member of a 4BBB pair, but there are other minor factors as well. Consideration is given to equal and unequal 4BBB pairs. The transitive law is seen to apply for 4BBB pairs. Thus, if pair A is better than pair B, and pair B is better than pair C, then pair A must be better than pair C. Correspondingly, if pair A is equal to pair B, and pair B is equal to pair C, then pair A is equal to pair C. Consideration is given to some strategic issues in 4BBB match-play golf. For example, the conditions under which a player should take a greater risk and have a higher probability of obtaining a bogey in order to achieve a higher probability of scoring a birdie, are determined. Also, the conditions under which a player, noting that his partner is about to have a 'bad' hole and score only a par or a bogey, should 'play safe', are determined. Thirdly, players who can interact in certain ways are seen to have an advantage over those pairs that cannot do this. Finally, one pair's optimal strategy when they see that their opponents are about to score a par or a bogey, but not a birdie, is analyzed.

Keywords: Strategies for four-ball-best-ball golf, selecting pairs of four-ball-best-ball players.

INTRODUCTION

In four-ball-best-ball match-play golf, two players (A and B) play as a pair against two other players (C and D) who also play as a pair. All players play in the usual way. On the first hole, pair (A, B)'s score is the minimum number of strokes A or B took to complete the hole. Correspondingly, pair (C, D)'s score on the first hole is the minimum number of strokes C or D took to complete the hole. One pair wins the hole if their minimum score is less than the other pair's minimum score, and 1 is added to their match score. The match ceases as soon as one pair has an unbeatable lead and/or 18 holes have been completed. If the match is tied or 'squared' after the 18th hole has been played, more holes may be played until a winner is determined by a lead by 1. This is called a 'play-off'.

In this paper several questions are addressed. What is a useful model for studying 4BBB golf? Is the scoring system fair? Can individual players be ranked even though they are playing as pairs? Can pairs of players be ranked or can cyclical relationships (pair A is better than pair B, who is better than pair C, who is better than pair A) exist between several pairs? Is it possible for pair A to be stronger than pair B at 4BBB golf, but weaker at other forms of the game? Are birdies particularly important in winning at 4BBB golf? Can two players make strategic decisions to interact in a useful way on a particular hole, and on any hole in general? Can such decisions be usefully made both statically and dynamically?

In a very recent paper, Hurley (2007) studied how a given set of players might be combined into pairs so as to form a very effective team against an identical opposing team. This problem is studied in a second paper by the authors (Pollard & Pollard, 2008).

METHODS

The four-ball-best-ball model

For PGA players playing 4BBB match-play, a pair's score on a hole is typically a birdie (-1), a par (0) or a bogey (+1). For top amateurs and A-grade players, quite often playing on easier courses, the same is true. Even for average players, typically playing on easier courses, this is also true. Thus, a useful model for 4BBB match-play golf is to denote pair A by $\{a_{-1}, a_0, a_{+1}\}$ where a_{-1} is the probability pair A gets a birdie, a_0 is the probability pair A gets a par, and a_{+1} is the probability they get a bogey on any hole. Correspondingly, pair B can be denoted by $\{b_{-1}, b_0, b_{+1}\}$. If we assume that pair A's score and pair B's score on any hole are independent, it can be seen that the probability pair A wins a hole against pair B, p_A, is given by $a_{-1}(1 - b_{-1}) + a_0 b_{+1}$, and the probability pair B wins a hole against pair A is given by $p_B = b_{-1}(1 - a_{-1}) + b_0 a_{+1}$. The probability that the hole is squared is given by $1 - p_A - p_B$. The probability that pair A wins (loses) an 18-hole 4BBB match-play event, P_A (P_B), can be evaluated using recurrence methods, and the probability that the match is squared is equal to $1 - P_A - P_B$. If a play-off is used when the 18 holes are squared, the probability pair A wins (loses) the match can be evaluated using the geometric distribution.

A further aspect of the model

We suppose players A_1 and A_2 obtain scores for pair A (birdie, par or bogey) that are independently distributed, and we denote player A_1 by $(p_{a1,-1}, p_{a1,0}, p_{a1,+1})$ and player A_2 by $(p_{a2,-1}, p_{a2,0}, p_{a2,+1})$. The pair A, playing independently, is denoted by $\{(p_{a1,-1}, p_{a1,0}, p_{a1,+1}), (p_{a2,-1}, p_{a2,0}, p_{a2,+1})\}$. A corresponding notation is used for pair B, and it is typically assumed that pair A's score on a hole is independent of pair B's score on the hole. For a hole for team A, $a_{-1} = p_{a1,-1} + p_{a2,-1} - p_{a1,-1}p_{a2,-1}$, $a_{+1} = p_{a1,+1}p_{a2,+1}$ and $a_0 = 1 - a_{-1} - a_{+1}$. Corresponding expressions apply for pair B.

'Fairness' of the four-ball-best-ball match-play scoring system

A minimal criterion for fairness of the 4BBB scoring system would appear to be the requirement that if pairs A and B are identical, the 'hole-probabilities' against each other p_A and p_B must be equal, and hence P_A must equal P_B. It is clear that this requirement is satisfied, since if a_{-1} is equal to b_{-1}, a_0 is equal to b_0 and a_{+1} is equal to b_{+1}, p_A must equal p_B and P_A must equal P_B. Further, it is clear that if player A_1 is identical to player B_1, and player A_2 identical to B_2, we must have $p_A = p_B$, and P_A must equal P_B.

Example 1. Noting that PGA Tour players score a birdie approximately 20% of the time on average, we consider in this example four identical players, each with a probability of 0.2 of scoring a birdie, a probability 0.6 of scoring a par, and a probability 0.2 of scoring a bogey. It would seem that these probabilities would also be applicable to top amateurs and top club players on somewhat easier courses. The probability pair A wins a hole against pair B is equal to $p_A = 0.2544$, the probability pair B wins a hole against pair A is equal to $p_B = 0.2544$, and the probability that the hole is halved is equal to 0.4912. Recurrence methods can be used to show that the probability pair A wins an 18-hole match is equal to 0.4346, the probability pair B wins it is also 0.4346, whilst the probability that the 18-holes match is squared is equal to 0.1309. After a play-off if necessary, the probability each pair wins is equal to 0.5. These results are shown in table 1, and this example demonstrates the 'fairness' of the 4BBB scoring system.

Example 2. In this example the pair A has characteristics {(0.25, 0.5, 0.25), (0.25, 0.5, 0.25)} and the pair B has characteristics {(0.2, 0.6, 0.2), (0.2, 0.6, 0.2)}. It can be seen from table 1 that pair A has a probability of 0.6417 of winning an 18-holes match (with play-off) against pair B. Pair A's increased probability of winning the match (compared with their value in example 1) is essentially due to the role of a_{-1} in the expression $p_A = a_{-1}(1 - b_{-1}) + a_0 b_{+1}$. Note that it can easily be argued that all four players in this example are equal golfers, as their expected scores on every hole are the same (and equal to 0). Pair A is, however, the better 4BBB pair. The 4BBB format of scoring has favoured the more variable pair, pair A. Note further that if we increased pair A's probability of a bogey (or decreased their probability of a birdie) very slightly and decreased (increased) their probability of a par accordingly, so that pair A's expected score was slightly less than that of pair B, we would still have a situation in which pair A was the better 4BBB pair. This might seem incongruous to believers in stroke play and other golf formats. Thus, it is clearly possible to have a pair that is better than another pair at 4BBB golf, whilst being weaker at other forms of the game.

Ranking individual four-ball-best-ball players

In this section we investigate whether it is possible to rank individual players as 4BBB players.

Example 3. Suppose y = (p, $1 - 2p$, p) and x = ($p - \varepsilon$, $1 - p + \varepsilon$, 0) are two types of players where $0 < \varepsilon < p < 0.5$. Consider a 4BBB match between pair A with characteristics {x, y} and pair B with characteristics {x, x}. It can be shown that pair A is always the better pair (better since $p_A > p_B$), and this therefore suggests that y is a better 4BBB player than x. However, when we consider a 4BBB match between pair A with characteristics {y, x} and pair B with characteristics {y, y}, we find that pair A is the better 4BBB pair provided ε is sufficiently small so as to satisfy the equation $p^4 - (2 + \varepsilon)p^3 + (\varepsilon + 1)p^2 + \varepsilon p - \varepsilon > 0$. For example, when $p = 0.2$, ε needs to be less than 1/30. Thus, provided ε is sufficiently small, this result suggests that x is a better 4BBB player than y. Therefore, we have, for small ε, one result suggesting that x is a better 4BBB player than y, and another result suggesting that y is a better 4BBB player than x. The truth is that, for small ε, x combines more successfully with y than does either x with x, or y with y. These two examples with $p = 0.2$ and $\varepsilon = 0.01$ appear in table 1 as examples 3a and 3b respectively.

Table 1: Examples of the probability pair A wins a hole and wins a match of 4BBB golf

Example	$p_{i,j}$ values A is the first pair B is the second	Probability pair A wins, squares, loses a hole	Probability pair A wins, squares, loses an 18-holes match	Probability pair A wins, loses an 18-holes match, after play-off
1	{(0.2,0.6,0.2),(0.2,0.6,0.2)} {(0.2,0.6,0.2),(0.2,0.6,0.2)}	0.2544, 0.4912, 0.2544	0.4346, 0.1309, 0.4346	0.5000, 0.5000
2	{(0.25,0.5,0.25),(0.25,0.5,0.25)} {(0.2,0.6,0.2),(0.2,0.6,0.2)}	0.3000, 0.4600, 0.2400	0.5752, 0.1197, 0.3051	0.6417, 0.3583
3a	{(0.19,0.81,0.0),(0.2,0.6,0.2)} {(0.19,0.81,0.0),(0.19,0.81,0.0)}	0.2309, 0.5462, 0.2228	0.4508, 0.1387, 0.4105	0.5214, 0.4786
3b	{(0.2,0.6,0.2),(0.19,0.81,0.0)} {(0.2,0.6,0.2),(0.2,0.6,0.2)}	0.2512, 0.5155, 0.2333	0.4765, 0.1334, 0.3901	0.5456, 0.4544
3c	{(0.21,0.58,0.21),(0.21,0.79,0.0)} {(0.19,0.81,0.0),(0.21,0.79,0.0)}	0.2405, 0.5347, 0.2247	0.4706, 0.1364, 0.3930	0.5412, 0.4588
3d	{(0.19,0.81,0.0),(0.2,0.6,0.2)} {(0.21,0.58,0.21),(0.2,0.6,0.2)}	0.2497, 0.5119, 0.2385	0.4603, 0.1334, 0.4063	0.5285, 0.4715
4a	{(0.3,0.5,0.2),(0.1,0.7,0.2)} {(0.2,0.6,0.2),(0.2,0.6,0.2)}	0.2604,0.4888, 0.2508	0.4574, 0.1306, 0.4123	0.5238, 0.4762
4b	{(0.3,0.6,0.1),(0.1,0.6,0.3)} {(0.2,0.6,0.2),(0.2,0.6,0.2)}	0.2608, 0.4944, 0.2448	0.4724, 0.1307, 0.3969	0.5398, 0.4602
4c	{(0.3,0.4,0.3),(0.1,0.8,0.1)} {(0.2,0.6,0.2),(0.2,0.6,0.2)}	0.2608, 0.4944, 0.2448	0.4724, 0.1307, 0.3969	0.5398, 0.4602
4d	{(0.3,0.6,0.1),(0.1,0.7,0.2)} {(0.2,0.6,0.2),(0.2,0.6,0.2)}	0.2612, 0.5000, 0.2388	0.4877, 0.1309, 0.3814	0.5560, 0.4440
4e	{(0.3,0.6,0.1),(0.1,0.8,0.1)} {(0.2,0.6,0.2),(0.2,0.6,0.2)}	0.2616, 0.5056, 0.2328	0.5031, 0.1309, 0.3660	0.5724, 0.4276
4f	{(0.3,0.6,0.1),(0.1,0.9,0.0)} {(0.2,0.6,0.2),(0.2,0.6,0.2)}	0.2620, 0.5112, 0.2268	0.5187, 0.1307, 0.3506	0.5888, 0.4112

As another example, suppose w = ($p + \varepsilon$, $1 - 2p - 2\varepsilon$, $p + \varepsilon$), x = ($p + \varepsilon$, $1 - p - \varepsilon$, 0), y = (p, $1 - 2p$, p) and z = ($p - \varepsilon$, $1 - p + \varepsilon$, 0) are four types of players where $0 < \varepsilon < p < 0.5$. Consider a 4BBB match between pair A with characteristics {w, x} and pair B with characteristics {z, x}. It can be shown that pair A is always the better pair, suggesting that w is a better 4BBB player than z. However, when we consider a 4BBB match between pair A with characteristics {z, y} and pair B with characteristics {w, y}, we find that pair A is the better 4BBB pair provided ε is sufficiently small. (For example, when $p = 0.2$, provided ε is less than about 0.017 (to 0.018)). Thus, provided ε is sufficiently small, this second result suggests that z is a better 4BBB player than w. So, again we have a contradiction when ε is small. For the case in which $p = 0.2$ and $\varepsilon = 0.01$, see examples 3c and 3d in table 1. Thus, it can be concluded that it is not possible to strictly/generally rank individual 4BBB players in an order from best to worst. We will see however that it is possible to rank 4BBB pairs.

An important characteristic of a four-ball-best-ball pair

It is clear from the above examples that the sum of the birdie probabilities of the players in a pair of 4BBB players tells us a great deal about the strength of that pair, but it is not the complete picture. The following example highlights this point.

Example 4. We start by assuming players B_1 and B_2 are the same as in example 1, and that pair A's characteristics are {(0.3, 0.5, 0.2), (0.1, 0.7, 0.2)}, rather than {(0.2, 0.6, 0.2), (0.2, 0.6, 0.2)}. Note that the sum of pair A's birdie probabilities are the same as in example 1, as are the sum of pair A's par probabilities, and the sum of pair A's bogey probabilities. However, the interaction between player A_1's and A_2's birdie outcomes is now smaller than before (0.03 instead of 0.04). This reduces the likelihood of players A_1 and A_2 having birdies at the same time, and this increases their probability of winning an 18-holes match (with play-off) from 0.5 to 0.5238, as can be seen in example 4a of table 1. If pair A's characteristics change to {(0.3, 0.6, 0.1), (0.1, 0.6, 0.3)}, their probability of winning an 18-holes match against pair B (with play-off) increases further to 0.5398. Note that the probability that both A_1 and A_2 have a bogey has decreased by 0.01 (from 0.04 to 0.03), as can be seen in example 4b of table 1. If pair A's characteristics are changed to {(0.3, 0.6, 0.1), (0.1, 0.7, 0.2)}, their probability of winning an 18-holes match against pair B (with play-off) increases further to 0.5560. The probability that both A_1 and A_2 have a bogey has decreased by 0.02 (from 0.04 to 0.02), as can be seen in example 4d in table 1. If pair A's characteristics are changed to {(0.3, 0.6, 0.1), (0.1, 0.8, 0.1)}, their probability of winning an 18-holes match (with play-off) increases further to 0.5724, as their probability of both having a bogey has decreased by a further amount 0.01 (from 0.02 to 0.01), as can be seen in example 4e of table 1. Finally, if pair A's characteristics are changed to {(0.3, 0.6, 0.1), (0.1, 0.9, 0.0)}, their probability of winning an 18-holes match (with play-off) increases further to 0.5888, as can be seen in example 4f of table 1.

Further aspects on ranking individual four-ball-best-ball players

Example 5. In this example we see that if the 'gap in the performance characteristics' between player A and player B is 'sufficiently large', player A can be a better 4BBB player than player B no matter what are the common characteristics of player A's and player B's partners. Supposing the two partners with common characteristics are denoted by X, pair {A, X_1} can be represented by $\{(p_{a,-1}, p_{a,0}, p_{a,+1}), (x_{-1}, x_0, x_{+1})\}$, and pair {B, X_2} can be represented by $\{(p_{b,-1}, p_{b,0}, p_{b,+1}), (x_{-1}, x_0, x_{+1})\}$. The probability pair {A, X_1} wins a hole against pair {B, X_2}, p_A, can be shown to be given by

$$p_A = (p_{a,-1} + x_{-1} - x_{-1}p_{a,-1})(1 - p_{b,-1} - x_{-1} + x_{-1}p_{b,-1}) + (1 - p_{a,-1} - x_{-1} + x_{-1}p_{a,-1} - p_{a,+1}x_{+1})p_{b,+1}x_{+1}$$

A corresponding expression can be written down for p_B, the probability pair {B, X_2} wins a hole against pair {A, X_1}. It can be shown that $p_A > p_B$ provided

$$(p_{a,-1} - p_{b,-1})/(p_{a,+1}(1 - p_{b,-1}) - p_{b,+1}(1 - p_{a,-1})) > x_{+1}$$

Noting that $x_{+1} \in [0, 1]$, it can be seen, for example, that player (0.3, 0.6, 0.1) is always a better 4BBB player than player (0.1, 0.8. 0.1), and that player (0.3, 0.4, 0.3) is a better 4BBB player than player (0.1, 0.9, 0.0) provided x_{+1} for the players with the common characteristics is less than 20/27.

Example 6a. As special case of example 5, we consider the important family of 'symmetric' players (p, $1-2p$, p). Suppose player A has $p = p_1$ and player B has $p = p_2$, where $p_1 > p_2 > 0$. Then it can be seen that player A is always a better 4BBB player than player B (unless $x_{+1} = 1$, which is the (irrelevant) case in which the partners with the common characteristics always score bogeys).

Example 6b. As another special case of example 5, we consider the family of 'equally non-symmetric' players $(\theta p, 1-(1+\theta)p, p)$, where $\theta \neq 1$ and $0 < \theta < p^{-1} - 1$. Supposing θ is fixed, player A has $p = p_1$ and player B has $p = p_2$, where $p_1 > p_2 > 0$, it can be shown that player A is always a better 4BBB player than B, provided $x_{+1} < \theta$ and $x_{-1} \neq 1$. Further, for all (fixed) θ such that $0 < \theta < p^{-1} - 1$ (i.e. the symmetric ($\theta = 1$) as well as the non-symmetric ($\theta \neq 1$)cases), if players C and D have p-values equal to p_3 and p_4 where $p_1 > p_2 > p_3 > p_4 > 0$, and $p_1 + p_4 = p_2 + p_3$, it can be shown that the pair (A, D) is always a better 4BBB pair than pair (B, C).

Equal 4BBB pairs

Definition. Two independent 4BBB pairs A and B (i.e. two 4BBB pairs A and B whose scores on a hole are independent) are said to be equal 4BBB pairs if $p_A = p_B$.

Theorem. If independent pairs A and B are equal 4BBB pairs, and if independent pairs B and C are also equal 4BBB pairs, it follows that independent pairs A and C are also equal 4BBB pairs.

Proof. Since pair A is equal to pair B,

$a_{-1}(1-b_{-1}) + a_0 b_{+1} = b_{-1}(1-a_{-1}) + b_0 a_{+1}$, and substituting

$a_0 = 1 - a_{-1} - a_{+1}$ and

$b_0 = 1 - b_{-1} - b_{+1}$, and subtracting 1 from each side of the equation, we obtain

$a_{-1} + b_{+1} - a_{-1}b_{+1} - 1 = b_{-1} + a_{+1} - a_{+1}b_{-1} - 1$,and hence

$(a_{-1} - 1)(1 - b_{+1}) = (b_{-1} - 1)(1 - a_{+1})$.

Correspondingly, since pair B is equal to pair C, we have

$(b_{-1} - 1)(1 - c_{+1}) = (c_{-1} - 1)(1 - b_{+1})$,and hence

$(a_{-1} - 1)(1 - c_{+1}) = ((b_{-1} - 1)(1 - a_{+1})/(1 - b_{+1}))((c_{-1} - 1)(1 - b_{+1})/(b_{-1} - 1))$,since $1 - b_{+1}$ and $b_{-1} - 1$ are not equal to zero, and hence

$(a_{-1} - 1)(1 - c_{+1}) = (c_{-1} - 1)(1 - a_{+1})$,and hence

$a_{-1}(1-c_{-1}) + a_0 c_{+1} = c_{-1}(1-a_{-1}) + c_0 a_{+1}$, and so pair A and pair C are equal, completing the proof.

Example 7. If independent pairs A and B are $\{(p_1, 1-2p_1, p_1), (p_2, 1-2p_2, p_2)\}$ and $\{(p_1, 1-p_1-p_2, p_2), (p_2, 1-p_1-p_2, p_1)\}$ respectively, it can be seen that $a_{-1} = b_{-1}$, $a_{+1} = b_{+1}$ and hence $a_0 = b_0$, and it follows that the pairs A and B are equal for all legitimate values of p_1 and p_2.

Example 8. The pair-wise independent pairs A, B and C given by {(0.25, 0.5, 0.25), (0.25, 0.5, 0.25)}, {(0.25, 0.5, 0.25), (0.3, 0.2, 0.5)} and {(0.25, 0.5, 0.25), (0.2, 0.8, 0)} can be seen to be pair-wise equal. Similarly, the pair-wise independent pairs D, E and F given by {(0.2, 0.6, 0.2), (0.2, 0.6, 0.2)}, {(0.2, 0.6, 0.2), (7/30, 11/30, 0.4)} and {(0.2, 0.6, 0.2), (1/6, 5/6, 0)} also can be seen to be equal.

Ranking 4BBB pairs

We have seen above that in general it is not strictly possible to rank individual 4BBB players. However, it is possible to rank pairs of players as a result of the following theorem.

Definition. Given two independent 4BBB pairs A and B, pair A is said to be the better pair if $p_A > p_B$.

Theorem. (transitive theorem) Given three pair-wise independent pairs A, B and C, with A better than B, and B better than C, it follows that A is better than C.

Proof. This theorem can be proved in an analogous manner to the theorem above.

Example 9. Here we reconsider example 7 with $p_1 = p + \partial$ and $p_2 = p - \partial$. Then, given $0.5 > p > \partial_2 > \partial_1 > 0$, it follows that pairs A and B with $\partial = \partial_2$ are better 4BBB pairs than pairs A and B with $\partial = \partial_1$.

Example 10. Here we return to example 8 and consider the pair-wise independent pairs G, H and J given by {(0.15, 0.7, 0.15), (0.15, 0.7, 0.15)}, {(0.15, 0.7, 0.15), (39/230, 61/115, 0.3)} and {(0.15, 0.7, 0.15), (3/23, 20/23, 0)} respectively. It can be verified that these three pairs are equal pairs. Then, assuming the nine pairs A, B, C, D, E, F, G, H and J are pair-wise independent, it can be seen that any of A, B and C, are better than any of D, E and F, who in turn are all better than any of G, H and J. Note however, for example, that although A, B and C are equal pairs, and D, E and F are also equal pairs, p_A, p_B and p_C take on 9 different values when playing D, E or F, as can be seen in table 2.

Which of two strategies might a player regularly use?

Example 11. Suppose pair A has characteristics $\{(p_1 + \varepsilon, 1 - p_1 - p_2 - \varepsilon - \partial, p_2 + \partial), (p_1, 1 - p_1 - p_2, p_2)\}$ and pair B has characteristics $\{(p_1, 1 - p_1 - p_2, p_2), (p_1, 1 - p_1 - p_2, p_2)\}$. In this example players A_2, B_1 and B_2 are identical players. Player A_1 is very similar to these three players except that he can, if he wishes, increase his probability of getting a birdie by $\varepsilon > 0$ by taking a strategic risk, but in the process his probability of getting a bogey increases by $\partial > 0$. It can be shown that pair A's probability of winning a hole is greater than pair B's provided $\partial < ((1 - p_2^2)/(p_2(1 - p_1)))\ \varepsilon$. For example when $p_1 = p_2 = 0.2$, this inequality is $\partial < 6\varepsilon$, indicating that for these values of p_1 and p_2, player A_1 can suffer a considerable increase in his bogey probability and it still is worthwhile taking the risk.

One player's strategy when his/her 4BBB partner is about to have a 'bad' hole

Example 12. Assume pairs A and B both have characteristics $\{(p_1, 1 - p_1 - p_2, p_2), (p_1, 1 - p_1 - p_2, p_2)\}$. However, we make an additional assumption concerning player A_1. It is assumed that player A_1 can detect on a particular hole when his partner, A_2, cannot possibly get a birdie, but will get a bogey with probability z and a par with probability $1\text{-}z$. (For example, if player A_2 always tees off first on every hole, such an event can sometimes be immediately clear to player A_1 if player A_2 plays a very bad first shot).

Table 2: Examples of the probability some pairs win a hole and win a match of 4BBB golf

Pairs	Probability first pair wins, squares, loses a hole	Probability first pair wins, squares, loses an 18 holes match	Probability first pair wins, loses an 18 holes match, after play-off
A versus D	0.3000, 0.4600, 0.2400	0.5752, 0.1197, 0.3051	0.6417, 0.3583
A versus E	0.3083, 0.4408, 0.2508	0.5683, 0.1183, 0.3134	0.6335, 0.3655
A versus F	0.2917, 0.4792, 0.2292	0.5823, 0.1210, 0.2966	0.6501, 0.3499
B versus D	0.3200, 0.4160, 0.2640	0.5636, 0.1163, 0.3202	0.6273, 0.3727
B versus E	0.3233, 0.4070, 0.2697	0.5580, 0.1159, 0.3261	0.6212, 0.3788
B versus F	0.3167, 0.4250, 0.2583	0.5692, 0.1167, 0.3141	0.6335, 0.3665
C versus D	0.2800, 0.5040, 0.2160	0.5879, 0.1233, 0.2888	0.6575, 0.3425
C versus E	0.2933, 0.4747, 0.2320	0.5793, 0.1209, 0.2998	0.6468, 0.3532
C versus F	0.2667, 0.5333, 0.2000	0.5970, 0.1259, 0.2771	0.6689, 0.3311

We assume player A_1 can take evasive action and reduce his own probability of a bogey by x, but in doing so he reduces his probability of scoring a birdie by an amount y. The question is whether player A_1 should, in this situation, take this evasive action. Let us suppose that he will take this evasive action if he can reduce pair B's probability of winning the hole. If he does not take the evasive action, pair B's probability of winning the hole is equal to $p_B = (2p_1 - p_1^2)(1 - p_1) + ((1 - p_1)^2 - p_2^2)zp_2$, whereas if he does take the evasive action $p_B = (2p_1 - p_1^2)(1 - p_1 + y) + ((1 - p_1)^2 - p_2^2)(p_2 - x)z$. It follows that he should take the evasive action provided $y < (((1\text{-}p_1)^2 - p_2^2)xz)/(2p_1 - p_1^2)$. For example, when $p_1 = p_2 = 0.2$, this inequality is $y < 5x/6$. Thus, it is worth taking this evasive action provided the associated decrease in his probability of scoring a birdie is only moderate. This result further demonstrates the importance of scoring birdies in 4BBB golf.

The benefits to a pair of 4BBB players of decreasing their interaction on a hole

Example 13. Suppose players A_1 and A_2 have the capacity to interact in some way that modifies the assumption of independence between player A_1's score and player A_2's score. This is demonstrated mathematically in table 3. For example, $i_{-1,-1}$ represents the probability interaction between players A_1 and A_2 when they both score birdies, and $i_{0,0}$ represents the interaction when they both score pars. The value of a_{-1} is now $p_{a1,-1} + p_{a2,-1} - p_{a1,-1}p_{a2,-1} - i_{-1,-1}$, and a_{+1} is now $p_{a1,+1}p_{a2,+1} + i_{-1,-1} + i_{-1,0} + i_{0,-1} + i_{0,0}$. Thus, assuming player B_1's and B_2's scores are independent, and pair A's score is independent of pair B's score, the probability pair A wins a hole is given by

$$P(A) = a_{-1}(1-b_{-1}) + (1-a_{-1}-a_{+1})b_{+1}$$

where b_{-1} and b_{+1} are given at the beginning of this paper. Substituting the above expressions for a_{-1} and a_{+1} in this equation, we have, for example, the following partial derivatives

$$\frac{\delta P(A)}{\delta p_{a1,-1}} = (1-p_{a2,-1})(1-b_{-1}-b_{+1})$$

$$\frac{\delta P(A)}{\delta p_{a2,-1}} = (1-p_{a1,-1})(1-b_{-1}-b_{+1})$$

$$\frac{\delta P(A)}{\delta i_{-1,-1}} = -(1-b_{-1})$$

For the p-values in example 1, these three partial derivatives are equal to 0.48, 0.48 and –0.64 respectively, making the third one of them the largest in absolute value (actually, it is always the largest in absolute value). Thus, and very interestingly, if pair A can decrease their interaction $i_{-1,-1}$ by an amount $\varepsilon > 0$, they get a greater return than they would by increasing either of their birdie probabilities $p_{a1,-1}$ or $p_{a2,-1}$ by the same amount ε. It would seem that this observation could be potentially very useful, but it is not clear how the two players could achieve this in practice. (An example of how something somewhat similar to this might be achieved in practice is if one of them is a particularly strong birdie chance on the short holes and the other is a particularly strong birdie chance on the long holes.) The application of this result, however, would appear to be limited.

Table 3: The joint distribution of player A_1's and player A_2's score, with interactions

		A_2's score			
Joint Prob		–1	0	+1	
	-1	$p_{a1,-1}p_{a2,-1} + i_{-1,-1}$	$p_{a1,-1}p_{a2,0} + i_{-1,0}$	$p_{a1,-1}p_{a2,+1} - i_{-1,-1} - i_{-1,0}$	$p_{a1,-1}$
A_1's score	0	$p_{a1,0}p_{a2,-1} + i_{0,-1}$	$p_{a1,0}p_{a2,0} + i_{0,0}$	$p_{a1,0}p_{a2,+1} - i_{0,-1} - i_{0,0}$	$p_{a1,0}$
	+1	$p_{a1,+1}p_{a2,-1} - i_{-1,-1} - i_{0,-1}$	$p_{a1,+1}p_{a2,0} - i_{-1,0} - i_{0,0}$	$p_{a1,+1}p_{a2,+1} + i_{-1,-1} + i_{-1,0} + i_{0,-1} + i_{0,0}$	$p_{a1,+1}$
		$p_{a2,-1}$	$p_{a2,0}$	$p_{a2,+1}$	1

Strategies when the opposing 4BBB pair is about to have a 'bad' hole

Example 14. Suppose that at some stage in a hole it becomes clear to pair A that pair B will not be able to get a birdie, but that they will have a par with probability p (assumed given or estimable) and a bogey with probability q ($p + q = 1$). Pair A's characteristics are typically $\{a_{-1}, a_0, a_{+1}\}$, but they have several other available strategies that can be represented by $\{a_{-1} + \varepsilon_i, a_0 - \varepsilon_i - \partial_i, a_{+1} + \partial_i\}$, $(i = 1, 2, \ldots, n)$, $(\partial_i > 0, \varepsilon_i > 0)$. If pair A use their typical or standard strategy, their probability of winning the hole is equal to $a_{-1} + qa_0$, given the additional information that pair B must have either a par with probability p or a bogey with probability q. If they use the ith available strategy above, their probability of winning the hole is given by $a_{-1} + qa_0 + p\varepsilon_i - q\partial_i$. Thus, the strategy which maximizes their probability of winning the hole is either the standard strategy or the one that maximizes $p\varepsilon_i - q\partial_i$ (provided this is positive). Alternatively, for example, if they are '1 ahead' at the 18th hole when this occurs, they may be interested in minimizing their probability of losing the hole, in which case they should use their standard strategy.

The effect of a modification of the 4BBB model

At the beginning of this section it was stated that the assumption that each player had a birdie, a par, or a bogey on each hole was a reasonable assumption, and not a restrictive one for 4BBB matchplay. This proposition is considered in the next example.

Example 15. We reconsider example 1 and assume that each of players B_1 and B_2 have a probability of 0.01 of scoring a double bogey, so that the probabilities for their scoring characteristics can be represented by (0.2, 0.6, 0.19, 0.01), using an obvious extension of the earlier notation. Using a further extension of the earlier notation, it can be seen that $b_{-1} = p_{b1,-1} + p_{b2,-1} - p_{b1,-1}p_{b2,-1} = 0.36$ as before, $b_{+2} = p_{b1,+2}p_{b2,+2} = 0.0001$, $b_{+1} = p_{b1,+1}p_{b2,+1} + p_{b1,+1}p_{b2,+2} + p_{b1,+2}p_{b2,+1} = 0.0399$, and $b_0 = 1 - b_{-1} - b_{+1} - b_{+2} = 0.6$, as before. The probability pair A wins a hole is now 0.254404 (very slightly greater than before), the probability they lose is 0.2544 (as before), and the probability of the hole being squared is 0.491196. Note that the only changed result or outcome in this example compared with that in example 1 is when players A_1 and A_2 both get bogeys whilst B_1 and B_2 both get double bogeys, and this outcome has probability $p_{a1,+1}p_{a2,+1}p_{b1,+2}p_{b2,+2} =$ 0.000004, a very small change compared with the changes in $p_{b1,+1}$ and $p_{b2,+1}$ of –0.01. Further, note that if only one (and not both) of the players B_1 and B_2 had a non-zero probability of getting a double bogey, the results in example 1 would be unchanged.

CONCLUSIONS

In this paper a four-ball-best-ball (4BBB) model for golf has been established. The 4BBB match-play scoring system has been shown to satisfy a basic requirement of fairness. It has been demonstrated that, in a strict sense, it is not possible, in general, to rank individual players as 4BBB players, as each player's performance interacts with the performance of his/her partner. Of course, if two players' characteristics are 'far enough apart', one player can be better than the other. A reasonably broad class of individual players has been identified, however, such that it is possible to rate them individually (within that class) as 4BBB players. The capacity of an individual to play birdies is seen to be a very important determinant in being a successful member of a 4BBB pair, but there are other minor factors as well. Equal and unequal 4BBB pairs have been studied, and the transitive law is seen to apply to 4BBB pairs. Thus, if pair A is better than pair B, and pair B is better than pair C, then pair A must be better than pair C. Correspondingly, if pair A is equal to pair B, and pair B is equal to pair C, then pair A is equal to pair C.

Some strategic issues in 4BBB match-play golf have been studied. For example, the conditions under which a player should take a greater risk and have a higher probability of obtaining a bogey in order to achieve a higher probability of scoring a birdie, have been determined. Also, the conditions under which a player, noting that his partner is about to have a 'bad' hole and score only a par or a bogey, should 'play safe', have been determined. Thirdly, players who can interact in certain ways have been shown to have an advantage over those pairs that cannot do this. Finally, one pair's optimal strategy when they see that their opponents are about to score a par or a bogey, but not a birdie, has been studied. The 4BBB model can be used to study other strategic considerations that might be of interest to golf enthusiasts.

References

Hurley, W. (2007) The Ryder Cup: Are balancing four-ball pairings optimal? *Journal of Quantitative Analysis in Sports.* **3(4):** Article 6, 1-18.

Pollard, G. and Pollard, G. (2008) Four ball best ball 2. In J. Hammond (ed.) *Proceedings of the Ninth Australiasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM) Coolangatta, Queensland, Australia (in press).

FOUR BALL BEST BALL 2

Pollard, Geoff [1] and Pollard, Graham [2]

[1] Faculty of life and Social Sciences, Swinburne University of Technology, Melbourne, Australia

[2] Faculty of Information Sciences and Engineering, University of Canberra, Australia

Paper Submitted for Review: 27 March 2008
Revision submitted and accepted: 11 June 2008

Abstract. In this paper the decision as to which four-ball-best-ball (4BBB) pair(s) should be selected by a club is considered. The proposition that a club should select its best pair(s) is considered. Although this is most often the case, it is shown to be not necessarily valid. The case in which the two pairs from club X play one match each against club Y is considered, as is the case with both pairs from club X playing both pairs from club Y ('reverse' matches). It is shown that the pairs best suited to represent the club can depend on the characteristics of the pairs in the other club. For example, pairs likely to perform best when playing a 'stronger' club can be different from pairs likely to perform best when playing a 'weaker' club. When playing a club of similar standard, it is shown that it is typically best to pair the players so as to 'balance' their 4BBB ability. Further, it is shown that, in order to maximize the likelihood of correctly identifying the better team, a match that is tied after 18 holes is best left as a tied match, rather than having a play-off. Statistics from the 2006 PGA Tour are used to identify four pairs of players who might have represented the US in 2006. The characteristics of these top players suggest that their performances on the par 5 holes can be more important than on other holes. Information on players' performances on par 3, par 4 and par 5 holes (along with those of their opponents) is sufficient to determine the better of two pairs against an opponent pair. That is, it is not necessary to do a full 18-holes analysis of such a match.

Keywords: characteristics of four-ball-best-ball golf, selecting four-ball-best-ball teams, Ryder Cup.

INTRODUCTION

The Ryder Cup is a competition played between the US and Europe every 2 years. There are 12 players in each team, and 8 out of the 28 matches are four-ball-best-ball (4BBB) matches. The captain of each team needs to select 8 players from the 12 in the team (on each of two days), and combine them into pairs. A paper on this problem of selecting a 4BBB team has only very recently been written (Hurley, 2007). By considering teams with identical players and using expected scores, he concluded that, given 8 players, it was best for teams to pair their best and worst players.

In this paper several questions are addressed. For example, if a team has two equal pairs, can it be advantageous to use one pair rather than the other against a particular team? Does this selection depend on whether the opposing team is stronger or weaker? Can there be occasions when it is better not to use the best pair in the team? Should play-offs be used in this team situation if a match is squared at the 18^{th} hole? How should players be combined to form a team, and does the combination depend on whether the opponents are stronger or weaker? Is a par 5 hole relatively more important than (say) a par 3 hole? Rather than doing a full analysis of 18 holes for each possible match, is a simpler approach possible in achieving answers to some of these questions? Finally, some analysis relevant to the 2006 Ryder Cup 4BBB selection is given.

METHODS

In the first of these two papers (Pollard & Pollard, 2008), a model for 4BBB golf was set-up. A pair's score on a hole is typically a birdie (-1), a par (0) or a bogey (+1), and so a useful model for 4BBB match-play golf is to denote pair A by $\{a_{-1}, a_0, a_{+1}\}$ where a_{-1} is the probability pair A gets a birdie, a_0 is the probability pair A gets a par, and a_{+1} is the probability they get a bogey on any hole. Correspondingly, pair B can be denoted by $\{b_{-1}, b_0, b_{+1}\}$. If we assume that pair A's score and pair B's score on any hole are independent, it can be seen that the probability pair A wins a hole against pair B, p_A, is given by $a_{-1}(1 - b_{-1}) + a_0b_{+1}$, and the probability pair B wins a hole against pair A is given by $p_B = b_{-1}(1 - a_{-1}) + b_0a_{+1}$. The probability that the hole is tied or squared is given by $1 - p_A - p_B$.

Suppose players A_1 and A_2 obtain scores for pair A (birdie, par or bogey) that are independently distributed, and we denote player A_1 by $(p_{a1,-1}, p_{a1,0}, p_{a1,+1})$ and player A_2 by $(p_{a2,-1}, p_{a2,0}, p_{a2,+1})$. Then the pair A, playing independently, is denoted by $\{(p_{a1,-1}, p_{a1,0}, p_{a1,+1}), (p_{a2,-1}, p_{a2,0}, p_{a2,+1})\}$. A corresponding notation is used for pair B, and it is typically assumed that pair A's score on a hole is independent of pair B's score on that hole. For a hole for team A, it can be seen that $a_{-1} = p_{a1,-1} + p_{a2,-1} - p_{a1,-1}p_{a2,-1}$, $a_{+1} = p_{a1,+1}p_{a2,+1}$ and $a_0 = 1 - a_{-1} - a_{+1}$. Corresponding expressions apply for pair B.

The probability that pair A wins (loses) an 18-hole 4BBB match-play event against pair B, P_A (P_B), can be evaluated using recurrence methods, and the probability that the match is squared is equal to $1 - P_A - P_B$. If a play-off is used when the 18 holes are squared, the probability pair A wins (loses) the match can be evaluated using the geometric distribution. Thus, if a club can be represented by one of two pairs, these recurrence methods can be used together with the respective probability structure for each hole in order to determine which pair is best to represent the club. Is it possible to decide which is the better pair to represent the club by using only the information on each hole rather than using the recurrence process for the full 18 holes? This question is considered.

The explicit analysis of an 18-holes competition is more complicated than the analysis of some other scoring systems, for example the W_n system, which is a system which is won by the first pair to be n holes ahead. It would appear reasonable to believe that the pair which is the best to represent a club under one 'stopping rule' for a contest is the same one as would be best to represent it under many other 'stopping rules'. We explore this proposition for the two systems, W_n and best of 18 holes.

For the W_n system, the playing of each hole (a 'module' of play) can be considered to be the realization of the steps in a general one-dimensional walk in discrete time. Using the approach and notation of Cox and Miller (1965, p. 46-58), the steps in the random walk, Z_i, are mutually independent random variables on the integers ….., -2, -1, 0, 1, 2,… and the moment generating function (m.g.f.) of Z_i is defined by

$$f^*(\theta) = \sum_{j=-\infty}^{\infty} \exp(-j\theta)P(Z_i = j)$$

If P(Q) represents the probability of absorption in states [a, infinity) ((-infinity, -b]), then, neglecting the excess over the barriers,

$$P/Q = (1 - \exp(\theta_0 b))/(\exp(-\theta_0 a) - 1) \text{ when } E(Z_i) \neq 0$$

where θ_0 is the non-zero solution of the equation $f^*(\theta) = 1$. For the W_n scoring system, we set a and b equal to n. It can be seen that P/Q (when using W_n) is always largest for the pair with the largest θ_0 value, and thus the proposition is that P/Q (for best of 18 holes of golf) is also largest for the pair with the largest θ_0 value. Thus, the proposition is that the selection of the best pair does not depend on the 'stopping rule' for the scoring system. We also make use of the result that, if the steps in the random walk are assumed to be normally distributed, θ_0 can be shown to equal the ratio, R, given by R = 2*mean/variance of Z_i. The value of R might be easier for some users to calculate.

Example 1. Suppose club X has three equal 4BBB pairs, A , B and C, and club Y also has three equal pairs of players, D, E and F. Suppose pairs A, B and C have characteristics {(0.25,0.5,0.25), (0.25,0.5,0.25)}, {(0.25,0.5,0.25), (0.3,0.2,0.5)} and {(0.25,0.5,0.25), (0.2,0.8,0.0)} respectively. Note that the three types of players in club X are very different from each other. Suppose pairs D, E and F have characteristics {(0.2,0.6,0.2), (0.2,0.6,0.2)}, {(0.2,0.6,0.2), (7/30,11/30,0.4)}, and {(0.2,0.6,0.2), (1/6,5/6,0.0)} respectively. Again, all of these three players in club Y are very different from each other. All of the pairs A, B and C are better 4BBB pairs than all of the pairs D, E and F. Each club needs to select one pair to play for it. Table 1 gives, for every possible match, the probability that the pair from club X wins, ties, loses a hole, the probability that the pair from club X wins, ties, loses an 18-holes 4BBB match, and the probability that it wins, and loses a match if there is a hole-by-hole play-off when the match is square at the 18th hole. Table 1 also gives the values of R and θ_0.

It can be seen from table 1 that the probability that club X wins an 18-holes 4BBB match with a hole-by-hole play-off if necessary, for example, ranges from 0.62 to 0.67 (a 'moderate' range), depending on each club's selection. It can be seen that club X's best strategy is to use pair C always, no matter which pair is used by club Y. Club Y's best strategy is to use pair E always, no matter which pair is used by club X. Thus, the 'better' club's best strategy is to make use of the very consistent player with characteristics (0.2, 0.8, 0.0), whilst the 'weaker' club's best strategy is to make use of their 'most variable' player with the highest birdie probability (and lowest expected value). The ordering from best to worst for club X is C, A, B, and for club Y it is E, D, F. Thus, it is better for the stronger club to combine its 'middle-birdie' player (0.25, 0.5, 0.25) with the very 'steady' player (0.2, 0.8, 0.0), rather than with the very variable player (0.3, 0.2, 0.5). The reverse is true for the weaker club. Note that the best selection for club X (club Y) has the highest (lowest) values in that column (row) for θ_0 and R. (It is noted here (and throughout this paper) that R is always very slightly greater than θ_0 although this is typically not observable when these values are expressed to two decimal points). Further, it can be seen that, in this example, it does not help one team to know in advance which pair is to be selected for the other club. Note that, using a continuity argument, a pair E* that is very slightly worse than E (and hence worse than both D and F) but with similar characteristics to E, would still be the best pair for club Y to select when playing club X. A corresponding statement can be made about club X with regard to selecting a pair C* that is very slightly worse than pair C. Thus, it can be concluded that, in the presence of wide divergence in player characteristics, a club need not necessarily select its best pair when playing another club, and that the strategy for the stronger club can be different to that of the weaker club. It should be noted, however, that the characteristics of pairs A, B and C are very different indeed from each other, as are those of pairs D, E and F. It would appear that differences of this magnitude and type are unlikely to occur in practice. Thus, it would appear that, in the absence of any major differences in the characteristics of their players, clubs should select their best pairs.

Table 1: The hole probabilities (X wins, tie, X loses), the match probabilities (X wins, tie, X loses), the match probabilities (X wins, X loses possibly after a play-off), and R and θ_0, for the nine cases when club X is represented by pair A, B or C, and club Y by pair D, E or F.

Hole probabilities Match probabilities 1 Match probabilities 2 R ; θ_0	D (0.2,0.6,0.2), (0.2,0.6,0.2)	E (0.2,0.6,0.2), (7/30,11/30,0.4)	F (0.2,0.6,0.2), (1/6,5/6,0.0)
A (0.25,0.5,0.25),	0.30, 0.46, 0.24	0.31, 0.44, 0.25	0.29, 0.48, 0.23
(0.25,0.5,0.25)	0.58, 0.12, 0.31	0.57, 0.12, 0.31	0.58, 0.12, 0.30
	0.64, 0.36	0.63, 0.37	0.65, 0.35
	0.22, 0.22	0.21, 0.21	0.24, 0.24
B (0.25,0.5,0.25),	0.32, 0.42, 0.26	0.32, 0.41, 0.27	0.32, 0.43, 0.26
(0.3,0.2,0.5)	0.56, 0.12, 0.32	0.56, 0.12, 0.33	0.57, 0.12, 0.31
	0.63, 0.37	0.62, 0.38	0.63, 0.37
	0.19, 0.19	0.18, 0.18	0.20, 0.20
C (0.25,0.5,0.25),	0.28, 0.50, 0.22	0.29, 0.47, 0.23	0.27, 0.53, 0.20
(0.2,0.8,0.0)	0.59, 0.12, 0.29	0.58, 0.12, 0.30	0.60, 0.13, 0.28
	0.66, 0.34	0.65, 0.35	0.67, 0.33
	0.26, 0.26	0.24, 0.23	0.29, 0.29

Example 2. It is assumed that each club in example 1 is required to select two pairs to play for it. Table 2 lists, for all possible ways in which the two clubs can meet, three distributions in each cell of the table. The first cell in the body of the matrix is for the case in which A plays D, and B plays E (i.e. two matches). The first row in that cell is the distribution of win (by 2)/draw/loss (by 2) to club X, given each individual match is a win or a loss (after a possible play-off) to club X. The second row in that cell is the distribution of win-by-2/win-by-1/draw/loss-by-1/loss-by-2 to club X, given each individual match is a win/draw/loss to club X. The third row in that cell in obtained from the second row, and is simply the corresponding win/draw/loss to club X. The values of R and θ_0 are for the convolution of the two modules of play (i.e. the sum of the respective Z_i variables). The other 17 cells in table 2 have a similar description.

Several observations can be made from table 2. Firstly, the pairs that are selected for each club, and the way in which the pairs meet, affect the probability that club X wins the overall match. Secondly, no matter who plays for club Y, club X's strategy of playing A and C is better than playing B and C, and this is better than playing A and B. Correspondingly for club Y, their results are best for D and E, followed by E and F, followed by D and F. These results are consistent with those in example 1.

In the first cell of table 3 there are the results for a match (consisting of two or four individual 18-holes matches) between pairs A and B, and pairs D and E. For the first 2 rows within that cell, the underlying win/draw/loss probabilities (for pair A or B) have been used. The first of these 2 rows gives the win/draw/loss outcome probabilities given the reverse matches were played (i.e. A versus D, and B versus E, followed by A versus E, and B versus D). In brackets is this distribution with the drawn result randomly allocated to each club. The second row gives the results for a match between these two clubs where pair A plays pair D or E at random, and pair B plays the other pair, there being just two 18-holes matches in this case. In brackets is this distribution with the drawn result randomly allocated to each club. Rows 3 and 4 within the first cell repeat the results of rows 1 and 2 using the underlying win/loss distribution for each 18-holes event with play-offs. The fifth row gives values for R and θ_0 for the convolution of the four relevant modules. It can be seen that, under all of the above formats (reversing the pairings or the pairs meeting at random; and using underlying win/draw/loss or win/loss outcomes) the best selection for club X is pairs C and A, and for club Y it is pairs E and D. For each of the two clubs, these pairs were the best two pairs to select in example 1. Note that the best selection for club X (club Y) has the highest (lowest) values in that column (row) of θ_0 and R. As can be seen elsewhere in this paper, the better club is more likely to be identified if play-offs are not used for individual matches. Again, it would appear that, in the absence of any major differences in the characteristics of their players, clubs should select their best pairs.

Table 2: The match probabilities (X wins by 2, draw, X loses by 2) given each individual match is a win or a loss to X after a possible play-off; the match probabilities (X wins by 2, X wins by 1, draw, X loses by 1, X loses by 2) given each individual match is a win, a tie or a loss to X; the match probabilities (X wins, draw, X loses) given each individual match is a win, a tie or a loss to X; R and θ_0, when clubs X and Y are each represented by two pairs.

3 Dist^{n}s R ; θ_0	D and E in order	E and F in order	D and F in order
A and B	0.40, 0.47, 0.14	0.40, 0.46, 0.13	0.41, 0.46, 0.13
in order	0.32, 0.13, 0.37, 0.07, 0.10	0.32, 0.13, 0.37, 0.07, 0.10	0.33, 0.14, 0.37, 0.07, 0.10
	0.45, 0.37, 0.17	0.46, 0.37, 0.17	0.46, 0.37, 0.17
	0.20; 0.20	0.21; 0.21	0.21; 0.21
B and A	0.40, 0.47, 0.14	0.40, 0.46, 0.13	0.41, 0.46, 0.13
in order	0.32, 0.13, 0.37, 0.07, 0.10	0.32, 0.14, 0.37, 0.07, 0.10	0.33, 0.14, 0.37, 0.07, 0.10
	0.45, 0.37, 0.17	0.46, 0.37, 0.17	0.46, 0.37, 0.17
	0.20; 0.20	0.21; 0.21	0.22; 0.22
B and C	0.41, 0.46, 0.13	0.42, 0.46, 0.13	0.42, 0.46, 0.12
in order	0.33, 0.14, 0.37, 0.07, 0.10	0.33, 0.14, 0.36, 0.07, 0.09	0.34, 0.14, 0.36, 0.07, 0.10
	0.46, 0.37, 0.17	0.47, 0.36, 0.16	0.48, 0.36, 0.16
	0.21; 0.21	0.23; 0.23	0.24; 0.23
C and B	0.41, 0.46, 0.13	0.41, 0.46, 0.13	0.42, 0.46, 0.13
in order	0.33, 0.14, 0.37, 0.07, 0.09	0.33, 0.14, 0.37, 0.07, 0.09	0.33, 0.14, 0.36, 0.07, 0.09
	0.46, 0.37, 0.17	0.47, 0.37, 0.17	0.47, 0.36, 0.16
	0.22; 0.22	0.22; 0.22	0.23; 0.23
A and C	0.42, 0.46, 0.13	0.42, 0.45, 0.12	0.43, 0.45, 0.12
in order	0.33, 0.14, 0.36, 0.07, 0.09	0.34, 0.14, 0.36, 0.07, 0.09	0.34, 0.14, 0.36, 0.07, 0.08
	0.47, 0.36, 0.16	0.48, 0.36, 0.16	0.49, 0.36, 0.16
	0.23; 0.23	0.24, 0.24	0.25; 0.25
C and A	0.42, 0.46, 0.13	0.42, 0.46, 0.12	0.43, 0.45, 0.12
in order	0.33, 0.14, 0.36, 0.07, 0.00	0.34, 0.14, 0.36, 0.07, 0.09	0.34, 0.14, 0.36, 0.07, 0.09
	0.47, 0.36, 0.16	0.48, 0.36, 0.16	0.49, 0.36, 0.16
	0.23; 0.23	0.24; 0.24	0.25; 0.25

Table 3: The match probabilities (X wins, draw, X loses) given the underlying matches are win/tie/loss and the reverse matches are played; the match probabilities (X wins, draw, X loses) given the underlying matches are win/tie/loss and reverse matches are not played; the match probabilities (X wins, draw, X loses) given the underlying matches are win/loss (after possible play-offs) and the reverse matches are played; the match probabilities (X wins, draw, X loses) given the underlying matches are win/loss and reverse matches are not played; R and θ_0, when clubs X and Y are each represented by two pairs. The probabilities in brackets are for (X wins, X loses), with the relevant draw randomly allocated to each team.

4 Dist^{n}s	D and E	E and F	D and F
R, θ_0			
(A and B),	0.59, 0.22, 0.19 (0.70, 0.30)	0.60, 0.22, 0.18 (0.71, 0.29)	0.60, 0.22, 0.18 (0.71, 0.29)
	0.45, 0.37, 0.17 (0.64, 0.36)	0.46, 0.37, 0.17 (0.64, 0.36)	0.46, 0.37, 0.17 (0.65, 0.35)
(B and A)	0.53, 0.33, 0.15 (0.69, 0.31)	0.54, 0.32, 0.14 (0.70, 0.30)	0.54, 0.32, 0.14 (0.70, 0.30)
	0.40, 0.47, 0.14 (0.63, 0.37)	0.40, 0.46, 0.13 (0.63, 0.37)	0.44, 0.46, 0.13 (0.64, 0.36)
	0.20, 0.20	0.21, 0.21	0.21, 0.21
(B and C),	0.60, 0.22, 0.18 (0.71, 0.29)	0.61, 0.21, 0.17 (0.72, 0.28)	0.62, 0.21, 0.17 (0.73, 0.28)
	0.46, 0.37, 0.17 (0.65, 0.35)	0.47, 0.37, 0.17 (0.65, 0.35)	0.48, 0.36, 0.16 (0.66, 0.34)
(C and B)	0.54, 0.32, 0.14 (0.70, 0.29)	0.55, 0.32, 0.13 (0.71, 0.29)	0.56, 0.31, 0.13 (0.71, 0.29)
	0.41, 0.46, 0.13 (0.64, 0.36)	0.41, 0.46, 0.13 (0.64, 0.36)	0.42, 0.46, 0.12 (0.65, 0.35)
	0.22, 0.21	0.22, 0.22	0.23, 0.23
(A and C),	0.62, 0.21, 0.17 (0.72, 0.28)	0.62, 0.21, 0.17 (0.73, 0.27)	0.64, 0.21, 0.16 (0.74, 0.26)
	0.47, 0.36, 0.16 (0.65, 0.35)	0.48, 0.36, 0.16 (0.66, 0.34)	0.49, 0.36, 0.16 (0.66, 0.34)
(C and A)	0.55, 0.31, 0.13 (0.71, 0.29)	0.56, 0.31, 0.13 (0.72, 0.28)	0.57, 0.31, 0.12 (0.72, 0.28)
	0.42, 0.46, 0.13 (0.64, 0.36)	0.42, 0.46, 0.12 (0.65, 0.35)	0.43, 0.45, 0.12 (0.65, 0.35)
	0.23, 0.23	0.24, 0.24	0.25, 0.25

Example 3. Suppose club X has 4 symmetric players A_i with characteristics (p_i, $1-2p_i$, p_i), (i = 1, 2, 3 and 4), where $p_1 = 0.25$, $p_2 = 0.2$, $p_3 = 0.15$ and $p_4 = 0.1$, and club Y has four players B_i with identical characteristics. Recall that player A_1 is a better 4BBB player than A_2, A_2 is a better 4BBB player than A_3, and A_3 is a better 4BBB player than A_4. The results for the various possible matches are given in table 4. The first row within the first cell gives the win/draw/loss probabilities for club X (after the reverse matches are played) when the underlying results are win/tie/loss for the individual 4 matches. The next row within the first cell gives the same probabilities for club X when there are only two matches played at random, and the underlying results are win/tie/loss for the individual 2 matches. Rows 3 and 4 within the first cell are a repeat of rows 1 and 2 when the underlying results are win/loss (possibly after play-offs).

It can be seen that club X's best strategy is to pair players A_1 and A_4 (and hence A_2 with A_3). Correspondingly, club Y's best strategy is to combine B_1 with B_4 (and B_2 with B_3). It can be seen by a continuity argument that if the parameters for club X and club Y were to change by a small amount, the above optimal combinations would remain optimal. More generally, when two 'roughly equal' clubs compete, it is typically best for each of them to 'balance' their players by playing their 'best 4BBB' player with their 'worst', etcetera. These strategies may no longer be optimal if the clubs are 'moderately unequal'.

An example of this 'unequal' case is considered in table 5. The values of p_3 and p_4 for club Y were modified to 0.05 and 0.00 respectively, whilst the other six p-values in table 4 were left unaltered. It can be seen (in this example and more generally) that, whilst the better club's best strategy is to 'balance' the individual players, the weaker club's best strategy (in terms of increasing the likelihood of a draw and decreasing the probability of the better team winning) can be to put their best players together, especially in the case when the pairs from the clubs meet at random.

Example 4: (Ryder Cup) Table 6 gives the performances of the 12 players in the 2006 US Ryder Cup team over the 2006 PGA Tour season (Hurley, 2007). The 'ranges' for the birdie statistics are from the 2nd biggest to the 2nd smallest value. This was done so that the ranges were not particularly dependent on any extreme values. Using the methods of Examples 3a and 3b in the first of these papers, these players can be ranked (approximately) from the best to the worst 4BBB player. This order is as they appear in Table 7.

Table 4: An example of 'balancing' the pairs within the club when two equal clubs meet. The win/draw/loss probabilities for club X (after the reverse matches are played) when the underlying results are win/tie/loss for the individual 4 matches; the win/draw/loss probabilities for club X when there are only two matches played at random, and the underlying results are win/tie/loss for the individual 2 matches, and these two sets of win/draw/loss probabilities when the underlying results are win/loss (possibly after play-offs).

	$\{(B_1,B_2),(B_3,B_4)\}$	$\{(B_1,B_3),(B_2,B_4)\}$	$\{(B_1,B_4),(B_2,B_3)\}$
$\{(A_1,A_2),(A_3,A_4)\}$	0.35,0.30,0.35	0.33,0.28,0.39	0.32,0.28,0.40
	0.25,0.49,0.25	0.25,0.46,0.29	0.25,0.45,0.30
	0.29,0.42,0.29	0.26,0.41,0.33	0.25,0.41,0.34
	0.20,0.60,0.30	0.20,0.57,0.24	0.19,0.56,0.25
$\{(A_1,A_3),(A_2,A_4)\}$	0.39,0.28,0.33	0.37,0.26,0.37	0.36,0.26,0.38
	0.29,0.46,0.25	0.29,0.42,0.29	0.29,0.41,0.30
	0.33,0.41,0.26	0.31,0.39,0.31	0.30,0.38,0.32
	0.24,0.57,0.20	0.24,0.53,0.24	0.24,0.52,0.25
$\{(A_1,A_4),(A_2,A_3)\}$	0.40,0.28,0.32	0.38,0.26,0.36	0.37,0.25,0.37
	0.30,0.45,0.25	0.30,0.41,0.29	0.30,0.39,0.39
	0.34,0.41,0.25	0.32,0.38,0.30	0.33,0.38,0.31
	0.25,0.56,0.19	0.25,0.52,0.24	0.25,0.50,0.25

Table 5: An example in which the weaker club is better not to 'balance' their players. The win/draw/loss probabilities for club X (after the reverse matches are played) when the underlying results are win/tie/loss for the individual 4 matches; the win/draw/loss probabilities for club X when there are only two matches played at random, and the underlying results are win/tie/loss for the individual 2 matches, and these two sets of win/draw/loss probabilities when the underlying results are win/loss (possibly after play-offs).

	$\{(B_1,B_2),(B_3,B_4)\}$	$\{(B_1,B_3),(B_2,B_4)\}$	$\{(B_1,B_4),(B_2,B_3)\}$
$\{(A_1,A_2),(A_3,A_4)\}$	0.63,0.31,0.05	0.58,0.25,0.18	0.57,0.25,0.19
	0.39,0.57,0.04	0.42,0.43,0.15	0.42,0.42,0.16
	0.56,0.41,0.03	0.50,0.36,0.13	0.49,0.36,0.14
	0.33,0.65,0.02	0.35,0.54,0.11	0.35,0.53,0.12
$\{(A_1,A_3),(A_2,A_4)\}$	0.64,0.32,0.04	0.63,0.21,0.16	0.62,0.21,0.17
	0.41,0.56,0.03	0.47,0.38,0.15	0.47,0.37,0.16
	0.56,0.42,0.02	0.56,0.32, 0.12	0.56,0.32,0.12
	0.34,0.64,0.01	0.41,0.49,0.11	0.41,0.47,0.12
$\{(A_1,A_4),(A_2,A_3)\}$	0.64,0.32,0.03	0.65,0.21,0.15	0.64,0.20,0.16
	0.41,0.56,0.03	0.49,0.37,0.14	0.49,0.35,0.16
	0.56,0.42,0.02	0.58,0.31,0.11	0.58,0.30,0.12
	0.35,0.64,0.01	0.43,0.47,0.11	0.43,0.45,0.12

The top ranked 4BBB players are 1. Woods 2. Mickelson 3=. Furyk, Wetterich 5. Taylor 6. Toms 7. Verplank 8. Cink, so, on the basis of the 2006 PGA Tour statistics, these 8 players would probably be the best to select as the 4BBB players against a team that was no better than the US. (It is noted that there are ${}^{12}C_8 = 495$ ways of selecting the 8 players who play on one day of 4BBB matches, and there is 105 ways of arranging those 8 players into 4 unordered pairs, making a total of 51,975 ways in which the US can field 4 unordered pairs of 4BBB pairs on a particular day.)

It can be seen from Table 7 that the top 2 players are somewhat ahead of the next 4 players who are somewhat grouped together, but who in turn are ahead of the seventh and eighth ranked players. In order to balance out these 8 players, Woods should play with Cink or (say) Verplank, and Mickelson play with Verplank or Cink respectively. It can be shown that Woods with Verplank against Mickelson and Cink is a more even match than when the pairings are reversed. Considering the four players ranked 3 to 6, it can be shown that the most evenly matched pairs are Furyk and Toms, and Wetterich and Taylor, as may have been anticipated. Thus, based on these statistics and this analysis, we have a balanced team.

Table 6: The US Ryder Cup players' performances over the 2006 PGA Tour season.

Rel frequency	Par-3-holes birdie,par,bogey	Par-4-holes birdie,par,bogey	Par-5-holes eagle,birdie,par
Tiger Woods	0.15, 0.68, 0.17	0.22, 0.64, 0.14	0.05, 0.53, 0.42
Phil Mickelson	0.16, 0.63, 0.21	0.19, 0.63, 0.18	0.04, 0.59, 0.37
Jim Furyk	0.16, 0.72, 0.13	0.18, 0.67, 0.15	0.00, 0.51, 0.49
Chad Campbell	0.13, 0.74, 0.14	0.16, 0.62, 0.22	0.03, 0.47, 0.50
David Toms	0.14, 0.67, 0.19	0.19, 0.60, 0.21	0.03, 0.43, 0.54
Chris DiMarco	0.14, 0.66, 0.20	0.16, 0.62, 0.22	0.01, 0.40, 0.58
Vaugn Taylor	0.13, 0.68, 0.20	0.19, 0.57, 0.24	0.03, 0.53, 0.44
JJ Henry	0.13, 0.69, 0.18	0.16, 0.63, 0.22	0.02, 0.44, 0.53
Zach Johnson	0.11, 0.71, 0.17	0.17, 0.63, 0.20	0.02, 0.37, 0.61
Brett Wetterich	0.13, 0.65, 0.22	0.19, 0.59, 0.22	0.03, 0.51, 0.46
Stewart Cink	0.08, 0.78, 0.15	0.18, 0.61, 0.21	0.04, 0.54, 0.43
Scott Verplank	0.16, 0.70, 0.15	0.17, 0.64, 0.19	0.02, 0.41, 0.57
Average	0.14, 0.69, 0.17	0.18, 0.62, 0.20	0.03, 0.48, 0.50
Birdie 'range'	0.16-0.11 = 0.05	0.19-0.16 = 0.03	0.54-0.40 = 0.13

Table 7: Calculation of the approximate individual player 4BBB rankings based on the PGA Tour data.

(A, A) Vs (A, B) and (B, B) Vs (B, A)	P('A' wins) / P(18 holes squared) / P('B' wins)
A is Woods; B is Mickelson	0.49/0.13/0.37 and 0.50/0.13/0.37
A is Mickelson; B is Furyk	0.47/0.14/0.40 and 0.48/0.14/0.38
A is Furyk; B is Wetterich	0.43/0.14/0.43 and 0.44/0.14/0.43
A is Wetterich; B is Taylor	0.44/0.14/0.43 and 0.44/0.14/0.43
A is Taylor, B is Toms	0.44/0.14/0.43 and 0.44/0.13/0.42
A is Toms, B is Verplank	0.45/0.13/0.41 and 0.45/0.14/0.41
A is Verplank; B is Cink	0.49/0.14/0.37 and 0.50/0.14/0.36
A is Cink, B is Campbell	0.44/0.14/0.41 and 0.45/0.14/0.41
A is Campbell; B is Henry	0.44/0.14/0.42 and 0.44/0.14/0.42
A is Henry, B is DiMarco	0.44/0.14/0.42 and 0.44/0.14/0.42
A is DiMarco, B is Johnson	0.44/0.14/0.43 and 0.44/0.14/0.42

Example 5: It can be seen that the 'range' on par 5 holes for the US Ryder Cup players was considerably larger than the 'range' for the par 3 and par 4 holes. Although there are typically only four par 5 holes whilst there are typically ten par 4 holes, the bigger 'range' for the par 5 holes gives them the potential to be particularly useful holes to those players who are relatively very good on the par 5s. To study this effect, we express all numbers to 4 decimal places (whereas in all the other tables, for the ease of the reader the numbers have been expressed to just 2 decimal places), and considered player A with par 3, 4 and 5 parameters (0.1466, 0.6788, 0.1746), (0.1784, 0.6210, 0.2006) and (0.0272, 0.4775, 0.4953), player B with respective parameters (0.1350, 0.6904, 0.1746), (0.1867, 0.6127, 0.2006), (0.0272, 0.4775, 0.4953), player C with parameters (0.1350, 0.6904, 0.1746), (0.1784, 0.6210, 0.2006), (0.0272, 0.5104, 0.4624), and player D with parameters (0.1350, 0.6904, 0.1746), (0.1784, 0.6210, 0.2006), (0.0272, 0.4775, 0.4953). Here player D has the average parameter values in table 6, player A has a larger than average par 3 birdie probability, player B has a larger than average par 4 birdie probability, and player C has a larger than average par 5 birdie probability. In each case the larger than average component was one-quarter of the 'range'. The win/tie/loss probabilities for A and D Vs D and D were 0.4363/0.1380/0.4256, for B and D Vs D and D were 0.4400/0.1380/0.4220, and for C and D Vs D and D were 0.4398/0.1374/0.4229, indicating that the four par 5 holes for 'above average' par 5 players can be almost as 'important' as the ten par 4 holes for 'above average' par 4 players. It is less valuable to be an 'above average' par 3 player, as the 'range' for par 3s is small and there are typically only four par 3 holes. Further, it can be shown that the 18th hole is more 'important' than the 17th, which is more important than the 16th, etcetera (Morris, 1977). Thus, an 18th hole which is also a par 5 hole is very 'important' indeed for player with statistics similar to player C above.

Example 6: It can be shown that when using the '(2*mean)/variance' or the 'θ_0 method' in the situation in which the par 3, par 4 and par 5 holes have different characteristics (for Z_i), the first 'module' criterion becomes '(2*$\sum$mean)/$\sum$variance', and the second is $\sum\theta_0/18$, where the summation is over the 18 holes.

CONCLUSIONS

The decision as to which four-ball-best-ball (4BBB) pairs should be selected in a team has been considered. The proposition that a club should select its best pairs has been considered. Although this is most often the case, it has been shown to be not necessarily valid. The case in which the two pairs from club X play one match each against club Y has been considered, as has the case with both pairs from club X playing both pairs from club Y ('reverse' matches). It has been shown that the pairs best suited to represent the club can depend on the characteristics of the pairs in the other club. For example, pairs likely to perform best when playing a 'stronger' club can be different from pairs likely to perform best when playing a 'weaker' club. When playing a club of similar standard, it has been shown that it is typically best to pair the players so as to 'balance' their 4BBB ability. Further, it has been shown that, in order to maximize the likelihood of correctly identifying the better team, a match that is tied after 18 holes is typically best left as a tied match, rather than having a play-off.

Statistics from the 2006 PGA Tour were used to identify four pairs of players who might have 'best' represented the US in the Ryder Cup in 2006. The characteristics of the twelve players in the 2006 US Ryder Cup team suggested that their performances on the par 5 holes in particular can be very important in selecting the pairs. Information on players' performances on par 3, par 4 and par 5 holes (along with those of their opponents) has been shown to be sufficient in determining the better of two pairs to represent a club against an opponent pair. That is, it is not necessary to do a full 18-holes analysis of such possible matches.

References

Hurley, W. (2007) The Ryder Cup: Are balancing four-ball pairings optimal? *Journal of Quantitative Analysis in Sports.* **3(4):** Article 6, 1-18.

Morris, C. (1977) The most important points in tennis. In S. Ladany and R. Machol (eds.) *Optimal Strategies in Sports.* Amsterdam: North-Holland, 131-140.

Pollard, G. and Pollard, G. (2008) Four ball best ball 1. In J. Hammond (ed.) *Proceedings of the Ninth Australiasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM) Coolangatta, Queensland, Australia (in press).

MOTION ANALYSIS IN SPORTS WITH JAVA SUN SPOTS EVALUATING TRAINING SUCCESS VERIFIED BY OPTICAL MOTION CAPTURING

(ABSTRACT ONLY)

Koehler, H. [1], Stein, T. [2], Fischer, A. [2], Schwameder, H. [2], Woerner, A. [1]

[1] Universitaet Karlsuhe (TH), Institute for Algorithms and Cognitive Systems, Karlsruhe, Germany

[2] Universitaet Karlsruhe (TH), Institute for Sports and Sport Science, Karlsruhe, Germany

Abstract Submitted: 31 January 2008
Presentation accepted: 28 March 2008

Due to sport analysis being an expensive and immobile high performance application made for professional sports, a permanently installed professional training lab is required for vision based motion analysis, today. At this background these tools are quite unusable for educational or leisure sports purpose, although available hardware sensors decrease rapidly in price while precision and quality increases.

To achieve an easy to use motion evaluation application we implemented a system based on Java Sun SPOT hardware. These are small light-weight sensor devices including a 3D accelerometer which is able to collect the related acceleration data and to send them via Bluetooth to a base station for evaluation and visualisation. The developed software is enabled to evaluate the captured data in two ways. The capturing, visualisation and storage of common training moves and kicks of a person as well as its statistical analysis including chronological relation and the technical interpretation is possible.

Additional kinematic data of experts are captured by an optical motion tracking system generating reference values representing an optimal motion trajectory. This trajectory can serve as an ideal representation of a certain movement. Therefore it can be used as a reference move to be compared directly with the movements of the learner by an additional simulation. The system is approved in the fields of pitches and strike movements. The application is evaluated by a specific test group representing typical beginners and advanced sportsmen as well as professional athletes.

It can be obtained from the results that exact measurement and evaluation still depends on the quality of the used hardware sensors and a well predefined training environment. Overall we are able to show the system robustness as well as the handling and affordances of the hardware and software enabling the application to support a wide range of educational levels in sports.