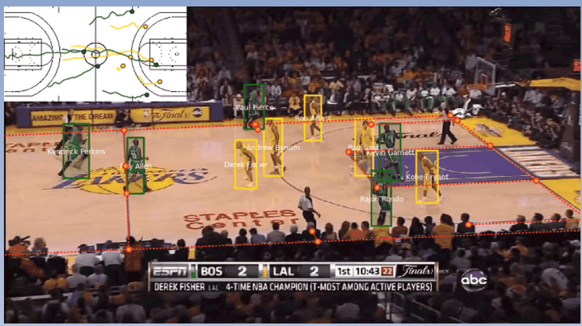
Proceedings of the 12th Australasian Conference on

MATHEMATICS AND COMPUTERS IN SPORT

[EDITED BY A BEDFORD & T HEAZLEWOOD]



June 25th-27th 2014
Darwin, Northern Territory

THE PROCEEDINGS OF THE 12TH AUSTRALIASIAN CONFERENCE ON MATHEMATICS AND COMPUTERS IN SPORT

Edited by

Anthony Bedford and Tim Heazlewood

JUNE 25 TH -27TH

DARWIN, NORTHERN TERRITORY, AUSTRALIA

Published by MathSport (ANZIAM)

ISBN: 978-0-9925475-1-6

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

CONFERENCE DIRECTOR'S REPORT

We welcome you to Darwin for the 12^{TH} Australasian conference on mathematics and computers in sport. With a variety of talks and papers ranging from scheduling, simulation, and ratings, to visual and applied sports science, it promises to be a stimulating conference.

We look forward to presentations for Australia's leading researchers, and thank our day one and two keynote speakers from the Australian Institute of Sport, Dr Nick Brown and Dr Stuart Morgan and on day three, Professor Ray Stefani.

We encourage all delegates to engage in the Panel Sessions as we steer towards delivering cutting edge outcomes for sport through our collective research pathways.

Thanks go to ANZIAM, the MathSport Executive, our team of peer reviewers, our hard working helpers and finally, the delegates.

Kindest Regards

Anthony and Tim

Conference Co-Chairs

CONTENTS

KEY ISSUES IN HIGH PERFORMANCE ANALYSIS IN SPORT

Nick Brown

Performance Science and Innovation, Australian Institute of Sport

Abstract

In this presentation, I will share examples of where the AIS currently utilises mathematics and computers for High Performance Sport. I will go over a number of important questions in high performance analysis in sport, and discuss the needs that the Australian Institute of Sport and High Performance sporting have in relation to mathematics, statistics, computers and analytics. I will go through the array of data currently collected and its current use in sport.

I will highlight key research possibilities with the AIS and potential co funding opportunities, and highlight the synergies needed to catalyse work amongst state bodies and sporting organisations.

Keywords: AIS, High Performance

A MOVEMENT SEQUENCING ANALYSIS OF TEAM-SPORT ATHLETE MATCH ACTIVITY PROFILE

Alice J Sweeting[123], Stuart Morgan[3], Stuart Cormack[4] and Robert J Aughey[1]

[1] Institute of Sport, Exercise and Active Living (ISEAL), Victoria University, Melbourne, Victoria, AUSTRALIA; [2] Netball Australia, Melbourne, Victoria, AUSTRALIA; [3] Performance Research Centre, Australian Institute of Sport, Belconnen, Australian Capital Territory, AUSTRALIA; [4] School of Exercise Science, Australian Catholic University, Melbourne, Victoria, AUSTRALIA.

Abstract

Traditional time-motion analysis of athlete physical output, or activity profile, during team-sport matches classifies movement according to pre-defined velocity thresholds. Comparing physical output using standardised thresholds is problematic given the differences in athlete chronological age, weight, playing position and standard. Limited research exists on the combination of velocity, acceleration and angular velocity movements completed by team-sport athletes. Athlete activity profiles were collected during one quarter of a junior-elite netball match, using radio-frequency (RF) tracking. Velocity, acceleration and angular velocity were calculated from raw, individual positional data. Each continuous variable was clustered using1d *k*-means and player movements were discretised with permutations of velocity, acceleration, and angular velocity, and assigned a unique alphabetic label. Continuous sequences of movement units were compared using the Levenshtein distance, and a hierarchical cluster analysis found groups of similar movement patterns. Common shared features in movement strings for each cluster were obtained by computing the longest common substring (LCS). The percentage of all movements represented by the LCS for each cluster was measured for various movement epoch sizes. Eighteen movement sequences were obtained over a 0.5 s epoch. Sprinting in a straight direction with neutral acceleration was a common feature for cluster 1. In contrast, over a 1.5 s epoch, sprinting and accelerating in a straight direction immediately followed by a sprint with neutral deceleration was a common feature for cluster 1. The most frequent combinations of velocity, acceleration and angular velocity movements were derived from empirical sequences of movement units. Future comparison across team-sport athlete playing position and standard, via the density of individual athlete movement features, could be achieved through this analysis and may assist with position-specific coaching and training strategies.

Keywords: k-means, Minimum Description Length,Levenshtein distance, Netball

1.INTRODUCTION

Netball is a predominantly female team sport with a large participation base within Commonwealth countries(Steele & Chad, 1991a).Matches consist of 15 minute quarters and are contested on a 30.5 m by 15.25 m court divided into equal thirds. Players are assigned one of seven positions which restrict movement to specific on-court areas(Woolford & Angove, 1992). The substitution of players is only permitted during quarter and half-time breaks or if an injury time-out is called. The objective of the game is to score a goal through a ring that is 3.05 m above the ground. Netball athletes are not permitted to move more than one step with the ball and when in possession, must pass to a teammate within three seconds.

Quantification of athlete physical movement, or activity profile, during matches is critical in understanding performance. Investigation into athlete match activity profiles can assist with sport-specific preparation and conditioning(Di Salvo et al., 2007; Mendez-Villanueva, Buchheit, Simpson, & Bourdon, 2013).(Di Salvo et al., 2007; Mendez-Villanueva et al., 2013).Examination of netball match-play reveals a combination of short, high intensity movement interspersed with periods of low intensity activity, including walking and jogging(Steele & Chad, 1991a). Early studies on netball activity profile investigated sub-elite athletes(Davidson & Trewartha, 2008; Loughran & O'Donoghue, 1999; Steele & Chad, 1991a; Steele & Chad, 1991b) and were conducted before rule changes to the current length of a match, currently

15 minute quarters(Otago, 1983). Positions were either grouped(Steele & Chad, 1991a), into defender, midcourter or goaler, or combined entirely (Davidson & Trewartha, 2008)in the analysis. Only two studies (Fox, Spittle, Otago, & Saunders, 2013; Otago, 1983) have examined elite netball match activity profile according to individual playing position, using video analysis.

Video analysis is commonly utilised in netball(Davidson & Trewartha, 2008; Fox et al., 2013; Otago, 1983) however, estimating short, high-intensity movement using inferences from visible movement types is error-prone. Micro-technology, including accelerometers(Boyd, Ball, & Aughey, 2011)and global positioning systems or GPS(Jennings, Cormack, Coutts, Boyd, & Aughey, 2010), allow quantification of athlete activity profiles according to physical capacity(Buchheit, Mendez-Villanueva, Simpson, & Bourdon, 2010), chronological age(Mendez-Villanueva et al., 2013), playing standard(Jennings, Cormack, Coutts, & Aughey, 2012) and position.(Mendez-Villanueva et al., 2013). Accelerometer load, as a measure of activity profile, can differentiate between netball playing standard at the sub-elite level (Cormack, Smith, Mooney, Young, & O'Brien, 2013) but remains to be investigated in an elite cohort. The validity and reliability of GPS to measure short high-intensity movements in confined spaces(Duffield, Reid, Baker, & Spratford, 2010) is likely insufficient for netball use (Duffield et al., 2010). Elite netball matches also take place indoors, where GPS is rendered inoperable. The lack of research on netball match activity profile in contemporary athletes, according to position and playing standard, may be attributed to the types of technologies previously available for this analysis.

Recognising the limitations of GPS and video-analysis, radio-frequency (RF) tracking has been developed to monitor athlete activity both indoors and outdoors. The validity and reliability of the method considered, the Wireless ad-hoc System for Positioning or WASP (Hedley et al., 2010), has been established indoors(Sathyan, Shuttleworth, Hedley, & Davids, 2012). At present, RF technology is yet to be deployed in competitive netball matches to quantify match activity profile.

Athlete activity profile is typically analysed using movement thresholds, including velocity bands(Aughey, 2010; Gabbett, Jenkins, & Abernethy, 2012) or arbitrary classifications(Fox et al., 2013). However, comparison between studies is difficult due to the multitude of inconsistent analysis techniques and movement definitions employed(Carling, 2013). Physical output expressed per minute of game time(Varley, Gabbett, & Aughey, 2013) or as a function of physiological capacity(Lovell & Abt, 2012) requires pre-determined parameters to be fitted to data. Using pre-defined thresholds to compare across and between groups is problematic given athlete mass(Gabbett, 2002), playing standard(Jennings et al., 2012), position(Macutkiewicz & Sunderland, 2011) and chronological age(Gastin, Fahrner, Meyer, Robinson, & Cook, 2013) may influence physical output.

Data mining is a problem-solving methodology that sources a logical or mathematical description of patterns and regularities in a data set(Fayyad, Piatetsky-Shapiro, & Smyth, 1996).Whilst data mining techniques can determine the tactical patterns of play during elite volleyball matches(Jäger & Schöllhorn, 2007), determine weight transfer during the golf swing(Ball & Best, 2007) and examine basketball match score outcome(Sampaio & Janeira, 2003), the analysis of athlete match activity profile, using data mining techniques, remains to be explored.

Clustering mines data according to similarity/ dissimilarity and groups items regarding these criteria. Cluster analysis discriminated between high and low inter-personal coordination between soccer players(Morgan & Williams, 2012). Utilised in analysing the performance qualities of elite track cycling athletes to ascertain riders best suited to the omnium event(Ofoghi, Zeleznikow, Dwyer, & Macmahon, 2013), clustering may assist with informing athlete selection, training and strategic planning. Clustering, via self-organising maps (SOM),can provide an objective method to explain movement patterning during basketball shooting(Lamb, Bartlett, & Robins, 2010).However, applying a clustering approach to athlete match activity profile, remains to be explored.

The aim of this study was to develop a movement sequencing technique that exploits the emergent movement characteristics of team-sport athletes. Specifically, to discover the most frequently recurring sequences and create insight into the temporal sequence of movement elements that are representative of netball athlete match activity profile.

2.METHODS

Activity profiles were collected from six female elite-junior netball athletes via RF tracking(Hedley et al., 2010) during a competitive international match. The clustering model was trained on five athletes and tested on the sixth, across the first quarter of play. The sampling rate of the RF system is 1000Hz, divisible by the number of tracking units used during match play. In our sample, 22 units were active (including players and substitutes from two teams), resulting in a sampling rate of approximately 45.5 Hz. Raw athlete position data were downloaded post-match via custom-built software (WhereIsBruce?, Australian Institute of Sport, Canberra, ACT, Australia) and exported into the R environment (R: A language and environment for statistical computing, Vienna, Austria). The elemental movement characteristics for each individual athlete over the first quarter (15 minutes in duration) were calculated in the following way:

Velocity for each player were derived from the position data

$$V_i = \frac{\sqrt{\Delta x^2 + \Delta y^2}}{\Delta t} \qquad (1)$$

Acceleration was derived from velocity.

$$A_i = \frac{V_i - V_{i-1}}{\Delta t} \qquad (2)$$

The angular displacement (θ_i) was calculated from the dot product of consecutive movement vectors, ***a*** and ***b***

$$\theta_i = \cos^{-1}\left[\frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|}\right] \qquad (3)$$

Next, angular velocity (rate of change in angular displacement) was calculated as follows.

$$\omega_i = \frac{\theta_i - \theta_{i-1}}{\Delta t} \qquad (4)$$

In each case, (for Equations 1, 2 and 4), *t* was equal to a time epoch that was varied between separate experimental trials where *t* = 0.5, 0.75, 1.0, 1.25 and 1.50 seconds respectively. The observations for each of these movement characteristics were classified into groups of arbitrary *n*-size using a one-dimensional *k*-means clustering algorithm (Wang & Song, 2011). Four velocity clusters (notionally Walk, Jog, Run, Sprint), three acceleration clusters (Accelerate, Neutral, Decelerate)and four angular velocity clusters (U-Turn, 90 degree turn, 45 degree turn, and Straight)were declared. Figure 1 illustrates the bandwidths represented by each cluster described above. Figure 2 illustrates the relative frequency of each representation in movement classification.

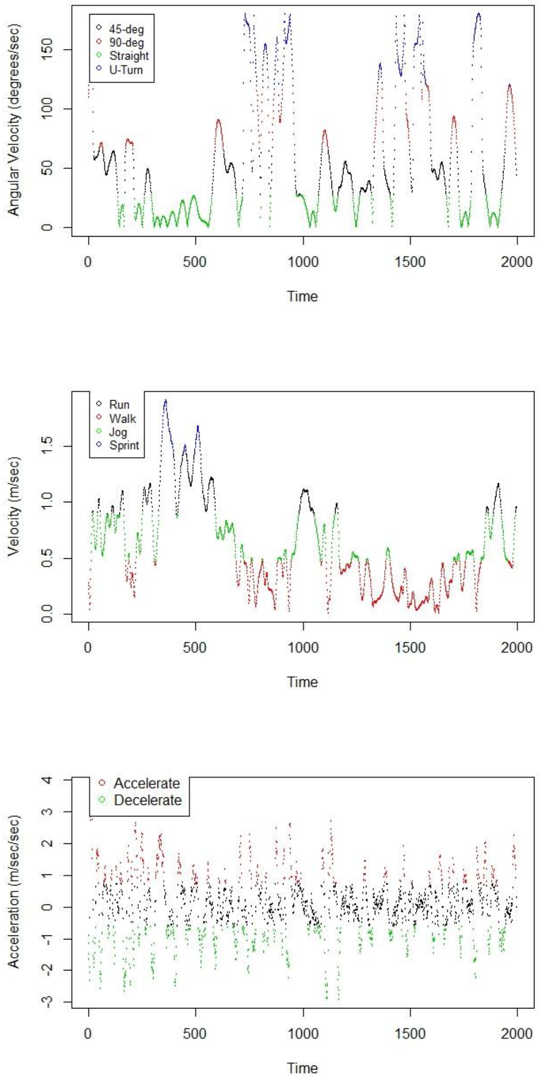
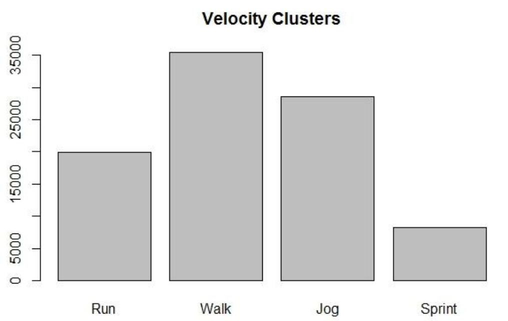


Figure 1. Classification bands representing movement clusters with exemplar data for a) Angular Velocity, b) Velocity, and c) Acceleration.



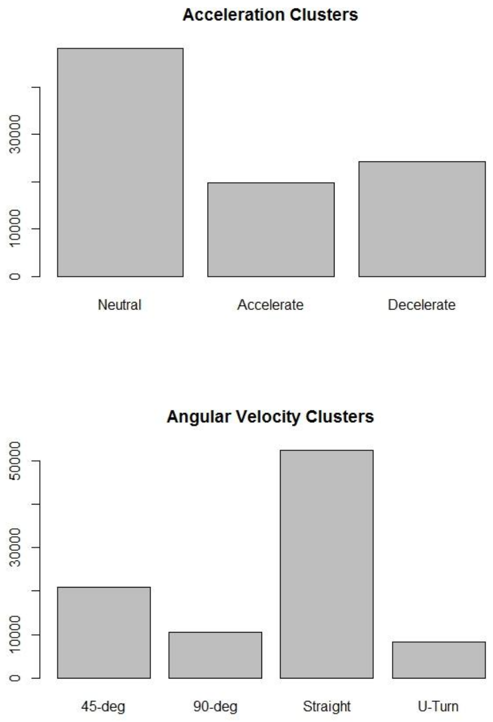


Figure 2. Relative frequency of clustered observations for Velocity, Acceleration and Angular Velocity.

This approach produced 48 permutations (4 x 4 x 3), each of which was described as a unique combination of velocity, acceleration and angular velocity. A permuted identification code (upper and lower case alphabet letters)was assigned to each unique combination of velocity, acceleration and angular velocity. Table 1 lists the specific alphabetic character assigned to each permutation of velocity, acceleration and angular velocity. We refer to these assignments as movement subunits. A frequency distribution of these movement subunits is displayed in Figure 3.

The characteristics of any continuous movement are then represented by a temporal sequence of movement subunits. We further describe any sequence of movement subunits as a discrete movement sequence. Any movement sequence is temporally discrete from other movement sequences where the athlete does not move for the duration equal to the time epoch *t*. In practice it is difficult to identify moments where athletes are motionless in competition, so we applied a movement threshold of 0.5 $m.s^{-1}$, to temporally discretise movement sequences. Additionally, any movement sequence must exceed the movement threshold for at least 1 second (note that this will occur by default where $t \geq 1.0$ seconds).

Character	Movement Subunit	Character	Movement Subunit
a	Run Neutral 45°	m	Run Neutral 90°
A	Run Decelerate Straight	M	Run Decelerate U-Turn
b	Run Accelerate 45°	n	Run Accelerate 90°
B	Walk Neutral Straight	N	Walk Neutral U-Turn
c	Run Decelerate 45°	o	Run Decelerate 90°
C	Walk Accelerate Straight	O	Walk Accelerate U-Turn
d	Walk Neutral 45°	p	Walk Neutral 90°
D	Walk Decelerate Straight	P	Walk Decelerate U-Turn
e	Walk Accelerate 45°	q	Walk Accelerate 90°
E	Jog Neutral Straight	Q	Jog Neutral U-Turn
f	Walk Decelerate 45°	r	Walk Decelerate 90°
F	Jog Accelerate Straight	R	Jog Accelerate U-Turn
g	Jog Neutral 45°	s	Jog Neutral 90°
G	Jog Decelerate Straight	S	Jog Decelerate U-Turn
h	Jog Accelerate 45°	t	Jog Accelerate 90°
H	Sprint Neutral Straight	T	Sprint Neutral U-Turn
i	Jog Decelerate 45°	u	Jog Decelerate 90°
I	Sprint Accelerate Straight	U	Sprint Accelerate U-Turn
j	Sprint Neutral 45°	V	Sprint Decelerate U-Turn
J	Sprint Decelerate Straight	w	Sprint Accelerate 90°
k	Sprint Accelerate 45°	x	Sprint Decelerate 90°
K	Run Neutral U-Turn	y	Run Neutral Straight
l	Sprint Decelerate 45°	z	Run Accelerate Straight
L	Run Accelerate U-Turn		

Table 1. Alphabetical characters for permuted movement subunits.

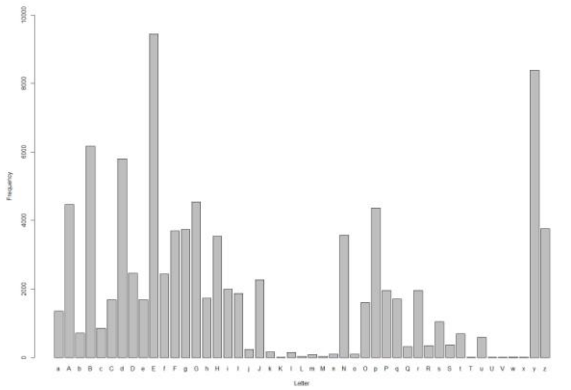


Figure 3. Frequency distribution of movement subunits.

Any period of player movement is now described as a set of movement sequences, where each subunit is characterised by an alphabetic character. Movement sequences were therefore represented by character strings of *k* length, where *k* is the number of composite subunits. It is also possible to quantify the similarity of movement sequences by comparing character strings using the Levenshtein distance(Levenshtein, 1966), which is a function of the minimum number of single-character edits (including insertions, deletions or substitutions) required to change one sequence into another.

3.RESULTS

The means of each of the four velocity clusters, for combined epochs, were 1.12, 0.67, 0.27 and 1.75 $m \cdot s^{-1}$(which we notionally referred to as running, jogging, walking and sprinting respectively). It is important to note that these labels are arbitrary, and in practice is might be better to simply refer to them in such a manner as slow, slow-moderate, moderate, and fast. The means of the three acceleration clusters were 1.41, 0.05 and -1.25 m·s⁻¹.These values are more clearly defined as accelerating, neutral, and decelerating. The means of the three angular velocity clusters were149.68, 11.15, 42.72 and 88.88 deg·s⁻¹.

Movement sequences were generated using strings of character values.We then conducted a cluster analysis using the Ward method (Ward Jr, 1963). All movement strings in our dataset are therefore grouped proximally according to the Levenshtein distance. A sequence analysis, using hierarchical clustering, revealed the most common clusters. A representative example, occurring with an epoch of 0.5 s, is displayed as a dendrogram in Figure 4 (attached).We identified 18 clusters using this method, and an algorithm to find the longest common substring (LCS) (Kuo & Cross, 1989) was utilised to find the longest string that is a substring of two or more strings, within each cluster. The two most common clusters include EEEEE and FEEEE, only one permuted subunit apart. Each cluster was iterated through to find the longest common substring, for each time epoch.The support value for each movement sequence was measured as the percentage of all movements represented by each example. These values were calculated for each of the epoch size.Thisdata is presented in Table 2.

Cluster Number	0.5 s		0.75 s		1 s		1.25 s		1.5 s	
	String	%	String	%	String	%	String	%	String	%
1	AzAy	2	zIycAy	1	yyA	11	IJJIJ	5	IIJJIJ	10
2	FJzyHAE	1	yyyy	19	yyy	28	IHHHJ	10	HHH	40
3	EEE	25	yy	50	HHJ	17	HHHJ	19	IJJ	40
4	EEE	25	yIJ	8	IJJIJy	2	HHHHH	14		
5	yyy	30	IJIJ	3	yyyyyyy	4	IJIJJJ	5		
6	yyIJy	1	yyyyyy	6	JHH	9				
7	hzIycAy	1	HJHH	3	IHHHJ	4				
8	IIJ	3	HHH	15	zAyy	4				
9	ggEE	2	zHA	3	HHA	11				
10	yyyJ	1	E	28	HHHH	11				
11	yyy	30	A	61	IJy	2				
12	EEE	25	zIHHHIIHHJ	1	HAyy	2				
13	zIHHH	1	HAyyy	1	IJIJJJ	2				
14	zIHHHIIHHJg	1	AyyyyyA	1						
15	zyyyyyyyyG	1								
16	aHHHHAz	1								
17	yyyyHHyHH	1								
18	AEEE	3								

Table 2. The most frequently reoccurring movements, per cluster, as a function of epoch stamp and support.

4. DISCUSSION

This study is the first generative contribution to the problem of robust athlete activity-profiling that is independent of age, gender, sport-related constraints, and other features of physical capacity. It is also the first work, to our knowledge, to attempt the development of a movement sequencing technique that can create insight into the temporal sequence of movement elements in sport. Traditional analyses focus on quantifying athlete movement as a function of arbitrary or commercially developed thresholds.

Using a one-dimensional *k*-means clustering algorithm, we were able to identify four velocity clusters, three acceleration clusters and four angular velocity clusters. By permuting elemental features of movement and characterising continuous athlete movement in the form of strings, the LCS sequence analysis approach revealed discrete and recurring combinations of athletic movement, representative of athlete activity typical in netball. In the 0.5 s epoch, running at a straight or 45° angle with neutral and acceleration components was a common feature for cluster 1. In contrast, the 1.5 s epoch showed sprinting and accelerating in a straight direction immediately followed by a sprint with deceleration was a common feature for cluster 1.

Obtaining the most frequently recurring movements of an athlete or a number of athletes grouped according to position or playing standard, may have application for coaching and conditioning purposes. Knowledge of the movements performed, angle of attack and acceleration qualities may assist with planning sport-specific training and conditioning practices. Sprinting, accelerating and decelerating components were a common feature across a 1.5 s epoch for the athlete tested. This data may be used to target specific training qualities within a program. Further analysis could focus on movements performed before a successful or unsuccessful attempt at goal, which may assist with tactical planning. A movement sequencing analysis of athletes according to chronological age, playing standard and position should be investigated in future analyses.

Eighteen clusters were obtained over a 0.5 s epoch in comparison to three clusters over a 1.5 s movement threshold, highlighting the importance of under-fitting versus over-fitting a model. The number of clusters to trim, or focus on, within a dendrogram is an important consideration when analysing athlete movement. For the purpose of this investigation (and the sport examined), we chose to trim at 25 clusters. Further investigation into epoch and trimming selection, dependent upon the sport considered, is warranted.

5.CONCLUSION

A movement sequencing technique was developed to analyse athlete activity profile. Using a one-dimensional *k*-means clustering algorithm, four velocity clusters, three acceleration clusters and four angular velocity clusters were identified. The LCS sequence analysis approach revealed discrete and recurring combinations of athletic movement, representative of athlete activity typical in netball. Eighteen clusters were obtained over a 0.5 s epoch, in contrast to three clusters over 1.5 s, highlighting the importance of under-fitting versus over-fitting a model. The three clusters over 1.5 s reveal a combination of sprinting, acceleration and deceleration qualities in a straight direction. Examining athlete activity profile using this movement sequencing technique, in contrast to traditional analyses, may assist with position specific training and conditioning practices.

REFERENCES

Aughey, R. J. (2010). Australian football player work rate: evidence of fatigue and pacing? *International Journal of Sports Physiology and Performance, 5*(3), 394.

Ball, K., & Best, R. (2007). Different centre of pressure patterns within the golf stroke I: Cluster analysis. *Journal of Sports Sciences, 25*(7), 757-770.

Boyd, L. J., Ball, K., & Aughey, R. J. (2011). The reliability of MinimaxX accelerometers for measuring physical activity in Australian football. *Int J Sports Physiol Perform, 6*(3), 311-321.

Buchheit, M., Mendez-Villanueva, A., Simpson, B., & Bourdon, P. (2010). Match running performance and fitness in youth soccer. *International Journal of Sports Medicine, 31*(11), 818-825.

Carling, C. (2013). Interpreting Physical Performance in Professional Soccer Match-Play: Should We be More Pragmatic in Our Approach? *Sports Medicine*, 1-9. doi: 10.1080/02640414.2011.652655

Cormack, S., Smith, R., Mooney, M., Young, W., & O'Brien, B. (2013). Accelerometer Load as a Measure of Activity Profile in Different Standards of Netball Match Play. *International Journal of Sports Physiology and Performance*.

Davidson, A., & Trewartha, G. (2008). Understanding the physiological demands of netball: A time-motion investigation. *International Journal of Performance Analysis in Sport, 8*(3), 1-17.

Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F., Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *International Journal of Sports Medicine, 28*(3), 222.

Duffield, R., Reid, M., Baker, J., & Spratford, W. (2010). Accuracy and reliability of GPS devices for measurement of movement patterns in confined spaces for court-based sports. *Journal of Science and Medicine in Sport, 13*(5), 523-525.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine, 17*(3), 37.

Fox, A., Spittle, M., Otago, L., & Saunders, N. (2013). Activity profiles of the Australian female netball team players during international competition: Implications for training practice. *Journal of Sports Sciences*. doi: 10.1080/02640414.2013.792943

Gabbett, T. (2002). Physiological characteristics of junior and senior rugby league players. *British Journal of Sports Medicine, 36*(5), 334-339.

Gabbett, T. J., Jenkins, D. G., & Abernethy, B. (2012). Physical demands of professional rugby league training and competition using microtechnology. *Journal of Science and Medicine in Sport, 15*(1), 80-86.

Gastin, P. B., Fahrner, B., Meyer, D., Robinson, D., & Cook, J. L. (2013). Influence of Physical Fitness, Age, Experience, and Weekly Training Load on Match Performance in Elite Australian Football. *The Journal of Strength & Conditioning Research, 27*(5), 1272-1279.

Hedley, M., Mackintosh, C., Shuttleworth, R., Humphrey, D., Sathyan, T., & Ho, P. (2010). Wireless tracking system for sports training indoors and outdoors. *Procedia Engineering, 2*(2), 2999-3004. doi: 10.1016/j.proeng.2010.04.101

Jäger, J. M., & Schöllhorn, W. I. (2007). Situation-orientated recognition of tactical patterns in volleyball. *Journal of Sports Sciences, 25*(12), 1345-1353.

Jennings, D., Cormack, S., Coutts, A. J., Boyd, L., & Aughey, R. J. (2010). The validity and reliability of GPS units for measuring distance in team sport specific running patterns. *International Journal of Sports Physiology and Performance, 5*(3), 328-341.

Jennings, D. H., Cormack, S. J., Coutts, A. J., & Aughey, R. J. (2012). International field hockey players perform more high-speed running than national-level counterparts. *Journal of Strength and Conditioning Research, 26*(4), 947-952.

Kuo, S., & Cross, G. R. (1989). *An improved algorithm to find the length of the longest common subsequence of two strings.* Paper presented at the ACM SIGIR Forum.

Lamb, P., Bartlett, R., & Robins, A. (2010). Self-Organising Maps: An Objective Method for Clustering Complex Human Movement. *International Journal of Computer Science in Sport, 9*(1).

Levenshtein, V. I. (1966). *Binary codes capable of correcting deletions, insertions and reversals.* Paper presented at the Soviet physics doklady.

Loughran, J., & O'Donoghue, P. (1999). Time-motion analysis of work-rate in club netball. *Journal of Human Movement Studies, 36*(1), 37.

Lovell, R., & Abt, G. (2012). Individualisation of Time-Motion Analysis: A Case-Cohort Example. *International Journal of Sports Physiology and Performance*.

Macutkiewicz, D., & Sunderland, C. (2011). The use of GPS to evaluate activity profiles of elite women hockey players during match-play. *Journal of Sports Sciences, 29*(9), 967-973.

Mendez-Villanueva, A., Buchheit, M., Simpson, B., & Bourdon, P. (2013). Match play intensity distribution in youth soccer. *International Journal of Sports Medicine, 34*(02), 101-110.

Morgan, S., & Williams, M. D. (2012). Using relative phase of position derivatives to quantify within-team player couplings.

Ofoghi, B., Zeleznikow, J., Dwyer, D., & Macmahon, C. (2013). Modelling and analysing track cycling Omnium performances using statistical and machine learning techniques. *Journal of Sports Sciences, 31*(9), 954-962.

Otago, L. (1983). A game analysis of the activity patterns of netball players. *Sports Coach, 7*(1), 24-28.

Sampaio, J., & Janeira, M. (2003). Statistical analyses of basketball team performance: understanding teams wins and losses according to a different index of ball possessions. *International Journal of Performance Analysis in Sport, 3*(1), 40-49.

Sathyan, T., Shuttleworth, R., Hedley, M., & Davids, K. (2012). Validity and reliability of a radio positioning system for tracking athletes in indoor and outdoor team sports. *Behavior Research Methods, 44*(4), 1108-1114. doi: 10.3758/s13428-012-0192-2

Steele, J., & Chad, K. (1991a). Relationship between movement patterns performed in match play and in training by skilled netball players. *Journal of Human Movement Studies, 20*(6), 249-278.

Steele, J. R., & Chad, K. E. (1991b). *An analysis of the movement patterns of netball players during match play: implications for designing training programs*: National Sports Research Centre.

Varley, M. C., Gabbett, T., & Aughey, R. J. (2013). Acceleration Profiles in Elite Australian Soccer. *International Journal of Sports Medicine, 34*(1), 34-39. doi: DOI 10.1055/s-0032-1316315

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association, 58*(301), 236-244.

Woolford, S., & Angove, M. (1992). Game intensities in elite level netball: position specific trends. *Sports Coach, 15*, 28-28.

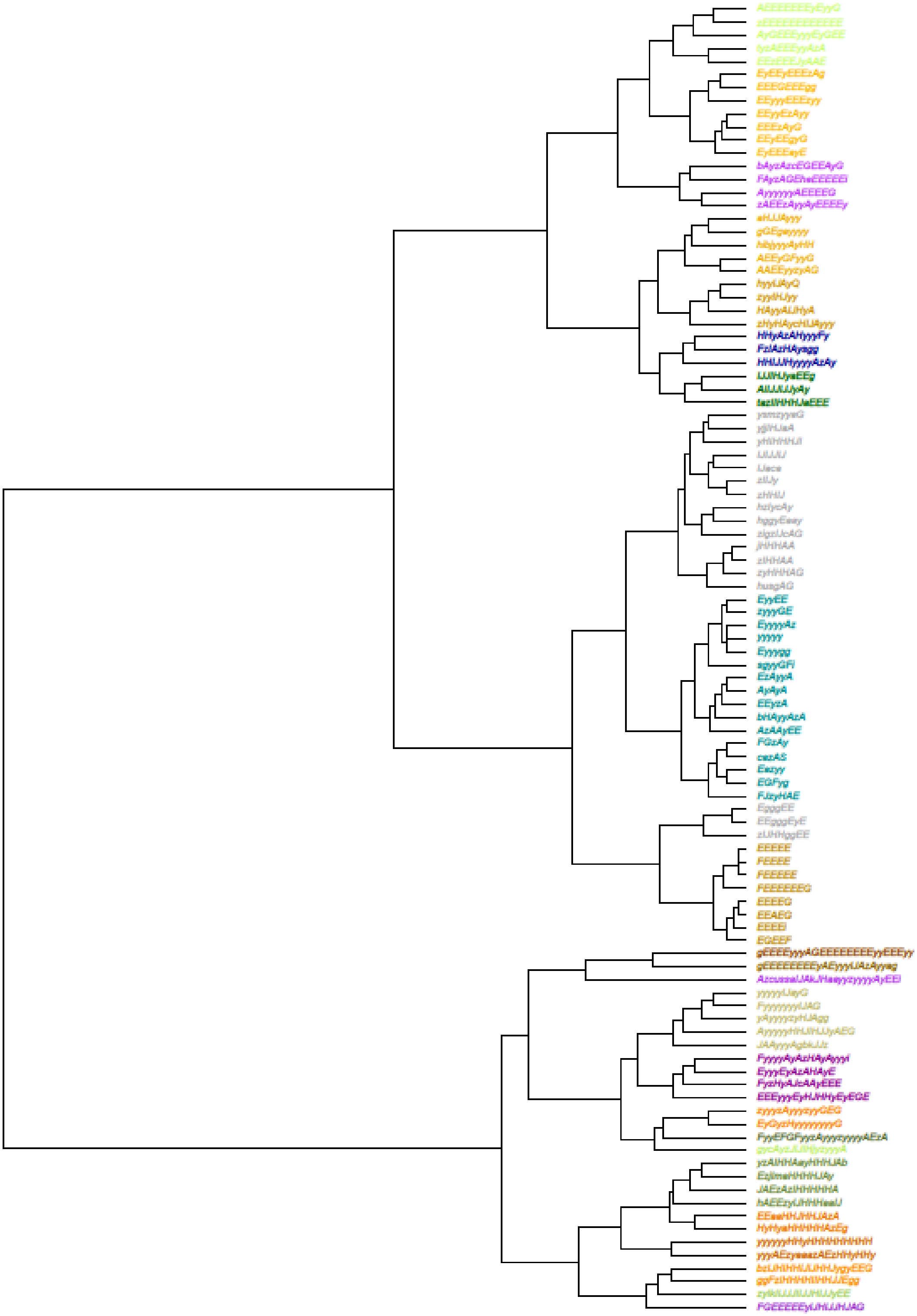


Figure 4. Coloured dendrogram of hierarchical clustering.

ALTERNATIVE NINE TEAM AFL FINALS SYSTEMS

Anthony Bedford and Andrew Gonn

School of Mathematical & Geospatial Sciences
RMIT University, Melbourne, Australia.

Abstract

In this paper we investigate a variety of possible systems for the AFL finals if they were to move to a nine team final series. A number of issues arise that hamper traditional style systems - the largest hindrance the amount of time the final series needs to be played. Further, nine team finals do not lend themselves to an elegant tree like structure. In this paper, we bring together a number of concepts and thrash them through simulations. We consider the results of the systems through the variation of parameters such as scoring and home advantages. We also look at pool type approaches and tree structure models.

Keywords: Finals, simulation, scheduling

1. INTRODUCTION

Prior to 1972 the most common finals system was a four team, three weeks structure. There were twelve teams in the VFL from 1944 until 1986 inclusive.
From 1972 to 1990 inclusive a final five was used, expanding to a final six for three years to 1993. The Macintyre Final 8 was adopted in 1994 and used up to and including 1999. This matched 1st and 8th against each other in Week 1, and 2nd against 7th etc. The two lowest ranked losers would be eliminated in the first week, meaning that individual matches' results did not have predetermined consequences.
Since the year 2000, a new final 8 system replaced the Macintyre Final 8. This was due to a number of deemed unsuitable scenarios in the preceding years. This is still in use in 2014. The current system will not be explained in this project, however the probabilities of teams winning the premiership will be referred to, as calculated by Lowe and Clarke (2000).
With the inclusion of two new clubs in recent years (Gold Coast Suns, Western Sydney Giants) the AFL are again considering adopting a new system, in which there would be nine or ten teams.

This paper focuses on nine team systems, in an attempt to improve the 'fairness' of the current system, which has issues inherent in the model. The criterion for measuring the fairness of any system was discussed by Monahan and Berger (1977) in regards to hockey, and this paper centres around three of their principles;

- Maximise the probability that the highest ranked team wins,
- Maximise the probability that the two highest ranking teams meet in the grand final,
- The probability of a team finishing in any position or higher should be greater than for any lower ranked team (the expected final positions should mirror the original rankings).

We also have the added constraint of a four week window, and the desire to maximise the number of matches to increase revenue potential and fan participation, whilst avoiding meaningless encounters. In this paper, we shall divide the work into the nine and the ten team approaches.

2. METHODS

For the case when all teams considered have equal probability of winning any game, the premiership probabilities for each team can easily be calculated mathematically, and has been done so for a couple of systems. It is common practice to test systems primarily employing the equal probability model. The theory behind it being that a model should be advantageous to higher ranking teams even when all teams are considered equal. But to comprehensively test the robustness of any particular model, a range of probabilities should be explored.

Calculating Premiership probabilities becomes exponentially more difficult when adding complexities such as a superior team, or even home ground advantage. Therefore to calculate the premiership probabilities, models were created on Microsoft Excel and simulated with Simulation 4.0. Two methods were employed for simulation, the first is score based, which generates scores from a given normal distribution, the second binomial. The methods employed can be used to test the fairness of all types of sporting competitions. It is similar to a decision tree, in that at each node only one path can be chosen, hence could be used in that regard, or any where an optimal decision is needed to be calculated.

2.1 Nine team system

2.1.1. The Score System

The purpose of the simulations is to incorporate the different strengths of the different teams to analyse the fairness of the system. For this reason, a number of different potential scores should be tested.
A score was generated for every team and their opponent, for every match, based upon a chosen normal distribution for that team. For the majority of simulations the standard deviation of scores was held constant at 10 points. Some of the systems were simulated with a standard deviation of 20, with the results giving similar distributions as when the standard deviation was 10, although less amplified

The first of two variables which determined the mean for any given simulations was an advantage given to the 'home' team (the team with a higher ranking), where the advantage is f = 0, 3, 6, 9, 12 or 15 points. The second was based on ladder position; the first simulation has all teams equal with a mean of 100, in further simulations the mean scores of the lower placed teams were lowered by equal increments (Table 1). For example, in the second simulation where the increment is half a point, the top team keeps a mean of 100, but the second team has a mean of 99.5 points, the third team 99 points, and so on. In the third simulation the decreasing increment was one point, giving the second team a mean of 99, the third 98, and so on. The increments were g = 0, 0.5, 1, 1.5, 2, 2.5, therefore a total of 36 score simulations were run for each model (f x g = 6 x 6 = 36).

Simulation:	1	2	3	4	5	6
Difference:	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**
team 1	100	100	100	100	100	100
team 2	100	99.5	99.0	98.5	98.0	97.5
team 3	100	99.0	98.0	97.0	96.0	95.0
team 4	100	98.5	97.0	95.5	94.0	92.5
team 5	100	98.0	96.0	94.0	92.0	90.0
team 6	100	97.5	95.0	92.5	90.0	87.5
team 7	100	97.0	94.0	91.0	88.0	85.0
team 8	100	96.5	93.0	89.5	86.0	82.5
team 9	100	96.0	92.0	88.0	84.0	80.0

Table1: Scores for the differing simulations based on ladder positions

2.1.2. Binomial Probability System

To properly test the fairness of the models, it is beneficial to produce an alternate method for simulating the matches. This second method devised was not able to be used on the Division models, as those models need a score in order for a percentage to be calculated (in the case of all teams winning one game, the team with the highest percentage will qualify).
The binomial method involves picking the winner based on chance and is similar to the equal probability model. This variable ranged from 0.5 to 0.75, in 0.05 increments, the advantage given to the 'home' team (the team with a higher ranking), resulting in six binomial simulations for each model.

2.2. Difference between Teams

It is thought that when teams are closely ranked their abilities are evenly matched, creating a close, and therefore more exciting game. In an effort to measure the inherent excitement of a system, the average difference between the rankings of matched up teams was calculated by the simulations.

2.3. Analysis

Analysis was completed using MS Excel. In most cases results will be probability distributions for either winning the Premiership or reaching the Grand Final in graph form. While the actual percentages may help in deciphering advantage, it is important to realise that the shape of the distribution is the most important aspect. For this the reason, the axes of most graphs have been lightened. In these cases the x axis will be the teams from first to ninth, the y axis will be percentage with range of 0-100%.

2.4. Number of Trials

To investigate the number of trials needed to produce accurate results, the heuristic probabilities for one of the systems were calculated and compared with simulation results to find a MAPE, the mean average percentage error. A MAPE of less than 0.01% for 50,000 trials was achieved - which is an acceptable level of error for the simulations.

3. RESULTS

The existing systems are details below:

Team	Final 4	Final 5	Final 6	R Final 6	MF8	RMF8
1	37.5	37.5	25	25	18.75	18.75
2	37.5	25	25	25	18.75	18.75
3	12.5	25	18.75	18.75	15.625	18.75
4	12.5	6.25	18.75	12.5	12.5	18.75
5		6.25	6.25	12.5	12.5	6.25
6			6.25	6.25	9.375	6.25
7					6.25	6.25
8					6.25	6.25

Table 2:McIntyre System Likelihoods (Clarke (1996))

The revised McIntyre final system is covered in detail by Lowe and Clarke (2000). The obvious flaws are the top4/bottom4 disparity. It is argued by some that the RMF8 is too 'top heavy', in that only the top four sides can possibly win. Table 3 lists every Premier since the models' inception, which seems to agree with that hypothesis – no team has won from outside the top four. The Minor Premier has only won 6 out of the 14 years, perhaps a result of the unfair nature of this model.

Year	Minor Premier	Position Finished	Premier	Premier Ladder Position
2000	**Essendon**	Premiers		
2001	Essendon	2nd	Brisbane	2nd
2002	Port Adelaide	3rd	Brisbane	2nd
2003	Port Adelaide	4th	Brisbane	3rd
2004	**Port Adelaide**	Premiers		
2005	Adelaide	4th	Sydney	3rd
2006	**West Coast**	Premiers		
2007	**Geelong**	Premiers		
2008	Geelong	2nd	Hawthorn	2nd
2009	St Kilda	2nd	Geelong	2nd
2010	**Collingwood**	Premiers		
2011	Collingwood	2nd	Geelong	2nd
2012	Hawthorn	2nd	Sydney	3rd
2013	**Hawthorn**	Premiers		

Table 3: Existing system results (2000-13)

We consider two broad systems – Divisions and Roll Overs.

3.1 Divisions

In a divisional system the qualifying teams are split into groups (divisions), whereby they play a round robin, usually playing one game against every other team in their division. Both the FIFA World Cup and the FIFA World Cup Finals are based on this model - the 32 World Cup Finals qualifying teams are divided into groups of four teams. After each team has played their three group members, the top two teams advance to the next round. It's clear that the probability of any one team advancing past the round robin stage is heavily dependent on which other teams they are grouped with. For this reason three different groupings were simulated for this paper, Divisions A, B and C.

A	B	C
1	2	3
4	5	6
7	8	9

Division A

A	B	C
1	2	3
6	5	4
9	8	7

Division B

A	B	C
1	2	3
6	5	4
7	8	9

Division C

There are obviously many more options for the groupings, so this paper focuses on holding Group B steady with second, fifth and eighth. Group A and C always includes first and third respectively, with fourth, sixth, seventh and ninth rotating. Holding first, second and third in groups A, B and C respectively, there are 90 combinations possible.

Note that this system differs from the World Cup Finals model where there are four teams competing in each group. As opposed to the World Cup which has three divisional games, in a nine team system with three divisions there are two matches per team. The loss of a game translates to a much higher importance being placed on each game played. Using the World Cup example who have three divisional games, a team could lose their first match, win the next two and still be well placed to proceed to the next round. With only two matches, a loss in the first would significantly decrease the chances of proceeding past the round robin stage, especially for the lower ranked teams.

Further to assigning teams into groups, there are many possible variations on any of these systems through the order of scheduling. Consider the hypothetical group of teams 1, 2, and 3 where the order of the matches is as follows; 1v2, 1v3, 2v3. If 1 wins both matches the last is a dead rubber; its outcome is meaningless (apart from improving percentage in order to gain the wildcard entry). Therefore the best two teams should meet last, as both are expected to beat the third ranked team. This gives the matchups: 1v3, 2v3, 1v2 or 2v3, 1v3, 1v2. In either case a dead rubber would eventuate if team

3 won both its games. So to avoid a dead rubber, the winner of the first match must always play in the third match.

There is also the problem of when to schedule the byes for each team. It is difficult to predict in which week any particular team would want to take their turn of sitting out. Would the top team want to take their bye in the second week to be fresh for the third week, or would they assume victory in these games and want to rest up in the third week for tougher opponents in later rounds? An interesting variation on these systems could be to give the top ranked team the choice of when to take their bye, either week 1 or week 2, but this is beyond the scope of this paper

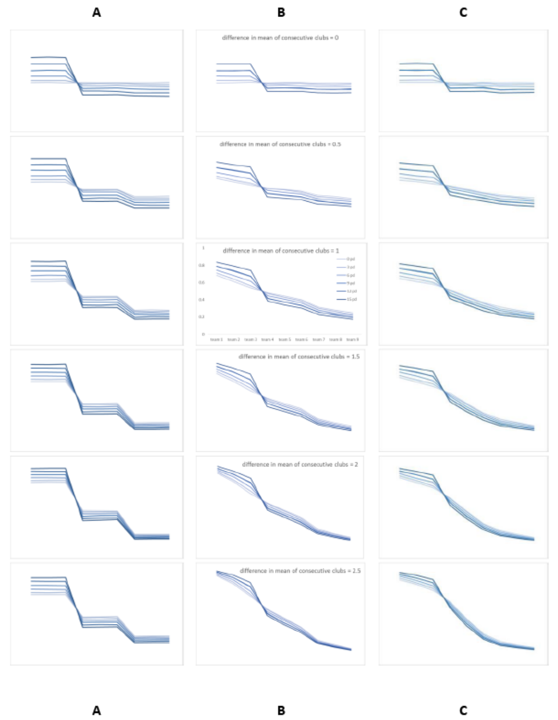


Figure 1: Desirable Shapes of the three division systems

Adapting a divisional model for the AFL finals system would be tricky for continuous reasons – if it were decided to move to a ten team system, a whole new system would need to be introduced, instead of modifying the current model. Changing the system may be a big enough change, so to change the type of system may create an overload of confusion. One of the major drawbacks of all Division models is that it would be impossible to complete the series in four weeks, failing one of the supplementary goals.

The variations available to this type of system make the possibilities for future research endless and exciting. As discussed, 90 combinations of divisions exists (holding 1, 2 & 3 constant), but realistically some of these combinations could be ruled out logically, possibly simplifying the experiment.

3.2 Roll Over

The Roll Over systems were originally created for this paper, named so because the matchups in the first week ‘roll down’ in ladder order. All Roll Over systems consist of four matches in Week 1 of the finals; one qualifying final, 2v3, and three elimination finals, 4v5, 6v7 and 8v9. First is awarded a first round bye in all systems. These systems were created with the primary purpose of decreasing premiership probabilities with decreasing ladder position.

3.2.1. Roll Over A

Figure 2 shows the structure map of Roll Over A. The Roll Overs all keep within the boundary of four weeks by allowing first place a massive advantage of a bye. The pathways shown on the map are colour coded – blue is for the winner, red for the loser. The second slot in match H is taken up by the highest ranked loser from the previous week, shown in a dotted red line.

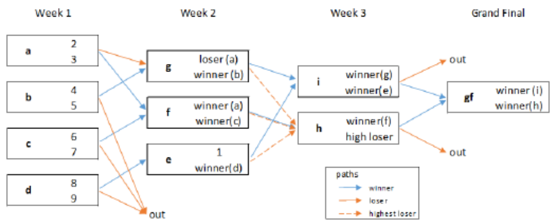


Figure 2 Roll Over A

The top ranking team receives massive benefits for securing the Minor Premiership with a highly advantageous draw. Granted the only bye of the series in the first round, in Week 2 they play the lowest ranked winner from Week 1, giving them a high probability of reaching Week 3 with little physical strain. But this directs to a massive flaw in Roll A, and a match fixing paradise; first can get smashed in Week 2 and be guaranteed to play in Week Three. Although the high significance of winning a home final for the next round should be sufficient to distinguish this course of action, in round 3 it is possible for their opponent not to be the lowest ranking team left. This could be seen as unfair, but it is a result of both fulfilling the goal criteria of the top two teams playing in the Grand Final, and refraining from scheduling repeat matches (second and third already play in Week 1).

Team	Final Position									EFP
	1	2	3	4	5	6	7	8	9	
1	25.11	24.81	50.07	0	0	0	0	0	0	**2.25**
2	18.65	18.71	24.94	12.45	25.25	0	0	0	0	**3.07**
3	15.39	15.66	12.51	18.93	24.76	12.74	0	0	0	**3.60**
4	7.92	7.75	3.08	12.58	12.38	6.30	49.98	0	0	**5.33**
5	7.75	7.78	3.15	12.54	12.49	6.27	50.02	0	0	**5.33**
6	6.42	6.21	3.20	9.32	12.65	12.39	0	49.81	0	**6.02**
7	6.35	6.32	3.04	9.21	12.47	12.43	0	50.19	0	**6.03**
8	6.15	6.22	0	12.41	0	24.79	0	0	50.43	**6.71**
9	6.25	6.53	0	12.57	0	25.08	0	0	49.57	**6.66**

Table 4 Roll Over A: *Percentage chance of team finishing in any position, EFP - Equal probability model*

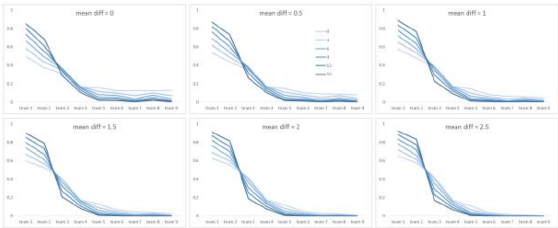


Figure 3 P(GF) by points diff and mean diff.

3.2.2 Roll Over B

In an effort to decrease the gap in advantage first has over the rest of the field in Roll Over A, first will now face a higher ranking team than the winner of Match A. Roll Over B is basically the same as Roll Over A, the only difference being the winner of Match A now plays the winner of Match D, and after their bye, first plays the winner of Match C

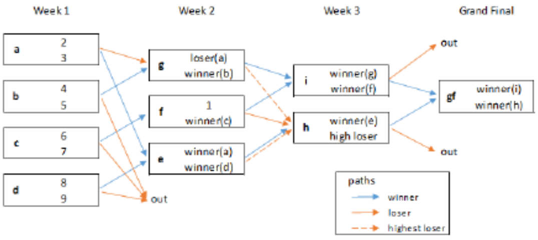


Figure 4 Roll Over B

The probability tables for Roll Overs A, B and D reveal almost identical probabilities for the case of teams being considered equal. This is due to the systems changing slightly in order to suit particular teams, but when there is the same chance of beating third as eighth, these subtle differences will not be seen. They are included for completeness.

Team	Final Position									EFP
	1	2	3	4	5	6	7	8	9	
1	24.76	25.17	50.07	0	0	0	0	0	0	**2.25**
2	18.93	18.82	24.91	12.52	24.81	0	0	0	0	**3.05**
3	15.60	15.64	12.47	18.63	25.33	12.33	0	0	0	**3.59**
4	7.89	7.67	3.06	12.42	12.49	6.37	50.10	0	0	**5.33**
5	7.81	7.91	3.23	12.63	12.44	6.08	49.90	0	0	**5.32**
6	6.14	6.28	3.03	9.28	12.69	12.47	0	50.10	0	**6.04**
7	6.23	6.16	3.22	9.26	12.24	12.99	0	49.90	0	**6.04**
8	6.22	6.05	0	12.72	0	24.85	0	0	50.16	**6.70**
9	6.42	6.29	0	12.54	0	24.91	0	0	49.84	**6.67**

Table 5 Roll Over B: *Percentage chance of team finishing in any position, EFP - Equal probability model*

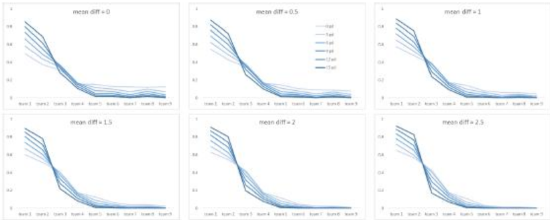


Figure 5 P(GF) by points diff and mean diff.

The simulations surprisingly showed that switching the games had no effect on either the Premiership probabilities or the Grand Final Probabilities, as Roll B gave almost exactly the same results as Roll A. Investigation revealed the results differed by roughly 0.04%.

3.2.3 Roll Over D

A big disadvantage in the previous Roll systems is that in Week 3, given favourites win, the matchups are 1v3 and 2v4. Roll D was designed with the goal of scheduling the more traditional matchups 1v4 and 2v3 for week 3, which is the major difference in Roll D from its predecessors. It is an adaption of Roll A, rather than Roll B.

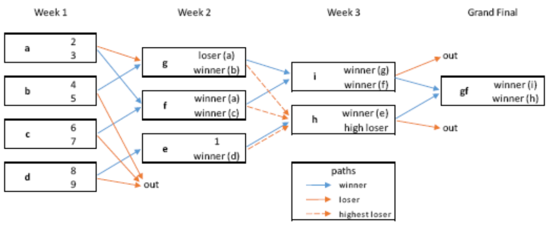


Figure 6 Roll Over D

Although this seems a fair system, it has some downfalls. In the unlikely event first lose their week 2 match, they will play that same team the following week. One major advantage of this system is the likely event of a second versus third matchup in both week 1 and 3. In a season where the top team is a clear favourite for the premiership (although not often true, or currently true, it's not an outrageous assumption), having these two teams battle twice will be a good test to see who gets the honour of playing in that years' Grand Final. It was not expected that the probability distributions for this system to be much better than previous systems, however as the matchups differ, simulations have been run.

Team	Final Position									EFP
	1	2	3	4	5	6	7	8	9	
1	25.12	24.95	49.93	0	0	0	0	0	0	**2.25**
2	18.69	18.61	24.84	12.54	25.32	0	0	0	0	**3.07**
3	15.44	15.91	12.55	18.42	24.99	12.68	0	0	0	**3.60**
4	7.98	7.60	3.15	12.46	12.55	6.26	50.01	0	0	**5.33**
5	8.00	7.74	3.07	12.43	12.51	6.25	49.99	0	0	**5.32**
6	6.11	6.29	3.19	9.30	12.49	12.53	0	50.09	0	**6.04**
7	6.40	6.34	3.27	9.40	12.14	12.53	0	49.91	0	**6.02**
8	6.14	6.21	0	12.77	0	25.03	0	0	49.85	**6.68**
9	6.13	6.34	0	12.67	0	24.71	0	0	50.15	**6.69**

Table 6: Roll Over D: *Percentage chance of team finishing in any position, EFP - Equal probability model*

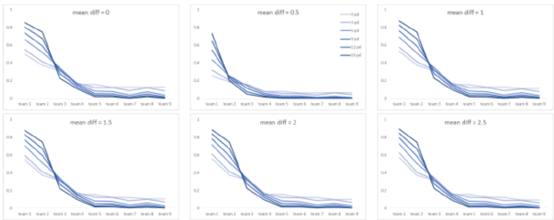


Figure 7 P(GF) by points diff and mean diff.

3.2.4. Roll Over C

There are multiple reasons, some financial and fan based, for the AFL desiring to schedule more matches. In an effort to achieve this goal, Roll Over C (Figure 8) was designed. Featuring a record breaking 11 games, where seven of the nine teams are guaranteed to play in two finals (the loser of match D: 8v9 is eliminated and first has a first week bye).

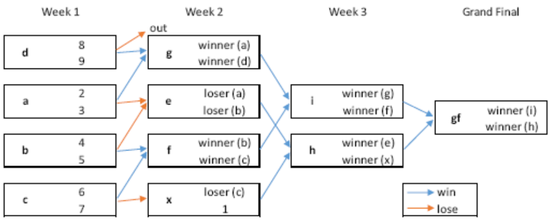


Figure 8 Roll Over C

While the first week matchups were devised for close games, the second and third week matchups were designed to favour the higher ranking teams. Assuming the favourites win, the week 2 matchups are 2v8, 4v6, 3v5 and 1v7, obviously fairer to the higher placed teams. In a perfect world the matchups might be 3v6 and 4v5, but that would repeat a first week match. The winner of match A is given the lowest ranked team in an attempt to lessen the massive advantage of first's first week bye. Again assuming the favourites win, the week 3 matchups are 1v4 and 2v3, as discussed previously, the most desired for that round.

Team	Final Position									EFP
	1	2	3	4	5	6	7	8	9	
1	12.76	12.55	24.87	0	49.83	0	0	0	0	**3.62**
2	12.47	12.60	18.88	6.13	25.15	24.77	0	0	0	**3.93**
3	12.28	12.70	18.65	6.14	12.56	25.18	12	0	0	**4.20**
4	12.53	12.19	15.61	9.25	9.37	22.21	15.65	3	0	**4.36**
5	12.42	12.65	9.30	15.75	3.09	15.44	22.00	9	0	**4.66**
6	12.50	12.39	6.45	18.68	0	12.40	24.98	12.61	0	**4.81**
7	12.57	12.50	6.25	18.74	0	0.00	24.88	25.06	0	**5.06**
8	6.26	6.26	0	12.55	0	0.00	0	25	50.15	**7.19**
9	6.22	6.17	0	12.76	0	0.00	0	25	49.85	**7.18**

Table 7: Roll Over C *Percentage chance of team finishing in any position, EFP - Equal probability model*

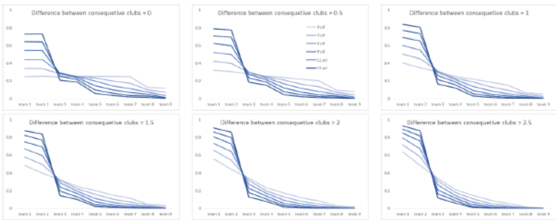


Figure 9 P(GF) by points diff and mean diff.

An advantage of this system is that when all teams have an equal chance of winning each game, teams One through Seven each have an equal chance of winning the Grand Final. However, when advantage is introduced the results are similar to the previous models, but the advantage is more even over the field. This means that if there is a season where the finals qualifiers are somewhat close in ability (such as the current year), it would be a very exciting series, with excellent first week matches (like all Roll Overs), and a very fair system for higher ranking teams in later weeks. The extra game this system has over its predecessors, and extra three over the RMF8, would bring in a considerable amount of spectators for the AFL.

4. DISCUSSION

To confirm the systems behave in the correct way, the advantage given to each team by the different systems was graphed individually, Figure 25 below. It seems RMF8 is a clear loser; second through fifth lose advantage, a definite drawback of the current system. Hardest done by is second, who gains little more advantage than the rest of the field. Roll Over C is arguably best. Although it gives first place a decent advantage, it is lower than the others; the difference is picked up by some of the lower ranks teams (fourth to seventh), a sizable percentage increase in some cases.

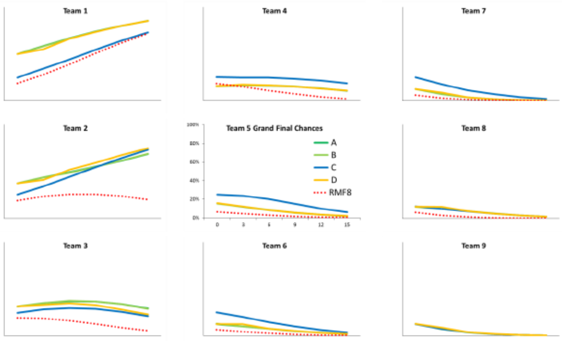


Table 8: The advantage given to each team. Difference between consecutive means = 0. The x-axis is advantage given in the form of points (0-15), the y-axis still percentage (0-100)

The following table displays a "×" if a certain goal was met. Added to the predetermined goals are "Difference" and "Games". Difference is the average difference of the competing teams, Games is the amount of games played in the system.

	Roll Over A	Roll Over B	Roll Over C	Roll Over D	RMF8
Maximise the probability that the highest ranked team wins	×	×	×	×	×
Maximise the probability that the two highest ranking teams meet in the grand final	×	×	×	×	×
The probability of a team finishing in any position of higher should be greater than for any lower ranked team			×		
Difference	2.31	2.33	2.46	2.48	**2.23**
Games	10	10	**11**	10	9

Table 9: Comparison of systems

If considering only the factors listed in the above table, again Roll Over C is clearly the best. Not only is it the only system which fulfils all criterion, but the most games are played under its structure. It is interesting to note the current system had the best Difference, but all are close enough to negate significance.

The mean and standard deviation of the AFL top nine teams in 2013 was calculated and is displayed in Table 10**.** These scores were used in a simulation for each of the systems shown if figure 10.

Pos.	Team	Home		Away	
		mean	s.d.	mean	s.d.
1	Hawthorn	111.45	22.98	117.91	25.34
2	Geelong	109.27	17.97	109.73	21.04
3	Fremantle	100.55	29.43	84.45	32.04
4	Sydney	104.73	29.73	99.27	29.82
5	Richmond	93.45	24.07	102.36	27.53
6	Collingwood	102.82	19.45	92.45	22.03
7	Port Adelaide	91.00	22.45	95.45	23.36
8	Carlton	101.18	22.21	92.00	22.41
9	North Melbourne	98.91	38.12	110.82	17.49

Table 10: Home and away mean and standard deviation of the top AFL teams of 2013.

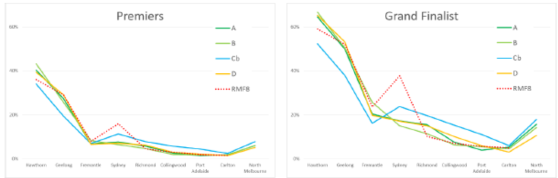


Figure 10: Simulation of all Roll Overs and RMF8

It was surprising that all the systems produced similar results. The most obvious thing to do with a prediction is compare it with the actual, and these models somewhat fail; Hawthorn beat Fremantle in the Grand Final, although Fremantle won the unwinnable game the week before in the Preliminary Final, beating arguably the best team of all time, Geelong, in Geelong. It is interesting to note the spike for North Melbourne, due mainly to their impressive away mean score. Even though they do have that high mean score, it often wasn't enough to upset the top sides, proving the strength of the models.

The expected position increases with increasing teams for all systems. RMF8 can again be seen to be the one which is least fair, due to the ever present 'equal for top four and bottom four' nature of that system. It will be interesting to see these graphs when advantage is introduced, as they may tell a different story.

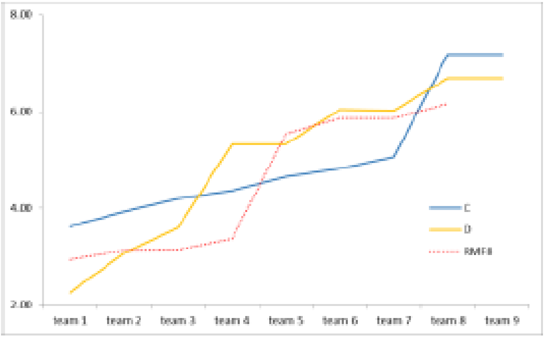


Figure 11: The EFP for the systems

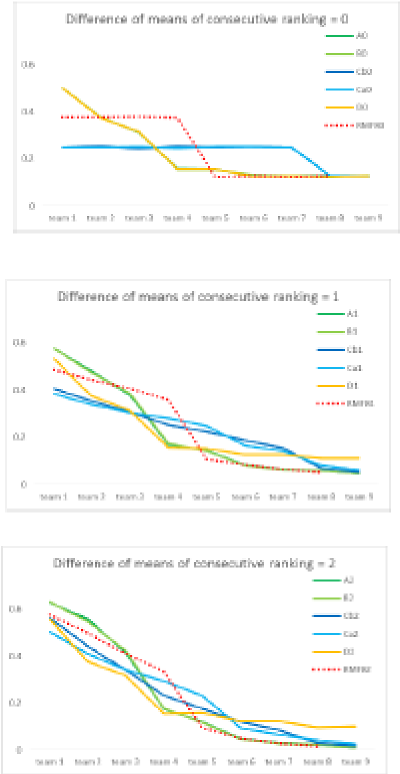


Figure 12: Mean Difference Comparisons

For the final analysis it was decided to compare the chance of reaching the Grand Final under all the Roll Overs and the current system. Figure 28 has these probabilities for when there is no home ground advantage, but the mean for consecutive teams changes slightly. The clear best of the Roll Overs up until this point has been C, but the left graph along with the probability tables shows its major weakness; when all teams are considered equal, first to seventh have equal chance of playing in the final. However, with slight advantage increase it improves dramatically. Again the worst seems to be RMF8, whose massive drop from fourth to fifth is mistakably prevalent, even in the rightmost graph.

5. CONCLUSION

All systems tested met most of the criteria and were shown to be fairer than the current system. Of the original systems Roll C was shown in many cases to have a fairer spread of winners while still giving first its due advantage, but has inherent flaws. The systems each have considerable benefits and drawbacks, the best being a matter of opinion. Although some unfavourable events may occur, to achieve the desired outcomes there may always be some unfairness inherent to the system. The division method is an exciting idea which deserves further study.

References

Christos, G. (2000) Variations in Random Pairing Finals Systems, *The 5th Australasian Conference on Mathematics and Computers in Sport.*

Clarke, S.R., 1996. Calculating premiership odds by computer: and analysis of the AFL final eight playoff system. *Asia-Pacific Journal of Operational Research*, 13, p89-104.

Holbrook, B.C, 1996. More probability models for the NCAA regional basketball tournaments. *The American Statistician*, 50.1, p34.

Lowe, J. & Clarke, S.R.(2000). An Analysis of the AFL Final Eight System. *The 5th Australasian Conference on Mathematics and Computers in Sport.*

Monahan, J.P. and P.D. Berger, 1977. Play-off structures in the National Hockey League. Optimal Strategies in Sports, North Holland, Amsterdam, p131-140

Ross, J. (Editor), 1996. 100 Years of Australian Football. Penguin Books, Australia

AFL Record Grand Final 1999 Souvenir MCG Edition

INTERCHANGE AND INJURY EFFECTS IN AFL FOOTBALL

Anthony Bedford[a,c], Adrian Schembri[a], Bradley O'Bree[a], Shane McCurry[b] and Patrick Clifton[b]

[a]*School of Mathematical & Geospatial Sciences*
RMIT University, Melbourne, AUSTRALIA
[b] *Australian Football League*
[c] *Corresponding author: anthony.bedford@rmit.edu.au*

Abstract

The increase in recent seasons in interchange rotations in the AFL has had a substantial impact on various facets of the game. These include the speed of general play, player exertion and intensity, and the incidence of collision and non-collision injuries. The aim of this research was to examine the longitudinal impact of one or more players being injured (to the extent that they are unable to return to the field) on scoring patterns, interchange rotations, and likelihood of either team winning the match. Differences in each of these outcome variables were analysed for matches where no injuries occurred, with these matches being compared with matches where one, two, and three injuries were sustained by either team. Results were compared across the 2007 to 2010 AFL seasons and indicated a substantial negative impact on a team's ability to score and rotate players on and off the interchange bench when at least one injury was sustained. The net effect was a decrease in the likelihood a team would win a match.

Keywords: AFL, injury, interchange rotation, injury effects

1.INTRODUCTION

Theincrease in recent seasons in interchange rotations in the AFL has had a substantial impact on various facets of the game. These include the speed of general play, player exertion and intensity, and the incidence of collision and non-collision injuries. The elevation in player injuries has been projected to further increase over the next five to ten years. Based on research conducted by the AFL and research associates, The AFL Laws of the Game Committee proposed a series of changes to the interchange system in order to curb the increase in player injuries. Suggested amendments included:

- Retaining four players on the interchange bench, and capping interchange rotations at 80 per team;
- Reducing the number of players on the interchange bench to three, and introducing one substitute player per team;
- Reducing the number of players on the interchange bench to two, and introducing two substitute players per team.

One consequence of the increase in general player injuries is the heightened incidence of player injuries during a match, and the subsequent limitations placed on teams who are unable to rotate all 22 players on and off the ground. To enhance our understanding of the impact of player injuries on the outcome of games played in the AFL, further analysis is required. Specifically, limited data is available on the impact of one or more players being unable to return to the field after sustaining an injury during a game. In effect, it is possible that the likelihood of a team winning a match following a player being injured is reduced, and this may be compounded by a team sustaining multiple injuries during a game.

In collaboration with the AFL, the RMIT University Sports Statistics Research Group examined the impact of player injuries on match outcome AFL matches. This was achieved by incorporating several variables into a detailed analysis. These variables included:

- In-play injury incidents for AFL matches played over the 2007 to 2010 seasons;

- In-play scoring patterns for teams prior to and following an injury;
- Number of interchange rotations prior to and following an injury;
- Win percentage of teams who sustain one or multiple injuries during a match.

The aim of this analysis was to examine the longitudinal impact of one or more players being injured (to the extent that they are unable to return to the field) on scoring patterns, interchange rotations, and likelihood of either team winning the match.

Examination of the effect of player injuries will be analysed at multiple levels, including the impact of player injury on points scored and points conceded, and the association between the margin at the time of the injury and the final margin at the end of the game. Finally, the effect of injury on interchange rotations will be reviewed, with respect to the team who sustained an injury as well as the opposition team who maintained a full complement of players to rotate for the remainder of the match.

2.METHODS

Multi-Phase Analysis

In order to address the aforementioned research aims, a multi-phase analysis was conducted. Analyses were completed on the association between single and multiple in-game injuries and:

- Scoring patterns;
- Interchange rotations;
- Points conceded;
- Differential scoring;
- Likelihood of winning the match.

Each of these analyses was conducted using statistics provided by Champion Data. This data incorporated all AFL matches played during the 2007, 2008, 2009, and 2010 seasons. Variables that were utilised for all games regardless of whether an injury was sustained included:

- Season;
- Round;
- Match code;
- Home and away team;
- Home and away team final score;
- Home and away team interchange rotations;
- Quarter length (seconds);
- Scores at each quarter break.

Additional variables that were incorporated for matches where at least one injury occurred:

- Name of the player who was injured;
- Quarter the player was injured;
- Time in the quarter the player was injured;
- Team scores at the time of injury;
- Team interchange rotations at the time of injury.

Research Constraints

Whilst all matches played from 2007 to 2010 were incorporated in the analysis, a constraint was placed on matches where multiple injuries occurred. These matches could be incorporated when only one team had sustained injuries (e.g., 2 injuries to the home team, and no injuries to the away team). This constraint was placed on the analysis given that findings would become considerably more ambiguous had games where injuries occurred for both teams been incorporated (e.g., one injury each or two injuries to one team and one injury to the other). In cases where injuries occur for both teams, each and every match situation is unique, given that injuries occur for each team at different times. For example, the home team may sustain an injury in the first quarter, whilst the away team sustains two injuries in the third quarter. Endeavouring to establish which team was more disadvantaged would require considerable speculation, and thus was removed from the analysis.

3.RESULTS

Points Scored

In the first phase of this analysis, an examination of the effect of injuries is undertaken with respect to scoring. Figure 1 presents the scoring trends of all AFL teams for the 2007 to 2010 seasons for matches where no players were injured, and either one, two or three players were injured over the course of the game. Injuries only refer to those players who left the ground and were unable to return for the remainder of the match.

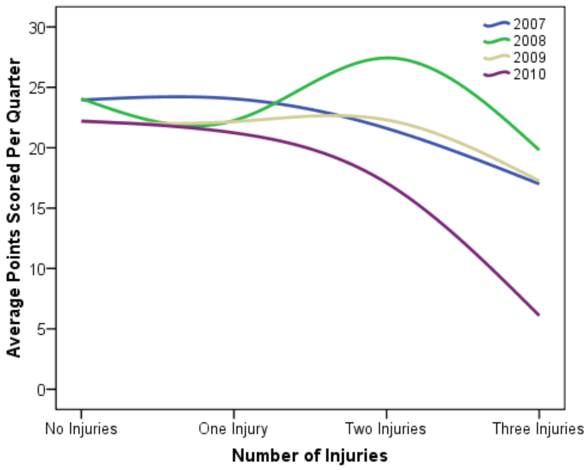


Figure 1. Scoring trends for the 2007-2010 AFL seasons for teams with no injuries and one, two and three injuries.

Results in Figure 1 indicate that injuries resulted in a general decline in a team's ability to score when players were injured. This was particularly the case when two or three players were injured in the one game. What is most striking is the steep reduction in scoring potential when two or three injuries occurred in the one match during the 2010 season. In effect, during 2010, two or three injuries resulted in a scoring rate of 5 and 16 points lower than equivalent matches where no injuries had occurred. The reduction in scoring potential when injuries occurred in 2010 may be related to the increase in interchange rotations. Specifically, the reduced potential to rotate players following an injury may be associated with the decline in scoring following multiple injuries.

Findings from Figure 1 can be further interpreted based on statistics presented in Table 1. This table provides data on the change in scoring potential following one, two or three injuries when compared with scoring potential when all 22 players are available, that is, 18 players on the field, and four on the interchange bench.

Season	Quarterly Scoring Rate When No Injury Was Sustained	Quarterly Scoring Rate When Compared With Matches Where No Injury Was Sustained		
		One Injury	Two Injuries	Three Injuries
2007	23.95	+0.1	-2.35	-6.95
2008	24.06	-1.81	+3.37	-4.26
2009	22.14	+0.02	+0.16	-4.89
2010	22.21	-0.98	-5.13	-16.11
2007 – 2010	23.27	-0.92	-1.82	-9.54

Table 1. Tabulated scoring trends for the 2007-2010 AFL seasons comparing scoring rates of teams with no injuries with teams with one, two and three injuries.

Table 1 indicates that in 2010, having one injury resulted in scoring 1 point less per quarter, two injuries resulted in scoring 5 points less per quarter, and three injuries resulted in scoring 16 points less per quarter.In 2007 and 2009, having one injury did not influence scoring potential of teams when compared with teams who had a full complement of players available.In 2007, 2008, and 2009, having three injuries resulted in reduced scoring potential of between four and seven points per quarter, however in 2010, this figure increased to over 16 points per quarter, which is the equivalent to over five goals per half of football.

When examining all data from 2007 to 2010 combined, having one injury resulted in scoring 1 point less per quarter, two injuries resulted in 2 points less per quarter, and three injuries resulted in 9.5 points less per quarter.

Interchange Rotations

It can be hypothesised that the greatest effect of in-game injuries will be on interchange rotations from that point in the match onwards. This has been a particularly salient issue in seasons considered, given the rapid increase in player rotations during a game at the time.

Analysis of the average interchange rotations for five minute periods throughout a match in 2010 sheds light on this increase. During 2010, teams averaged approximately five interchange rotations for every five minute period in a match, which equates to one interchange rotation per minute. Based on theinterchange resources that are required to maintain this level of rotations, it can be predicted that in-game injuries will have a substantial effect on the ability of teams to rotate their players at the frequency that has been evident during 2010.

In this analysis, the frequency of player rotations was examined for the 2007 to 2010 seasons. Specifically, average rotations were computed for teams who had sustained no injuries, and those teams who had sustained one, two or three injuries, refer to Figure 2.

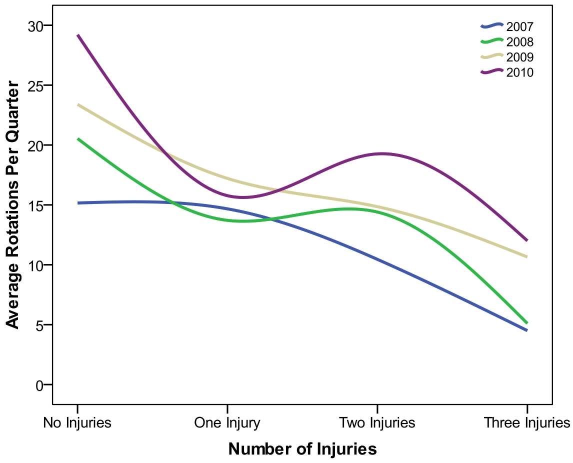


Figure 2. Interchange rotation rates for the 2007-2010 AFL Seasons for teams with no injuries and one, two and three injuries.

As can be observed in this figure, player injuries had a substantial impact on the capacity of teams to rotate their players during every season over the four seasons considered. The steepest decline in interchange rotations was evident in 2010, when a team sustained a single injury. Two injuries was associated with between five and ten less rotations per quarter across each of the four seasons, whilst three injuries was associated with a steep decline in the capacity to rotate, and this was particularly evident in 2008 and 2010.

From this analysis, it can be identified thatthe rotation rates that were evident in 2010 are not tenable when a team suffers a single injury. This was not the case in 2007, as results in Figure 2 indicate that interchange rotations were consistent regardless of whether a team had no injuries or one injury during the match. An alternate depiction of the data in Figure 2 is presented in Table 3.

Season	Quarterly Rotation Rate When No Injury Was Sustained	Quarterly Rotation Rate When Compared With Matches Where No Injury Was Sustained		
		One Injury	Two Injuries	Three Injuries
2007	15.16	-0.49	-4.71	-10.66
2008	20.54	-6.83	-6.14	-15.44
2009	23.39	-6.18	-8.54	-12.74
2010	29.21	-13.44	-9.95	-17.21
2007 – 2010	21.47	-5.89	-6.33	-13.04

Table 3. Tabulated rotation rates for the 2007-2010 AFL seasons comparing interchange rotations of teams with no injuries with teams with one, two and three injuries.

Table 3 indicates that in 2010, having one injury resulted in an average of 14 less interchangerotations per quarter, two injuries resulted in 11 less interchange rotations per quarter, and three injuries resulted in 18 less interchange rotations per quarter.The impediment that one injury placed on interchange rotations in 2010 is double that of 2008 and 2009, and 13 times that observed in 2007. This finding indicates that under the then current circumstances, interchange rotation rates recorded in 2010 are not sustainable when less than four places are available on the interchange bench.

Overall, between 2007 and 2010, having one injury resulted in an average of 6 less interchange rotations per quarter, two injuries resulted in 6.5 less interchange rotations per quarter, and three injuries resulted in 13 less interchange rotations.

Points Conceded

Points conceded refers to the number of points scored by the opposition, which for the purposes of this research, refers to the team who does not have any injured players. Again, the number of points scored is examined with respect to quarter of football, and thus an analysis of the number of points conceded per quarter was reviewed when a team had one, two, or three injuries within the one game. Figure 3 presents the average number of points conceded in 30 minutes of football when teams had no injuries, or one, two or three injuries.

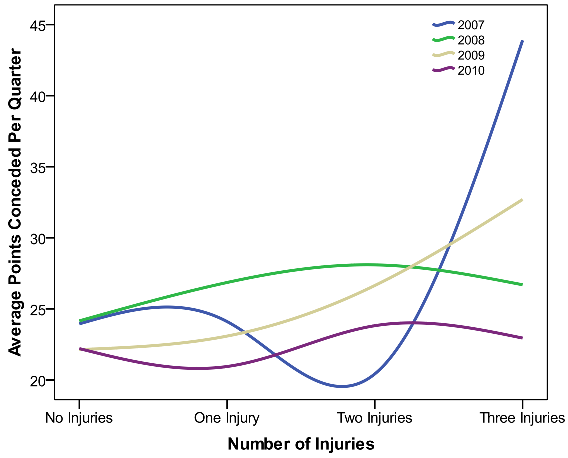


Figure 3. Average points conceded for the 2007-2010 AFL seasons for teams with no injuries and one, two and three injuries.

Sustaining one injury had only minor detrimental effects on points conceded during 2007 and 2009, and of note, teams conceded slightly fewer points after sustaining one injury in 2007 and 2010. This finding must be interpreted in the context of the points scored analysis undertaken in the previous section. In effect, despite a mild decrease in points conceded in 2010, teams were scoring at a slower rate after sustaining an injury in this season, thus the net result remains negative for teams who sustain an injury.

Sustaining two injuries resulted in teams conceding more points during 2008, 2009, and 2010. Whilst this was not the case in 2007, teams who sustained three injuries during this season conceded an average of 43 points per 30 minutes of football when only one player was available on the interchange bench.

The general trend in 2010 was that injuries did not adversely affect the number points conceded for the remainder of the match, however as stated, teams were scoring fewer points after sustaining injuries, and thus the net result remains negative.

Table 4 presents data on the average points conceded after sustaining one, two or three injuries, when compared with teams who did not sustain any injuries for the duration of matches.

Season	Quarterly Points Conceded When No Injury Was Sustained	Quarterly Points Conceded When Compared With Matches Where No Injury Was Sustained		
		One Injury	Two Injuries	Three Injuries
2007	23.95	+0.16	-3.52	+19.95
2008	24.06	+2.08	+4.04	+2.64
2009	22.14	+0.95	+4.51	+10.56
2010	22.21	-1.26	+1.61	+0.74
2007 – 2010	23.95	+0.24	+1.30	+8.22

Table 4. Tabulated points conceded for the 2007-2010 AFL seasons comparing points conceded for teams with no injuries and teams with one, two and three injuries.

Results in Table 4 indicate sustaining one injury had a minor impact on points conceded, with points conceded fluctuating by between 1 and 2 points per quarterduring 2007, 2009, and 2010 after one injury was sustained.Sustaining two injuries had the greatest effect on points conceded over the past three seasons, particularly during 2008 and 2009, with four additional points being conceded for every 30 minutes of football.

Over the past four seasons, when three injuries were sustained, AFL teams conceded considerably more points when compared with teams who had a full complement of players to rotate through the interchange bench. This finding was most salientduring 2007 and 2009. It should be noted that the variability in points conceded following three injuries may be the result of sample size, given that teams infrequently sustain three injuries in the one match.

When considering all four seasons combined, results indicated that having one injury was associated with conceding the same number of points per quarter, two injuries resulted in conceding 1 point more per quarter, three injuries resulted in conceding 7 points more per quarter, and four injuries resulted in conceding 16 points more per quarter. It should be noted that four injuries occurred on one occasion only.

Overall Scoring Differential

Based on the analysis of scoring rate and points conceded, a final analysis was conducted to examine the overall scoring differential when teams sustained one, two or three injuries. Scoring differential can be calculated by subtracting points conceded from points scored, and thus a positive score indicates that the team with injuries has scored more than their opponent whilst a negative score indicates that the team with injuries has scored less than their opponent. Figure 4 displays the average score differential for matches where one, two or three injuries were sustained.

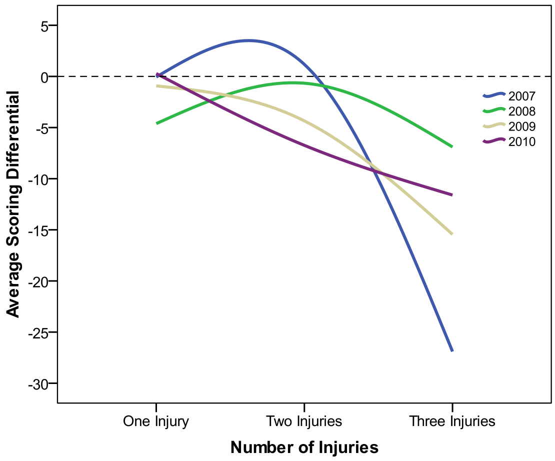


Figure 4. Scoring differential for the 2007-2010 AFL seasons for teams with one, two and three injuries.

Results displayed in Figure 4 indicate that teams who sustain one injury maintained an overall scoringdifferential of approximately zero in 2007, 2009, and 2010, however teams conceded four points more than they scored each quarter during 2008.

In 2009 and 2010, sustaining two injuries resulted in a net loss of between 4 and 5 points each quarter respectively, which equates to a total of 16 to 20 points of the course of the game. When teams sustained three injuries in 2009 and 2010, the net score differential was between 11 and 15 points, thus teams were much more likely to concede considerably more points than were scored for each quarter that was played with three injured players on the interchange bench. Whether this finding has resulted from the increase in player rotations and subsequent elevations in game speed and intensity warrants further research.

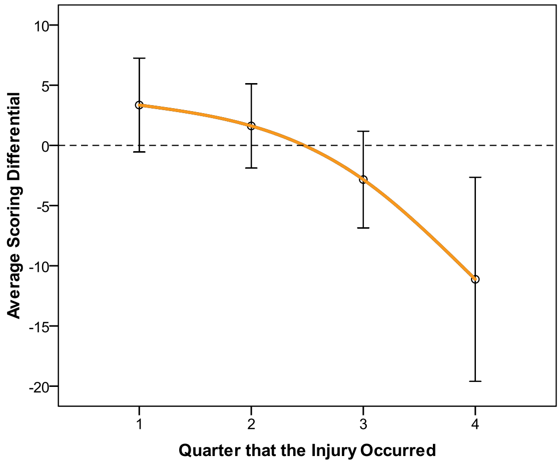


Figure 5. Average scoring differential for the 2007-2010 AFL seasons for injuries sustained in the first, second, third and fourth quarters.

Figure 5 displays the average score differential for injuries sustained in the first, second, third and fourth quarters. This analysis examines the difference between points for and score concededfor teams who sustained an injury in the 2007, 2008, 2009, or 2010 seasons.Results in Figure 5 indicate that the scoring pattern of teams (relative to their opposition's scoring) is not impeded by an injury that occurs in the first or second quarters. However, when an injury occurs in the third quarter, teams have a 2 point deficit (on average) for each quarter for the remainder of the match. In addition, when an injury occurs in the fourth quarter, teams on average, have a 10 point scoring differential, which indicates that they concede 10 points more than they score if the injury is sustained at the beginning of the final quarter. If the injury is sustained at the 15 minute mark of the final, the results in Figure 5 suggest that this will result in a five point deficit by the end of the match.

Likelihood of Winning the Match

In addition to the examination of the effect that in-game injuries have on scoring patterns, an aim of this analysis was to examine the ability of teams to win after sustaining one, two or three injuries in a match. Table 5 presents the percentage of matches won by teams who had one, two or three injuries during a match in the 2007 to 2010 AFL seasons.

Season	Percentage of Matches Won By Team with Players Injured		
	One Injury	Two Injuries	Three Injuries
2007	49.0%	62.5%	-[a]
2008	47.8%	37.5%	-[a]
2009	45.5%	30.0%	-[a]
2010	46.4%	41.7%	-[a]
2007–2010	46.9%	42.1%	25.0%

[a] Insufficient data to compute a win percentage.

Table 5. Percentage of matches won by teams who had one, two or three injuries during a match in the 2007-2010 AFL seasons.

Results in Table 5 indicate that when a single injury was sustained during the match, the likelihood of winning during the 2007 to 2010 seasons was less than 50%, and at times as low as 45%. This value was lowest during the 2009 and 2010 seasons respectively, which may be indicative of the increase in player rotations evident in these two seasons. Findings for 2010 provide furtherevidence of this contention, given that teams who sustained two injuries won only 41.7% of matches, whilst no team won a match after sustaining three injuries during a match in 2010. With the exception of the 2007 season, sustaining two injuries within a match resulted in a considerable decrease in win percentage, with teams only winning 30% of matches during 2009 when two injuries were sustained.

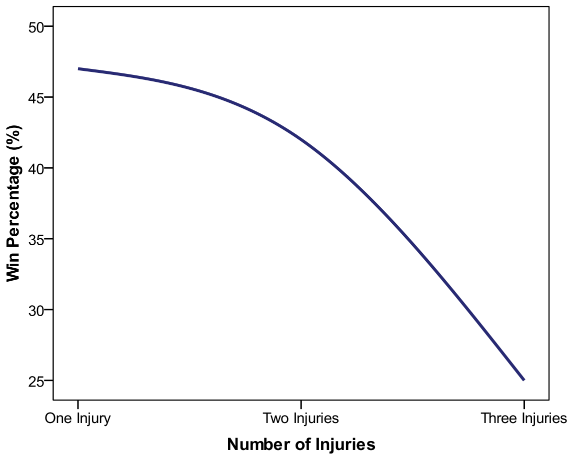


Figure 6. Percentage of matches won by teams who had one, two or three injuries during a match in the 2007-2010 AFL seasons.

Figure 6 presents data on all four seasons from 2007 to 2010 combined. A combined analysis was undertaken due to the small sample size of matches where three injuries occurred. This analysis provides some insight into the effect of multiple injuries on the outcome of matches. When considering all four seasons combined, between 2007 and 2010, teams with one injury won 47% of their matches, teams with two injuries won 42%, teams with three injuries won 25%, and teams with four injuries did not win a match between 2007 and 2010 when their opposition team had 22 available players throughout the match.

An examination of the likelihood of winning a match when an injury was sustained in the first, second, third or fourth quarter was also undertaken. This analysis incorporated those matches where only one player was injured over the course of the game. Findings revealed only minor variations in the likelihood of winning, irrespective of the quarter of the injury, or the season that was analysed (e.g., 2007, 2008, 2009, or, 2010). When considering all four seasons combined, the likelihood of winning was lowest (38%) when an injury was sustained in the third quarter.

4. DISCUSSION

Analysis showed sustaining multiple injuries resulted in a general decline in a team's ability to score when compared with teams who had not sustained injuries.In 2010, having one injury resulted in scoring 1 point less per quarter, two injuries resulted in scoring 5 points less per quarter, and three injuries resulted in scoring 16 points less per quarter. The effects of injuries in 2010 were markedly greater than in previous seasons.

Over the past four seasons, player injuries have had a substantial impact on the capacity of teams to rotate their players. The steepest decline in interchange rotations was evident during 2010.The impediment that one injury placed on interchange rotations in 2010 is double that of 2008 and 2009, and 13 times that observed in 2007.

Sustaining one injury had only minor detrimental effects on points conceded during 2007 and 2009, and of note, teams conceded slightly less points after sustaining one injury in 2007 and 2010.Sustaining two injuries had the greatest effect on points conceded over the past three seasons, particularly during 2008 and 2009, with four additional points being conceded for every 30 minutes of football.The general trend in 2010 was that injuries did not adversely affect the number of points conceded for the remainder of the match, however as stated, teams were scoring fewer points after sustaining injuries, and thus the net result remained negative.

In 2009 and 2010, sustaining two injuries resulted in a net loss of between 4 and 5 points each quarter respectively, which equates to a total of 16 to 20 points over the course of the game.The scoring pattern of teams is not impeded by an injury that occurs in the first or second quarters. However, when an injury occurs in the third quarter, teams have a 2 point deficit (on average) for each quarter for the remainder of the match. In addition, when an injury occurs in the fourth quarter, teams on average, have a 10 point deficit, which in practice, indicates that they concede ten points more than they score if the injury is sustained at the beginning of the final quarter.

When a single injury was sustained during the match, the likelihood of winning during the 2007 to 2010 seasons was less than 50%, and at times as low as 45%. This value was lowest during the 2009 and 2010 seasons respectively, which may be indicative of the increase in player rotations that has been evident in these two seasons.

The results of these analyses do not counteract intuition, that is, it is expected sustaining injuries would be detrimental to team performance, and this research bears evidence in favour of this contention. It is interesting however that the effect worsens, in scoring terms, the later in the match the injury occurs. Such a finding suggests teams are able to better manage injuries early in a match by perhaps modifying their game plan. Also of interest is the finding that detrimental effects impact on points conceded more than points scored. This suggests the defensive component of a team's game plan is hindered to a greater degree by injury than the offensive component.

Limitations

Several limitations exist in the current analysis. Firstly, it was only possible to examine matches where only one team sustained one or more injuries. Whilst an analysis of matches where both teams sustained injuries may yield some fruitful information, the inclusion of this data would likely increase ambiguity in the current findings.

A second limitation that should be noted is the limited sample size that was available for certain injury categories, particularly matches where a team had sustained three or four injuries. Given that teams seldom suffer three or four injuries in the one match, the sample size for these categories was limited. In each analysis, every effort was made to avoid misinterpreting the data, and therefore, an analysis was not undertaken when an insufficient sample size was available.

5.CONCLUSION

Based on the findings of this analysis, it can be concluded that sustaining one or more injuries during a match has a substantial impact on a team's ability to score and rotate players on and off the interchange bench. This was particularly evident in 2010, with a marked decrease in both scoring potential and player rotations when injuries were sustained. Of note, single injuries did not increase the number of points conceded, however multiple injuries resulted in a slightly higher number of points being conceded each quarter. Overall, sustaining injuries impacts on the likelihood of winning the match, with two injuries resulting in a win percentage of approximately 40%.

Based on these findings, it is evident that injuries have had a greater impact on scoring potential and player rotations during 2010 when compared with previous seasons. Whilst a progressive increase in the effect of injuries on scoring patterns and match outcome is evident, it is clear that interchange rotations reached a critical mass in 2010, and thus corresponding rotation rates cannot sustain a single injury during the match.

Acknowledgements

The authors would like to acknowledge the assistance, guidance and expert knowledge provided by Shane McCurry and Patrick Clifton through a variety of communication mediums. Without their contributions this analysis would not have been possible.

BEYOND AVERAGES – IMPROVED INDIVIDUAL PERFORMANCE METRICS IN LIMITED OVERS CRICKET

Steven E. Stern [a,b]

[a]The Queensland University of Technology, Brisbane, Australia
[b]Corresponding author: steven.stern@qut.edu.au

Abstract

The simplest, and most common, measures of individual performance in limited overs cricket are the batting and bowling average. Along with strike rates and economy rates, which are equally easy to construct, they form the traditional framework around which a cricketer's assessment is based. However, it is well known that these measures, while easy to calculate, can be misleading with regard to the true value of an individual to his team's success or failure. In this paper, we extend the work of Lewis (2005) to develop a measure of the actual relative contribution of each batsman and bowler to the final scores in a limited overs match. By so doing, we can develop better metrics of performance which avoid many of the pitfalls of the standard measures. Based on a new performance metric, the adjusted net runs attributable (aNRA), we rank the best performers in both batting and bowling overthe Indian Premier League seasons from 2010 to 2013 and examine the relationship between aNRA and the official Man of the Match (MotM) awards. In addition, we use the new metric to assess the outcomes of the 2014 player auction, wherein teams bid for the services of the players for upcoming seasons.

Keywords: Duckworth-Lewis method; Performance metrics; Player rankings

1. INTRODUCTION

Assessing individual performers in team sporting arenas is a fundamental activity of both fans and administrators alike. Typically such assessments are largely subjective or else based on objective measurements which owe their prominence to their ready availability and simplicity of calculation, but may not be the most directly applicable measures with respect to the most important aspect of an individual's performance; namely, their contribution to the success of the team. In the case of cricketers, as for many other individual sportspersons in team sports, the most common and traditional objective measures of performance, such as averages and aggregates, strike rates and economy, are used largely due to their ease of construction and ubiquitousness. However, these statistics do not directly measure a player's contribution to the most important aspect of a match, its outcome. Players may accumulate impressive statistical performances in lost causes or easy victories, while others may have their match-changing, though not voluminous in terms of the usual measures, performances under-valued. As in anything, the output of a participant needs to be judged in proper context.

A limited overs cricket match proceeds in two innings, each continuing until either the completion of a prescribed number of deliveries, the loss of 10 wickets or, in the case of the second innings, the game is won. The first innings sees one team score as many runs as they can, using their available resources (i.e., deliveries and wickets). Then, in the second innings the other team attempts to score more than their opponent. As in many sporting contests, it is often the case that a limited overs cricket match's outcome is clear well before it concludes. For a cricketer, then, when runs are scored or wickets taken, and the circumstances of the match under which these events occur, are at least as important as their mere number in assessing their contributions to the team cause. Indeed, it has long been accepted that the simple, and most common, measures of performance, the batting and bowling averages, are often a misleading indicator of a player's true ability and worth. Perhaps more insidiously, use of simple averages as a key performance indicator may actually encourage players to undertake strategies which prioritise personal statistics over team goals. As one example, a batsman may choose to accumulate runs slowly (and safely) to pad his personal tally at the expense of under-utilising the team's finite available resources and thus not leaving his team enough to actually win the match. Further, the importance of "not outs" in batting average calculations provides strong incentive for batters to preserve their own wicket at the expense of seeking risky runs which might more directly benefit the team cause.

A range of researchers have attempted to better account for the true performance of cricketers by using statistical measures beyond the ones in most common usage; namely batting averages and strike rates (runs per delivery) for batsmen and bowling averages and economy rates (runs conceded per over bowled) for bowlers. For example, Croucher (2000) introduced the "batting index" metric, which is the product of a batsman's average and strike rate. Alternatively, the average plus strike rate (per 100 deliveries), or APS, is becoming a frequently quoted measure of batsman's capabilities, particularly in the shortest version of the game, Twenty20 cricket. Other approaches include detailed multivariate analysis of scorecards (e.g.,Barr and Kantor, 2004). These approaches, though,use only aggregate match information (i.e., how many runs were scored or wickets taken, but not when during the match they occurred). Indeed, the official International Cricket Council (ICC)player ranking methodology, while calculated using "a sophisticated moving average…based on various circumstances of the match,"(http://www.reliancemobileiccrankings.com/) , uses solely information available on a match scorecard. While such approaches have the benefit of ease of implementation, as no detailed (and,typically, difficult to obtain) information is needed, they, like the simple averages they replace, tend to ignore crucial contextual information contained in the timing of when runs are scored or wickets taken.

One reason no early attempts were made to add match context to performance metrics was that, until recently, there was no definitive quantitative measure of "match situation" to incorporate. An early attempt to assess the net contribution of individual players using contextual information was investigated by Johnston et al. (1993) employing a dynamic programming approach to assessing expected versus observed outcomes on individual deliveries. However, with the development of the Duckworth-Lewis (D/L) methodology (Duckworth and Lewis 1998, 2004), which determines the relative importance of each ball bowled in a limited overs cricket match by calculating the proportion of the final total score which would have been expected to be scored, given the match situation at the time (i.e., how many balls remain in the innings and how many wickets have already been lost). A number of authors (for instance, Clarke and Allsop, 1993; de Silva et al, 2001; and, Stern, 2008) used the D/L-defined notion of "scoring resources" in assessing match outcomes (margin of victory and team performance ranks). In addition, Beaudoin and Swartz (2003) defineda player's*Runs per Match* (though perhaps a more accurate name would be*Runs per Resources Utilised*) as a potential replacementfor the common averages.

To better account for the true value of runs scored and conceded, Lewis (2005) suggests that player performance in a match is sensibly calibrated using a measure of the *net runs attributable* (NRA) to them. While we leave the details of the calculations to the following sections, we note that the underlying philosophy of this approach is to assess not only the number of runs scored or conceded, but to also contextualise their importance. For instance, a batter who scores a large number of runs, but does so slowly and utilising excessive resources, will find that his NRA is far lower than his actual run total, and may even be negative.

As such, Lewis (2005) suggests the contribution of any player can be assessed by comparing their actual output with what would have been expected to occur during the period of the match to which they contributed. In this paper, we continue this train of development by extending the idea of a player's NRA in any given match to include an assessment of not just the timing of their performance but also the relative quality of the opposition they faced. To do so, we proceed by using the fundamental construct of determining what would have been expected to happen had the player being evaluated been absent from the match and instead been replaced by a player with an "average" contribution. In this respect, our newly proposed adjusted NRA (aNRA) for cricketers is akin to the concepts which have become staples of the famous "sabermetrics" movement in American Major League Baseball (popularised in the famous book *Moneyball* by Michael Lewis).

2. NET RUNS ATTRIBUTABLE

Lewis (2005) suggested that a batsman's or bowler's net contribution to his side could be calculated by assessing how many runs he actually scored or conceded, respectively, in relation to the number of runs he would have been expected to score or concede given the proportion of his team's resources he utilised. For example, if a batsman accumulated a large personal score, but in order to do so utilised an excessive amount of his team's available resources, then his contribution would be appropriately down-weighted. Further, this gives a method to assess the true contribution of batsmen at different spots in the order, since early batsmen must weigh the risks of losing wickets differently than those batting at the end of the innings.

2.1. Lewis' Net Runs Attributable for Batsman

Specifically, Lewis (2005) suggested that the *net runs attributable* (NRA) to the *i*th batsman in a given side for a given match should be defined as:

$$N_i = \sum_{k \in K_i} (\sigma_k - \varepsilon_k)$$

whereK_i is the set of indices of balls faced by batsman i, σ_k is the number of runs scored by the batsman (i.e., excluding extras) on the *k*th ball and ε_k is the expected number of runs scored on the *k*th ball. To determine ε_k, Lewis (2005) suggested employing the Duckworth-Lewis (D/L) methodology, so that $\varepsilon_k = G_{50}\rho_k$, where ρ_k is the D/L resources associated with the kth ball and G_{50} is the global average score for 50-over matches of the appropriate level (e.g., for men's international matches at present G_{50} = 245).

While this approach definition provides stability and makes sense over the long-run, we note that the value of G_{50} has gradually increased over the years, which means that recent player ratings would not be directly comparable to historic ones. Moreover, using the global average value of G_{50} will mean that player performances will not be calibrated to the match-specific conditions which, in the short to medium term, will mean, for instance, that batsmen who tend to play on batting-friendly pitches will have their assessment measures overstate their actual performance when compared to batsmen who tend to play on bowling-friendly pitches.

Instead, then, we might choose to define $\varepsilon_k = U\rho_k$, where U is a match-specific resource utilisation rate based on the observed scoring rate in the specific match in which the players' performances took place. There are several possible choices for U. We might use the innings-specific utilisation rate associated with the innings in which batsman i participated, so that $U = S/R$ where S is the final score (of runs off the bat) of the innings in which the player being evaluated participated and R is the total resources available in that innings (e.g., a full 50-over innings would have R = 1, meaning $U = S$). Doing so, however, will tend to damp performances, since a large innings score will translate into a large base utilisation rate and thus mean that individual performances gauged against this baseline will not appear as impressive as they actually were.

As an alternative, we can define an overall match-specific utilisation rate, so that

$$U = \frac{S_1+S_2}{R_1+R_2} \qquad (1)$$

where S_1 and S_2 are the total runs scored off the bat in the first and second innings, respectively, and R_1 and R_2 are the associated total innings resources. In this way, we use all available match-specific information to assess the appropriate baseline for comparison, meaning that if one innings score is much larger than the other, the batsmen who scored those runs will get appropriate credit (and the bowlers of the opposition will also be adequately held accountable).

Even using the overall match scoring rate (1), though, has an issue. Specifically, if we are to assess the performance of a player accurately, we should assess their performance against an expected rate calculated from the observed performace in the match with their own contribution removed. Otherwise, an extremely good (or bad) performance will noticeably affect the overall match-specific resource utilisation rate and make individual performances seem less pronounced than they actually were. Thus, we define the adjusted baseline resource utilisation rate for the ith batsman:

$$U_i = \frac{S_1+S_2-s_i}{R_1+R_2-r_i} \qquad (2)$$

where $s_i = \sum_{k\in K_i}\sigma_k$ and $r_i = \sum_{k\in K_i}\rho_k$ represent the runs scored and the resources used by batsman I, respectively. Using (2), we then set the net runs attributable to batsman i as:

$$N_i = \sum_{k\in K_i}(\sigma_k - U_i\rho_k).$$

2.2. Net Runs Attributable for Bowlers

The calculation of the corresponding performance measure for the jth bowler follows essentially identical lines, but must take account of the fact that wides and no balls are counted against a bowler. Also, more runs means a worse performance for a bowler as opposed to a better one. So we define our measure as a subtraction of expectation minus observation, as opposed to the reverse as we did for batsmen. As a result, the net runs attributable to a bowler should be interpreted as the net runs he was able to prevent when compared to the average rate of runs conceded by the other bowlers in the match.

Thus, the net runs attributable to bowler j is:

$$M_j = \sum_{k\in L_j}(\epsilon_k - \sigma_k) + \sum_{k\in L_j}(V_j\rho_k - \omega_k) = m_j + V_j r_j - w_j$$

where L_j is the set of indices of balls bowled by bowler j, ω_k is the number of wides and no balls tallied on the kth ball, $w_j = \sum_{k\in L_j}\omega_k$ is the total number of wides and no balls delivered by bowler j, $r_j = \sum_{k\in L_j}\rho_k$ is the resources associated with the deliveries of bowler j, $m_j = \sum_{k\in L_j}(\epsilon_k - \sigma_k)$ is the net runs off the bat attributable to the bowler (a quantity we shall find useful in the next section), $\epsilon_k = \mathcal{U}_{j(k)}\rho_k$ is the expected runs off the bat for ball k, the function $j(k)$ represents the index number of the bowler who delivered ball k, and we define the baseline rates of runs scored off the bat and runs from no balls and wides relevant for bowler j as

$$\mathcal{U}_j = \frac{S_1+S_2-m_j}{R_1+R_2-r_j} \text{and } V_j = \frac{W_1+W_2-w_j}{R_1+R_2-r_j},$$

where W_1 and W_2 are the total wides and no balls bowled in the first and second innings, respectively.

2.3. Opponent-Adjusted Net Runs Attributable

The net runs attributable is a major advancement on the more commonly used performance measures such as batting and bowling averages and strike or economy rates. However, NRA does not adjust directly for the ability of the individual opponents faced (they do, of course, account for the overall ability of the opponents as a team, but not for the individual fluctuations of ability within the opponent team). Clearly, if a batsman faces most of his deliveries from the opponent's best bowler, then his net contribution will be expected to be lower than it would have been had he faced the lesser bowlers.

To account for this, we can compare a player's runs scored or conceded not to a "match averaged" expectation, but to an adjusted version which accounts for which opponents were faced. In particular, we can define adjusted versions of N_i and M_j by augmenting the expected runs values, $\varepsilon_k$and ϵ_k, with an adjustment factor to account for the match performance of the opponent faced on that delivery. Specifically, we will define the *adjusted net runs attributable* (aNRA) to a batsman as:

$$\mathcal{N}_i = N_i + \alpha \sum_{j \in J_i} \sum_{k \in K_i \cap L_j} m_j \pi_{jk}$$
$$= \sum_{j \in J_i} \sum_{k \in K_i \cap L_j} \{\sigma_k - (U_i r_j - \alpha m_j)\pi_{jk}\}$$

whereJ_i is the set of indices of bowlers faced by batsman i, α is a tuning parameter (set to 0.01 in what follows, though further work is needed to assess an optimal value) and $\pi_{jk} = \rho_k / r_j$ is the proportion of the resources associated with the jth bowler's deliveries that the kth ball comprises.

Similarly

$$\mathcal{M}_j = M_j + \alpha \sum_{i \in I_j} \sum_{k \in K_i \cap L_j} N_i \varpi_{ik}$$
$$= \sum_{i \in I_j} \sum_{k \in K_i \cap L_j} \left[\{(\mathcal{U}_j + V_j) r_j - \alpha N_i\}\varpi_{ik} - (\sigma_k + \omega_k)\right]$$

whereI_j is the set of indices of batsmen that bowler j faced and $\varpi_{ik} = \rho_k / r_i$ is the proportion of the resources associated with the ith batsmen's deliveries that the kth ball comprises.

This adjustment allows for opponent performance by simply replacing the average expectation for any given ball by an amount which is modified according to the average performance (within the match) of the specific opponent. For instance, for a batsman's calculation we compare their ball-by-ball scores, σ_k, to an expected outcome which comprises the overall average expection, ε_k, modified by a proportion of the resource-weighted average amount of net runs attributable to that ball for the opposing bowler, though we are careful in this case to adjust according to m_j, the actual runs off the bat attributable to bowler j and not M_j, as a batsman's performance does not include wides and no balls. Also, note that the use of the proportionality constant, α, allows for this process to be both iterated and damped so that the adjustment does not become "circular".

3. USES OF NET RUNS ATTRIBUTABLE

Using ball-by-ball data compiled from commentary of all completed IPL matches between 2010 and 2013, we now examine the top performers in terms of aNRA. In particular, we examine both average performances across matches played, as well as individual performances within single matches. The former investigation allows us to assess the overall rating and ability of players, smoothing out the vagaries of performances within individual matches. The latter investigation, however, indicates which players most contributed to their team's performance on the day. As will be discussed, such individual match investigations will allow us to determine the extent to which a player's performance in a given match had a direct influence on its final outcome.

3.1 Player rating and ranking

Table 1 shows the top 20 average aNRA values for batsmen and bowlers who contributed (i.e., actually did bat or bowl, as opposed to just being in the side) in at least 10 matches over the entire period. In addition, each batsman's average and strike rate (runs per 100 deliveries faced) is included and each bowler's average and economy rate (runs conceded per over bowled) along with the rank of each of these values among the 125 batsman and 106 bowlers who played in the IPL between 2010 and 2013 and contributed statistically to at least 10 matches.

For those who follow cricket, most of the names in Table 1 are both familiar and not unexpected, as they are also the players who top lists of the more conventional statistical measures. Indeed, 7 of the 10 highest batting averages and 9 of the 10 lowest bowling averages belong to players in Table 1, as do 7 of the 10 highest batting strike rates and 9 of the 10 lowest bowling economy rates. It is no surprise, then, that the Pearson and Spearman correlations between average aNRA and the more commonly used statistics are reasonably high, as Table 2 shows (note that the correlation values are negative for the bowling statistics since for the usual measures lower values indicate better performance, while aNRA has been defined so that larger values indicate better performance for both batsman and bowlers).

Nevertheless, there are some interesting omissions and inclusions in Table 1. As a notable example, Sachin Tendulkar, perhaps the most accomplished batsman of his generation, does not appear. In part, this may be attributed to the fact that he is reaching the end of a long career. However, there are also suggestions that, while he has scored numerous runs (over the four seasons, Tendulkar scored 1,782 runs at an average of 35.64, which ranked 10th), he often does so at a rate which is potentially detrimental to his team. Of course, care must be taken in making such an interpretation, but it is interesting that despite his very high batting average, Sachin's average aNRA for the 4 seasons under study is only 1.70.

Other notable bastman missing from Table 1 are Michael Hussey, whose batting average of 42.31 was the fourth highest but his performances only translated to an average aNRA of 1.18 (36th ranked); and ViratKohli, an up and coming Indian player whose batting average of 37.24 was 8th highest but whose average aNRA of –0.26 was only ranked 66th. Furthermore, neither Kumar Sangakkara, MahelaJayawardene nor Rahul Dravid are among the top 20, despite each having very high batting averages and aggregate runs totals during the period.

Table 1: Top 20 IPL Batting and Bowling Performers by average aNRA between 2010 and 2013

Batsmen				Bowlers			
Player	Average aNRA	Batting Average (rank)	Strike Rate (rank)	Player	Average aNRA	Bowling Average (rank)	Economy Rate (rank)
1. CH Gayle	11.89	55.74 (2)	164.28 (1)	1. SP Narine	9.36	14.65 (2)	5.47 (1)
2. SR Watson	8.26	37.51 (7)	143.50 (11)	2. MM Sharma	6.29	16.30 (5)	6.43 (7)
3. KP Pietersen	7.53	60.11 (1)	148.63 (5)	3. A Kumble	6.06	23.94 (28)	6.43 (6)
4. V Sehwag	6.74	30.19 (27)	157.31 (3)	4. R Rampaul	5.98	20.08 (10)	6.93 (20)
5. DA Miller	5.67	51.60 (3)	156.84 (4)	5. DL Vettori	5.84	33.84 (74)	6.54 (8)
6. G Gambhir	5.15	31.75 (22)	128.18 (42)	6. A Chandila	5.74	22.00 (20)	6.21 (2)
7. MS Dhoni	4.81	36.51 (9)	147.05 (8)	7. JP Faulkner	5.54	15.88 (3)	6.94 (21)
8. RN ten Doeschate	4.54	30.43 (26)	135.67 (18)	8. DW Steyn	5.26	20.82 (12)	6.27 (3)
9. S Sohal	3.74	24.90 (50)	130.37 (32)	9. R Ashwin	5.07	24.23 (30)	6.40 (5)
10. SK Raina	3.30	35.40 (11)	141.19 (15)	10. SL Malinga	4.56	18.08 (6)	6.60 (10)
11. KA Pollard	2.85	25.83 (43)	147.08 (7)	11. M Muralitharan	3.69	25.91 (36)	6.98 (23)
12. SPD Smith	2.74	40.08 (5)	130.58 (30)	12. RE van der Merwe	3.69	20.83 (13)	6.28 (4)
13. STR Binny	2.56	28.53 (34)	141.72 (13)	13. GB Hogg	3.67	28.40 (51)	7.22 (32)
14. YK Pathan	2.50	26.56 (41)	140.81 (16)	14. S Nadeem	3.48	34.24 (76)	6.66 (12)
15. RG Sharma	2.46	32.96 (20)	129.41 (36)	15. MG Johnson	2.76	19.13 (8)	7.17 (28)
16. DA Warner	2.45	29.58 (29)	134.46 (21)	16. DE Bollinger	2.74	18.73 (7)	7.22 (31)
17. SE Marsh	2.28	37.85 (6)	130.40 (31)	17. SK Warne	2.69	28.25 (49)	6.99 (24)
18. RV Uthappa	2.21	27.87 (36)	129.45 (35)	18. Shakib Al Hasan	2.59	16.09 (4)	6.67 (13)
19. Harbhajan Singh	2.09	19.64 (84)	147.44 (6)	19. BJ Hodge	2.54	14.20 (1)	7.47 (39)
20. RA Jadeja	2.05	23.28 (59)	131.32 (26)	20. B Kumar	2.38	29.96 (58)	6.71 (15)

Table 2: Correlations between aNRA and Common Measures

Batsmen (n = 125)			**Bowlers** (n = 106)		
Statistic	Pearson	Spearman	Statistic	Pearson	Spearman
Batting Average	0.665	0.606	Bowling Average	–0.431	–0.564
Strike Rate	0.767	0.870	Economy Rate	–0.858	–0.868

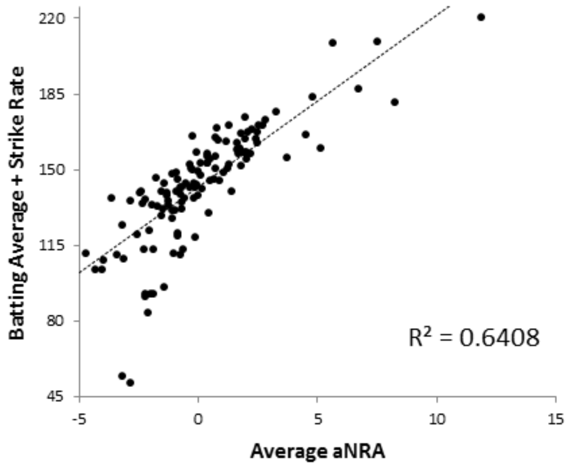
On the bowling side there are notable omissions as well, including Zaheer Kahn (though he is ranked 25th), the brothers Morne and AlbieMorkel (ranked 41st and 71st, respectively) and Dwayne Bravo (ranked 50th). Another key name which does not appear in Table 1 is Jacques Kallis, widely regarded as one of the best all-rounders of his generation: in 90 IPL matches between 2008 and 2013 he score 2,276 runs (6th most all time as of the close of 2013) and took 61 wickets (18th most all time, again as of the close of 2013). Nevertheless, many critics have been of the view that his batting style is not suited to the requirements of Twenty20 cricket. His aNRA tends to support this view (over the 2010–3 period, his batting aNRA was –2.41 which ranked 110th among those with at least 10 innings during that timeframe). By contrast, the current all-rounder seen as on a par with Kallis, Shane Watson, appears in Table 1 on the batting side and his bowling aNRA is ranked 29th. Interestingly, critics of Watson have proferred views that his batting style is not suited to the traditional longer form of the game (and a basic comparison of Test career statistics for Kallis and Watson clearly support this position).

Of equal interest to the omissions, some of the inclusions in Table 1 show that having a high batting average or low bowling average is not necessary to make an important contribution to the team score. For instance, Kieron Pollard adds an average of nearly 3 runs per match as measured by aNRA but has only the 43rd highest batting average at 25.83, and Harbhajan Singh, primarily a bowler, makes the top 20 list as a batsman despite his batting average of only 19.64. In addition, Brad Hodge, primarily a batsman and only a part-time spin bowling option adds over 2.5 runs per match according to aNRA, despite his high economy rate of nearly 7.5 runs per over (ranked 39th). In part, this may be explained by batsmen taking unwarranted risks off his bowling as he is not a top-line bowler and this view is only enhanced by noting that Hodge's bowling average of 14.20 is the lowest among all 106 bowlers with at least 10 bowling performances.

Overall, the pattern of included and excluded players suggests that aNRA rewards players who contribute quality not quantity. In addition, it recognises that being not out, for batsman, is not necessarily of huge importance, unlike the case for batting averages, where lower order batsman often gain the benefit of increased batting averages due to a large proportion of not out innings. Similarly, aNRA recognises that taking wickets is only directly important insofar as it helps keep scoring rates down. Thus, taking wickets late in matches, when batsmen are playing in a high risk manner in search of quick runs, will aid the bowling average greatly, but not the average aNRA. Of course, none of these observations directly validate or invalidateaverage aNRA as a rating measure. However, the underlying D/L structure gives aNRA a solid foundation. Moreover, we note that a recent *ad hoc* measure that has been proposed for batsmen is the so-called APS (average plus strike rate) and Figure 1 displays the relationship between this statistic and the average aNRA for the 125 batsmen who had at least 10 innings during the four IPL seasons between 2010 and 2013.

The correlation between average aNRA and APS is 0.8, higher than for either batting average or strike

Figure 1: average aNRA versus Batting Average plus Strike Rate (APS) for 125 IPL Batsmen



rate individually (see Table 2), and gives a degree of external validation for aNRA as a rating tool for batsmen, given that APS is generally agreed by experts to give a good generic assessment of a batsman's capabilities in the Twenty20 format. The benefit of aNRA over APS lies in its methodological underpinnings, avoiding the somewhat arbitrary simple summation of the two basic statistical measures of batting performance. For example, there is no solid rationale as to why the two statistics are equally weighted in the APS combination.

To further examine the alignment of aNRA values and "expert" opinion, we now investigate aNRA values from individual matches.

3.2. Man of the Match Award calculations

The worth of individual performances can be measured in many ways; however, we focus here on the player(s) with the highest aNRA in the match. Doing so immediately raises the question of comparison of batting and bowling aNRA values. Given that each are ostensibly measured in terms of runs attributed to an individual, it seems reasonable (and indeed is a strength of the metric itself) that simple additive combination of batting and bowling values gives a sensible measure of the overall contribution of an individual player to a single match (of course, it does not measure contributions made via the third major discipline of the game; namely fielding). We refer to this value as the combined aNRA, or cNRA.

Assessing the reasonableness of this combined aNRA as a measure of performance, we examine the frequency with which players with high combined aNRA are deemed to be the "Man of the Match (MoM)", an award given in each game played by a pre-determined (though often different for each match) panel of "expert" assessors. Table 3 gives a breakdown of the correspondence between players with the high combined aNRA and the MoM award.

Man of the Match awardee had: (n = 282)		
Criteria	*Count*	*Cumulative Percentage*
Highest combined aNRA in match	147	52.1%
or Highest combined aNRA on winning side	163	57.8%
or Highest individual aNRA in match	167	59.2%
or Highest batting aNRA value in match	186	66.0%
or Highest bowling aNRA value in match	202	71.6%
or Highest individual aNRA on winning side	205	72.7%
or Highest batting aNRA value on winning side	211	74.8%
or Highest bowling aNRA value on winning side	218	77.3%
or Combined aNRA among top 3 in match	237	84.0%

Table 3: Man of the Match Awards in IPL 2010 to 2013

So, in just over half the matches, the winner of the MoM award also had the highest combined aNRA of any player in the match. Further, in nearly 60% of the matches, the MoM award went to the player with either the highest combined aNRA overall or else the highest combined aNRA on the winning side, where MoM awardees come from well over 95% of the time (of the 282 IPL matches completed between 2010 and 2013, only 7 MoM awardees played on the losing side). If we broaden our scope, Table 3 indicates that over three-quarters of MoM award winners had either the highest combined aNRA value or the highest individual aNRA value in one discipline (perhaps restricted to scores from the winning side). Given this, we conclude that the aNRA metric is reasonably well in line with what experts deem to be the "best" performance of the match. Indeed, as Table 3 further shows, when we include combined aNRA values in the top 3 for each match, we capture 84% of all MoM award winners.

While the degree of matching between high aNRA values and MoM award winners gives some degree of validity to the use of aNRA as an appropriate rating metric, it is equally instructive to investigate the remaining 45 (16%) matches in which the MoM award winner did not have one of higher aNRA values. Table 4 breaks down these 45 matches according to some simple criteria.

Table 4: Breakdown of Man of the Match with low aNRA

Man of the Match awardee with low aNRA had: (n = 45)		
Criteria	*Count*	*Percentage*
Highest individual score in match	16	35.6%
Highest individual score on winning side	9	20.0%
Highest wicket tally in match	12	26.7%
Highest wicket tally on winning side	2	4.4%
TOTAL:	**39**	**86.7%**

This casts the correspondence of Table 3 in a somewhat different light, as we see the expert panel assessment skews toward the traditional metrics of quantity of runs and/or wickets (indeed, in 2 of the 6 matches not covered in Table 4, the MoM winner had what would typically be deemed to be the best “quantitative all-round” performance in terms of combined runs scored and wickets taken). While high aNRA values will frequently align with the best “quantitative” performances (which explains the high correspondence in Table 3), the underlying focus of aNRA is quality instead of quantity, and explains the discrepancy in the matches investigated in Table 4. Just finally, though, it should be pointed out that in 2 of the 6 matches not covered by the criteria in Table 4, the MoM winner’s performance fell in to the category of “match winner”; that is, a performance which was not quantitatively the largest, but was clearly key in taking the match position from a potentially losing one to a winning one (e.g., in one such instance, eventual MoM winner DR Smith, having scored just 10 runs from the 6 deliveries he had faced and needing 14 runs for victory from the 3 remaining deliveries, a seemingly hopeless position, proceeded hit a six and two boundaries to grasp victory from the jaws of defeat, though not accumulating many net runs attributable).

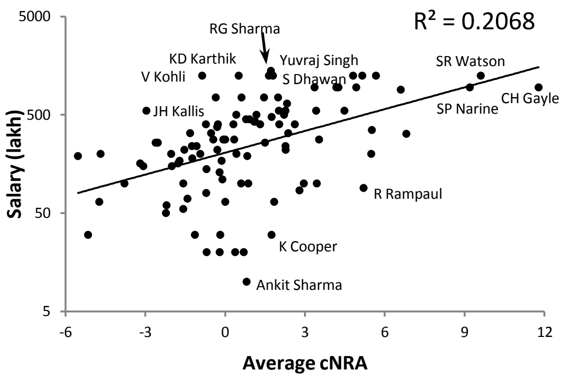
3.3. Player salaries

Finally, we investigate the relationship between cNRA and the 2014 salary of players. It should be noted that the 2014 salary of players is determined in two possible ways. Players whose contracts are not complete may be “retained” by their current club at their previous salaries. Otherwise, the remaining players (i.e., players either not retained or out of contract) have their salaries determined at auction.

Figure 2 shows the relationship between 2014 salary and the players average cNRA over the previous 4 seasons. Clearly, there is some connection between ability and salary, but there are also other factors at play. In particular, while winning matches is the driving incentive for players, the team owners are generally interested in profit. Of course, having a winning team is a good way of generating profit, as fans tend not to flock to watch losing teams. However, there are other factors which determine attendances and profits. Specifically, name recognition is important in bringing in large crowds. As such, we note that many of the apparently over-valued players (i.e., those whose salary is large relative to their ability, as measured by average cNRA) are well-known Indian players, such as ViratKohli and Yuvraj Singh, who would tend to have loyal followings.

Alternatively, salaries may also be driven by international economics. Indeed, many of the apparently under-valed players are from Pakistan or the West Indies. In these countries, income for cricketers is limited, and thus they will likely be more willing to play for (relatively) lower wages. Furthermore, domestic players without any international experience have their salaries capped.

Figure 2: average cNRA versus 2014 Salary for IPL Players



Finally, though, we note that there is some connection between skill and salary, and thus there may well be interest among team owners in trying to determine whether they are under- or over-paying their players. Doing so might allow them to construct a team more likely to be successful at a fixed salary level.

4. CONCLUSION

In this paper, we have introduced extensions to Lewis’ NRA measure of individual contribution to a limited-overs cricket match. The extensions include improved relative comparisons by employing expected results which account for scoring rates of the other players in the match as well as the relative proficiency of actual opponents faced.

We have seen that the newly derived performance metrics, aNRA and cNRA, have a reasonable correlation with more traditional statistical measures; however, given their use of contextual information via the D/L methodology, we believe they provide a more appropriate and interpretable measure of value. Nevertheless, we should clearly note that these measures do still have various deficiencies. In particular, they cannot account for fielding, nor do they account for the potential importance of partnerships in determining valuable contribution to the team outcome. For instance, it may be that a batsman playing a “sheet anchor” role will be extremely valuable to a team’s performance even though on its own his innings may seem to be scoring at a relatively low resource utilisation rate. Similarly, strike bowlers are well-known to have generally high economy (and thus resource utilisation) rates, but the potential psychological factors that their inclusion in the team brings may lead to other bowlers achieving greater success than they otherwise would have.

To some extent, some of these shortcomings can be ameliorated if we adjust our metric to focus not just on the performance of individuals but on the outcome of a match. For example, a batsman playing the "sheet anchor" role in a losing side may well be blamed for batting too slowly, while in a winning side his contribution is clearer. To this end, further work on extending aNRA and cNRA to include indicators of whether the individual player in question was on the winning or losing side is warranted and is currently ongoing.

Acknowledgements

As always, my heartfelt thanks goesto Frank Duckworth and Tony Lewis for their long collaboration on all things D/L. In addition, thanks to Professor UjwalKayande of the Melbourne Business School for ongoing collaboration and insightful discussion of all things cricket.

References

Barr, G.D.I. and Kantor, B.S. (2004) A criterion for comparing and selecting batsmen in limited overs cricket.*Journal of the Operational Research Society*55,1266-1274.

Beaudoin, D. and Swartz, T.B. (2003) The best batsmen and bowlers in one day cricket. *South African Statistical Journal* 37,203–222

Croucher, JS (2000) Player ratings in one-day cricket. In: *Proceedings of the Fifth Australian Conference on Mathematics and Computers in Sport*. Eds: Cohen, G. and Langtry, T. Sydney University of Technology, NSW. 95-106

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

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

Duckworth, F. C. & Lewis, A. J. (2004).A successful operational research intervention in one-day cricket. The Journal of the Operational Research Society, 55, 749-759.Johnston, M.I., Clarke, S.R. and Noble, D.H. (1993) Assessing player performance in one-day cricket using dynamic programming. *Asia-Pacific Journal of Operational Research* 10, 45-55.

Lewis (2005)Towards fairer measures of player performance in one-day cricket. Journal of the Operational Research Society 56, 804-815.

Stern (2008)Rankinginternational limited-overs cricket teams using a weighted, heteroskedastic logistic regression with beta distributed outcomes. *Proceedings of the 9th Annual Australasian Conference on Mathematics and Computers in Sport*. (ed. John Hammond). ANZIAM.

PROFITING FROM THE RUN CHASE IN 50-OVER CRICKET

Jonathan Sargent

School of Mathematical & Geospatial Sciences
RMIT University, Melbourne, Australia.

Abstract
This paper investigates fitting Weibull probability distributions to "required runs" data during the second innings of one-day international cricket matches (ODI), to derive a profit from wagering. Each innings consists of a maximum of 50 overs, which decay at a constant rate as the match progresses, and 10 wickets, which decay as the bowling team dismisses each opposition batsman. Any intersection of these resources in the second innings—defined in this research as the "match state"—yields the required runs variable (R) or the difference between first-innings team (p) runs plus one runs (target) and observed second-innings team (q) runs at match state, t. Historical match states were populated using "ball-by-ball" data and Weibull distributions with optimised scale parameters fit to the R samples of sufficient size. Bootstrapping was applied to generate relevant statistics in match states where parametric assumptions were violated. The probability density function produced the likelihood of q defeating p, given any match state in the second innings and the team strengths, which were determined by a betting agency's head-to-head odds offered immediately prior to the commencement of the first innings. The probabilities were converted to decimal odds and compared with the betting agency's odds of q defeating p, simultaneously offered at match state, t. A stratified betting strategy with a fixed amount wager on the author's head-to-head favourite at selected match states produced an attractive return on investment.

Keywords: Weibull distribution; probability density function; return on investment

1. INTRODUCTION

One-day international cricket (ODI) is a bat and ball sport comprising a maximum of 300 legitimate independent trials, or deliveries from the bowling team to the batting team, over two innings. Each team needs to accumulate as many runs as possible for a maximum of 50 "overs" (one innings) or until 10 of the 11 batsmen in the batting team are dismissed, or "out". The team batting first is declared the winner if one of these terminal points is reached in the second innings with the second batting team victorious if they surpass the first innings team's total with wickets and/or overs remaining. [1] The discrete composition of limited-overs cricket, in comparison to test cricket which lasts a maximum of five days with each team allotted two innings each to score their runs, has provided statisticians and mathematicians with countless research opportunities through the game's lifetime; Lewis (2005) described the game of cricket as a "sporting statistician's dream". Statistical modelling of runs scored for predictive purposes has been of particular interest with work as early as Elderton (1945) and Wood (1945) proving the geometric distribution to be an adequate fit for test match cricket runs. In the 50-over game, Clarke and Allsopp (2001) and de Silva et al (2001) made use of the Duckworth-Lewis rain interruption rules (Duckworth and Lewis, 1998) to project a second innings winning score, after the match's completion, to calculate a true margin of victory with respect to runs, not just wickets. The online publishing of "ball-by-ball" data in recent times has facilitated statistical modelling of matches in progress, or "in-play". Swartz et al (2006) applied a log-linear model to simulate runs scored during any stage of an ODI match for a proposed batting order while Sargent and Bedford (2012) simulated in-play outcomes through conditional probability distributions where the likelihood of a run(s) or a dismissal was

[1] The winning runs may be struck from the final delivery of the innings.

estimated prior to any delivery. The huge volumes of money wagered on ODI matches, as well as 50-over matches at the domestic level, have whet the appetites of researchers attempting to exploit betting market inefficiencies. Bailey and Clarke (2004) designed strategies to maximise profits derived from wagering on one batsman outscoring another during the 2003 ODI World Cup. Easton and Uylangco (2007) were even able to provide some evidence of the ability of market odds to predict the outcomes of impending deliveries in ODI matches. The research detailed in this paper was encouraged by fluctuations in in-play betting market odds which, at certain match stages, may over- or undervalue the likelihood of a team winning the match. By generating optimal betting moments with respect to the "match state" and the strength of the competing teams, it was anticipated that significant profits could be derived from the inefficiencies in these head-to-head market offerings.

The match state in either of the first or second innings describes the evolution of an ODI match through a finite number of intersections of overs—sets of six deliveries from the bowling team—and wickets—the number of times the bowling team has dismissed a member of the batting team. Overs decay at a constant rate as the match progresses while wickets decay as the bowling team dismisses each opposition batsman. Duckworth and Lewis (1998) modelled the joint decay of these two resources when setting revised second innings run targets for rain-interrupted matches. To win the match, the team (q) batting in the second innings of an ODI attempts to eclipse the first innings team's (p) aggregate runs; this is termed the "run chase". Archival records suggest the highest ODI run chase was achieved by South Africa in March, 2006, surpassing Australia's record first innings total of 434 with one delivery remaining in the match. The lowest chase was achieved by Sri Lanka in April, 2004, chasing down Zimbabwe's paltry total of 35 within the first ten overs of the match. With a sample of ball-by-ball run chase data from completed ODI matches going back to 2005, all possible match states were populated by team q's required runs for victory (R) at any stage in the second innings. Following some descriptive analysis, Weibull distributions, with optimised α and β parameters, were found to be adequate fits for most of these R samples with the probability density functions housing the likelihood of team q surpassing team p's run total at any match state, $t_{i=2}$. The Weibull distribution was especially relevant given its common use in testing failure rates (achieving the run target) over a specified time span (50 overs) (Smith, 1993). In-play head-to-head market odds of q defeating p were logged in close intervals during a series of recent ODI matches, assigned as the training set, then paired with the Weibull odds from the same t. Pre-match head-to-head market odds were also recorded as team strength indicators. A stratified betting strategy was devised where in-play betting moments were conditional on the match state with the best fitting R distributions, the pre-match market favourite and states with positive overlay. Return on investment increased as each stratum was added to the training sample, reaching in excess of 20%.

2. METHODS

i. Match State

An ODI match state, $t_{v,w}$ was defined in this research as any intersection between the number of elapsed overs ($v_i = 1,\ldots,50$) and the number of dismissals ($w_i = 1,\ldots,10$) during innings $i = [1, 2]$. Arriving at these unique match states required ball-by-ball data which was scraped from a cricket website then formatted and cleansed so over, wicket and run aggregates for every innings matched the "scorecard" (summary) data retrieved from the same site. The foundation variables were innings, delivery number and dismissals—from which to calculate the match state—first innings runs (the target) and runs scored off each delivery in the second innings to calculate the runs required for victory from any t. The sample included every delivery from ODI matches played by the recognized cricket nations (Australia, England, India, New Zealand, Pakistan, South Africa, Sri Lanka and the West Indies) dating back to 2005. Matches which were abandoned after the commencement of play or shortened due to rain interruption were excluded from the sample because of the erroneous effects of reduced targets. A total of 439 match states from all games were recognized within the sample, excluding terminal points where $w_2 = 10$. To ensure adequate distribution fits, only match states with statistically large R samples were selected.

Any match state during innings 2 accommodates the random variable, runs required for team q's victory, or:

$$R_t = target - r_{qt} \tag{1}$$

where R_t are required runs at match state t, $target = r_{pT} + 1$ where r_{pT} are aggregate runs achieved at the termination of the first innings by team p, where $T_p = \{t_{50,w<10} \cup t_{v<50,10}\}$, and r_q is aggregate runs achieved by team q at match state t ($r_q = 0$ at $t_{0,0}$). The terminal state for i=2 is:

$$T_q = \begin{cases} t_{v \leq 50,10} \bigcup t_{50,w \leq 10} & \text{if } R_t > 0 \\ t_{v \leq 50,w<10} & \text{if } R_t \leq 0 \end{cases} \quad (2)$$

where $R_t \leq 0$ is victory to team q where $v \leq 50$ is inclusive of victory from the final delivery of the match ($v_2 = 50$), and $R_t > 0$ is victory to team p at T_q. If at T_q, $R_t = 1$, $r_p = r_q$ and the match is declared a draw. In this scenario, if the teams are playing in a series, they share the points on offer; however, punters predicting a p or q victory relinquish their wager.

In limited-overs cricket, if team q achieves *target*, they are said to have won by the number of wicket resources remaining. In a famous match in 1996 against the West Indies, Australia, batting second, required four runs for victory from the final delivery with one wicket remaining ($t_{50,9}$). Michael Bevan was able to hit four runs from the final delivery to hand Australia an unlikely victory; Australia won by one wicket with $R_t = 0$ at T_q. In the case where there are over resources remaining at $R_t \leq 0$, rather than a team winning by the number of wicket resources remaining, a margin of victory with respect to runs can be calculated by rearranging the Duckworth-Lewis (D/L) formula for resetting a run target (due to rain interruption) (Duckworth and Lewis, 1998), as demonstrated by Clarke and Allsopp (2001) and de Silva et al (2001). Say team q eclipses a modest first innings total of 150 by 2 runs with ample wickets and overs remaining, and their D/L run projection is 275 runs by the end of the 50th over, their margin of victory would be 275 - (150 + 2) = 123 runs. In defeat, team q's run projection remains at r_{qT} and p is said to have won by R_t. For this research, where $R_t \leq 0$, run projections were retrospectively calculated for all r_{qt}, then required runs at t recalculated as:

$$\hat{R}_t = proj - r_{qt} \quad (3)$$

where *proj* is the D/L projected run aggregate for team q, replacing *target*. Equation (3) was deemed a fairer reflection of q's ability than (1) as it reflects how much further q would have progressed into their innings had they kept batting after surpassing the target. Figure 1 reveals the spread of the recalculated R mean at every match state through each over (top) and wicket (bottom). The mean of R at each match state in the sample has the lowest interquartile ranges at the start and end of the match. The start can be explained by a dismissal effect: frequent wickets diminish *proj* and accrued team runs—because batsmen become more defensive—and very few wickets fall in the opening overs (for v=2, max(w)=2). The lesser interquartile range towards match-end is logical because of the combination of teams who win, or are closing in on the target in the 50th over and teams who will not reach the target by match-end. The diminishing R in the wickets boxplot is reflective of a proportional relationship with over rate - the further a match progresses, the fewer the required runs are likely to be but the more likely wickets are to have fallen.

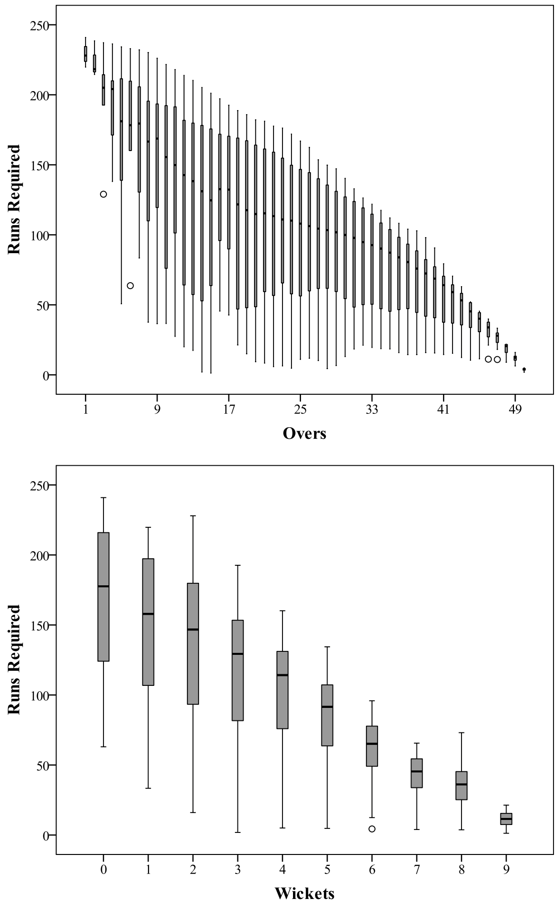


Figure 1. Dispersion of mean runs required through overs (top) and wickets (bottom).

ii. Distribution Fit

The two-parameter Weibull distribution is commonly used in analysing lifetime data where a measured quantity, X is set a "failure rate", proportional to some time period (Smith, 1993), for example failure of electronic equipment over time. Such properties were found to be applicable to an ODI run chase where team q could be set such a failure (or in this case, success) rate for achieving R through 50 overs. Furthermore, the negative skew in the majority of the R distributions demanded a flexible distribution, with a wide variety of shapes, such as the Weibull. While it is difficult to locate literature on the use of Weibull distributions in cricket, its application in estimating goal distributions in Association Football (Hamilton, 2011) and time between goals scored in the NHL (Thomas, 2007) is interesting.

Once the appropriate match states had been selected, Weibull distributions were fit to the R samples using the probability density function (*pdf*), denoted as:

$$f(\hat{R}, \lambda, k) = \begin{cases} \frac{k}{\lambda^k} \hat{R}^{k-1} e^{-(\hat{R}/\lambda)^k} & , if\ \hat{R} > 0 \\ 0 & , otherwise \end{cases} \tag{4}$$

where k and λ are the shape and scale parameters, respectively. These parameters were optimised through each sample to minimize the Pearson's chi-squared statistic, χ^2, using the observed (O) and expected (E) frequencies from Equation (1), or:

$$\min \chi^2 = \sum_{r=0}^{n} \frac{(O-E)^2}{E} \tag{5}$$

s.t. $\lambda > 0$ and $k = 5$.

where k was fixed, i) to address the existence of left and right tails in these distributions, ii) because $k > 1$ indicates the failure rate is proportional to time, such as in a run chase. The mean and variance of the Weibull distribution (see Table 1) are as follows:

$$\mathrm{E}(R) = \lambda \Gamma\left(1 + \frac{1}{k}\right) \tag{6}$$

$$\mathrm{var}(R) = \lambda^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2 \right] \tag{7}$$

where Γ is the gamma function. Table 1 reveals the ten match states with the lowest errors (in ascending order) as determined by Equation (5). Of the 127 match states selected, the maximum overs bowled in any state was 41 and maximum dismissals was 6, suggesting small and/or volatile samples as the second innings approaches its termination. The lowest error was at $t_{37,4}$ averaging 88.19 runs for victory, meaning, q required just over 88 runs in 13 overs; a run rate of 88/13 = 6.76 per over or just over a run every delivery. Table 1 confirms the discussion in Section 2i that mean and variance of R decrease as the second innings progresses and team q acquire their runs.

State(t)	E(R)	Var(R)	χ^2
(37, 4)	88.19	20.20	0.1104
(2, 0)	242.35	55.51	0.1429
(1, 0)	245.65	56.26	0.1467
(3, 0)	242.83	55.62	0.1468
(26, 3)	135.93	31.13	0.1729
(29, 4)	118.52	27.14	0.1823
(5, 1)	216.90	49.68	0.1855
(12, 2)	186.53	42.72	0.1862
(7, 1)	213.71	48.95	0.2208
(39, 5)	77.99	17.86	0.2500

Table 1. Statistics for match states with lowest error.

In Figure 2, the evolution of $f(R_t)$ as a match progresses is revealed through selected distributions (with optimised parameters). The right-most distribution is at $t_{1,0}$, the most populated match state. The long tails are reflective of the early match stage where very few runs have been scored and no wickets lost so a wide range of runs falls into the one wicket sub-sample (w=0), closely reflecting the match state prior to the commencement of the innings. The left-most distribution (v=37 and w=4) is taller with shorter tails; as the overs decay, required runs progressively diminish and fall into progressively more wicket sub-samples.

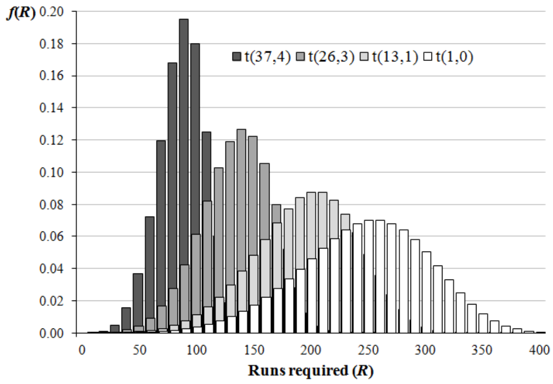


Figure 2. Weibull distributions at four different match states.

From the Weibull *pdf*, the probability of team q scoring the required runs, R with respect to t is:

$$P(r_q \geq R_t) = e^{-(R/\lambda)^k} \tag{8}$$

where $k = 5$, $\lambda > 0$. A training sample was generated which included a series of recently played ODI matches between the recognised cricket nations. The probabilities from Equation (8) were calculated at the completion of each over and where match state samples were large. Prior to investigating the efficiency of the probabilities generated from the Weibull *pdf* in the in-play betting market, the model success rate was tested by calculating the percentage of correctly predicted team q victories from all valid match scenarios, $mean(\theta_l)$, where:

$$\theta_l = \begin{cases} 1 & if \ P > 0.5, obs = 1 \\ 1 & if \ P < 0.5, obs = 0 \\ 0 & otherwise \end{cases} \qquad (9)$$

where θ_l is the binary outcome of match scenario l, P is $P(r_q \geq R_t)$ and obs is the observed team q match outcome:

$$obs = \begin{cases} 1 & \text{if } q \text{ defeats } p \\ 0 & \text{if } q \text{ loses to/draws with } p \end{cases} \qquad (10)$$

The baseline model predicted 67.1% of team q's victories from any l, irrespective of important modelling considerations such as team strength and home ground effects. This success rate, whilst modest, prompted an examination of the probabilities' performance in in-play wagering (see Section 3).

iii. Betting Strategy

After the development of team q victory likelihoods, a stratified betting strategy was investigated to maximise the return on investment (ROI) in the head-to-head ODI in-play markets. The first betting stratum was generated by indentifying the statistically significant match states (see Section 2i); Table 1 offers the ten best fitting distributions. This is the baseline model where fixed amount wagers, b at selected t were defined by:

$$b_t = \begin{cases} \$100 & if \ exp_q < \$2.00 \\ \$0 & if \ exp_q \geq \$2.00 \end{cases} \qquad (11)$$

where exp is the Weibull pdf expectation of q defeating p expressed in decimal odds. Profit (π) generated from each b_t for exp < \$2.00 was calculated with:

$$\pi_t = obs(100 m_q) - b_t \qquad (12)$$

where m_q is the in-play market price offered for q to defeat p at t and obs is from Equation (10). The second predetermined stratum was a team strength effect, that is, betting on q when q was a stronger side than p, as determined by pre-match market decimal odds, so $\varphi_q < \varphi_p$, logged just prior to the first delivery of i=1. A third stratum was included, which identified exp with positive overlay, simply calculated as $(exp - m_q)/exp$. After locating and matching reliable in-play market odds, profits at the various strata could be calculated.

3. RESULTS

Table 2 shows the baseline model to be unprofitable (ROI=-8.62%) due to heavy losses between \$1.01 and \$1.39. The baseline model produces a 17.3% ROI when betting on exp between \$1.40 and \$2.00; however, this observation was made with the benefit of hindsight. Such a trend would have to be monitored in future matches to become reliable.

Interval	AvePick	Wagers	SumProfit	AveProfit
1.00-1.10	80.95%	21	-$287.00	-$13.67
1.10-1.20	50.00%	12	-$517.00	-$43.08
1.20-1.30	63.16%	19	-$449.00	-$23.63
1.30-1.40	54.29%	35	-$950.00	-$27.14
1.40-1.50	71.43%	21	$100.00	$4.76
1.50-1.60	93.33%	15	$608.00	$40.53
1.60-1.70	88.89%	9	$455.00	$50.56
1.70-1.80	0.00%	2	-$200.00	-$100.00
1.80-1.90	50.00%	4	-$5.00	-$1.25
1.90-2.00	60.00%	5	$12.00	$2.40
Grand Total	**67.13%**	**143**	**-$1,233.00**	**-$8.62**

Table 2. First stratum wagering profit - match state only.

Triggering the other stratum (betting on q when they are the stronger side and with a positive overlay), although reducing the number of wagers, significantly increases ROI to over 22% (Table 3) which is a considerable profit. The prediction success rate (AvePick) of nearly 85% is also encouraging.

Favourite	y
Overlay	+

Interval	AvePick	Wagers	SumProfit	AveProfit
1.00-1.10	100.00%	16	$98.00	$6.13
1.20-1.30	50.00%	2	-$50.00	-$25.00
1.30-1.40	63.64%	11	-$52.00	-$4.73
1.40-1.50	80.00%	5	$176.00	$35.20
1.50-1.60	100.00%	6	$432.00	$72.00
1.60-1.70	100.00%	4	$391.00	$97.75
1.70-1.80	0.00%	1	-$100.00	-$100.00
1.80-1.90	100.00%	1	$138.00	$138.00
Grand Total	**84.78%**	**46**	**$1,033.00**	**$22.46**

Table 3. Third stratum wagering profit - match state, favourite and positive overlay.

In Table 4, results from Table 3 were filtered so betting was triggered when prices were at or above \$1.40—the observation from Table 2—further reducing the quantity of wagers, however, markedly increasing the ROI, suggesting profits earned when match outcomes are highly probable (\$1.00 to \$1.40, or between 71% and 100%), do not adequately cover losses in the same intervals. This is evident in Table 2 where a success rate of 81% in the \$1.00-\$1.10 interval translates to a -13.7% ROI.

Market inefficiencies were also located in intervals of the first stratum, match state (Table 5). By wagering on the stronger q in the first five overs of innings 2, without any dismissals, ROI was 31% suggesting that the markets become more efficient as wickets and overs progress, that is, as the match outcome becomes more predictable.

Favourite	y
Overlay	+

Interval	AvePick	Wagers	SumProfit	AveProfit
1.40-1.50	80.00%	5	$176.00	$35.20
1.50-1.60	100.00%	6	$432.00	$72.00
1.60-1.70	100.00%	4	$391.00	$97.75
1.70-1.80	0.00%	1	-$100.00	-$100.00
1.80-1.90	100.00%	1	$138.00	$138.00
Grand Total	**88.24%**	**17**	**$1,037.00**	**$61.00**

Table 4. Third stratum wagering profit - selected intervals

Favourite	y

State	AvePick	Wagers	SumProfit	AveProfit
01_0	75.00%	7	$208.00	$29.71
02_0	75.00%	8	$50.00	$6.25
03_0	77.78%	7	$300.00	$42.86
04_0	63.64%	4	$159.00	$39.75
05_0	72.73%	4	$222.00	$55.50
Grand Total	**73.68%**	**30**	**$939.00**	**$31.30**

Table 5. Second stratum wagering profit - match state intervals

4. DISCUSSION

ODI home ground advantage was not investigated in this stage of the research but remains a critical factor, particularly for sides like India where the pitches are, arguably, the most unique in the world. Along with team strength, included as a post-hoc consideration in this research, such a factor might be best addressed as adjustments to the generated probabilities, rather than to the betting strategy. Team strength was deliberately kept from the *R* samples at this stage to preserve sample size, however, it is anticipated that augmenting the Weibull probabilities in Equation (8) would be an effective solution, rather than further segmenting the samples. The research almost certainly stands to benefit from a mathematical approach to wagering, rather the categorical one outlined in this paper. A Kelly system, for example, where optimum wager amounts are determined by a mathematical system, would be a suitable starting point. Timelines prevented this from being feasible but is now a high priority. Extending the methodology outlined in this paper to the 20-over cricket game would be worthwhile because of volatile in-play markets.

5. CONCLUSION

The Weibull distribution is an interesting application to in-play quantitative analysis as it is concerned with a failure rate, that is, the chance of winning a match, over some function of time, that is, the match length. In-play likelihoods of victory, drawn from a Weibull probability density function were not only a good predictor of the team batting second chasing down their target in ODI matches, but a fine indicator of when to lay a bet on that team doing so. Selective betting moments in the match, defined by elapsed overs and number of dismissals, the pre-match favourite and a positive overlay, produced a return on investment in excess of 20% which is an excellent result. The only barrier a punter in Australia may face, armed with such a tool, is the inability to place a bet on the internet. The beauty of cricket is that time between overs should be sufficient to make a telephone call and lay the bet.

References

Bailey, M.J. & Clarke, S.R. (2004). Market inefficiencies in player head-to-head betting on the 2003 cricket world cup. Economics, Management and Optimization in Sports (P.Pardalos & S. Butenko eds.) pp. 185-201, New York: Springer.

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

Clarke, S.R. & Allsopp, P. (2001). Fair measures of performance: the World Cup of cricket. Journal of the Operational Research Society, 52, pp. 471-479.

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

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

Easton, S. & Uylangco, K. (2007). An examination of in-play sports betting using one-day cricket matches. Journal of Prediction Markets, Vol. 1, Iss. 2, pp. 93-109.

Elderton, W.P. (1945). Cricket scores and some skew correlation distribution. Journal of the Royal Statistical Society, Series A. 108, pp. 1-11.

Hamilton, H. (2011). An Extension of the Pythagorean Expectation for Association Football. Journal of Quantitative Analysis in Sports. 7, 2, pp.1-18.

Lewis, A.J. (2005). Towards fairer measures of player performance in one-day cricket. Journal of the Operational Research Society, 56, 7, pp. 804-815.

Sargent, J. and Bedford, A. (2012). Using Conditional Estimates to Simulate In-Play Outcomes in Limited Overs Cricket. Journal of Quantitative Analysis in Sports. 8, 2, pp. 1-20.

Smith, P. (1993). Into Statistics. Thomas Nelson Australia.

Swartz, T.B., Gill, P.G., Beaudoin, D. & deSilva, B.M. (2006). Optimal batting orders in one-day cricket. Computers & Operations Research, 33, pp. 1939-1950.

Thomas, A.C. (2007). Inter-arrival Times of Goals in Ice Hockey. Journal of Quantitative Analysis in Sports, 3, 3, pp. x-y.

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

THE DEVELOPMENT OF A PERFORMANCE BASED RATING SYSTEM FOR LIMITED OVERS CRICKET

Paul J. Bracewell[a], Marcus C. F. Downs[a] and Jarred W. Sewell[a]

[a]*CricHQ, Wellington, New Zealand*

Abstract

Methods for rating teams in cricket are hampered by the way match results are recorded. If the team that batted first wins, then the margin of victory is expressed in terms of the difference in runs between the team batting first and the team batting second. However, if the team batting second wins, the margin of victory is then expressed as the number of remaining wickets for the team batting second. As there is no meaningful mapping function between these two forms of margin of victory, team rating systems in cricket default to mechanisms based on win\loss records over a defined time frame. This paper outlines a method for creating performance based team ratings for cricket for application with limited overs cricket, utilising a margin of victory that is solely runs based.

The challenge lies in the implementation of a method for calculating the margin of victory for the team batting second. This is resolved by estimating the number of runs that would have been scored had the team batting second continued until their resources were exhausted. In this instance, resources refer to the number of wicket and balls that are available. The underlying approach is similar to the Duckworth-Lewis method (Duckworth & Lewis, 1998) for resetting the target in rain interrupted matches. The consequence is a more meaningful way of comparing results which is useful for coaching and development purposes.

To create a meaningful rating, the algorithm outlined by Bracewell *et al.* (2009) is implemented. This method has been shown to produce robust ratings based on the relative performances of the competing teams across a wide range of different types of team sports. The resultant ratings are validated using an existing rating system.

Keywords: cricket, team ratings, Duckworth-Lewis

1. INTRODUCTION

There is a large amount of interest surrounding the statistics of sports, evidenced by the adaptation of Michael Lewis' (2004) book, Moneyball, into a movie (Miller, 2011), describing the use of statistics in baseball. This interest comes from different people for different reasons. Teams and coaches critically analyse the performance of individuals, combinations and teams to derive insights for enhancing performance and creating strategies; for this they need as much information as possible. Fans seek to know how their team, or favourite player, is performing relative to other teams or players. In both these examples, it is necessary that suitable statistics are simple to understand and interpret, but embody as much important information as possible to satisfy their desires.

Cricket is a data rich sport characterised by distinct events between a batsman and a bowler, which are all recorded as a minimum standard as defined by the laws of the game. However, the way that the margin of victory is defined complicates the ability to create meaningful team ratings. Considering the limited overs example only, where are set number of overs are specified and each team is permitted one innings, the winner is the team that has the most runs at the completion of both teams turn batting. Crucially, the innings is completed either when the allotted overs have been bowled, the team is dismissed, or, for the team batting second, the target total has been reached. If the team that batted first wins, then the margin of victory is expressed in terms of the difference in runs between the team batting first and the team batting second. However, if the team batting second wins, the margin of victory is then expressed as the number of remaining wickets for the team batting second. As there is no meaningful mapping function between these two forms of margin of victory, team rating systems in cricket tend to default to mechanisms which tend to use competition points, win percentage or net run rates over a defined time frame.

The aim of this paper is to create a meaningful statistic for cricket that quantifies a team's performance using a method that is more informative than the current methods for assessment. This aim is achieved by the introduction of a method for creating performance based cricket team ratings for limited overs cricket, utilising a framework that provides a margin of victory that is solely runs based. The development of a consistent method for quantifying the magnitude of victory with a standard measure enables standard rating methods to assess the relative performance of teams competing within a closed competition.

This approach provides greater depth of detail in team assessment than winner-takes-all type mechanisms, such as competition points or win percentage, which are fundamentally binary, or at best, ordinal (ICC, 2014). These approaches ignore close losses. Net run rate is an accepted method for ranking teams within a competition, as it is used for breaking ties in limited overs tournaments (ESPN, 2014). The net run rate for a team is calculated by subtracting the average runs per over scored against that team throughout the competition from the average runs per over scored by that team throughout the competition. However, net run rate does not take into account the number of wickets that have fallen when a chasing team wins a match. The impact of this is most clearly seen in low scoring games. If for example the team batting first is bowled out for 100 in 25 overs in a 50-over game. Then the team chasing wins by scoring 101 in 20 overs, their net run rate will be 101/20-100/50=3.05. A net run rate that large would indicate a very large win, but if the winning team was nine wickets down when they reached the target you would consider the win as anything but convincing.

Through the development of a margin of victory metric which is consistent, irrespective of whether the team batting first won or lost, enables a meaningful rating to be created which encompasses relative team performance. Importantly, this enables interested parties to review a single statistic to understand how well a team has performed compared to their competitors. In addition, the statistic is an indication of the team's "form"; hence when two teams play each other it provides an indication of not only the result (winner/loser), but also how evenly matched the contest will be.

2. TEAM RATING SCORE COMPONENTS

The challenge for creating a suitable team rating for cricket lies in the implementation of a method for calculating the margin of victory for when the team batting second wins. This is resolved by estimating the number of runs that would have been scored had the team batting second continued until their resources were exhausted. That is, had a team chasing 200 runs to win in 50 overs reached that total in just 42 overs, what would they have scored had they batted out their allotted overs?

2.1 Methods for forecasting cricket scores

The underlying philosophy we adopt is the same as that used by Duckworth and Lewis (1998), where the basis of their method is recognition that the batting team has two resources at its disposal from which to make its total score; it has overs to face and it has wickets in hand. Duckworth and Lewis acknowledge the influence of Clarke (1988) in highlighting potential approaches for setting a fair target in rain interrupted one-day matches. An alternative approach for resetting targets, deployed in the now defunct Indian Cricket League, was derived by Jayadevan (2002). Importantly, both methods determine what proportion of a team's runs it is expected to have scored, based upon the number of overs faced and the number of wickets lost. This approach enables the method to be deployed into different match conditions as there is an inherent adaptation to the run rate in the match.

However, these methods are used to adjust rain affected targets. In order to generate a margin of victory, the method needs to extrapolate a second innings score, rather than interpolate.

2.2 Rating Methods

Daud and Muhammad (2013) use adaptations of two algorithms: PageRank, designed to rank websites in search engine results, and h-index, designed to measure the impact of a scientist's cumulative contributions, for ranking teams. The intent was to give more weight to a team defeating stronger teams by considering the number of runs and wickets. This approach derives a rating by considering a number of recently played matches and assesses the total number of wickets taken and runs scored in those matches. This approach avoids the issue of defining a margin of victory, but does not consider the runs and wickets scored in each match independently.

There are numerous publications describing the development and improvement of sport ratings systems. Stefani (2011) provides a detailed review of methods for officially recognised international sports rating systems and is an excellent resource for evaluating the strengths and weaknesses of various systems.

Bracewell, Forbes, Jowett, and Kitson (2009) introduced a method for quantifying the relative performances of teams, which used score ratios rather than scores (or differences), which simplified the calculation of the Team Lodeings, L, for the Tth team to:

$$L_T = \frac{2\sum_{f=1}^{h_T} p_{T,f} \sum_{g=1}^{a_T} q_{T,g}}{h_T a_T}, \qquad (1)$$

where h_T and a_T are the respective number of home and away games played by the *Tth* team, $p_{t,f}$ is the ratio of victory for the *Tth* Team in the *fth* match at home and $q_{t,f}$ is one minus the ratio of victory for the *Tth* Team in the *gth* away match. The ratio of victory is calculated for each match as the normalized points scored by the home team divided by the sum of the normalized points scored by the home and away teams. Team ratings are calculated for a specified time frame, t, (typically either weeks or rounds, where t ≥ hT + aT) enabling team performance to be rated and changes in performance to be quantified. This is useful for match prediction. Higher team ratings are associated with better performed teams. The ratings are bound by 0 and 1. Importantly, this algorithm is suitable for use within cricket as the use of score ratios enables the impact of extraneous factors to be limited (boundary size, pace of wicket and state of outfield). The score ratio method is preferred as it provides a fairer assessment of the performance of both teams in the result. Consider rugby-type results 13:3 and 40:30. Both have a difference in scores of 10. However, in the second case it appears the game was much more even than in the first instance. This is reflected in the score ratios which are 0.81 and 0.57, respectively. In addition, this approach was shown to be robust across many different sports, which is a useful property when considering different cricket match formats (T20, 40 over & 50 over). Finally, the emphasis of this paper is the consistent definition of the margin of victory in cricket, meaning that the ratings method is a secondary consideration.

Whilst we have chosen to use the ratings algorithm described above in this paper, an area for future research is the assessment of other algorithms.

3. CONSTRUCTING A PERFORMANCE BASED RATING SYSTEM FOR LIMITED OVERS CRICKET

There are three key components in the development on a performance based rating system for limited overs cricket. The first component is the data. The second component, and the most important contribution of this paper, is the extrapolation a chasing team's total after the total has been reached so that a margin of victory in terms of runs can be extracted. The second stage is the application of a ratings algorithm to summarise the relative performance of teams.

3.1 Processing Data

Data was extracted from CricHQ's source systems (www.crichq.com). CricHQ is a cricket technology industry pioneer with headquarters in Wellington, New Zealand. CricHQ's scoring and competition administration software collects cricket data from all around the world over numerous levels of competition. The data used for this project included final score data from five premier T20 competitions around the world (included Indian Premier League, HRV Cup and more), over the past one to five years, depending on the length of history of that competition). We also used data from One Day International matches between the top ten ranked sides dating back to 2000. For simplicity, we assess matches that were not affected by rain. Whilst the Duckworth-Lewis method can be used to adjust the totals in rain affected matches, we have omitted those games from our initial analyses. Subsequent applications use an adjustment based on the Duckworth-Lewis method.

3.2 Extrapolating a Chasing Team's Total

In the previous sections it was stated that the main obstacle preventing cricket from being processed using typical team rating algorithms, is due to the margin of victory problem. When team one, batting first, wins a game of cricket the result will be stated as ‘team one wins by *x* runs’. However, when team two, batting second, wins the result will be stated as ‘team two wins by *y* wickets’. The reason for this is that team two stops batting as soon as their score is greater than team ones score (the target score). Consequently, we cannot report how many runs team two won by because they may not have used up all their resources (balls or wickets), and hence could have scored more than they actually did.

To resolve this issue, we seek to produce a projection of the score team two would have got too had they not stopped batting as a consequence of winning the match. The intent is that after this calculation, we will have a margin of victory for every game in terms of runs. The process employed to produce the projections is based on a proprietary algorithm created for CricHQ which includes a generic score and probability of winning projection model. The output is available via the CricHQ platform (accessed via app stores for a variety of technology platforms). The model considers the resources available to team two at the completion of the game. In this instance, resources refer to the number of wicket and balls that are available similar to the Duckworth-Lewis method (Duckworth & Lewis, 1998) for resetting the target in rain interrupted matches. If either of team two's resources have been exhausted at the completion of the game then the projection will simply be team two's actual score. However, if team two still has resources remaining, but the game is finished, then a score projection is calculated. The difference between team two's forecasted total and their actual total is positively related to the amount of resources still available to the team at the completion of the game. This encompasses the wickets and balls resources, along with a value derived from the relative team totals, to produce a metric that represents the proportion of total resources used by team two at the completion of the game. Again, this approach is similar to that used by Jayadevan (2002), Duckworth and Lewis (1998). We then divide team two's actual score (C_2) by this proportion (R_2) to get a projection (T_2).

$$T_2 = \frac{C_2}{R_2} \qquad (2)$$

This result leaves us with a margin of victory for every match regardless of which team won. The model is a generic model, and hence allows us to produce projections and win margins for multiple forms of cricket. However, the focus of this paper is limited overs cricket (T20 and 50 Over matches). From a practical perspective, this is useful for CricHQ as it enables the deployment of ratings for teams playing in different age-groups, divisions and competitions across different formats.

3.3 Score Transformations

The projections of the previous section provide us with a win margin in terms of runs for all games irrespective of whether the team batting second won or lost. These are necessary because in creating the team rating we will be using the score ratio which is defined as the final total of the team batting first (T_1) divided by the total number of runs scored in the match (T_1 + T_2); where T_2 is the final adjusted total of the team batting second. The final adjusted total for the chasing team is simply the final total when the chasing team loses or wins from the last ball on the innings.

However, the raw scores and projections discussed in the previous section are not suitable for immediate using in a rating algorithm. The nature of cricket is such that the scores of both teams tend to be large numbers (compared with sports like soccer, hockey and rugby union), typically between one and two hundred for T20 cricket. This is not ideal because when the score ratios are calculated based on the raw scores there is not a good spread of the resultant output between zero and one, because the margin of victory is relatively small compared to the team's totals. Instead our ratios will be heavily concentrated around 0.5, typically between 0.4 and 0.6. In addition, the raw team totals are also far from being a normally distributed variable. We perform two transformations: to first produce a normally distributed variable, and then change the scale of this variable so that it has a mean and standard deviation that will result in ratios that cover a good range between zero and one.

A log transformation was used to turn the raw scores distribution, less a constant value, to an approximately normally distribution. The purpose for this is to reduce the impact of outlying scores and mitigate the impact when chasing teams are dismissed cheaply under ideal batting conditions. The constant used was determined heuristically and chosen to minimise the kurtosis of the distribution.

3.4 Assessing the Validity of Extrapolation

Before we are able to proceed with deriving the score ratios, the validity of the score extrapolation for when the chasing team wins needs to be assess. To assess this we compare the margins of victory in games where the team batting first wins compared to when the team batting second wins. Ideally, the winning margin of victory will be distributed similarly from the perspective of batting first or second. As the margin of victory when team two wins without using all their resource is based on an extrapolation we need to ensure that the projections have not produced margins of victory that are significantly different from those that are produced when the team batting first wins. This is to ensure the rating does not tend to unfairly favour either side. The extrapolated total is based on remaining resources. However, one of the strongest indications of the scoring potential is the number of balls remaining. Consequently, it is expected that by plotting the derived margin of victory against the number of balls remaining should display a symmetric function, with a turning point near zero, to be a suitable fit.

Figure 1. below shows the transformed margin of victory for T20 games (defined as the transformed

team setting score minus the transformed team two chasing score) on the horizontal axis with the number of balls remaining in the second innings shown on the vertical axis. This means the negative x values correspond to when the team batting first (setting) lost to the team batting second (chasing).

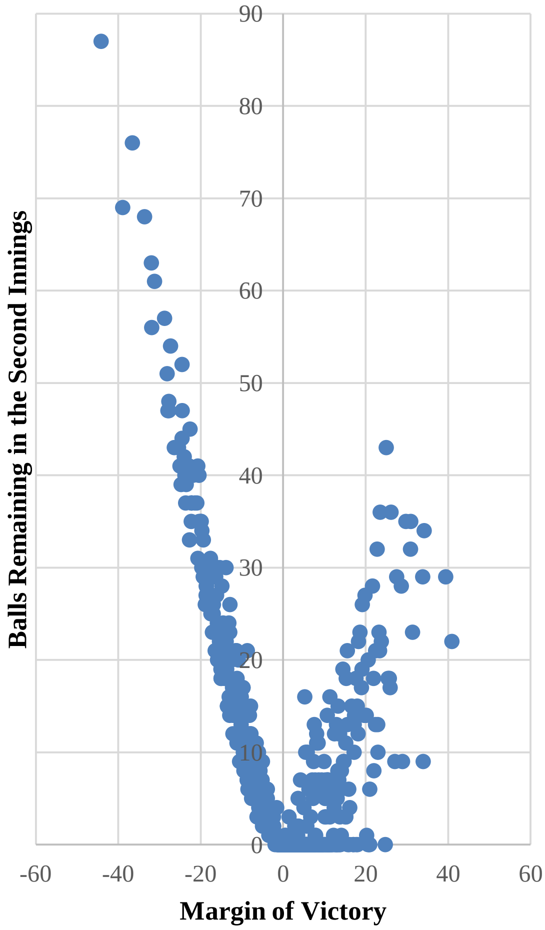


Figure 1: Transformed margin of victory for T20 matches versus balls remaining in the second innings.

There are several key outcomes from the above graph. It can be seen that that the margins of victory do not appear to be drastically different, with the majority of observations being concentrated between plus and minus thirty units and appearing to exhibit symmetry. These key features can be confirmed statistically. A quadratic function adequately explains the relationship between transformed margin of victory and balls remaining (r-sq = 0.8694), highlighting the symmetry. The resultant function:

$$\hat{B} = 0.0436M^2 - 0.4999M + 2.1931 \qquad (3)$$

where $\hat{B}$ is the estimated balls left and M is the transformed margin, indicates that the obtained margin of victory is indeed symmetrical and approximately near 0. To confirm that the distributions of the margin of victory below and above zero are relatively similar an F-test for sample variances was performed. The variance of the transformed margin of victory when the team bats first win (52.04, n=454) is not statistically significantly different (p=0.18) to the variance of the transformed margin of victory when the team batting second wins (56.73, n=426).

Consequently, the inability to reject the null hypothesis, that the variance of the transformed margin of victory when team one wins is the same as the variance of the transformed margin of victory when team two wins, confirms that the extrapolation of a winning chasing team total has not introduced any significant bias.

In addition, a t-test for sample means assuming equal variances was performed to test if the margins of victory have a similar mean irrespective of whether the team setting or chasing wins. For this test we use the absolute value of the margin of victory if team two wins so all margins are positive. The mean of the transformed margin of victory when the team bats first win (8.98, n=454) is statistically significantly different (p=0.04) to the mean of the transformed margin of victory when the team batting second wins (7.94, n=426). However, practically, the difference is only 1 unit, which equates to just 2.8 runs.

Based on these results, there is sufficient statistical evidence to conclude that the method for projecting totals when the chasing team wins has not produced margins of victory that are unreasonable when the team batting second wins.

These tests have been focused on T20 data that has been projected and transformed. However, our projection model is produced as a generic projection model, therefore we would expect that the same interpretation would apply to 50 over cricket.

As with the T20 matches, a quadratic function adequately explains the relationship between transformed margin of victory and balls remaining (r-sq = 0.6873), highlighting the symmetry in 50 over matches. The resultant function:

$$\hat{B} = 0.1149M^2 - 0.1407M + 12.2810 \qquad (4)$$

where $\hat{B}$ is the estimated balls left and M is the transformed margin, indicates that the obtained margin of victory is indeed symmetrical and approximately centred around 0. To confirm that the distributions of the margin of victory for 50 over cricket below and above zero are relatively similar, an F-test for sample variances was performed. The variance of the transformed margin of victory when the that team bats first wins (56.57, n=1136) is not statistically significantly different (p=0.12) to the variance of the transformed margin of victory when the team batting second wins (60.60, n=1217).

Consequently, the inability to reject the null hypothesis, that the variance of the transformed margin of victory when team one wins is the same as the variance of the transformed margin of victory when team two wins, confirms that the extrapolation of a winning chasing team total has not introduced any significant bias.

In addition, a t-test for sample means assuming equal variances was performed to test if the margins of victory have a similar mean irrespective of whether the team setting or chasing wins. For this test we use the absolute value of the transformed margin of victory if team two wins so all margins are positive. The mean of the transformed margin of victory when the team that bats first won (8.93, n=1136) is not statistically significantly different (p=0.89) to the mean of the transformed margin of victory when the team batting second wins (8.89, n=1217).

Based on these results focused on 50 over cricket, there is sufficient statistical insight to conclude that the method for projecting totals when the chasing team wins has not produced margins of victory that are unreasonable when the team batting second wins.

The results from this validation of the margin of victory on two different limited over match formats are critical. The method for extrapolating the total for the team batting second if their innings had continued until all resources were consumed generates a margin of victory in runs that is equivalent to the margin of victory obtained when the team batting first wins.

This sub-section has shown that we are in the position to apply a margin of victory in different situations, and hence create cricket team ratings for a number of different competitions using runs (observed and adjusted) as a performance metric.

3.5 Quantifying Relative Team Performance

The process for deriving the ratings is simple. Firstly, the framework for obtaining a margin of victory in terms of runs is deployed, enabling meaningful score ratios to be obtained. When the team batting first wins, or the team batting second wins on the last ball of the innings, the margin of victory is simply the difference between the total runs scored by each team (setting team total minus the chasing team total). If the team batting second wins with one or more balls remaining in the innings, then an adjusted final total is required. The projections are based on the resources, wickets and overs, that the chasing team has left at the completion of their innings. The data to derive these resource variables are readily available from scorecards which are readily available from several online data sources, for example: www.crichq.com, www.espncricinfo.com & www.cricketarchive.com. This information is used to calculate the proportion of resource consumed at the time of victory using a proprietary algorithm. The adjusted final total is then the observed final total divided by the resources consumed, which is our projection of what the chasing team would have scored had victory not ended the match.

The process described by Bracewell *et al.* (2009) is used to quantify the relative performances between competing teams. First, the totals are transformed to rugby-type scores. This is achieved by subtracting a constant, then applying a natural log transformation with linear scaling applied. This process is data driven to cater for different competitions and formats. Then the score ratio for each match is obtained, calculated as the final transformed total of the team batting first (T_1) divided by the sum of transformed runs scored in the match ($T_1 + T_2$); where T_2 is the transformed final adjusted total of the team batting second.

These ratios are then input into the rating algorithm. For domestic cricket, previous results from approximately one year and one month are considered (380 days). For international cricket where matches are more sparse, almost twice this range is examined (approximately 2 years and 2 months or 800 days).

In a departure from the rating algorithm described above, the raw ratings obtained were regressed against the winning percentages in order to derive a linear transformation that would increase the spread of the ratings between 0 and 1. The other consequence of this step was to create a rating that had a more natural interpretation, in that a team with a transformed rating of 0.7 tended to have a winning percentage of 70%. An index type measure was created by multiplying the transformed rating by 1000.

4. VALIDATING RATING PERFORMANCE

To validate the performance based rating system for limited overs cricket the team ratings are compared to the ICC ratings, produced by cricket's governing body.

4.1 Comparison with ICC ODI Ratings

Cricket's governing body, the International Cricket Council (ICC) regularly updates and publishes ratings for the three major forms of international cricket (T20, ODI and Test). A detailed explanation of the ratings can be found at several sources (e.g. ICC, 2014; Daud, 2013). In the graph below we compare the month end ICC ODI Ratings from January 2000 to March 2014 for the top 10 teams plotted against our corresponding team rating (Australia, England, New Zealand, India, Sri Lanka, West Indies, Pakistan, Bangladesh, Zimbabwe and South Africa). We can see clearly that the two ratings are correlated (r=0.91), indicating our approach is valid. As the ICC ratings are credible, it is important that our rating is highly correlated. We would not expect the relationship to be perfect because of the difference in methods, but the general movements of teams rating over time should be similar.

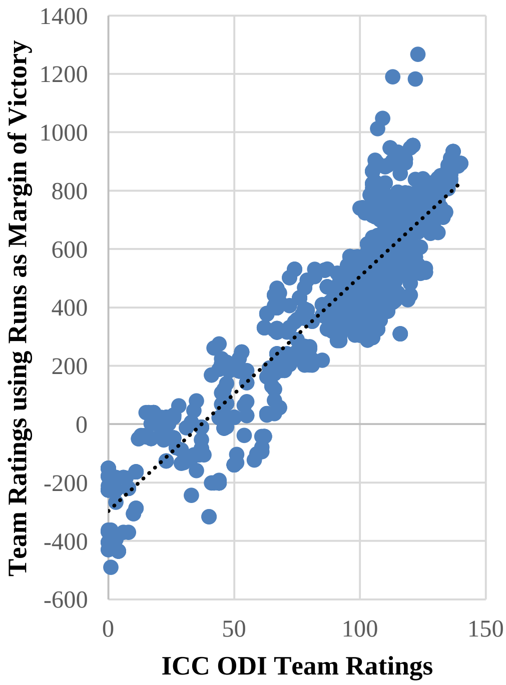


Figure 2: Scatter Plot of ICC ODI team ratings versus performance based team ratings.

The evolution of ratings of time is shown in the graph below. Figure 3 shows the ICC ratings over time for Australia as well as the team performance rating, which has been arbitrarily scaled for display purposes.

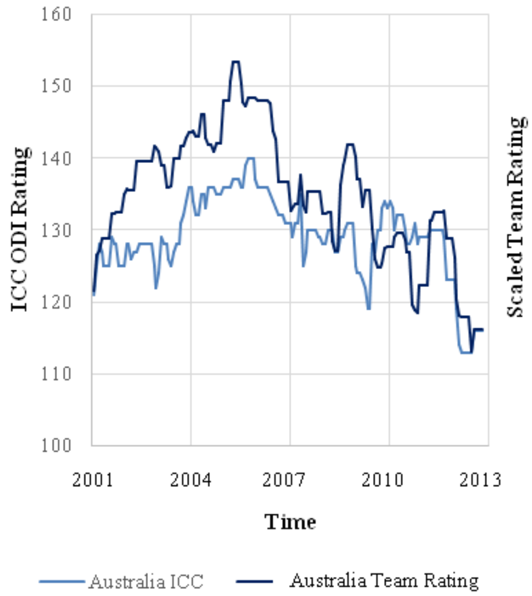


Figure 3: Line plot of ICC ODI team ratings overlaid with performance based team ratings for January 2000 to March 2014.

In the line plot above the two lines which represent the Australian cricket team's rating over the last decade appears to follow the same general slow upward or downward movement. The team performance rating is more variable, which is likely to be a result of the different methods of calculation for our rating compared to the ICC rating. The ICC rating also only drops games once a year (since 2012 start of every May, previously August), where they drop the oldest year worth of data and then begin adding all the new games as they happen. Consequently, for most of the year the monthly ICC rating will be based on a larger number of games than the previous month and therefore, it will be less variable. The team performance ratings are based on a set number of days. For comparison purposes, this means at the end of each month the oldest data month of data is dropped, and the most recent month of data added.

5. CONCLUSION

This paper has outlined a method for creating performance based team ratings by adapting approaches that have previously been explored in the sport statistics literature. The two key components for this were: a method for forecasting runs, and an algorithm for assigning ratings to team performance. The major contribution of this paper is the evaluation of victory in a cricket match as runs, irrespective if the team batted first or second.

A proprietary algorithm was used to forecast what a winning chasing team may have scored if they had continued to bat on once victory had been achieved; However, the approach used is philosophically the same as Duckworth and Lewis (1998) and Jayadevan (2002) where forecasted score totals are based on batting resources, wickets and overs, remaining. The distribution of margin of victory results for victorious chasing teams was not statistically significantly different from the corresponding distribution when teams batting first won. This indicated that no systematic bias, based on batting first or second, was introduced to our evaluation of the magnitude of victory.

A ratings algorithm that used score ratios (Bracewell *et al*., 2009) was then used to calculate ratings for defined time frames. An area for future research is investigating the applicability of different types of rating methods.

The performance of the ratings was validated by comparing with the ratings produced by cricket's governing body, the ICC. A correlation of 0.91 indicated that the team ratings created by the proposed performance based rating system for limited overs cricket is valid. Importantly, a range of limited over formats are covered by this approach. Additionally, the high correlation between the ICC ratings and the performance based team ratings indicate that the extrapolation of runs in the second innings provides meaningful results and that the rating algorithm used is suitable.

The consequence is a more meaningful way of comparing and tracking results which is useful for coaching and development purposes which can extend to assess player impacts on results.

References

Bracewell, P.J., Forbes, D.G.R., Jowett, C.A. and Kitson, H.I.J. (2009). Determining the evenness of domestic sporting competition using a generic rating engine. Journal of Quantitative Analysis in Sports **5** (1).

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

Daud, A. and Muhammad, F. (2013). Ranking cricket teams through runs and wickets. Active Media Technology 9th International Conference Proceedings (T. Yoshida, G. Kou, A. Skowron, J. Cao, H. Hacid & N. Zhong eds.) pp. 156-165. Switzerland: Springer International Publishing.

Duckworth, F.C. and Lewis, A.J. (1998). A fair method for resetting the target in interrupted one-day cricket matches. Journal of the Operational Research Society **49** (3): pp. 220-227.

ESPN (2014). Net Run Rate. Retrieved 20 June, 2014, from www.espncricinfo.com/ci/content/page/429305.html.

ICC (2014). Reliance ICC one-day international team rankings -frequently asked questions. Retrieved 20 June, 2014, from http://icc-live.s3.amazonaws.com/cms/media/about_docs/536b1a48c16e5-Reliance%20ICC%20ODI%20Team%20Rankings%20FAQs%202014.pdf.

Lewis, M. (2004). Moneyball. New York: W.W. Norton & Comapny, Inc.

Miller B. (Director) (2011). Moneyball. [Motion Picture]. United States: Columbia Pictures.

Jayadevan, V.A. (2002). New Method for the Computation of Target Scores in Interrupted, Limited-Over. Cricket Matches. Current Science **83** (5): 577–586.

Stefani, R. (2011). The methodology of officially recognized international sports rating systems. Journal of Quantitative Analysis in Sports **7** (4).

Acknowledgment

The authors appreciate the efforts of Dr. Steven Bowe of Swinburne University of Technology, Melbourne, in developing this paper.

MATHEMATICAL MODELS THAT PREDICT ATHLETIC PERFORMANCE IN THE MEN'S THROWING EVENTS AT THE OLYMPIC GAMES

Ian Tim Heazlewood[a] and Joe Walsh[a]

[a] *Charles Darwin University*
Corresponding author: Ian.Heazlewood@cdu.edu.au

Abstract

The prediction of future athletic performance is a recurring theme as sports scientists strive to understand the predicted limits of sports performance. Predictive models based on Olympic data for athletics have derived some accurate predictions of performance in the 2000, 2004, 2008 and 2012 Olympic Games. The aim of this research was to develop predictive models using performance data of the first three athletes competing in the finals of the men's shot put, discus, hammer and javelin at the Summer Olympic Games from Berlin 1936 to London 2012. The approach utilised regression-curve estimation using IBM SPSS Statistics Version 22 statistical software and by evaluating fit to linear, logarithmic, inverse, quadratic, cubic, compound, power, sigmoidal, growth exponential and logistic functions. The mathematical models varied represented very good predictors of past, current future throws performance in the four field events based on R^2 (0.850 - 0.972), p-values (<.001) and unstandardized residuals or error. The non-linear function of best fit for events was the cubic function, which indicated a decrease in performance in recent Olympics and predicted this performance decline would occur at the 2016 Olympic Games in Rio de Janeiro. The reasons for the current and predicted declines were more vigilance concerning drugs in sport and therefore dampening the enhanced performance effect of anabolic androgenic hormones, fewer athletes are undertaking the throwing events as a completive sport and changes in the source population providing the sample of potential throws athletes in Australia in terms of motor fitness abilities are getting smaller in terms of motor fitness abilities and thus fewer capable athletes exist to select from within source population. The good predictive models may be due to a longer timeframe data set to develop substantive predictive models, a timeframe able to detect phylogenetic trends in human athletic performance. The predictions may indicate a slightly modified Olympic motto from citius, altius, fortius to citius, altius and infirmius or "faster, higher and weaker?"

Keywords: Olympic Games, Throwing Event, curve estimation, nonlinear regression, predictive mathematical modelling

1. INTRODUCTION

The prediction of future athletic performance by athletes at the Olympic Games is a recurring theme as well as forming the basis for some stimulating discussions on the limits of human performance. Mathematics in sport and exercise and sports science are based on the principles of description and more importantly prediction. The ability to make substantive and accurate predictions of future elite level sports performance indicates that such approaches reflect substantive sport science. Often these predictions are purely speculative and are not based upon any substantial evidence, rather they are based on the belief that records are made to be broken and that performances based on past experiences must continue to improve over time. The accessibility of data in the form of results from Olympic Games, world records and world best performances in a specific year allows the analysis of performances in any number of events. From these analyses, changes in performance over time can be observed and predictions of future performance can be made utilising the process of mathematical extrapolation and interpolation.

A number of researchers have attempted to predict future performances by deriving and applying a number of mathematical statistical models based on past performances in athletics. Prendergast (1990) applied the average speeds of world record times to determine a mathematical model for world records. The records or data used in the analysis spanned a 10 year period. Following his analysis, Prendergast (1990) raised the question of whether any further improvements can be expected or if the limits of human performance have been reached. The sports of athletics (Heazlewood and Lackey, 1996; Heazlewood, 2011, 2013a, 2013b) and swimming (Lackey and Heazlewood, 1998) have been addressed in this manner. The knowledge of future levels of sporting performance has been identified by Banister and Calvert (1980) as beneficial in the areas of talent identification, both long and short term goal setting, and training program development based on the next level of expected future performance. In addition, expected levels of future performance are often used in the selection of national representative teams where performance criteria are explicitly stated in terms of athletics times and distances for example as required entry standards at Olympic Games (International Olympic Committee (IOC), 2014).

Péronnet and Thibault (1989) postulate that some performances, such as the men's 100m sprint is limited to the low 9 seconds, whereas, Seiler (referred to by Hopkins, 2000) envisages no limits on improvements based on data reflecting progression of records over the last 50 years. According to Seiler improvements per decade have been approximately 1% for sprinting, 1.5% for distance running, 2-3% for jumping, 5% for pole vault, 5% for swimming and 10% for skiing for male athletes, whereas female sprint times may have already peaked. The differences for males and females it is thought to reflect the impact of successful drugs in sport testing on females. Previous derived curve estimations that significantly fit the data have also displayed interesting findings as no one curve fits all the data sets. Different events displayed different curves or mathematical functions (Lackey & Heazlewood, 1998) of best fit. In swimming the men's 50m freestyle was inverse, 100m freestyle compound, 200m sigmoidal, and the 400m and 1500m freestyle cubic.

In athletics for the men's events the mathematical functions (Heazlewood and Lackey, 1996) were 100m inverse, 400m sigmoidal, long jump cubic and the high jump displayed four functions (compound, logistic, exponential and growth). The curves that fit the data have also displayed interesting findings as no one curve fits all the data sets. This may indicate that different events are dependent upon different factors that are being trained differently or factors underpinning performance evolving in slightly different ways. This has resulted in different curves or mathematical functions that reflect these improvements in training or phylogenetic changes over time. However, at some point in the future when time catches-up with the actual performance, then how accurately the predictive models reflect reality can be assessed.

However, the ability to predict performances at the 2000, 2004, 2008 and 2012 Olympic Games for the men's 100m, 400m, long jump and high jump, based on the 1924 to 2012 data, was very accurate with low percentage error. The International Olympic Committee (IOC, 2014) has produced descriptive data and descriptive graphical analysis to indicate trends in World records for the men's throwing events however not based on mathematical predictive modelling.

The dominant and important research question is, can mathematical models based on nonlinear curve estimation, which have proven to be very successful in fitting and predicting past, current and future Olympic performances for event finals in athletics and swimming for men display equal effectiveness in predicting athletic performances in the men's throwing events of the shot put, discus, hammer and javelin at the Olympic Games, which are past, current and future performances?

2. METHODS

The mean score for the first three placed finalist in the men's throwing events which consist of the shot put, discus, hammer and javelin. The data were selected from 1936 to 2012 Summer Olympic Games and provided by the International Olympic Committee (IOC, 2014) served as the exemplars for performance in each event for each competitive year. The data covered eighteen Summer Olympics. These scores served as the data set to derive and test predictive models based on curve estimations and the distances in each event was recorded metres to the nearest centimetre.

This is a similar method used previously (Heazlewood & Lackey, 1998; Heazlewood, 2006, 2008, 2013a, 2013b) to curve fit Olympic data for swimming and athletic events. According to Garson (2010) curve estimation is an exploratory tool in model building and model selection, where the best mathematical model or function is selected to represent quantitative relationships between an independent/predictor variable and a dependent/response variable. The mathematical solutions and curve estimations were derived using the IBM SPSS Statistics Version 22 statistical software (SPSS Inc. 2014).

The most common curve estimation or model fit approaches are based on the following mathematical functions (Garson, 2010) and these are linear, logarithmic, inverse, quadratic, cubic, power, compound, S-curve, logistic, growth, and exponential models. In terms of statistical approach model fit indices are then applied to test the quality of the model and the general method of determining the appropriate regression models is represented by the following steps.

1. Commence with an initial estimated value for each variable in the equation.
2. Generate the curve defined by the initial values. Calculate the sum-of-squares (the sum of the squares of the vertical distances of the points from the curve).
3. Adjust the variables to make the curve come closer to the data points. There are several algorithms for adjusting the variables. The most commonly used method was derived by Levenberg and Marquardt (often called simply the Marquardt method). Adjust the variables again so that the curve comes even closer to the points. Keep adjusting the variables until the adjustments make virtually no difference in the sum-of-squares.
4. Report the best-fit results and then the precise values you obtain will depend in part on the initial values chosen in step 1 and the stopping criteria of step 5. This means that repeat analyses of the same data will not always give exactly the same results. To investigate the hypotheses of model fit and prediction, the eleven regression models were individually applied to each of the athletic events. The regression equation that produced the best fit for each event, that is, produced the highest coefficient of determination (abbreviated as R^2), was then determined from these eleven equations.
5. The specific criteria to select the regression equation of best were the magnitude of R^2, the significance of the coefficient of determination (R^2) is a measure of accuracy of the model used. A coefficient of determination of 1.00 indicates a perfectly fitting model where the predicted values match the actual values for each independent variable (Garson, 2010; Hair et al., 2006; Norušis, 1993). Where more than one model was able to be selected due to an equal R^2, the simplest model was used under the principle of parsimony, that is, the avoidance of waste and following the simplest explanatory model, as well as the statistical significance of the analysis of variance, the alpha or p-value and size of residuals or error in predictions.
6. Some caution is required to not over interpret a high R^2 as it does not mean that the researcher has chosen the equation that best describes the data. It also does not mean that the fit is unique - other values of the variables may generate a curve that fits just as well.

It should be noted the men's. Shot, discus and hammer have not changed dramatically in specification concerning weight, area and volume making comparisons across Olympic year possible. The men's javelin specification was changed in 1 April 1986; the men's javelin 800 grams was redesigned by the IAAF Technical Committee (International Association of Athletics Federations (IAAF), 2014) where the centre of gravity was moved 4 cm forward and the surface areas in front of, and behind the centre of gravity were reduced and increased, respectively. The effect was to reduce lift and increase the downward pitching moment resulting in bringing the nose down earlier and reducing flight distance by around 10% but causing the javelin head to stick in the ground more consistently.

3. RESULTS

The results as illustrated in table 1 indicates each men's throwing events, best-fit functions, r-square, p-value and equations of best fit. Note all the eleven mathematical functions that were tested for model fit. Specifically, these are linear, logarithmic, inverse, quadratic, cubic, power, compound, growth, S-curve, logistic and exponential models. It can be observed that all nonlinear functions of best fit were for the cubic function across all men's throwing events. This indicated a decline in performance in the more recent Olympic Games. Specifically, from 1988 onwards for the shot put, 2000 for discus, 1988 for hammer and 1976-1980 for javelin. The decline in javelin performance actually commenced prior to the re-specified javelin entering competition at the 1988 Olympic.

Table 1. Men's Throwing Events, Best-fit Functions, R-square, P-value and Equations.

Event and Weight	Function	R-Square	p-value	Cubic Equation of Best Fit: Constant	b1	b2	b3
Shot Put 7.26kg	Cubic	.965	<.001	-3.336	3.137	-.131	.002
Discus 2kg	Cubic	.972	<.001	5.211	6.746	-.236	.003
Hammer 7.26kg	Cubic	.922	<.001	30.876	1.704	.141	-.005
Javelin 800g	Cubic	.850	<.001	-14.604	13.944	-.617	.009

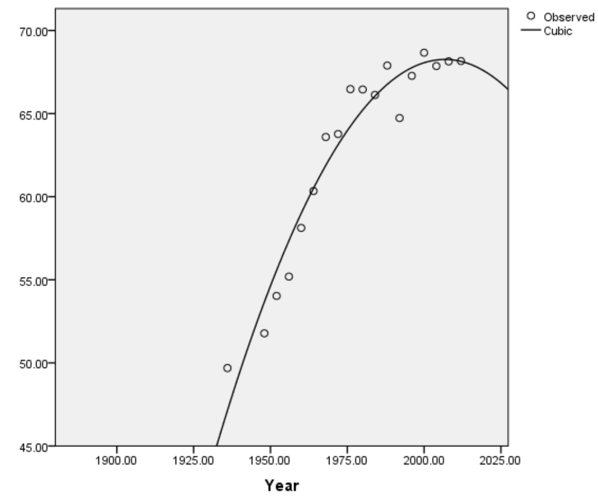


Figure 1. The cubic function line of best fit and actual data point for men's discus. Note the X-axis in years and Y-axis in metres.

To highlight the trends the cubic function line of best fit and actual data points for men's discus from 19936-2012 are displayed in figure 1, which covers eighteen Summer Olympic Games. Table 2 indicates the model fit based on each Olympic year, actual performance, predicted performance and residual error.

Table 2. Men's Discus Data in terms of Olympic year, actual performance, predicted performance and residual error.

Year	Performance (m)	Predicted (m)	Residual Error (m)
1936	49.69	47.14	2.54
1948	51.78	53.66	-1.88
1952	54.03	55.57	-1.54
1956	55.19	57.35	-2.16
1960	58.12	58.99	-.87
1964	60.34	60.50	-.16
1968	63.59	61.88	1.70
1972	63.77	63.12	.64
1976	66.47	64.23	2.23
1980	66.45	65.21	1.23
1984	66.12	66.05	.06
1988	67.89	66.75	1.13
1992	64.73	67.32	-2.59
1996	67.27	67.76	-.49
2000	68.66	68.06	.59
2004	67.86	68.22	-.36
2008	68.14	68.25	-.11
2012	68.16	68.14	.019

4. DISCUSSION

The best fit mathematical functions for men's throwing events were the cubic function with high R-square values (0.850-.972), very significant p-values (<.001) resulting in good model prediction and low residual error. The shot put, discus and hammer where there were no re-specifications for the implements predicted performance decrements for these events and which had started to occur from 1988 onwards for the shot put, 2000 for discus, 1988 for hammer and 1976-1980 for javelin. It is interesting to note the world records for the shot put occurred in 1990 at 23.12m, discus 1986 at 74.08m, hammer in 1986 at 86.74m and javelin in 1996 at 98.48m. The change in specifications of the men's javelin in 1986 would suggest a significant decline in performances however the decline as suggested by this analysis commenced in 1976-1980 and prior the new javelin being introduced?

So what plausible explanations can be theorized for the current trend of declining throwing performance? What will occur in 2016 Rio de Janeiro Summer Olympic Games in Brazil and are we getting citius, altius and fortius or "faster, higher, and stronger?" if the throwing events are based on strength, force and power production the future trends based on table 3 indicates reductions in all men's throwing events and might this indicate citius, altius and infirmius or "faster, higher, and weaker?"

Table 3. The predicted trends for the shot put, discus, hammer and javelin performances for the 2016 Summer Olympic Games.

Event and Weight	Actual/Predicted 2012	Predicted Performance 2016
Shot Put 7.26kg	21.66m/21.29m	21.27m
Discus 2kg	68.16m/68.14m	67.89m
Hammer 7.26kg	79.55m/80.27m	79.51m
Javelin 800g	84.40m/84.97m	83.67m

Three predominant explanations can provide some insights as to these declining trends in the men's throwing events. Specifically,

1. More vigilance concerning drugs in sport therefore dampening the enhanced performance effect of anabolic androgenic hormones. Drug testing by the IAAF is now undertaken in competition testing and out of competition testing and to have a world record ratified by the IAAF the application form has a section indicating the athlete was drug tested at the time of competition and that the athlete has passed the test, that is no adverse findings (IAAF, 2014). The normal suspension/ban for first offence anabolic androgenic hormones is two year. A second offence of this kind results in a life ban. If found positive all performances, monies paid by IAAF, awards and prizes are forfeit so the punitive outcomes can very significant and are thought to act as a strong deterrent to taking prohibited substances on the World Anti-Doping Agency (WADA, 2014) list.
2. Fewer athletes are undertaking the throwing events. In Australia fewer athletes are taking up throwing events as other sports compete for the limited talent pool based on Australia's small population. The Athletics Australia data of permit competitions and ranked athletes performances to be evaluated each Athletic season and indicate that in Australia this is a problem Athletics Australia (2014).
3. The source population providing the sample of potential throws athletes in Australia in terms of motor fitness abilities is getting smaller in terms of motor fitness abilities and thus fewer capable athletes to select from within source population. This appears to be a result of Australia's increasing overweight and obesity epidemic for males 18 years and over, which is currently estimated at 2011-12 to be 70% and females at 56% (ABS, 2014). The age of male high performance throwers is usually between 20-35 years. These overweight/obesity rates have increased by five and six percent respectively, when compared to the 1995 results. "People being overweight or obese may have significant health, social and economic impacts, and is closely related to lack of exercise and to diet," (ABS, 2014). The carry-over effect is reduced motor fitness of which strength is a component and as a consequence reduced performances.

5. CONCLUSIONS

The predictions may indicate a slightly modified Olympic motto from citius, altius and fortius to citius, altius and infirmius or "faster, higher and weaker?" Predicting performances of athletes at future Olympic Games based on past Olympic Game performances for the four Athletic throwing events indicates good predictions, if somewhat disconcerting in term of the predicted declines in performance across all these event. It is important to highlight these declines might be attributable to three factors, that are, more vigilant drug testing, punitive sanctions for adverse drug tests, the decline in participation in the throwing events and finally a decline in the general motor fitness, especially strength in the source population from which throwing athletes will emerge.

References

Athletics Australia (2014). Rankings Alltime List. Athletics-Australia. http://www.athletics.com.au/Results/Rankings

Australian Bureau of Statistics (ABS) (2014) Overweight/ Obesity (Measured and Self-Reported Body Mass Index (BMI) 2013 - Age standardised 18 years and Over. Australian Bureau of Statistics. http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4125.0~Jan%202013~Main%20Features~Overweight%20and%20obesity~3330

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

FINA. (2010) Shanghai 2011 - Registration and Qualification of Competitors. Retrieved 10/11/2010 from http://www.fina.org/H2O/index.php?option=com_content&view=article&id=1572&Itemid=721

Garson, D.G. (2010) Curve estimation. *Statnotes: Topics in Multivariate Analysis.* Retrieved 1/12/2010 from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.

Hair, J.E., Block, W., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006) *Multivariate Data Analysis*. (6th Ed.). Pearson - Prentice Hall.

Heazlewood, I. (2006) Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. *Journal of Sports Science and Medicine* 5, 541-547.

Heazlewood, I. (2008) Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Oral presentation to *Australian Statistical Society (NSW Branch*).University of Sydney. 24/9/2008.

Heazlewood, I. & Lackey, G. (1996) The Use of Mathematical Models to Predict Elite Athletic Performance at the Olympic Games. In *Proc. Third Conference on Mathematics and Computers in Sport* (deMestre, N, Ed.), Bond University, pp. 185-206.

Heazlewood, I. (2013a). The Congruence Between Mathematical Models Predicting 2012 London Olympic Games Swimming and Athletic

Performances and Actual Performance. 9th International Symposium on Computer Science in Sport (IACSS2013). June 19-22, 2013. Istanbul, Turkey. Oral presentation

Heazlewood, I. (2013b). The Congruence Between Mathematical Models Predicting 2011 Athletic World Championship Performances And Actual Performance in the 100m, 400m, Long Jump and High Jump. 9th International Symposium on Computer Science in Sport (IACSS2013). June 19-22, 2013. Istanbul, Turkey. Oral presentation.

Hopkins, W. (2000) Limits to performance. *SportScience* **4**, Retrieved 3/4/2006 from http://www.sportsci.org/jour/0002/inbrief.html#limits.

IAAF. (2008) The XXIX Olympic Games Beijing, China 8 - 24 August 2008 Qualification Standards – Men and Women retrieved on 5/5/2008 from http://www.iaaf.org/statistics/standards/newsid=43175 .html

IAAF. (2010) Qualification System – Games of the XXX Olympiad – London 2012. Retrieved 10/11/2010 from http://www.iaaf.org/mm/Document/Statistics/Standards/05/90/83/20110119112034_httppostedfile_QualificationSystem-Athletics-IAAF-21112010_23249.pdf

International Association of Athletic Federations – IAAF. (2010) Event Entry Standards and Entry Rules Men and Women IAAF World Championships in Athletics - Daegu, Korea - 27 August - 4 September 2011. Retrieved 1/6/2010 from http://www.iaaf.org/wch11/standards/index.html.

International Association of Athletics Federations (2014) IAAF Technical Committee. International Association of Athletics Federations. http://www.iaaf.org/about-iaaf/documents/rules-regulations#rules

International Amateur Athletic Federation (IAAF) (2014) Current List Sanctioned Athletes. http://www.iaaf.org/about-iaaf/medical-anti-doping

Lackey, G. & Heazlewood, I. (1998) The Use of Mathematical Models to Predict Elite Swimming Performance. In *Proc. Fourth Conference on Mathematics and Computers in Sport* (deMestre, Ed.), Bond University, pp. 79-110.

Péronnet. F. & Thibault. G. (1989) Mathematical analysis of running performance and World running records. *Journal of Applied Physiology* **6**, 453-465.

Prendergas, T, K. (1990) What do world running records tell us? *Modern Athlete and Coach* **28**, 33-36.

SPSS INC. (2014) IBM SPSS Statistics Version 22 statistical software (SPSS Inc. 2014). *Core System Users' Guide*. SPSS Inc.

World Anti-Doping Agency (WADA) 2014. List of Prohibited Substances and Methods. http://www.wada-ama.org/en/world-anti-doping-program/sports-and-anti-doping-organizations/international-standards/prohibited-list/

MATHEMATICAL MODELS THAT PREDICT ATHLETIC PERFORMANCE IN THE WOMEN'S THROWING EVENTS AT THE OLYMPIC GAMES

Ian Tim Heazlewood[a] and Joe Walsh[a]

[a] *Charles Darwin University*
Corresponding author: Ian.Heazlewood@cdu.edu.au

Abstract

The prediction of future athletic performance is a recurring theme as sports scientists strive to understand the predicted limits of sports performance. Predictive models based on Olympic data for athletics have derived some accurate predictions of performance in the 2000, 2004, 2008 and 2012 Olympic Games. The aim of this research was to develop predictive models using performance data of the first three athletes competing in the finals of the women's shot put, discus, hammer and javelin at the Summer Olympic Games from Berlin 1936 to London 2012. The approach utilised regression-curve estimation using IBM SPSS Statistics Version 22 statistical software and by evaluating fit to linear, logarithmic, inverse, quadratic, cubic, compound, power, sigmoidal, growth exponential and logistic functions. The mathematical models varied represented very good predictors of past, current future throws performance in the four field events based on R^2 (0.850 - 0.972), p-values (<.001) and unstandardized residuals or error. The non-linear function of best fit for events was the cubic function, which indicated a decrease in performance in recent Olympics and predicted this performance decline would occur at the 2016 Olympic Games in Rio de Janeiro. The reasons for the current and predicted declines were more vigilance concerning drugs in sport and therefore dampening the enhanced performance effect of anabolic androgenic hormones, fewer athletes are undertaking the throwing events as a completive sport and changes in the source population providing the sample of potential throws athletes in Australia in terms of motor fitness abilities are getting smaller in terms of motor fitness abilities and thus fewer capable athletes exist to select from within source population. The good predictive models may be due to a longer timeframe data set to develop substantive predictive models, a timeframe able to detect phylogenetic trends in human athletic performance. The predictions may indicate a slightly modified Olympic motto from citius, altius, fortius to citius, altius and infirmius or "faster, higher and weaker?"

Keywords: Olympic Games, Throwing Event, curve estimation, nonlinear regression, predictive mathematical modelling

---

1. INTRODUCTION

The prediction of future athletic performance by athletes at the Olympic Games is a recurring theme as well as forming the basis for some stimulating discussions on the limits of human performance. Mathematics in sport and exercise and sports science are based on the principles of description and more importantly prediction. The ability to make substantive and accurate predictions of future elite level sports performance indicates that such approaches reflect substantive sport science. Often these predictions are purely speculative and are not based upon any substantial evidence, rather they are based on the belief that records are made to be broken and that performances based on past experiences must continue to improve over time. The accessibility of data in the form of results from Olympic Games, world records and world best performances in a specific year allows the analysis of performances in any number of events. From these analyses, changes in performance over time can be observed and predictions of future performance can be made utilising the process of mathematical extrapolation and interpolation.

A number of researchers have attempted to predict future performances by deriving and applying a number of mathematical statistical models based on past performances in athletics. Prendergast (1990) applied the average speeds of world record times to determine a mathematical model for world records. The records or data used in the analysis spanned a 10 year period. Following his analysis, Prendergast (1990) raised the question of whether any further improvements can be expected or if the limits of human performance have been reached. The sports of athletics (Heazlewood and Lackey, 1996; Heazlewood, 2011, 2013a, 2013b) and swimming (Lackey and Heazlewood, 1998) have been addressed in this manner. The knowledge of future levels of sporting performance has been identified by Banister and Calvert (1980) as beneficial in the areas of talent identification, both long and short term goal setting, and training program development based on the next level of expected future performance. In addition, expected levels of future performance are often used in the selection of national representative teams where performance criteria are explicitly stated in terms of athletics times and distances for example as required entry standards at Olympic Games (International Olympic Committee (IOC), 2014).

Péronnet and Thibault (1989) postulate that some performances, such as the men's 100m sprint is limited to the low 9 seconds, whereas, Seiler (referred to by Hopkins, 2000) envisages no limits on improvements based on data reflecting progression of records over the last 50 years. According to Seiler improvements per decade have been approximately 1% for sprinting, 1.5% for distance running, 2-3% for jumping, 5% for pole vault, 5% for swimming and 10% for skiing for male athletes, whereas female sprint times may have already peaked. The differences for males and females it is thought to reflect the impact of successful drugs in sport testing on females. Previous derived curve estimations that significantly fit the data have also displayed interesting findings as no one curve fits all the data sets. Different events displayed different curves or mathematical functions (Lackey & Heazlewood, 1998) of best fit. For the women’s freestyle events the 50m was inverse, 100m cubic, 200m sigmoidal, 400m cubic and 800m sigmoidal.

In the women’s events the mathematical functions were 100m cubic, 400m sigmoidal, long jump inverse and high jump displayed four functions (compound, logistic, exponential and growth). The curves that fit the data have also displayed interesting findings as no one curve fits all the data sets. This may indicate that different events are dependent upon different factors that are being trained differently or factors underpinning performance evolving in slightly different ways. This has resulted in different curves or mathematical functions that reflect these improvements in training or phylogenetic changes over time. However, at some point in the future when time catches-up with the actual performance, then how accurately the predictive models reflect reality can be assessed.

However, the ability to predict performances at the 2000, 2004, 2008 and 2012 Olympic Games for the men’s and women’s 100m, 400m, long jump and high jump, based on the 1924 to 2012 data, was very accurate with low percentage error. The International Olympic Committee (IOC, 2014) has produced descriptive data and descriptive graphical analysis to indicate trends in World records for the women’s throwing events however not based on mathematical predictive modelling.

The dominant and important research question is, can mathematical models based on nonlinear curve estimation, which have proven to be very successful in fitting and predicting past, current and future Olympic performances for event finals in athletics and swimming for women display equal effectiveness in predicting athletic performances in the women’s throwing events of the shot put, discus, hammer and javelin at the Olympic Games, which are past, current and future performances?

2. METHODS

The mean score for the first three placed finalist in the women's throwing events which consist of the shot put, discus, hammer and javelin. The data were selected from 1936 to 2012 Summer Olympic Games and provided by the International Olympic Committee (IOC, 2014) served as the exemplars for performance in each event for each competitive year. The data covered eighteen Summer Olympics for the women's discus and javelin. The introduction of women's throwing at the Summer Olympic Games events lagged behind the men. Specifically, the discus and javelin were introduced in 1932, the shot put in 1948 and the hammer as late as 2000. These scores served as the data set to derive and test the predictive models based on curve estimations and the distances in each event was recorded metres to the nearest centimetre. The analysis of the women's hammer data proved problematic as it has only been competed at four Olympics, whereas the shot put has a longer completion history at seventeen Olympics.

This is a similar method used previously (Heazlewood & Lackey, 1998; Heazlewood, 2006, 2008, 2013a, 2013b) to curve fit Olympic data for swimming and athletic events. According to Garson (2010) curve estimation is an exploratory tool in model building and model selection, where the best mathematical model or function is selected to represent quantitative relationships between an independent/predictor variable and a dependent/response variable. The mathematical solutions and curve estimations were derived using the IBM SPSS Statistics Version 22 statistical software (SPSS Inc. 2014).

The most common curve estimation or model fit approaches are based on the following mathematical functions (Garson, 2010) and these are linear, logarithmic, inverse, quadratic, cubic, power, compound, S-curve, logistic, growth, and exponential models. In terms of statistical approach model fit indices are then applied to test the quality of the model and the general method of determining the appropriate regression models is represented by the following steps.

1. Commence with an initial estimated value for each variable in the equation.
2. Generate the curve defined by the initial values. Calculate the sum-of-squares (the sum of the squares of the vertical distances of the points from the curve).
3. Adjust the variables to make the curve come closer to the data points. There are several algorithms for adjusting the variables. The most commonly used method was derived by Levenberg and Marquardt (often called simply the Marquardt method). Adjust the variables again so that the curve comes even closer to the points. Keep adjusting the variables until the adjustments make virtually no difference in the sum-of-squares.
4. Report the best-fit results and then the precise values you obtain will depend in part on the initial values chosen in step 1 and the stopping criteria of step 5. This means that repeat analyses of the same data will not always give exactly the same results. To investigate the hypotheses of model fit and prediction, the eleven regression models were individually applied to each of the athletic events. The regression equation that produced the best fit for each event, that is, produced the highest coefficient of determination (abbreviated as R^2), was then determined from these eleven equations.
5. The specific criteria to select the regression equation of best were the magnitude of R^2, the significance of the coefficient of determination (R^2) is a measure of accuracy of the model used. A coefficient of determination of 1.00 indicates a perfectly fitting model where the predicted values match the actual values for each independent variable (Garson, 2010; Hair et al., 2006; Norušis, 1993). Where more than one model was able to be selected due to an equal R^2, the simplest model was used under the principle of parsimony, that is, the avoidance of waste and following the simplest explanatory model, as well as the statistical significance of the analysis of variance, the alpha or p-value and size of residuals or error in predictions.
6. Some caution is required to not over interpret a high R^2 as it does not mean that the researcher has chosen the equation that best describes the data. It also does not mean that the fit is unique - other values of the variables may generate a curve that fits just as well.

It should be noted the women's shot put, discus and hammer have not changed dramatically in specification concerning weight, area and volume making comparisons across Olympic year possible. Specifically, the shot is 4kg, discus 1kg, hammer 4kg and javelin 600g.

Similar to the men's javelin, which was respecified in 1986, the women's javelin was respecified in 1999 to change its aerodynamics to reduce lift, increase the downward pitching moment, and to bring the nose down earlier, which reduced the flight distance by around 10%. Although it should be emphasised that prior to the 1999 respecification women were only throwing in the 66 - 71 metre range and not the 100 metre plus throws exhibited by the top male javelin throwers.

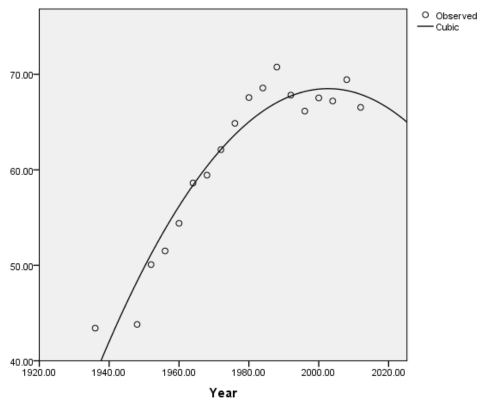
3. RESULTS

The results as illustrated in table 1 indicates each women's throwing events, best-fit functions, r-square, p-value and equations of best fit. Note all the eleven mathematical functions that were tested for model fit. It can be observed that all nonlinear functions of best fit were for the cubic function across all women's throwing events with a significant quadratic function for the javelin and a non-significant value for the hammer due to the low number of cases. This indicated a decline in performance in the Olympic Games from 1988 onwards for the women's shot put, discus and javelin. The hammer performances, like many new events, continue to improve.

Table 1. Women's Throwing Events, Best-fit Functions, R-square, P-value and Equations.

Event and Weight	Function	R-Square	p-value	Cubic Equation of Best Fit Constant b1 b2 b3
Shot Put 4kg	Cubic	.918	<.001	-6.609 2.653 -.064 .000
Discus 1kg	Cubic	.925	<.001	-1.064 5.209 -.035 -.003
Hammer 4kg	Cubic Quadratic	.973 .972	Ns n=4	-200.5 20.052 -.360 .00
Javelin 600g	Cubic Quadratic	.983 .983	<.001	-.647 5.491 -.070 - .002

Figure 1. The cubic function line of best fit and actual data point for women's javelin. Note the X-axis in years and Y-axis in metres.



To highlight the trends the cubic function line of best fit and actual data points for women's javelin from 19936-2012 are displayed in figure 1, which covers eighteen Summer Olympic Games. Table 2 indicates the model fit based on each Olympic year, actual performance, predicted performance and residual error.

Table 2. Women's Discus Data in terms of Olympic year, actual performance, predicted performance and residual error.

Year	Performance (m)	Predicted (m)	Residual Error (m)
1936	43.42	38.55	4.87
1948	43.81	48.32	-4.51
1952	50.08	51.16	-1.08
1956	51.51	53.78	-2.27
1960	54.40	56.19	-1.79
1964	58.62	58.39	.23
1968	59.44	60.37	-.93
1972	62.12	62.13	-.01
1976	64.87	63.68	1.19
1980	67.57	65.02	2.55
1984	68.57	66.14	2.43
1988	70.76	67.04	3.72
1992	67.82	67.73	.09
1996	66.15	68.20	-2.05
2000	67.53	68.45	-.92
2004	67.21	68.49	-1.28
2008	69.44	68.30	1.14
2012	66.54	67.90	-1.36

4. DISCUSSION

The best fit mathematical functions for women's throwing events were the cubic function with high R-square values (0.918 - .983), very significant p-values (<.001) resulting in good model prediction and low residual error, except for the new women's hammer was nonsignificant due to low n of cases. The shot put and discus where there were no re-specifications for the implements predicted performance decrements for these events. The respecified javelin also indicted performance decline. The declining performances started to occur from 1988 for shot, discus and javelin. The new event hammer has displayed continuous improvement from 2000-2012 and continued improvement is predicted into the future. It is interesting to note the world records for the shot put occurred in 1987 at 22.63m, discus in 1988 at 76.80m, hammer in 2011 at 79.42m and expectantly in the javelin in 2008 at 72.28m. The change in specifications of the women's javelin in 1999 would suggest a significant decline in performances however the decline as suggested by this analysis commenced in 1988 at the Olympic level and prior the new javelin being introduced.

So what plausible explanations can be theorized for the current trend of declining throwing performance? What will occur in 2016 Rio de Janeiro Summer Olympic Games in Brazil and are we getting citius, altius and fortius or "faster, higher, and stronger?" if the throwing events are based on strength, force and power production the future trends based on table 3 indicates reductions in all men's throwing events and might this indicate citius, altius and infirmius or "faster, higher, and weaker except for the hammer?"

Table 3. The predicted trends for the women's shot put, discus, hammer and javelin performances for the 2016 Summer Olympic Games.

Event and Weight	Actual/Predicted 2012	Predicted Performance 2016
Shot Put 4kg	20.47m/19.31m	18.59m
Discus 1kg	67.96m/64.54m	62.35m
Hammer 4kg	77.63m/77.46m	78.44m
Javelin 600g	66.54m/66.24m	64.61m

Three predominant explanations can provide some insights as to these declining trends in the men's throwing events. Specifically,

1. More vigilance concerning drugs in sport therefore dampening the enhanced performance effect of anabolic androgenic hormones. Drug testing by the IAAF is now undertaken in competition testing and out of competition testing and to have a world record ratified by the IAAF the application form has a section indicating the athlete was drug tested at the time of competition and that the athlete has passed the test, that is no adverse findings (IAAF, 2014). The normal suspension/ban for first offence anabolic androgenic hormones is two year. A second offence of this kind results in a life ban. If found positive all performances, monies paid by IAAF, awards and prizes are forfeit so the punitive outcomes can very significant and are thought to act as a strong deterrent to taking prohibited substances on the World Anti-Doping Agency (WADA, 2014) list. It is interesting to emphasise the actual winner of the women's shot put in London 2012 tested positive to an anabolic androgenic steroid and was subsequently disqualified.
2. Fewer athletes are undertaking the throwing events. In Australia fewer athletes are taking up throwing events as other sports compete for the limited talent pool based on Australia's small population. The Athletics Australia data of permit competitions and ranked athletes performances to be evaluated each Athletic season and indicate that in Australia this is a problem Athletics Australia (2014).
3. The source population providing the sample of potential throws athletes in Australia in terms of motor fitness abilities is getting smaller in terms of motor fitness abilities and thus fewer capable athletes to select from within source population. This appears to be a result of Australia's increasing overweight and obesity epidemic for males 18 years and over, which is currently estimated at 2011-12 to be 70% and females at 56% (ABS, 2014). The age of male high performance throwers is usually between 20-35 years. These overweight/obesity rates for women have increased by five and six percent respectively, when compared to the 1995 results. "People being overweight or obese may have significant health, social and economic impacts, and is closely related to lack of exercise and to diet," (ABS, 2014). The carry-over effect is reduced motor fitness of which strength is a component and as a consequence reduced performances.
4. The continued improvement in the women's hammer normally occurs when a new event is introduced as athletes improve their technical understanding of the event combined with more event specific training. This observation is relevant when you consider the rapid improvements in the women's pole vault when it was a new event.

5. CONCLUSIONS

The predictions may indicate a slightly modified Olympic motto from citius, altius and fortius to citius, altius and infirmius or "faster, higher and weaker" when interpreting the trends in the women's shot put, discus and javelin. Predicting performances of athletes at future Olympic Games based on past Olympic Game performances for the shot put, discus and javelin Athletic throwing events indicates good predictions, if somewhat disconcerting in term of the predicted declines in performance across all these three event. Note the 2008 world record in the women's javelin might be a data outlier and not representing the overall trend. It is important to highlight these declines might be attributable to three factors, that are, more vigilant drug testing, punitive sanctions for adverse drug tests, the decline in participation in the throwing events and finally a decline in the general motor fitness, especially strength in the source population from which throwing athletes will emerge. The increasing performance in the hammer probably indicates athletes are progressively mastering the technical and specific training demands of the new event.

References

Athletics Australia (2014). Rankings Alltime List. Athletics-Australia. http://www.athletics.com.au/Results/Rankings

Australian Bureau of Statistics (ABS) (2014) Overweight/ Obesity (Measured and Self-Reported Body Mass Index (BMI) 2013 - Age standardised 18 years and Over. Australian Bureau of Statistics.
http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4125.0~Jan%202013~Main%20Features~Overweight%20and%20obesity~3330

Banister, E.W. & Calvert, T.W. (1980) Planning for future performance: Implications for long term training.

Canadian Journal of Applied Sports Science 5, 170-176.

FINA. (2010) Shanghai 2011 - Registration and Qualification of Competitors. Retrieved 10/11/2010 from http://www.fina.org/H2O/index.php?option=com_content&view=article&id=1572&Itemid=721

Garson, D.G. (2010) Curve estimation. *Statnotes: Topics in Multivariate Analysis.* Retrieved 1/12/2010 from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.

Hair, J.E., Block, W., Babin, B.J., Anderson, R.E. & Tatham, R.L. (2006) *Multivariate Data Analysis*. (6th Ed.). Pearson - Prentice Hall.

Heazlewood, I. (2006) Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. *Journal of Sports Science and Medicine* 5, 541-547.

Heazlewood, I. (2008) Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games. Oral presentation to *Australian Statistical Society (NSW Branch*).University of Sydney. 24/9/2008.

Heazlewood, I. & Lackey, G. (1996) The Use of Mathematical Models to Predict Elite Athletic Performance at the Olympic Games. In *Proc. Third Conference on Mathematics and Computers in Sport* (deMestre, N, Ed.), Bond University, pp. 185-206.

Heazlewood, I. (2013a). The Congruence Between Mathematical Models Predicting 2012 London Olympic Games Swimming and Athletic Performances and Actual Performance. 9th International Symposium on Computer Science in Sport (IACSS2013). June 19-22, 2013. Istanbul, Turkey. Oral presentation

Heazlewood, I. (2013b). The Congruence Between Mathematical Models Predicting 2011 Athletic World Championship Performances And Actual Performance in the 100m, 400m, Long Jump and High Jump. 9th International Symposium on Computer Science in Sport (IACSS2013). June 19-22, 2013. Istanbul, Turkey. Oral presentation.

Hopkins, W. (2000) Limits to performance. *SportScience* **4**, Retrieved 3/4/2006 from http://www.sportsci.org/jour/0002/inbrief.html#limits.

IAAF. (2008) The XXIX Olympic Games Beijing, China 8 - 24 August 2008 Qualification Standards – Men and Women retrieved on 5/5/2008 from http://www.iaaf.org/statistics/standards/newsid=43175.html

IAAF. (2010) Qualification System – Games of the XXX Olympiad – London 2012. Retrieved 10/11/2010 from http://www.iaaf.org/mm/Document/Statistics/Standards/05/90/83/20110119112034_httppostedfile_QualificationSystem-Athletics-IAAF-21112010_23249.pdf

International Association of Athletic Federations – IAAF. (2010) Event Entry Standards and Entry Rules Men and Women IAAF World Championships in Athletics - Daegu, Korea - 27 August - 4 September 2011. Retrieved 1/6/2010 from http://www.iaaf.org/wch11/standards/index.html.

International Association of Athletics Federations (2014) IAAF Technical Committee. International Association of Athletics Federations. http://www.iaaf.org/about-iaaf/documents/rules-regulations#rules

International Amateur Athletic Federation (IAAF) (2014) Current List Sanctioned Athletes. http://www.iaaf.org/about-iaaf/medical-anti-doping

Lackey, G. & Heazlewood, I. (1998) The Use of Mathematical Models to Predict Elite Swimming Performance. In *Proc. Fourth Conference on Mathematics and Computers in Sport* (deMestre, Ed.), Bond University, pp. 79-110.

Péronnet. F. & Thibault. G. (1989) Mathematical analysis of running performance and World running records. *Journal of Applied Physiology* **6**, 453-465.

Prendergas, T, K. (1990) What do world running records tell us? *Modern Athlete and Coach* **28**, 33-36.

SPSS INC. (2014) IBM SPSS Statistics Version 22 statistical software (SPSS Inc. 2014). *Core System Users' Guide*. SPSS Inc.

World Anti-Doping Agency (WADA) 2014. List of Prohibited Substances and Methods. http://www.wada-ama.org/en/world-anti-doping-program/sports-and-anti-doping-organizations/international-standards/prohibited-list/

AN ALTERNATIVE DRAFT SYSTEM FOR THE ALLOCATION OF PLAYER DRAFT SELECTIONS IN THE NATIONAL BASKETBALL ASSOCIATION

Michael de Lorenzo[a] and Anthony Bedford[a, b]

[a]*School of Mathematical & Geospatial Sciences*
RMIT University, Melbourne, AUSTRALIA
[b]*anthony.bedford@rmit.edu.au*

Abstract
The issue of 'tanking' is a growing trend in the National Basketball Association (NBA) and is often the topic of debate. 'Tanking' is when a team lose games on purpose in order to improve their draft position. Previous research has been conducted on the incentives teams have to win games under the current NBA weighted-lottery system. The research has suggested that non-playoff bound teams do not have an incentive to win games under the current draft system, leading to the system being heavily criticised. With this in mind, the aim of this research was to implement a draft system that would reward non-playoff bound teams for winning games that are deemed to be 'unimportant' with respect to making the playoffs. The calculation of 'unimportance' is based on the probability of a team making the playoffs after the completion of each game of the season. A variety of approaches were then examined to help derive a score for 'unimportant' and unlikely wins. The results found evidence that non-playoff bound teams would have an incentive to win games late in the season under this alternative draft system. As well as exploring incentives to win, we will explore the attractiveness of this system as well as evaluating past draft picks to examine the reward teams could have obtained under this alternative draft system.

Keywords: NBA, Draft, Probability, Importance, Incentives

1.INTRODUCTION

The NBA draft system is designed to assist low ranked teams who are unable to qualify for the playoffs. The draft enables these teams to have the best opportunity to improve their roster by allowing them first choice of players entering the league. The current draft system uses a lottery to determine the first three draft selections with the remaining draft order being determined by the inverse final season standings. The fourteen teams who do not qualify for the playoffs are allocated a probabilistic chance of receiving the number one draft pick based on their end of season standing. This process is known as a weighted-lottery system.

Whilst this process has been used since 1990, it has not always been met with great support from the public and media as the issue of 'tanking' has arisen. 'Tanking' is when a team loses games on purpose in order to improve their lottery odds for the draft. This issue has arisen from the fact that a team must finish with the worst overall record in order to receive the highest odds in the lottery, and therefore the best opportunity of obtaining the number one draft selection.

It is often a topic of debate about whether or not tanking is the only way to rebuild a team (Michael, 2013; Spencer, 2013; Ziller, 2013), and in recent years has led the public and media to speculate about which teams could possibly tank during a season (Burton, 2013). However, Kertes (2003) explained how the Detroit Pistons used smart trading to bring in players who would assist the team in winning the 2004 NBA Championship. Yet, since it may be difficult for teams to complete trades, teams may turn their attention back to the draft and thus the incentive to lose games becomes a factor.

Research has been conducted in the past on the effect that tanking has on the league. Taylor and Trogdon (2002) assessed the performance of teams following changes to the draft system by the NBA. They found that under the current weighted-lottery system, non-playoff bound teams are two times more likely to lose games compared to teams headed for the playoffs. Walter and Williams (2012) assessed the issue of tanking to determine if teams do tank and the benefit they may gain. By analysing games played between non-playoff teams late in the season, they found that non-playoff bound teams were more inclined to tank in order to improve their draft position. They also found that if a team wins the lottery and gains the first draft selection, the attendance will increase by 5% for a period of five years after the draft. The incentive to tank has also been explored by Tuck and Whitten (2012), who found evidence to suggest that teams will tank when the weighted-lottery system is used.

An alternative draft system that has been suggested was first described by Gold (2010). Gold explained that it may be possible to create a draft system that eliminates the needs for teams to tank whilst giving fans a reason to continue to cheer for their team. The system used mathematical elimination to determine the draft order in the National Hockey League (NHL). After a team has been mathematically eliminated from the playoffs, their win/loss record was recorded and used to determine the selection order for the upcoming draft. The results found that supporters would have a reason to continue to cheer as their team was still competing for the number one draft late in the season.

The issue of tanking is routinely discussed in the Australian Football League (AFL). This was identified by Bedford and Schembri (2006), who suggested a probabilistic model to give teams who have been eliminated from finals contention an incentive to win games late in the season. This system was based on measuring the unimportance of a game by altering the *importance of a point* formula described in Morris (1977). The system would allocate a Draft Point Reward (*DPR*) to non-finals teams who would win games that were deemed to be unimportant. The cumulative sum of the *DPR*, known as the *DScore*, was then used to determine the final draft order. The research concluded that non-finals bound teams had an incentive to win games late in the season as they were competing for the number one draft selection. Taking this into consideration, the focus of this paper is to implement the *DScore* system into the NBA and evaluate the incentives that non-playoff bound teams have to win games late in the season.

2.METHODS

In this section, the methodology behind the *DScore* system will be explained. In order for the system to be successfully implemented into the NBA, some adjustments were required. These adjustments are explained throughout this section.

At its heart, the *DScore* system is designed to reward non-playoff bound teams for winning games late in the season that are deem to be ‘unimportant’. As mentioned before, the reward is in the form of a Draft Point Reward (*DPR*) with the team who finishes the season with the highest cumulative *DPR*, known as the *DScore*, receiving the number one draft selection.

Like the original *DScore* system, the NBA system is based on Carl Morris’ work on the most important point in tennis (Morris, 1977). Morris defined the most important point in tennis as the difference between two conditional probabilities: the probability of a server winning a game given they win the next point, minus the probability of a server winning the game given they lose the next point. However, we are interested in games in a season instead of points in a game. The unimportance of a game is then found after calculating the importance of a game.

The NBA *DScore* system has similar characteristics to the original system. These include no *DPR* being given in defeat so teams must win games in order to receive a reward; the *DPR* being awarded for the entire season minus the first game of the season; and teams that have qualified for the playoffs are ineligible to receive a reward. However, in the original system, the *DPR* for each team was calculated at the conclusion of each round of the season. This is not the case for the NBA system as there are no actual rounds in the season. Instead, the *DPR* is calculated for each team at the conclusion of each game g.

In order for the *DPR* to be calculated, there are a number of features that have to be determined first. The process includes determining the conference standing after each game g, determining the required number of wins for each team to make the playoffs, calculating the probability of each team making the playoffs after each game, and finally determining the importance and unimportance of the games.

i. Conference standings

The original AFL *DScore* system used the ladder at the completion of round r to determine various features of the probabilistic model. As explained in the previous section, there are no actual rounds in the NBA so there are no conference standings available for all teams at the conclusion of game g. Instead, there are two conferences, the East and the West, with each conference having their own standings available at the conclusion of each day of the season. To rectify this, hypothetical standings at the conclusion of game *g* for each conference were used. These standings were constructed similar to the normal standings but the total points differential was used as a tie-breaker for when two or more teams had the same win/loss record.

ii. Number of wins to make the playoffs

In the NBA, the top eight teams from each conference at the end of the season advance to the playoffs where they compete for the NBA championship. Since there is a top eight from each conference, the *DScore* system has to be split to accommodate this. Therefore, each conference would have its own required number of wins to make the playoffs. Bedford and Schembri (2006) explained that there were two possible ways of calculating the required number of wins to make the playoffs. This was either by using the final season's required wins and imposing them retrospectively onto the completed season, or using a projected requirement during the season. Since the attraction of the system is that teams will know the reward of winning, the projected wins, or *ParWins*, is used during the season. The required number of wins for team *i* after game *g* is defined as $ParWins_i\ (g)$. This is shown in (1).

$$ParWins_i(g) = max\left(\left\|\frac{\left(TW_{8^{th}\, ranked\ team}(g)\right)\times 82}{g}\right\| - TW_i(g), 0\right) \quad (1)$$

The total number of wins for team *i* after the completion of game *g* is defined as $TW_i(g)$. Equation 1 will return a result equal to zero if it finds that team *i* has already qualified for the playoffs. If this is not the case, the equation will return a positive number. It should be noted that this equation incorporates a rounding function which will round the projected total wins of the eighth placed team up or down to the nearest integer.

iii. Probability of making the playoffs

The basis behind determining the probability of team *i* making the playoffs is the binomial distribution. The probability of making the playoffs for team *i* after game *g* is defined as $PR_i\ (Playoffs\ |g)$. Incorporating the cumulative binomial distribution, with *x*= number of successes, *n* = number of trials and *p* = probability of success, the probability of making the playoffs is defined as:

$$Pr_i(Playoffs|g) = 1_{\{ParWins_i(g)=0\}} + 1_{\{(ParWins_i(g)>0)\cap(ParWins_i(g)\leq 82-g)\}}[1 - B(ParWins_i(g) - 1; 82 - g, p)] \quad (2)$$

Note(2) includes an indicator function $1_{\{a\}}$ which takes the value of 1 if condition *a* is true and 0 if condition *a* is not true. To keep the system simple, the probability of success used throughout this paper is equal to 0.5.

iv. Unimportance of a game

As stated previously, the unimportance of a game can be found after calculating the importance. The importance of a game is defined as the difference between two conditional probabilities: the probability of team *i* making the playoffs given they win the next game, minus the probability of team *i* making the playoffs given they lose the next game. This difference can be found by manipulating equation 2. The unimportance for team *i* after game *g*, defined as $U_i(g)$, is then calculated by one minus the importance of the game. A detailed breakdown of the importance calculations can be found in Bedford and Schembri (2006).

v. Moderator variable

In the original AFL *DScore* system, it was concluded that there were some weaknesses to the model. These weaknesses included the top draft pick being frequently awarded to teams who won a string of games late in the season. This meant that teams who finished in positions ninth to twelfth received a majority of the high draft picks. To counteract this, Bedford and Schembri (2010) introduced a moderator variable to improve the system. This variable was a scaling factor, which would allocate the full *DPR* to teams positioned at the bottom of the conference standings. The conference standing for team *i* after the completion of game *g* is defined as $CS_i(g)$. The moderator variable is then defined as the following:

$$\gamma_i(g) = 1_{\{CS_i(g)\geq 9\}}\left[\frac{(CS_i(g)-8)}{7}\right] \quad (3)$$

This equation will return a value equal to zero if it finds that team *i* is positioned in the top eight of their conference. If this is not the case, then it will return a value equal to team *i*'s conference standing minus eight divided by the number of teams not in the top eight. For example, if team *i* is positioned eleventh in their conference, then their scaling factor γ will be equal to $\frac{(11-8)}{7}$ = 0.429. This would then be multiplied onto the *DPR* for game *g*.

vi. Calculating the Draft Point Reward

The Draft Point Reward (*DPR*) is simply the probability of not making the playoffs multiplied by the unimportance of the game and the moderator variable. As mentioned previously, teams have to win in order to receive the reward so an indicator function is included in the equation. The *DPR* for team *i* at game *g* can then calculated using the following equation:

$$DPR_i(g) = 1_{\{i\ wins\ game\ g\}} \cdot U_i(g) \cdot (1 - Pr_i(Playoffs|g)) \cdot \gamma_i(g) \quad (4)$$

The total Draft Score (*DScore*) for team *i* after game *g* is then defined as the sum of all *DPR*:

$$DScore_i(g) = \sum_{k=2}^{g} DPR_i(k),\ g \in \{2, \dots, 82\} \quad (5)$$

3.RESULTS

The *DScore* system was trialled on seven seasons between 2005 and 2012. The results will examine findings from these seasons and assess how teams performed during the 2008-09 season. The importance of games will be examined as well as the competition for the number one draft selection.
As mentioned previously, the equal probability model (*p*=0.5) was used in the calculations of the *DScores*. Since the *DScores* from the two conferences had to be merged at the conclusion of the season to determine the draft order, it had to be confirmed that the two groups were equal. It was found that there was no significant difference between the two conferences (*P*=0.154).
The results across the seasons showed that lower ranked teams generally benefit from the system as they receive a majority of the top five draft selections. This suggests that the system is working correctly and only allocating a reward to low ranked teams who win games of unimportance. The allocation of top five draft selections is shown in Figure 1.

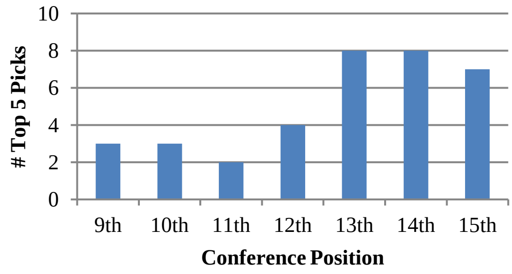


Figure 1: Distribution of Top 5 picks vs. Conference Position

The ideal situation for the NBA *DScore* system would be to have lower ranked teams competing for the number one draft selection right up to the conclusion of the season. This would result in the teams continuing to have an incentive to win as many games as possible despite being eliminated from playoff contention. This competition for the top draft selection is shown in Table 1. In two of the seasons, there were at least two teams still competing for the first draft selection with one game remaining. However, in all seven seasons, there are at least two teams competing with five games remaining. This competition suggests that teams would have an incentive to continue to win games late in the season as there is a valuable reward available.

	1 Game Remaining			5 Games Remaining		
	East	West	Total	East	West	Total
2005-06	0	2	2	0	3	3
2006-07	0	1	1	0	2	2
2007-08	0	1	1	0	2	2
2008-09	1	1	2	1	2	3
2009-10	0	1	1	0	2	2
2010-11	0	1	1	0	3	3
2011-12	0	1	1	0	2	2

Table 1: Teams in Competition for Top Pick with 1 and 5 games remaining

i. East vs. West

An interesting result that arose was that the highest *DScore* in each of the seven seasons came from a team from the Western conference. Of the 35 top five picks, only ten were awarded to teams from the East. This would suggest that the system is giving the Western conference a distinct advantage over the Eastern conference. However, this may not be the case when the total wins of each conference are assessed. In all but one season, teams from the West won more combined games than the East. In Table 2, the inter-conference wins (wins against teams from the other conference) for each conference across the seven seasons are shown. It can be seen that the West won more inter-conference games in all but one season. This result would explain why the West were awarded a majority of the picks as the *DScore* system is designed to reward winning and teams from the West are winning more games than teams from the East.

	East	West
2005-06	198	252
2006-07	193	257
2007-08	192	258
2008-09	231	219
2009-10	204	246
2010-11	189	261
2011-12	114	156

Table 2: Total number of inter-conference wins

ii. 2008-09 NBA season

The 2008-09 season presented the best possible example of the *DScore* system working successfully in the NBA. In this season, there were two teams competing for the number one draft selection with one game remaining. The teams were the Washington Wizards from the East and the Oklahoma City Thunder from the West. The game-by-game results for the top six *DScore* teams can be found in Figure 2.

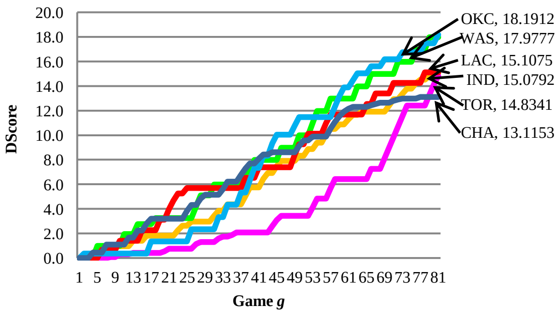


Figure 2: Competition for top pick during the 2008-09 NBA season

Figure 2 shows the number one draft selection changing teams numerous times throughout the season, particularly at the beginning. At the conclusion of the season, the competition came down to Washington and Oklahoma City. There was also strong competition for the third overall pick, with just 0.2734 separating picks three to five. These results are similar to the original AFL *DScore* system, which showed at least two teams competing for the top pick with a number of teams also in competition for the third draft selection.

iii. Importance

An interesting result that arose from Bedford and Schembri (2006) was that teams who finished in the top two or bottom three had their most important games in the early rounds of the season. Of interest to us was whether or not this result would replicate in the NBA. In particular, of interest was comparing the importance between the two conferences. To complete this, the maximum importance was found for each team in each of the seven seasons. The game in which the maximum importance occurred was recorded along with each team's end of season conference standing. The importance was then sorted according to the conference standing and the mean and standard deviation were found. The results are shown in Figure 3.

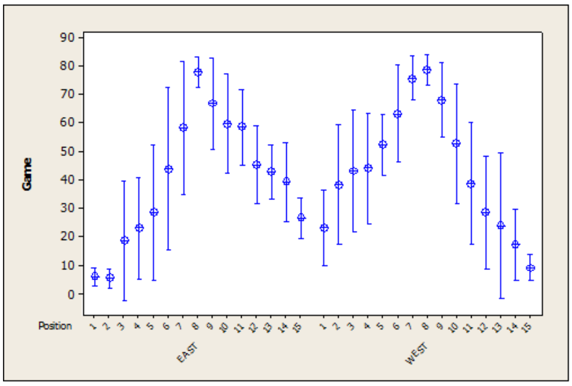


Figure 3: Error bars for maximum game of importance by conference position

A notable result from Figure 3 was that the top two teams in the East had their most important games within the first ten games of the season. This is different to the two top teams from the West, who had their most important games towards the halfway mark of the season. A more intriguing result is that the bottom teams in the West had their most important games at the commencement of the season compared to teams from the East, who's occurred at the halfway mark of the season. These results suggest that if a team from the West loses too many games early in the season, then their probability of making the playoffs decreases dramatically as the required number of wins (*ParWins*) increases quickly as teams above them continue to win. It also suggests that bottom teams in the West had more unimportant games than the East throughout the season. This would provide an explanation as to why teams from the West received a majority of the top five picks as they have more unimportant games and are winning more games than the East. It was also confirmed that the positions between the two conferences are significantly different (P<0.0000).

4. DISCUSSION

Thought it is difficult to measure the overall effect of the *DScore* system on the NBA, there is evidence to suggest that the system does eliminate the incentive for teams to tank. The results showed that there is competition for the number one draft selection in all seven seasons assessed, meaning that teams have to continue to strive for success even if they have been eliminated from playoff contention. However, it is difficult to determine how teams may have performed in later seasons if they had in fact been awarded an alternative draft selection under the *DScore* system. If players were selected in the same order as the 2008 NBA draft, then it could hypothetically be suggested that Derrick Rose now plays for Minnesota or Kevin Love now plays for Memphis. However, we cannot say this with confidence as we are unsure of how this system would have affected a team's draft preparation and how it would affect which played is selected.

One of the results that arose from the system was that there were multiple teams who would finish a season with a *DScore* equal to zero. This was because those teams spent all, or a majority, of the season in the top eight of their conference. Whilst this seems reasonable, it means that when determining the draft order, there were multiple teams with the same *DScore*. One of the ways that this could be amended is by using the inverse final season standings as a tie-breaker with the team who finished higher in the standings receiving the higher draft pick. An alternative tie-breaker would be to implement the current weighted-lottery system onto the fourteen highest *DScore* teams.

i. Incentive to win

The key idea about the implementation of the *DScore* system into the NBA was to eliminate the incentive to tank and replace it with an incentive to win. The results from the research have shown that the system creates an incentive for teams to continue to win despite having a reduced probability of making the playoffs. As explained in the results section of this paper, the number one draft pick was still undecided late in the 2008-09 season. Going into the final game of the season, Washington held a narrow lead over Oklahoma City. On the last day of the season, Oklahoma City defeated the Los Angeles Clippers by 41 points to claim the number one draft selection. Washington had an opportunity to reclaim the number one pick but lost to Boston by eight points. A win under the current weighted-lottery system would have awarded Washington the fourth highest lottery odds whilst a win under the *DScore* system would have awarded them the first draft selection. This further emphasises the point that non-playoff teams do not have an incentive to win under the current NBA draft system whereas the teams would under the *DScore* system.

Results from the implementation of the *DScore* system into the NBA have provided evidence that games played between lower-ranked teams late in the season can be as competitive as games played between playoff bound teams. This is evident by the strong competition for the third overall pick draft in the 2008-09 season. If teams had known that they were in competition with each other, then it is reasonable to assume that the games played between the teams would have had more meaning to them. Therefore, whilst teams at the top of the standings were competing for their playoff position, teams anchored at the bottom of the standings would be competing for a high draft position.

ii. Criticism and future work

A criticism that can be direct at the NBA *DScore* system is that teams from the Western conference appear to benefit more than teams from the Eastern conference. The results showed that teams from the West received all the number one draft selections available across the seasons. Whilst this appears to support the idea that the West is favoured by the system, it may not be the case when the total number of wins for each conference is assessed. In the results section, it was explained that teams from the West won more total games and more inter-conference games than the East in six of the seven seasons. Lower-ranked teams from the West also had their most important games at the commencement of the season which results in the teams having more unimportant games. These two results coupled together suggest that the lower-ranked teams from the West were winning more games of unimportance, and therefore receiving a high cumulative draft score than teams from the East. Whilst it is not ideal to have all the number one draft selections being awarded to the West, this may simply be a result of the NBA going through a period of time where the Western conference is stronger than the Eastern conference.

Throughout the methodology, when the cumulative binomial distribution was used, the probability of success was equal to 0.5 as this was seen as keeping the system simple. However, this may not accurately measure the probability that team *i* has to win game *g*. One way to improve this would be to alter the probability based on how team *i* is performing throughout the season. This could be done through prior probabilities, as described in Stefani and Clarke (1992). The use of prior probabilities could lead to an improvement in determining the probability of team *i* making the playoffs as the probability of success would reflect the relative skill of team *i*.

In order to completely understand which teams truly benefit from the *DScore* system in the NBA, some simulation could be conducted in the future, as completed by Bedford and Schembri (2010). Simulating results for a season 100,000 times may give us compelling results about which conference position benefits the most from the system. The simulation could also provide us with information about if one conference has an advantage over the other. However, in order for this simulation to work correctly, the hypothetical standings would have to be adjusted to better reflect the actual standings. This would mean that tie-breakers such as division leaders or conference record would be used before the total points differential is assessed.

5.CONCLUSION

In this paper, an alternative draft system for the allocation of player draft selections in the NBA has been presented with the aim of eliminating the incentive for teams to tank. The draft system, known as the *DScore* system, was first created by Bedford and Schembri (2006) and was first trialled in the Australian Football League (AFL). The system is designed to reward non-playoff bound teams for winning games late in the season that are deemed to be ‘unimportant’. The unimportance of a game was measured by evaluating the probability of a team making the playoffs, given they win or lose their next match. Teams with a reduced probability of making the playoffs received a higher reward, known as the Draft Point Reward. The cumulative sum of the Draft Point Reward, known as the *DScore*, was then used to determine the final draft order.

The implementation of the *DScore* system into the NBA provided promising results which included showing that teams would have an incentive to continue to win games despite being eliminated from playoff contention. Evidence was also found that games played between lower-ranked teams could be as competitive as games played between playoff-bound teams late in the season. This is due to the fact that teams would be competing against each other late in the season for the highest *DScore* and, therefore, the number one draft selection. This competition creates the incentive for teams to win as many games as possible throughout the season. With this incentive, it means that teams have a reason to continue to strive for success and also provides supporters with a reason to continue cheering on their team despite being eliminated from playoff contention.

References

Bedford, A. & Schembri, A. (2006). A probability based approach for the allocation of player draft selections in Australian Rules Football. *Journal of Sports Science and Medicine, 5,* 509-516

Bedford, A. & Schembri, A. (2010). Revisiting the draft: A revised probability based approach for allocating draft selections in Australian Rules Football. In A. Bedford, & M. Owens (Eds.), *Proceedings of the Australasian Conference on Mathematics and Computers in Sport (M&CS), 10,* 103-112. Australia

Burton, R. (2013, July 13). *NBA teams set to tank next season for shot at Andrew Wiggins.* Retrieved from http://bleacherreport.com/articles/1706530-nba-teams-set-to-tank-next-season-for-shot-at-andrew-wiggins

Gold, A. (2010). NHL draft order based on mathematical elimination. *Journal of Quantitative Analytics in Sports, 6(1),* 6. doi: 10.2202/1559-0410.1203

Kertes, T. (2003, February). Rebuilding on the fly. *Basketball Digest*, p. 46. Retrieved from RMIT Library database.

Michael, P. (2013, March 12). *NBA Debate: To Tank or Not to Tank?* Retrieved from http://sports.yahoo.com/news/nba-debate-tank-not-tank-183100761--nba.html

Morris, C. (1977). The most important points in tennis. In Ladany, S.P. and Machol, R.E. (Eds.), *Optimal strategies in sport.* North-Holland Publishing Company: Amsterdam. 131-140.

Spencer, J. (2013, August 5). *Is Tanking the Only Successful Path to Rebuilding an NBA Team?* Retrieved from http://bleacherreport.com/articles/1728199-is-tanking-the-only-successful-path-to-rebuilding-an-nba-team

Stefani, R. & Clarke, S. (1992). Predictions and home advantage for Australian Rules football. *Journal of Applied Statistics. 19,* 251-261

Taylor, B. & Trogdon, J. (2002). Losing to win: Tournament incentives in the National Basketball

Association. *Journal of Labour Economics, 20*, 23-41

Tuck, G. N., & Whitten, A. R. (2012). Who's driving the tank? A dynamic non-equilibrium simulation model to test alternative amateur draft systems for major sporting leagues. *Proceedings from the Australasian Conference of Mathematics and Computers in Sports (M&CS), 11*, 8-19. Melbourne, Australia

Walters, C., & Williams, T. (2012, March). *To Tank or Not To Tank? Evidence from the NBA.* Paper presented at the 2012 MIT Sloan Sports Analytics Conference, Boston, USA.

Ziller, T. (2013, July 3). *Why 'rebuilding' in the NBA is not the same thing as 'tanking'.* Retrieved from http://www.sbnation.com/nba/2013/7/3/4489712/nba-free-agency-2013-rebuilding-tanking

RECENT ADVANCES IN VIDEO TRACKING TECHNOLOGY

Stuart Morgan

[a] *Performance Research, Australian Institute of Sport*

Abstract
It is said that a picture tells a thousand words. At an average of 25 frames per second, and given the ubiquitous nature of affordable and portable video cameras, video in sport should therefore be able to tell an incredible amount about performance. Yet, in spite of the fact that a number of commercial providers have employed video-based ball and player tracking solutions for several years, the use of video tracking technologies remains in its infancy in Australia, and globally for most non-professional sports. Furthermore, tracking players and balls in large stadia, equipped with sophisticated camera arrays, represents arguably the simplest environment in which to track moving objects. Even so, continuous tracking of athletes and balls in team sports remains a significant challenge.

Further to that, in applied sports science, video sources are frequently suboptimal as they may be from hand-held cameras in competition, where other objects may occlude athletes, or visibility may be poor. Dealing with noisy or challenging video data sets is a current area of focus in the broader field of computer vision, and our work makes use of state of the art methods for localizing athlete positions "in the wild". Our ambitious aim is to leverage useful performance profiling data from a wide range of unconstrained data sources. In collaboration with a range of university partners, we have developed a suite of computer vision solutions for observing and measuring features of sport performance that are progressively becoming more robust, adaptable, mobile, timely, and ultimately informative.

In this presentation, we will demonstrate the evolution in our video tracking methods from simple player position estimates using arrays of fixed cameras, to detecting swimmer stroke rates using hand-held and unconstrained video sources, as well as some emerging themes in action recognition and performance profiling.

SINGLE-SENSOR CLASSIFICATION OF SPORTING MOVEMENTS THROUGH FEATURE EXTRACTION USING R

Casey Josman [a], Ritu Gupta [a], Dan Wundersitz [b] and Sam Robertson [b]

[a] Department of Mathematics and Statistics, Curtin University, Perth, Western Australia
[b] Centre for Exercise & Sports Sciences, School of Exercise & Nutrition Sciences, Deakin University,
caseyjosman@gmail.com

Abstract
Current methods for the classification of human movement rely on data collected from multiple sensors placed on different areas of the subject's body, such as the wrist or hip. These methodologies, whilst proven to be accurate, are impractical in real world sporting applications as such placement of common wearable sensors is cumbersome and can introduce additional risk of injury. To overcome this we introduce a classification system based on data obtained from a single sensor worn between the shoulder blades. Tri-axial accelerometer and gyroscope data was collected from 76 participants at a frequency of 100 Hz. Each participant performed a total of eight distinct movements along a circuit, with brief pauses taken between movements to aid movement distinction. In this paper we use the statistical package R to explore a variety of extractable features which may be used in the classification process and evaluate them on their efficacy by using a combination of ANOVA and Lasso. Some of the features tested included; time domain features such as amplitude maxima and minima, as well as frequency domain features such as bandwidth and spectral density which are extracted by applying the Fourier Fast Transform (FFT). The selected features were then extracted from a sample of the data which were pre-processed using 0.5 second, 1 second, and 1.5 second sliding windows respectively, and then classified by using algorithms such as Random Forest, Support Vector Machines and Logistic Model Tree. The results of these classifications were then compared, on accuracy of classification and computation time. We will present the basis for our classification model including selection criteria for feature extraction as well as the results generated from said classifications.

Keywords: Sport, Movement Classification, Tri-Axial Accelerometer, Feature Extraction, Fourier Fast Transform, Random Forest, Logistic Regression Tree, Support Vector Machine

1. INTRODUCTION

Objective measurement of human movement is essential for understanding the physical and technical demands related to sports performance (Aughey and Falloon, 2010). It is also important in evaluating the effectiveness of training programs designed to increase sports performance as well as those targeting both the prevention and rehabilitation of injury (Neville et al., 2010). Fundamental to furthering this understanding is the need to accurately collect specific information relating to the type, intensity and frequency of movements performed (Carling et al., 2009). Consequently, techniques for undertaking movement analysis in sports have improved substantially in recent years.

A reason at least partially responsible for these improvements relates to the considerable developments that have occurred in wearable tracking device technologies. Wearable tracking devices that integrate multiple sensors (global positioning system (GPS), heart rate, accelerometer, gyroscope and magnetometer) into a single, versatile unit worn on the body are now readily available (Carling et al., 2009). To date, the majority of research has focused on the GPS sensors contained within these devices and their ability to measure basic components of human movement, such as speed, distance travelled, and the number of high-intensity efforts (i.e., accelerations) (Cummins et al., 2013). However, more recently it has been shown that a more detailed analysis of human movement can be obtained using the accelerometer sensor

(Ermes et al., 2008). Specifically, different types of movements can be classified and distinguished based on their accelerometer features.

Previous research have positioned these devices on different areas of the body, including the wrist (Long et al., 2009), upper back (Mitchell et al., 2013), chest (Ermes et al., 2008), shin (Muscillo et al., 2010), hip (Jeong et al., 2007), waist (Mathie et al., 2004), lower back (Bonomi et al., 2009) and feet (Zhang et al., 2003). However, in the majority of contact-based team sports, the upper-back is the only appropriate location an accelerometer can be positioned (when contained within a wearable tracking device). Specifically, a device worn on other parts of the body may have the potential to cause injury to the user or indeed other participants.

Mitchell et al. (2013) recently proposed a method using a single accelerometer contained within a smartphone worn on the upper-back, with the aim of identifying seven different sporting movements (stationary, walking, jogging, sprinting, hitting a ball, standing tackle, dribbling a ball). In their study, an overall movement classification success rate of 75% was achieved using classification approaches that included Support Vector Machine, Logistic Model Tree, and range of Neural Network/Optimization type classifier.

However, with the aim of achieving higher accuracy rates, multiple inputs (i.e., both accelerometer and gyroscope) have also been considered in the literature (rather than a single accelerometer input alone). As the data acquired through the gyroscope provides essential information pertaining to the position of the body during human movement, it is not surprising to see both inputs combined to good effect previously Leutheuser et al., (2013).

A range of different analysis approaches have also been used previously in accelerometer studies, with varying levels of success. Three analysis approaches of particular interest in this study are i) Logistic Model Tree (LMT), ii) Random Forest (RF) and iii) Support Vector Machine (SVM). Logistic Model Trees is a commonly used classification algorithm, which performs competitively with other classifiers and is easier to interpret (Landwehr et al., 2005). The LMT combines two complementary classification techniques: tree induction and linear regression (Hornik et al., 2009). Random Forest is a classification algorithm, which in its application grows multiple classification trees and builds upon them until each tree is at its largest (Breiman and Cutler, 2001). The RF has various useful features including high efficiency with large data sets, built in ensemble classifiers, and an inability to overfit models (Breiman and Cutler, 2001). Support Vector Machine is a classification algorithm, which attempts to find the best separating vector between two groups within a set of descriptors (Bennett and Bredensteiner, 2000). For classification of data with more than two groups the original problem is split into multiple binary problems which are then classified and compared, with the problem having the most votes per instance being assigned as the classifier (Meyer et al., 2014).

The aim of this study was to determine whether data obtained from a wearable tracking device (specifically, gyroscope and accelerometer) can be used to identify team sport-related movements. Guided by practical considerations and current literature herein we focus on classifying difficult movements, explore classification methods that have been successful, consider mix of time and frequency domain features with varying window length and implement a simple ANOVA based methodology for filtering features.

2. METHODS

i. Participants

Seventy-six (*n*=76) recreationally active, healthy male participants (age 24.4 ± 3.3 years; height 181.8 ± 7.5 m; mass 77.4 ± 11.6 kg; mean ± SD) were recruited for participation in the study. All participants were regular competitors in one or more contact-based team sport events per week at the time of testing (Singh et al., 2010). All participants gave informed consent following full disclosure of the study protocol and procedures.

ii. Experimental Design

Participants were required to perform a range of movements commonly undertaken in contact-based team sports in a simulated team sport circuit. The research design allowed for the assessment of multiple team-sport specific movements in a confined space. Details of the circuit design are presented in the next sub-section.

During each trial participants wore a single, wearable tracking device (Minimax S4, Catapult Innovations, Australia), which contained a 100 Hz tri-axial accelerometer, gyroscope, and other devices not utilised for this study. The device weighed 67 grams, was 88 × 50 × 19 mm in dimension and was worn in a tightly fitted manufacturer supplied harness with the units located below the neck, in-line with the spine (superior to the scapulae). Each participant completed the circuit 6 times, with only data collected from each the third trial used in this study

iii. Simulated Team Sport Circuit

The simulated team sport circuit involved a modified version of the circuit developed by Singh et al. (2010). Each circuit included three counter-movement jumps, an eight meter jog, an eight meter agility section, two jumps for distance, a 10 m sprint, seven meters of walking, and a tackle bag to be taken to ground with maximum force. After each movement finished with the participant standing stationary for one second before commencing the next (i.e., three counter-movement jumps were performed in a row then a one second pause occurred). All movements were restricted within an optimal 8 × 8 m capture volume. Each individual circuit took approximately 45 seconds to complete, allowing 15 seconds to rest before the next circuit (on 1 minute) with six circuits completed in total (n = 456). All participants performed an active warm-up prior to commencing the full protocol, which involved five minutes of jogging followed by six laps of the circuit, during which time the experimenter explained all requirements of the circuit.

iv. Data Processing

The data gathered comprised of the accelerometer (3 axes and the resultant vector) and gyroscope (3 axes) readings for the duration of the circuit. For each of the eight movements of interest (counter movement jump (CMJ), change of direction (COD), jog, run and jump, sprint, stationary, tackle, and walk), the corresponding data was extracted and processed to generate features of interest. Figure 1 below gives an overview of both the feature extraction and classification process.

Features were extracted from the data using three different window lengths of 0.5, 1, and 1.5 seconds respectively, each with a 50% overlap. The window lengths were chosen in such a way as to capture the peak force of most activities whilst also being long enough to accurately capture a descriptive segment of each activity. Further, the use of a 50% overlap has been proven successful in previous movement classification studies (Bao and Intille, 2004).

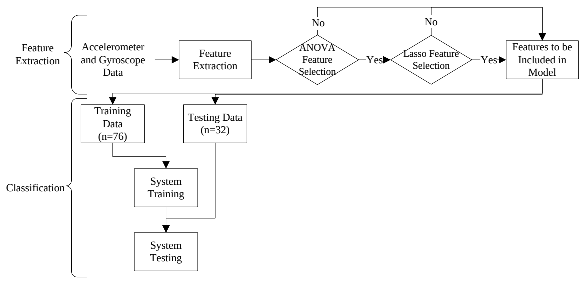


Figure 1: Overview of the feature extraction and classification process.

Features were extracted from the data using three different window lengths of 0.5, 1, and 1.5 seconds respectively, each with a 50% overlap. The window lengths were chosen in such a way as to capture the peak force of most activities whilst also being long enough to accurately capture a descriptive segment of each activity. Further, the use of a 50% overlap has been proven successful in previous movement classification studies (Bao and Intille, 2004).

For each variable associated with the accelerometer (3 axes and the resultant vector) and gyroscope (3 axes), a total of 59 features were calculated (7 total inputs). Those features being:

- Minimum amplitude.
- Maximum amplitude.
- Mean amplitude.
- Variance of amplitude.
- Spectral centroid.
- Bandwidth.
- Energy for each sensor (accelerometer and gyroscope).
- Percentiles (.25, 0.75, interquartile range [IQR])

The minimum, maximum, mean, and variance of the amplitude provide important descriptors of the input variables time domain and thus there were obtained for each input (Leutheuser et al., 2013). The spectral density and bandwidth provided via the use of the FFT represent important descriptors relating to the central mass and frequency domain of the input variables (Leutheuser et al., 2013). Additionally an energy feature was calculated for both the accelerometer and gyroscope (Leutheuser et al., 2013), as well as percentiles and IQR for each input variable (Liu et al., 2012). The energy feature is defined as

$$E = \frac{\left(\frac{\sum_{i=1}^{3} a_i{}^2}{3}\right)}{n}$$

Where a_i are the axes corresponding to either the accelerometer or gyroscope and n is the number of observations per axis.

iv. Analysis

The three classification algorithms (LMT, RF and SVM) were employed to classify the seven movements of interest. However as both computational and data collection burdens have been considered, analysis was conducted in two phases.

In phase one, a data collection burden was assessed. This was achieved by, for each of the three window lengths extracting the features as follows

- In case one all variables were considered (59 features).
- In case two only the accelerometer and resultant vector variables were considered (33 features).
- In case three only the accelerometer variables were considered (26 features).

In phase two a data processing burden was assessed. This was achieved by, for each of the three window lengths, extracting features as follows

- In case one, all 59 features were considered.
- In case two, features were reduced to 42 by using ANOVA.
- In case three, features were reduced to 37 (0.5 second window) and 38 (1 second and 1.5 second windows) using a combination of ANOVA and lasso regression.

Under phase 2, features with significant results for ANOVA at 5% level of significance across classification groups were retained for classification purpose. Next under case three, all features which that were accepted through the case two were passed for screening under lasso regression. Under this screening a feature was selected based on the combinations of (Mallow's Cp (Cp), residual sum of squares (RSS), and coefficient of determination (R^2))

For each subject group (n=76) from the computed set of feature data a single activity was randomly chosen (with equivalent probability) and assigned to the classification training set. A random sample (with equivalent probability) of 32 activities was then taken from the remaining set of feature data and assigned to the classification testing set. The above process was repeated 10 times. Each model was then validated with classification accuracy defined as percent of correctly classified cases.

3. RESULTS

Results were obtained using an analysis routine written in the statistical package R (R Core Team, 2013) which makes use of the following packages; e1071 (Meyer et al., 2014), lars (Hastie and Efron, 2013), randomForest (Liaw and Wiener, 2002), and RWeka (Hornik et al., 2009; Witten and Frank, 2005).

i. Data Collection Burden

Table 1 presents the classification accuracies and standard deviations for all three variable and window combinations. Throughout all classification iterations LMT greatly outperforms both RF and SVM classifiers, obtaining classification accuracies over 85%.

ii. Processing Burden

Table 2 presents the processing times for all three variable and window combinations for a given subject. The computational times are for an Intel® Core™ i7-2670QM CPU with 8 GB RAM. From this it can be seen that the processing time (extraction and classification) for a feature reduced model reduced using ANOVA is approximately 15% faster than classification of the full model, while the

Table 1: Mean (SD) accuracy of classifiers for each of the input variable variations and window lengths after 10-fold cross-validation.

Input	Window	0.5	1	1.5
Accelerometer, Gyroscope, and Resultant Vector	RF	0.28 (0.14)	0.47 (0.12)	0.28 (0.13)
	LMT	**0.90** (0.12)	**0.90** (0.05)	**0.88** (0.14)
	SVM	0.39 (0.18)	0.55 (0.12)	0.41 (0.12)
Accelerometer and Resultant Vector	RF	0.22 (0.11)	0.26 (0.14)	0.25 (0.16)
	LMT	**0.87** (0.13)	**0.88** (0.13)	**0.88** (0.12)
	SVM	0.41 (0.13)	0.46 (0.13)	0.46 (0.10)
Accelerometer Only	RF	0.26 (0.15)	0.25 (0.13)	0.29 (0.17)
	LMT	**0.85** (0.16)	**0.88** (0.13)	**0.85** (0.13)
	SVM	0.33 (0.09)	0.38 (0.13)	0.31 (0.11)

Highest mean accuracy per variable variation and window length in bold combined ANOVA and lasso based feature reduction model is between 1% and 3% slower than the pure ANOVA model. Table 2 presents the classification accuracies and standard deviations for all three model selection methods and window combinations. Once again throughout all classification iterations LMT greatly outperforms both RF and SVM classifiers, obtaining classification accuracies over 85%.

Table 2: Mean (SD) accuracy and [Processing Time (in seconds)] of classifiers for each of the model selection variations and window lengths after 10-fold cross-validation.

	Window	0.5	1	1.5
Input	Extraction	[236.25]	[230.98]	[228.87]
Accelerometer, Gyroscope, and Resultant Vector	RF	0.28 (0.14) [7.54]	0.47 (0.12) [7.58]	0.28 (0.13) [7.72]
	LMT	**0.90** (0.12) [53.47]	**0.90** (0.05) [53.66]	**0.88** (0.14) [56.66]
	SVM	0.39 (0.18) [104.55]	0.55 (0.12) [102.84]	0.41 (0.12) [105.98]
Accelerometer and Resultant Vector	RF	0.28 (0.12) [6.6]	0.28 (0.14) [7.47]	0.29 (0.14) [6.92]
	LMT	**0.86** (0.13) [42.51]	**0.85** (0.13) [44.94]	**0.85** (0.14) [44.38]
	SVM	0.48 (0.13) [77.14]	0.43 (0.13) [80.55]	0.46 (0.12) [81.46]
Accelerometer Only	RF	0.26 (0.11) [6.95]	0.28 (0.15) [6.83]	0.22 (0.10) [7.08]
	LMT	**0.86** (0.15) [44.16]	**0.86** (0.12) [45.21]	**0.85** (0.14) [46.71]
	SVM	0.46 (0.16) [77.57]	0.42 (0.15) [76.71]	0.42 (0.14) [83.15]

Highest mean accuracy per variable variation and window length in bold

4. DISCUSSION

The aim of this study was to determine whether data obtained from a commercially available accelerometer and gyroscope could be used to identify team sport related movements. Optimal results were obtained using LMT method and window length of 0.5 and 1 sec, with an overall accuracy of 90%. Largely the classification inaccuracy resulted from cases where Jog was being classified as Run&Jump or Sprint and Tackle being classified as COD or Sprint.

For reduced burden of data collection, using only accelerometer and resultant or accelerometer alone, the accuracy rate drops down to 87 and 85 % respectively. Such a model would require approximately 230 sec of time for feature extractions and further 54 seconds for classification of movements using LMT model.

Comparable results are obtained in literature with much larger volumes of data accumulation and smaller number of classification groups. Nathan et al. (2012) using accelerometer and GPS data gathering over 750000 measurements and achieved accuracy of over 84% for RF and SVM classifiers. Leutheuser et al. (2013) also using the subsequently large dataset and pre-clustering the activities, achieved an accuracy of 87% for SMV method. Mitchell et al. (2013) reported a similar trend in classification with LMT method, on much smaller frequency of data accumulation, being the best performing classifications.

The window lengths selected for feature extraction influence classification accuracy, and processing time. The classification accuracy across the explored window lengths and processing time are not hugely different. Nevertheless the shorter window lengths would be preferred with ability to capture all movement types. For example the average length of time taken to complete a tackle is 3.5 seconds with a peak force being experienced over 0.5 seconds and 40% of the tackle being captured over 1.5 seconds. To this end, our similar classification rates to previous studies using a lesser number of measurements is encouraging.

It can be seen through this research that LMT is highly effective at classifying sporting activity using a single accelerometer and gyroscope with only minimal data gathered.

5. CONCLUSION

In any sporting scenarios sports it is practical to obtain the accelerometer and gyroscope data using a single sensor worn between the shoulder blades. Our results indicate that LMT is highly effective at classifying sporting activities with approximate accuracy of 90% and classification time of 230 seconds.

References

Aughey, R. J., and Falloon, C. (2010). Real-Time Versus Post-Game Gps Data in Team Sports. *Journal of Science and Medicine in Sport*, 13 (3): 348-349.

Bao, L., and Intille, S. (2004). Activity Recognition from User-Annotated Acceleration Data. In *Pervasive Computing*, eds Alois Ferscha and Friedemann Mattern, 1-17. Berlin: Springer.

Bennett, K. P., and Bredensteiner, E. J. (2000). Duality and Geometry in Svm Classifiers *ICML*, pp. 57-64

Bonomi, A. G., Goris, A., Yin, B., and Westerterp, K. R. (2009). Detection of Type, Duration, and Intensity of Physical Activity Using an Accelerometer. *Med Sci Sports Exerc*, 41 (9): 1770-1777.

Breiman, L., and Cutler, A. (2001). Random Forests. Accessed 01/05/2014, http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.

Carling, C., Reilly, T., and Williams, A. M. (2009). *Performance Assessment for Field Sports*. New York: Routledge.

Cummins, C., Orr, R., O'Connor, H., and West, C. (2013). Global Positioning Systems (Gps) and Microtechnology Sensors in Team Sports: A Systematic Review. *Sports medicine*, 43 (10): 1025-1042.

Ermes, M., Parkka, J., Mantyjarvi, J., and Korhonen, I. (2008). Detection of Daily Activities and Sports with Wearable Sensors in Controlled and Uncontrolled Conditions. *IEEE Transactions on Information Technology in Biomedicine*, 12 (1): 20-26.

Hastie, T., and Efron, B. (2013). *Lars: Least Angle Regression, Lasso and Forward Stagewise*.

Hornik, K., Buchta, C., and Zeileis, A. (2009). Open-Source Machine Learning: R Meets Weka. *Computational Statistics*, 24 (2): 225-232. doi: 10.1007/s00180-008-0119-7.

Jeong, D.-U., Kim, S.-J., and Chung, W.-Y. (2007). Classification of Posture and Movement Using a 3-Axis Accelerometer *International Conference on Convergence Information Technology, 2007.*, pp. 837-844: IEEE.

Landwehr, N., Hall, M., and Frank, E. (2005). Logistic Model Trees. *Machine Learning*, 59 (1-2): 161-205.

Leutheuser, H., Schuldhaus, D., and Eskofier, B. M. (2013). Hierarchical, Multi-Sensor Based Classification of Daily Life Activities: Comparison with State-of-the-Art Algorithms Using a Benchmark Dataset. *PLoS One*, 8 (10). doi: http://dx.doi.org/10.1371/journal.pone.0075196.

Liaw, A., and Wiener, M. (2002). Classification and Regression by Randomforest. *R News*, 2 (3): 18-22.

Liu, S., Gao, R. X., John, D., Staudenmayer, J. W., and Freedson, P. S. (2012). Multisensor Data Fusion for Physical Activity Assessment. *IEEE Transactions on Biomedical Engineering*, 59 (3): 687-696.

Long, X., Yin, B., and Aarts, R. M. (2009). Single-Accelerometer-Based Daily Physical Activity Classification *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009. EMBC 2009.*, pp. 6107-6110: IEEE.

Mathie, M., Celler, B. G., Lovell, N. H., and Coster, A. (2004). Classification of Basic Daily Movements Using a Triaxial Accelerometer. *Medical and Biological Engineering and Computing*, 42 (5): 679-687.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2014). *E1071: Misc Functions of the Department of Statistics (E1071), Tu Wien*.

Mitchell, E., Monaghan, D., and O'Connor, N. E. (2013). Classification of Sporting Activities Using Smartphone Accelerometers. *Sensors (Basel, Switzerland)*, 13 (4): 5317-5337. doi: http://dx.doi.org/10.3390/s130405317.

Muscillo, R., Schmid, M., Conforto, S., and D'Alessio, T. (2010). An Adaptive Kalman-Based Bayes Estimation Technique to Classify Locomotor Activities in Young and Elderly Adults through Accelerometers. *Medical engineering & physics*, 32 (8): 849-859.

Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., and Getz, W. M. (2012). Using Tri-Axial Acceleration Data to Identify Behavioral Modes of Free-Ranging Animals: General Concepts and Tools Illustrated for Griffon Vultures. *The Journal of experimental biology*, 215: 986-996. doi: http://dx.doi.org/10.1242/jeb.058602.

Neville, J., Wixted, A., Rowlands, D., and James, D. (2010). Accelerometers: An Underutilized Resource in Sports Monitoring *Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2010*, pp. 287-290: IEEE.

R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Singh, T. K., Guelfi, K. J., Landers, G., Dawson, B., and Bishop, D. (2010). Reliability of a Contact and Non-Contact Simulated Team Game Circuit. *Journal of sports science & medicine*, 9 (4): 638.

Witten, I. H., and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*: Morgan Kaufmann.

Zhang, K., Werner, P., Sun, M., Pi-Sunyer, F. X., and Boozer, C. N. (2003). Measurement of Human Daily Physical Activity. *Obesity research*, 11 (1): 33-40.

WINDOWS OF OPPORTUNITY: USING A SIMULATION MODEL TO EXPLORE THE DURATION OF SUCCESS OF SPORTING TEAMS

G.N. Tuck [a], A.R. Whitten [b]

[a] *CSIRO Computational Informatics and the CSIRO Digital Productivity and Services Flagship, GPO Box 1538 Hobart, Tasmania 7000, Australia*

[b] *Mezo Research, Carlton North, Victoria 3054, Australia*
[a] *Corresponding author: geoff.tuck@csiro.au*

Abstract

The duration over which teams in sporting leagues have an enhanced likelihood of winning a premiership is called the premiership, championship or title window. Defining when a team enters their premiership window is subjective and often based upon current or projected team strength, the recent trajectory of the team (i.e. whether the team is moving up the ladder), and perhaps the coaching staff and club culture. Winning a premiership is a clear indicator of success, but for a team to have entered their premiership window, the definition is much broader and can include years of making finals or playoffs. Periods of success can be fleeting, with a single outstanding year surrounded by years of mediocrity, or can be enduring. Likewise, a lack of achievement can be brief or persist for decades, much to the frustration of clubs and supporters alike. A critical component of success is the ability of a club to acquire a rare grouping of players that, in combination, improve team strength to the point that success can be achieved. Trading aside, clubs obtain new players through an annual draft. One of the more common draft systems is the reverse-order draft, whereby clubs select players in reverse-order to their finishing position. This system can enable chronically poor-performing teams the ability to 'stock-pile' high draft pick players that have the potential to become a future collective of champion players; and open the team's premiership window.

In this paper we use a simulation model, Sports Synthesis, to explore how characteristics of the draft can influence team success and failure. We find that some teams consistently cycle between success and failure, while others can become stuck in mid-ranks, with minimal success over extended periods of time. A critical determinant of the duration of the premiership window is the degree to which clubs are able to resolve the ability of players in the draft, namely the draft choice error.

Keywords: Amateur draft, sporting success, player productivity, simulation model

A HOLE BY HOLE SIMULATION MODEL FOR MEN'S PGA GOLF

Bradley O'Bree[a]&Anthony Bedford [a,b]

[a]*School of Mathematical and Geospatial Sciences, RMIT University*
[b] *Corresponding author: anthony.bedford@rmit.edu.au*

Abstract

Golf is widely regarded as a difficult sport to model.The focus of this paper is to build on previous research in modelling PGA golf tournaments through simulation. The aim is to identify whether there is a use for hole based modelling in tournament simulation. Tournaments have previously been simulated round by round, with player scores sampled from one of multiple round score distributions.Distributions were chosen by chance based on the likelihood a player would have a certain rank following each round dependent on their current tournament score. In this research, we introduce a simple player ratings system, and a means of characterising each hole on any given tournament course.These hole characteristics were used to create a score distribution for each hole. Player ratings were used to create hole score distributions for each player for holes of the same par. Bayesian Inference was employed to combine both these distributions, creating player dependent hole score distributions. Such distributions facilitate hole by hole tournament simulation. The 2014 US Masters tournament was used as a case study to compare the previous round based simulation model with the new holebased simulation model. Analysis of results for predicted final rankings indicated the hole based model was better than the round based model when simulation included the first three tournament rounds. This result was reversed when only the last round was simulated. Findings suggest the benefits of the more computationally complex method of holebased simulation are reduced as the tournament progresses.However, a comparison of simulation outputs across multiple tournaments would be required for this to be concluded.

Keywords: Golf, Bayesian, Simulation, Hole by Hole, PGA

1. INTRODUCTION

1.1 Simulation in Professional Golf

The majority of modelling in professional golf is not simulation based. Most modelling of score and rankings take the form of correlational analysis of longitudinal performance statistics with scoring average (O'Bree& Bedford, 2012). The benefit simulation provides is a method of dynamically modelling what it known to be a difficult and highly complex sport to model. Tournament playing fields typically number 150, and with 18 holes per round and four rounds per tournament, performance modelling needs to account for the ample time available for variations in player performance. Tournament simulation provides a means of measuring the potential for variations in performance and scores, not just in the tournament as a whole, but by round scores and rankings, and potentially hole scores and rankings.

1.1.1 Round based Modelling

Round based models allow player performance to be assessed at the break points in the tournament. It is not unreasonable to think player performance will vary day by day. An issue however in round based modelling is that the current form of the player can only be evaluated three times throughout the tournament. Given the natural difficulty in modelling scores, this is not ideal.

It has been shown that tournament outcome predictions through round based simulation can be completed reasonably well using onlyhistoric round scores and their corresponding rankings (O'Bree& Bedford, 2012). In this research, tournaments were simulated by generating round scores for each player prior to each round. Historic round scores were standardised by the corresponding course par so round scores from courses with different course par scores could be combined, creating the new score variable Par Percentage. To account for a round effect, so as to say control for differences in player performance between say round one and three for example, round score distributions were generated for each using only scores from the corresponding round in previous tournaments. This result was a different score distribution for each round of the tournament. These score distributions were seen to be normally distributed, and due to the independent variable Par Percentage being continuous, were approximated using binomial distributions. This not only simplified computation but made the score distributions discrete.

Each round score distribution was split by some result outcome for the player who achieved the score. For rounds one and two, the score distributions were split by whether or not the player made the cut. For rounds three and four, the split was by whether or not the player was ranked in the final top 10. Multiple distributions for each round were introduced to account for differences in player ability.

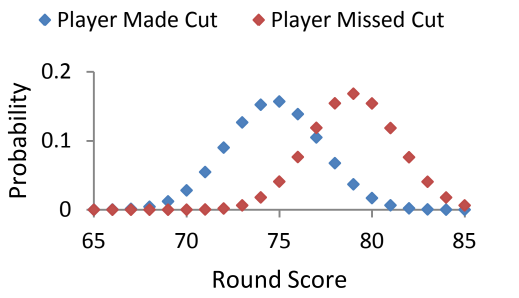


Figure 1.Example of a split round score distribution by making the cut.

Scores were randomly sampled from a round score distribution based on each player's likelihood of either making the cut or finishing in the top 10 from their current tournament score. The probability of each outcome based on current score was calculated from the distribution of historic round scores, and served as the prior distribution throughout simulation. The prior distribution itself was combinedwith individual player outcome likelihoods, taken as the ratio of tournaments where the player made the cut or finished top 10 to the number of tournaments played. The posterior distribution of the likelihood of sampling from either round score distribution was found using Bayes' Rule.

$$P(\theta_x|l) = \frac{P(l|\theta_x)P(\theta_x)}{P(l|\theta_x)P(\theta_x) + P(l|\theta_x')P(\theta_x')}$$

Where round score distribution θ is sampled dependent on current score x with probability $P(\theta_x)$ conditional on player outcome likelihood l.

1.1.2 Hole based Modelling

Hole based modelling has the potential to improve a simulation model's prediction performance because it can measure player performance at a more detailed level. Characteristics of holes can be included in calculations, and depending on detail of inputs, score prediction can be made specific to both the hole and the player. An accurate measure of a player's ability based on hole characteristics has the potential to capture variations in performance better than a round based model because multiple scores are modelled instead of just one.

A method for evaluating player performance (Stern, 2012) at the hole level used a semi-parametric Bradley-Terry type strength estimation model. This research provides a comparison between observed rankings and a measure of underlying player strength. Further, factors that affect performance such as tee times and hole difficulty can be quantified.

1.2 Aims

The aim of this research is to extend the round based model (O'Bree& Bedford, 2012) from previous research into a hole based model. Using the 2014 Masters as a case study, the two models will be used to simulate the tournament with prediction accuracies compared following each round. Analysis of results will indicate whether simulation at the hole level provides benefit in terms of prediction over the round based model. Should the hole based model prove to be of benefit, or at least practical in its current form, model inputs could be expanded to include results from research into factors not yet considered. The inclusion of a better measure of hole difficulty, for example, would like improve the accuracy of prediction by the hole based model.

2. METHODS

2.1 Data and Software

A database of hole by hole round scorecards was utilised to build score distributions and facilitate tournament simulation. A player's scorecard contains the hole and par score for each hole, as well as tournament characteristics such as date, location and year.These scorecards were sourced primarily from sports.yahoo.com/golf, using Visual Basic for Applications (VBA) macro procedures in Microsoft Excel.Player ratings and hole score distributions were generated using data from every official PGA tour event spanning the first tournament of the 2007 tour til the last tournament played before the 2014 Masters.All simulations were carried out using Microsoft Excel 2010. A more comprehensive description of software mechanisms is available in the proceedings of the International Association of Computer Science in Sport (IACSS) 2014 Conference (O'Bree& Bedford, 2014).

2.2Player Ratings

Player ratings were created dependent on hole scores from recently completed rounds. As such, these ratings are seen as a measure of the current form of a player.

For any given complete round, we can determine a ratio score for each player based on hole par. This ratio is taken as the sum of the hole scores divided by the average total across the playing field. The ratio score is calculated for each of par 3, 4 and 5 holes, giving three ratio scores for each player of each completed round. The rating itself is taken as the average of the eight most previously completed rounds. In the event less than eight rounds of data are available for a player, the average of any available ratios are used. It should be noted that this parameter has not been optimised. This was an arbitrary value chosen because it ensured that for any given player at least two recently played tournaments are included in calculation, given most tournaments consist of four rounds.

$$R_{t,i} = \frac{1}{8}\sum_{j=t-8}^{t-1} Ratio_{j,i}$$

Where $R_{t,i}$ is the rating prior to round *t* for holes of par *i*.

2.3 Hole Score Distributions

Conditional score distributions incorporating player ratings and hole characteristics are used to simulate hole scores, and as a result entire tournaments. Bayesian Inference is used to update probability distributions through the introduction of additional information using Bayes' Rule. In this case, we want to update hole score probability distributions based on the ability of a player.

2.3.1 The Prior Distribution

A prior score distribution for each hole was determined by considering two simple characteristics, the hole par and a measure of the hole's difficulty, average score. The hole par is defined as the expected number of shots a professional player should take to reach the green plus two putts. Typically, this makes the hole par dependent on its distance from tee to green. The difficulty of a hole can be gauged by comparing the average score for the hole with its corresponding par. This is of course neglecting any impact from intermittent, dynamic factors during a round such as wind, and assuming the quality of a tournament playing fields are equal across tournaments. Using these characteristics, we can create a probability distribution of observed hole scores based on the hole par and the inferred difficulty of the hole.

$$P(x|\{i,\bar{x}\})_{Prior} \sim \{Par\ i, Average\ Score\ \bar{x}\}$$

For the purposes of this research, the average score of a hole from the most recently completed tournament round is used. The average score is rounded to one significant decimal place.

2.3.2 The Posterior Distribution

The posterior hole score distribution is an updated version of the prior distribution using what is referred to as the likelihood function. The likelihood function simply contains additional information. The aim is to tailor the score distribution to individual players. In the same way a prior score distribution was determined, we can generate score distributions based on the hole par and observed scores from players with equal player ratings.

$$P(x|\{i,r\})_{Likelihood} \sim \{Par\ i, Player\ Rating\ r\}$$

For the purposes of this research, player ratings are rounded to one significant decimal place.Bayes' Rule is used to combine the two hole score distributions, creating the posterior distribution.

In the event no scorecard data was available for a player, a uniform likelihood distribution was assumed, meaning the posterior distribution is the same as the prior distribution, and the hole score distribution reflects the historic distributions of scores for that hole. Such a scenario is not unusual given the playing list for The Masters is created on an invite basis and tournament winners from non-PGA tours are invited to compete. There were four such instances in the 2014 Masters tournament.

Note that no data smoothing has been used when generating any of the hole score distributions.

3. RESULTS

Both the round and hole based models were be used to simulate the 2014 Masters tournament. A total of four simulations were completed for each model, where simulation took place prior to the start of each day's play. Each simulation comprised of 30,000 iterations.

Simulation results focussed on the accuracy of predicted rankings for each player, primarily the final rankings. Due to the cut taking place after round two, simulations of rounds prior to round three also looked at predicting players who will make the cut. Analysis of predictions was in the form of measures of classification, specificity, sensitivity and accuracy, and correlation coefficients. Specificity is a measure of a model's ability to correctly exclude cases, meaning to correctly classifying a player as missing the cut given they do in fact miss the cut. Sensitivity is a measure of a model's ability to correctly include cases, meaning to correctly classifying a player as making the cut when in fact they do.

A total of four models were analysed. The round based and hole based simulation models are labelled as RMITa and RMITb respectively. Also included is a naïve model, which essentially projects current rankings at any point as final rankings, labelled Naïve. The Naïve model provides reference for what can be seen as a measure of volatility in rankings. The final model comes in the form of ranking publicly released outright tournament win market prices from bet365.com, labelled Bet365.Predictions from this model assumed players who have been judged more likely to win the tournament, by having a lower market price, are more likely to have a better final ranking in general. As such, market prices are ranked to infer final rankings. The inclusion of the additional two models is to provide a point of reference with regards to prediction accuracy; particularly, the Naïve model which indicates if any modelling benefit exists from simulation.

3.1 Predicting the Players Making the Cut

Model	Players Correctly Predicted to Make the Cut#	
	Pre Round One	Pre Round Two
Naive	NA	39
Bet365*	31	35
RMITa	20	38
RMITb	30	38

Total of 51 players made the cut

Table 1: Number of Players Correctly Predicted to Make the Cut

Measures of Classification for Players Predicted to Make the Cut								
Measure	Pre Round One				Pre Round Two			
	Naive	Bet365#	RMITa	RMIT2b	Naive	Bet365#	RMITa	RMIT2b
Specificity	NA	0.489	0.326	0.500	0.717	0.320	0.717	0.717
Sensitivity	NA	0.620	0.392	0.588	0.765	0.745	0.745	0.745
Accuracy	NA	0.558	0.361	0.546	0.742	0.597	0.732	0.732
Samples		95	97	97	97	72	97	97

Bet365 Top 51converted from outright win market lines

Table 2: Measures of Players Correctly Predicted to Make the Cut following Round Two

Table 1 displays the number of players correctly predicted to make the cut. Prior to the tournament commencement the Bet365 and RMITb models correctly assigned 31 and 30 players respectively, while RMITa correctly assigned 20 players. As would be expected, all models improved following round one. The Naïve model outperformed the others, however only slightly in the case of both RMIT models. Table 2 displays similar trends in predictive success with classification measures, with both RMIT models outperforming the Bet365 model, and only slightly less successful than the Naïve model. These results indicate it is easier to correctly classify a player as missing the cut than to classify a player as making the cut.

3.2 Predicting the Top 10

In total, 13 players ranked in the final top 10. This is due to six players being tied ranking eighth.

Model	Players Correctly Placed in the final Top 10#	
	Pre Round Three	Pre Round Four
Naive	7	10
Bet365*	8	8
RMITa	6	9
RMITb	6	9

Total of 13 players placed in the final Top 10
* Bet365 Top 10 converted from outright win market lines

Table 3: Number of Players Correctly Predicted to Finish Top 10

Measure	Pre Round Three				Pre Round Four			
	Naive	Bet365#	RMITa	RMITb	Naive	Bet365#	RMITa	RMITb
Specificity	0.816	0.939	0.895	0.895	0.921	0.857	0.974	0.974
Sensitivity	0.538	0.615	0.462	0.462	0.769	0.667	0.692	0.692
Accuracy	0.745	0.848	0.784	0.784	0.882	0.769	0.902	0.902
Samples	51	45	51	51	51	26	51	51

Measures of Classification for Top 10 Placed Players

Bet365 Top 10 converted from outright win market lines

Table 4: Number of players correctly predicted to Place in the Top 10

Table 3 displays the number of players correctly predicted to finish in the top 10. The Bet365 model was the only model not to improve between rounds, with the other models all improving by 3 correct classifications. Note though the smaller sample size in market prices. As was the case when predicting the players who would make the cut, it appears it is easier to identify players who won't finish top 10 than to identify those that will. In this case, measures of classification were the same between RMIT models, despite differences in predicted rankings between the models for the same players.

3.3 Predicting the Final Rankings

Model	Final Rankings Correlation Coefficient		
	Pre Round Two	Pre Round Three	Pre Round Four
Naive	.389*	.454*	.825*
RMITa	.194	.436*	.828*
RMITb	.357*	.496*	.796*

* Correlation is significant at the 0.01 level

Table 5: Spearman Correlation Coefficient for Predicted and Observed Final Rankings

Absolute Error (<=)		0	1	2	3	4	5	6	7	8	9	10
Pre Round Three	Naïve	0.04	0.17	0.21	0.21	0.25	0.29	0.42	0.54	0.58	0.58	0.63
	RMITa	0.08	0.08	0.13	0.13	0.21	0.33	0.38	0.42	0.42	0.46	0.54
	RMITb	0.04	0.13	0.17	0.21	0.25	0.33	0.38	0.46	0.54	0.54	0.58
Pre Round Four	Naïve	0.13	0.29	0.54	0.54	0.58	0.58	0.63	0.67	0.67	0.71	0.88
	RMITa	0.13	0.33	0.50	0.54	0.58	0.67	0.67	0.67	0.67	0.71	0.75
	RMITb	0.13	0.29	0.46	0.50	0.63	0.63	0.63	0.63	0.67	0.71	0.71

Cumulative Proportion of Absolute Errors in Predicted Final Ranking

Table 6: Cumulative proportion of absolute errors in predicted final rankings

Table 5 displays correlation coefficients for observed and predicted final rankings. Correlations between the Naïve and RMITb models are essentially the same throughout simulations.

Table 6 displays absolute errors in predicted final rankings as a proportion of all predictions. The better the predicted rankings across the spectrum of the player list the greater the proportion with smaller absolute errors. Predictions prior to round three saw the Naïve model have the greatest proportion as error increased. Predictions prior to round four saw similar measurements between the three models until the absolute errors reached a value of 10, at which point error are quite large.

4. DISCUSSION

When models were used to determine the players who would make the cut following round two, there were varying results between models when predictions were made at different times.

Prior to round one, no observed scores or ranks can be used, so predicted ranks rely only on measures of form for each player that each model uses. It is no surprise the round model RMITa, which does not include such measures, is much less effective in classifying players. The Bet365 and RMITb models were essentially equal in success. While the Bet365 model uses publicly released market prices which are subject to influence from the public opinion, the similarity in success suggests the player ratings measure utilised by RMITb is a suitable model input.

Predictions in the players who would make the cut following round one have observed scores and rankings available. The Naïve model was most effective in inferring the players who would make the cut, which tends to suggest that a player's current position is a more important factor in determining final result than the player's form. This is further evidenced by the substantial improvement in the round based RMITa model, putting it on level pegging with the hole based RMITb model. Note these results relate to one tournament only, and such a finding needs to be measured in more tournaments to be concluded. The Bet365 model showed improvement but to a lesser degree. One would expect wagering patterns would somewhat reflect the presence of well-placed champion players and past winners over current scores, particularly in the early rounds of a tournament. Again, this may be explained by a tendency to favour some aspect of a player's history or current form over their current tournament position.

When predicting players who would finish with a final rank in the top 10, we saw a drop in classification accuracy for the Bet365 model between pre and post round three calculations. This tends to suggest that a decline in the quality of performance by bigger name players was influential in this tournament. Take for example Adam Scott, the pre-tournament favourite. His final ranking was tied 14th scoring 289 for the tournament. He scored 76 in the third round, four shots more than his next highest round score at 72. A third round score of 72 would have seen his final score be 286 and tied for fifth. A third round score of 75 would have seen him tied eighth. The difference between making and missing the top 10for Scott was one shot, and given his worst score was in the third round, it is reasonable to attribute the drop in classification accuracy to these differences in performances like these.

It is interesting to see the classification measures were the same between the RMITmodels when predicting the final top 10. The similarity suggests the worth of the features of the models are the same – at least following the end of round two. The RMITa model uses the likelihood a player will finish in the top 10 given they made the cut throughout the seven seasons of data used and their current score for simulation inputs. The RMITb model uses a measure of hole difficulty in the most previous average score and a player rating for current form to generate score distributions. Given results were the same, it may be the case that using this particular measure for player ratings does not provide better round score accuracy when holes are evaluated individually. Should this be the case, a measure of current form is no better than an historic measure at providing insight into performance in the last two rounds of the tournament, given inferences from sampling round scores were the same as from sampling hole scores and totalling round scores.

Analysing the absolute errors in predicted final rankings and correlation coefficients with observed final rankings indicated that prior to round three the RMITb model had slightly better predicted rankings, while post round three the RMITa model had better predicted rankings. In each instance, the better of the RMIT models had a better correlation coefficient than the Naïve model for the corresponding round. As absolute error in predicted ranking increased, the model with the greatest proportion of predictions varied. Such findings indicate the while variations in the distribution of absolute errors were present for each model, the difference in the RMIT models was able to tease out an improvement over the Naïve model following round two.

6. FUTURE RESEARCH

Future work will look to confirm findings in this research. Mainly, that current position is a more important factor to consider when simulating tournaments than current player form as the tournament progresses.

6. CONCLUSIONS

The aim of this research was to find out if there was benefit to simulating golf tournaments at the hole level when compared to the round level. The two models analysed performed well overall when compared with two other predictive models. Both models were seen to perform the best at certain times during the tournament, such that there would be value in including both round based and hole based simulation models for tournament simulation weighting results accordingly.

7. ACKNOWLEDGEMENTS

The authors would like to acknowledge the collaborative efforts and guidance provided by Dr James Baglin.

8. REFERENCES

O'Bree, B & Bedford, A, 2012, 'An Application of Bayesian Inference and Simulation in Golf', Proceedings of the eleventh Australasian Conference on Mathematics and Computers in Sport, A Bedford &A J Schembri (Eds), pp.194-199.

O'Bree, B & Bedford, A, 2014, 'Developing a database system for modelling of golf using Microsoft Excel', Proceedings of the 2014 International Association of Computer Science in Sport Conference, I Heazelwood (Eds).

Stern, S, 2012, 'Who really won the 2012 Masters? Ranking player performance and hole difficulty in a medal play golf tournament', Proceedings of the eleventh Australasian Conference on Mathematics and Computers in Sport, A Bedford &A J Schembri (Eds), pp.2-7.

MODELING SERVICE PLACEMENT BY PENETRATION POINTS OF RETURN PLAYER'S ACTION PLANE

O. Kolbinger & M. Lames

Chair for Performance Analysis and Computer Science in Sport, TU München, Germany

Aim of the study

The advent of 3D ball trajectory data in tennis has brought into reach new options of modeling relevant behavior in tennis by mathematical processing of these data. In previous work, placement of services was characterized by their bouncing points in the service box (Loffing et al., 2009). Much more relevant in terms of tennis tactics is the penetration point of the ball trajectory of the service through the action plane of the return player. The location of this point in vertical and lateral direction imposes problems for a return. With 3D ball trajectory data available mathematics for calculating penetration points becomes quite easy.

Methods

A vertical plane three feet in front of the receiver's baseline is assumed to be the action plane of the return. Ball trajectories are obtained by image detection methods. The part of the trajectory after the bounce of the service is given by a three-dimensional cubic polynomial. A MATLAB (The Mathworks, Inc.) procedure was programmed for extracting the y-z-coordinates of the ball's penetration point.

In order to investigate the distribution of tennis serves, a virtual plane with 24 slots (5 in z-direction, 8 in y-direction) was built for each of the four serve categories: deuce court – first serve, ad court – first serve, deuce court – second serve and ad court – second serve.

Data were obtained from 10418 serves of 53 right-handed male players during international tournaments on hard court.

Results

Due to different spin and speed, service bounces in the service box are quite dissimilar to penetration points in return action plane.

The main tactical plans associated with first and second services from deuce and advantage side could be found in hitting vertical and lateral slots in return action plane. Statistically significant differences proved different tactical behaviors in the respective situations.

Conclusions and outlook

This study demonstrates that data provided by new technologies allow for mathematical models more adequately describing behaviors in sport. (Simple) Mathematics allows revealing structures of performance that were formerly not accessible to performance analysis.

There may be still more adequate ways to model the return player's action plane, e.g. in a curve-shaped way. Also, the development of mathematical models for new performance indicators for the placement of services can now be addressed at a new level of evidence.

References

Loffing, F., Hagemann, N. and Strauss, B. (2009), The Serve in Professional Men's Tennis: Effects of Players' Handness, *International Journal of Performance Analysis in Sport, 9*(2), 255-274.

THE SURFACE EFFECT IN ATP TENNIS

Jaclyn Yap, Anthony Bedford* and Michelle Viney

RMIT University, Melbourne
**Anthony.bedford@rmit.edu.au*

Abstract

This paper focuses on the surface effect in men's tennis from both a performance perspective and a wagering angle. Through analysis of data from 2002-2013, we investigate a number of effects distilled by surface. We find that significant variation in both the predictability of outcomes exists across surfaces. We also look at the efficiency of the markets by surface, and how, in light of diminishing over-round, how surface plays a part. We also investigate the volatility of results by surface, controlling for the level of the tournament (grand slam down to ATP250), and the number of times there are in-game shifts in lead, for example, changes in lead (ie. 6-3 to 3-6, and by underdog against favourite).
We also consider the length of matches via games played as a function of potential matches possible (ie in 3 or 5 set tiebreak/non tiebreak). Whilst not necessarily a precise measure of length, we consider this as a reasonable pseudo measure.

Keywords: tennis, surface, ATP.

1. INTRODUCTION

Tennis is one of the few sports where the game can be played on different court surfaces. On the professional circuit players compete on grass, clay, carpet, indoor, acrylic and synthetic hard court. Each court surface has its own characteristics which can have a positive or negative effect on a player's style of play and consequently performance (Clarke and Dyte, 2000). For example, Pete Sampras won fourteen grand slams yet failed to win the French Open.

The most common court surface is hard court. Hard court is known to produce medium to fast courts which the ball tends to bounce more at contact. Players with big serves and forehands tend to perform better on hard court. Clay courts are considered to be slower courts where at contact with the surface the ball tends to bounce up or sit rather than skidding, where a greater number of bad bounces occur. Grass courts suits a serve-and-volley style of play as they are the fastest court to play on. When the ball contacts the grass surface the ball tends to stay low and skid.

Various research has been performed in analysing the effect of court surface as there is an expectation that all players win a higher percentage of serves on grass than other surface (Barnett and Clarke, 2005).

Barnett and Clarke (2005) and Barnett, O'Shaughnessy and Bedford (2011)applied the effects of court surface to calculate the probability of winning a point on serve using overall percentage of points won on serve for that tournament and player's returning and serving statistics for each surface.Del Corral and Prieto-Rodriguez (2010) found that a significant court effect regarding higher-ranked player victories in Wimbledon and Australian women's tournament. Koning (2011) found that performance advantage can be partially attributed to the familiarity with the court surface. McHale and Morton (2011) found that surface, time and ease of win is an important measure to assess player's quality and produce more accurate forecasts. They concluded that clay should be regarded as a separate entity in a forecasting model.

Thus the aim of this research is to investigate the surface effect in men's tennis from both a performance perspective and a wagering angle. Analysing eleven years of data we determine whether an effect of court surface was present, and how it can be considered when modelling ATP tennis.

2 METHODOLOGY

To examine our data, we utilised a number of sources. To assist we utilised both individual player's in-game data and post match results. We also required knowledge of the ATP tournaments, including the surface and points. In this way were able to isolate various facets of the effect. Tennis Insight (tennisinsight.com) was most useful in utilising player based data. Stevegtennis was also used – and has been around since the dawn of the internet. Our analysis methods were simple data extraction methods and cross tabulation. We also utilised two sample z-test of proportions and standard confidence intervals where needed.

A number of terms are used that the reader may not be familiar with, so we define these as follows:
Efficiency – this term is used in terms of a market being efficient, that is, that it predicts to expectation. This is a common term used to evaluate systems. A model's efficiency is used in that way.
Over-round is the amount of additional probability attributed to a market by a bookmaker. For example, if a market is framed as follows:

Roger Federer $1.73
Rafael Nadal $2.00

Then we Over-round $= (1/1.73) + (1/2) - 1 = 0.58 + 0.5 - 1 = .08$ or 8%.

Chi-square goodness of fit tests were also used to evaluate outcomes of results against expectation. Lower chi-square values indicate markets or models that are efficient.

We shall provide you with a variety of tables and figures that tear apart surface against a variety of measures.To be able to achieve this we can present the efficiency by tables.

3 RESULTS

We shall present our look at surface through consideration of a number of factors. We shall firstly consider surface in general for our data set, and how it compares by year.

3.1 Preliminary Look

To begin we looked at 2013 in terms of the service, breaks, aces, games and sets played. It is notable that such data is now available at this micro-level at ATP, WTA, and most challenger tournaments. It is feasible to obtain these statistics very easily via Tennis Insight. This data is navigable back to 2008.

3.1.1 Match Effects by surface: 2013

To consider the recency of performance, we investigated 2013 in detail. We isolated the three major surfaces, grass, clay and hard court. There were1443 clay matches, 1661 hard court and418 grass matches in our set of data played at the ATP level on these surfaces. We found that Clay had clearly the lower quantity of games and sets per match, and the lowest number of tie breaks. Table 1 exhibits the details.

	Grass	Clay	Hard
Sets/Match	2.64	2.43	2.54
Games/Match	26.18	23.07	24.54
TieBreak/Match	0.53	0.33	0.39
TieBreak/Set	0.20	0.13	0.15

Table 1: Rate of Match Statistics by Surface, 2013

These results tend to indicate that Clay may well be a more predictable surface, and Grass less so. More games and sets are played on grass, and the numbers of tie-breaks are significantly higher than the other two surfaces. Inverting the Tie Break statistics gives us a raw estimation of a price for each likelihood, without consideration of over-round, or players, as in Table 2.

	Grass	Clay	Hard
TieBreak/Match	$1.90	$3.10	$2.60
TieBreak/Set	$5	$7.70	$6.70

Table 2: Empirical Price by Surface.

Considering the Service specifically reveals some interesting findings. Whilst we infer that from Table 1 Clay has ashorter match in terms of games, and possibly more predictable, it is certainly not through server dominance. Table 3 shows the service statistics by surface for 2013.

Surface	Grass	Clay	Hard	Sig.
Aces per game	.57	.34	.49	SS
1st Serve In	.625	.615	.596	NS(GvC);SS
1st Serve Win	.730	.680	.713	SS
2nd Serve Win	.512	.499	.502	NS(CvH); SS
Serve Hold	.814	.744	.777	SS
BreakChances per game	.50	.62	.56	SS
Breaks Win per game	.19	.26	.22	SS

Table 3: Service Statistics by Surface for 2013

Test of proportions on all pairings yielded significance (SS) with the exception of 1st serve in for grass and clay. Service breaks are more possible and obtained on clay.

3.1.2 Case Study: Rafael Nadal and Novak Djokovic – 2011.

For interest, we isolated the matches of two of the best players of recent time. We selected 2011 as it was the year when both players had comparable exposure to both grass and clay.

Let us first consider the Clay performance, as shown in Table 4. Considering the match statistics, both players have a similar success ratio, even down to the game level.

Clay	Nadal	Djokovic
Match W/L	92% (24-2)	94% (17-1)
Set W/L	85% (55-10)	83% (38-8)
Gm W/L	63% (386-222)	63% (262-154)
Grass	Nadal	Djokovic
Match W/L	80% (8-2)	100% (7-0)
Set W/L	73% (24-9)	84% (21-4)
Gm W/L	56% (180-139)	61% (142-89)

Table 4: Match statistics by surface

Looking at Table 5, we see the serving power of Djokovic, with his higher rate of aces per game on both surfaces. Notably, clay yields a much lower rate of aces than hard, as expected. No difference exists in double faults. Despite Nadal's renowned dominance on clay, Djokovic holds serve with a higher rate. Grass overall provides higher rates of service holds than Clay.

Clay	Nadal	Djokovic
Aces per Gm	0.19	0.3
DFs per Gm	0.1	0.12
1st Serve %	70.30%	65.70%
1st Serve W%	70.50%	74.00%
2nd Serve W%	56.70%	59.10%
Service Pts W%	**66.40%**	**68.90%**
Service Hold %	**83.40%**	**88.10%**
Grass	Nadal	Djokovic
Aces per Gm	0.43	0.52
DFs per Gm	0.1	0.14
1st Serve %	69.50%	68.50%
1st Serve W%	74.10%	77.40%
2nd Serve W%	59.00%	59.00%
Service Pts W%	**69.50%**	**71.60%**
Service Hold %	**86.60%**	**90.60%**

Table 5: Service statistics by surface. 2011.

When considering Nadal's return game, we see Nadal's strength come through. Table 6 isolates this data.

Clay	Nadal	Djokovic
Opp. 1st Serve %	61.40%	63.40%
1st Return W%	38.60%	36.00%
2nd Return W%	61.30%	57.70%
Return Pts W%	**47.30%**	**43.90%**
BPs Won per Gm	0.44	0.37
BP Chances per Gm	0.9	0.81
Break Pt W%	**48.50%**	**46.00%**
Opp Hold %	**56.10%**	**62.90%**
Grass	Nadal	Djokovic
Opp. 1st Serve %	64.20%	63.30%
1st Return W%	27.60%	32.90%
2nd Return W%	55.60%	55.30%
Return Pts W%	**37.60%**	**41.10%**
BPs Won per Gm	0.25	0.31
BP Chances per Gm	0.54	0.66
Break Pt W%	**46.40%**	**47.30%**
Opp Hold %	**75.20%**	**68.80%**

Table 6: Return statistics by surface, 2011.

Nadal's ability to win on return on clay is his edge, with dominance on all statistics (over Djokovic) clear. He wins close to 50%.

3.2 Longitudinal Approach

We now consider the data set from 2003-2013 inclusive. Figure 1 outlines the sample size.

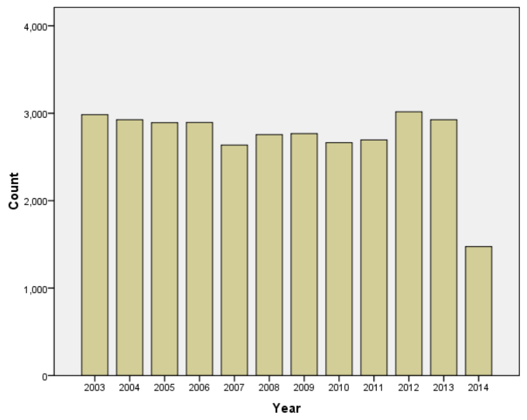


Figure 1: Matches in data set by year

Our set of data reveal a fairly uniform amount of matches for this period, excluding 2014, and we now look at how the market modelled this period.

3.2.1 Over-round

Firstly, we look at the over-round for this time frame. Figure 2 below provides the 95% C.I. for the mean over-round per year.

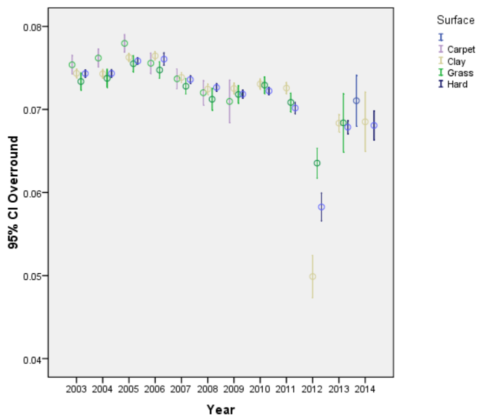


Figure 2: 95%CI for over-round by year by surface

A notable trend is the decline in over round, suggesting a more accurate framing of the market, or a higher turnover allowing reduced margins for the bookmaker. Notably grass generally yielded the lowest over round.

3.2.2 Upsets

We now consider the favourite and underdog winning performance by year. Table 7 shows the wins by year by status. Here, a tied start price is defined as a Tie, and Favourite the player winning at a shorter start price than their opponent.

Year	Fave	Tie	UDog
2003	66.6%	5.4%	28.1%
2004	66.1%	3.7%	30.3%
2005	69.1%	2.7%	28.1%
2006	67.6%	2.7%	29.7%
2007	70.6%	2.1%	27.3%
2008	69.4%	1.6%	29.0%
2009	69.8%	1.4%	28.8%
2010	70.1%	1.4%	28.5%
2011	72.5%	1.3%	26.3%
2012	69.3%	1.0%	29.7%
2013	68.6%	1.3%	30.1%

Table 7: Wins by status by year

What is evident in an improvement in the favourite winning, with the win % bubbling around 69% post 2004. There is a consistent reduction in matches starting at even money, and surprisingly 2013 yielding the highest win rate for underdogs since 2004.

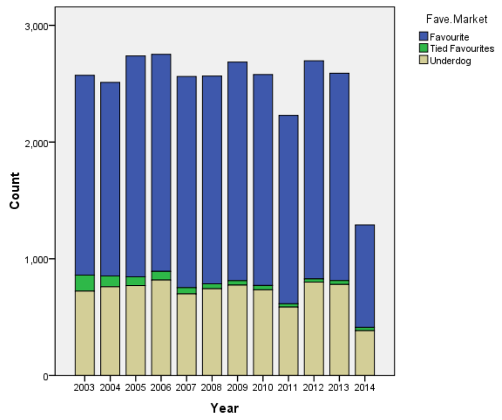


Figure 3: Favourite/Tied/Underdogs by year

Figure 3 shows the counts to provide some scale to these outcomes, with the evident reduction in even money start prices.

We now wish to consider if the surface has an impact on underdogs winning. We found for 2013 that clay may yield more unpredictable results.

Surface	Fave	Tie	Underdog
Clay	69.06%	2.06%	28.88%
Grass	71.55%	2.06%	26.93%
Hard	68.30%	2.28%	28.77%

Table 8: Overall Favourite results by Surface

Table 8 shows that both hard and clay yield more unpredictable results over the entire period than Grass.Splitting by surface by year, we now consider each of the winners.

Firstly, we look at the pre-match favourite by year by surface.

Year	Grass	Hard	Clay
2003	68.00%	66.00%	68.10%
2004	70.00%	64.40%	68.40%
2005	72.80%	68.60%	68.60%
2006	70.40%	67.80%	66.10%
2007	74.10%	71.00%	69.30%
2008	72.10%	68.70%	70.50%
2009	72.80%	70.60%	67.50%
2010	68.10%	69.10%	72.80%
2011	74.90%	73.20%	70.40%
2012	70.50%	68.80%	69.80%
2013	69.50%	68.30%	68.80%

Table 9: Favourites win % by year by surface

Table 9 above shows that Grass yielded the lowest number of proportional winners in only one year, 2010. The squares indicate the lowest proportional wins by the favourite for that year. Hard and Clay have five each. Grass yields the most predictable results from the public price perspective.

3.2 Winning from behind

We considered all surfaces by comeback – that is – when a player was down one set within a match. There is little variation in the comeback factor – with Clay the lowest and Hard the highest. So, one in five matches yields a comeback.

	Wins from<=1 set down		
Surface	Win	Loss	%
Clay	1858	7826	19.2%
Grass	666	2683	19.9%
Hard	3013	12253	19.7%
All	5768	23845	19.5%

Table 10: wins from at least one set down

Of greater interest is the underdog comeback. So we consider how often an underdog wins when going behind. Interestingly, the favourite wins 17.4% of the time they go behind, yet the underdog wins more often, recording a win 24.8% of the time they lead then go behind.

3.3 Market Efficiency

Through banding of the winners pre-match probability of winning, we can ascertain possible inefficiencies in the market. We see in Table 11 below all prices in our set by surface – i.e. not differentiated by year.

Banding-surface	Clay	Grass	Hard
0.00-0.05	2.1%	9.4%	0.0%
0.05-0.10	4.6%	4.6%	4.4%
0.10-0.15	6.8%	10.3%	9.5%
0.15-0.20	16.0%	11.0%	13.7%
0.20-0.25	20.8%	17.5%	17.9%
0.25-0.30	21.8%	24.4%	25.0%
0.30-0.35	29.2%	26.4%	30.1%
0.35-0.40	32.7%	31.3%	34.4%
0.40-0.45	40.5%	38.8%	40.0%
0.45-0.50	43.3%	41.5%	43.6%
0.50-0.55	48.2%	48.3%	48.4%
0.55-0.60	52.8%	53.7%	53.6%
0.60-0.65	58.1%	58.6%	57.3%
0.65-0.70	62.9%	65.9%	61.6%
0.70-0.75	69.3%	71.0%	68.2%
0.75-0.80	73.7%	74.0%	73.0%
0.80-0.85	80.0%	76.1%	78.3%
0.85-0.90	80.1%	84.9%	83.6%
0.90-0.95	90.1%	88.8%	87.7%
0.95-1.00	95.4%	95.0%	96.1%

Table 11: Banding by Surface

If we classify each cell's value as below (L) inside (O) or above (H) we note that Clay yields the most within band results, with Hard court the least. This is seen in Table 12 below. Across all analyses we see typically under-target strike rates and this can be attributed to, in part, the unadjusted allocation of the over-round. So to address this, we utilise the equal-distribution methodology to consider the over-round results.

Banding-surface	Clay	Grass	Hard
0-0.05	O	H	L
0.05-0.10	L	L	L
0.10-0.15	L	O	L
0.15-0.20	O	L	L
0.20-0.25	O	L	L
0.25-0.3	L	L	L
0.30-0.35	L	L	O
0.35-0.40	L	L	L
0.4-0.45	O	L	L
0.45-0.5	L	L	L
0.5-0.55	L	L	L
0.55-0.6	L	L	L
0.6-0.65	L	L	L
0.65-0.7	L	O	L
0.7-0.75	L	O	L
0.75-0.8	L	L	L
0.8-0.85	L	L	L
0.85-0.9	L	L	L
0.9-0.95	O	L	L
0.95-1	O	L	O

Table 12: Classification of Banding

If we were to use the equal distribution approach to redistribution of over-round (ie addition half of the over round to each player), we would see this improve to the results as seem in Table 13 below.

Banding-surface	Clay	Grass	Hard
0-0.05	H	H	O
0.05-0.10	O	O	O
0.10-0.15	L	O	O
0.15-0.20	O	L	O
0.20-0.25	O	O	O
0.25-0.3	L	O	O
0.30-0.35	O	L	O
0.35-0.40	O	L	O
0.4-0.45	O	O	O
0.45-0.5	O	L	O
0.5-0.55	O	O	O
0.55-0.6	O	O	O
0.6-0.65	O	O	O
0.65-0.7	O	O	L
0.7-0.75	O	O	O
0.75-0.8	O	O	O
0.8-0.85	O	L	O
0.85-0.9	L	O	O
0.9-0.95	O	O	O
0.95-1	O	O	O

Table 13: Inclusion of over-round (equal distribution)

The redistribution of over-round now indicates a much different picture of efficiency, as seen in Table 13. Grass yields the least within band results; Hard court the most. However, overall the picture is far better than that shown in Table*.

3.5 Predictability in differing tournament levels

An interesting comparison is that of the favourites in ATP tournaments of lesser value. We define that as Non-Masters/Grand Slams- as at the time of publishing this equates to ATO 500 or less.

	Clay		Grass		Hard	
	M+	NM	M+	NM	M+	NM
Fave	71.7	67.5	76.5	67.0	70.0	67.7
Tie	2.0	2.2	1.8	2.3	2.0	2.6
Underdog	26.3	30.3	21.7	30.7	28.0	29.7

Table 14: Masters and higher vs lower level tournaments

What is clear is that the market yields a greater deal of predictability in Masters/Grand Slams results than smaller events. There is a systematically reduced predictability about the non-masters events.

4 CONCLUSION

In this paper, we found a number of prevailing factors. Firstly, clay has the lowest amount of games per match, and the lowest amount of tie-breaks. Grass has significantly higher rates. Clay yields the lowest service success. It also has the highest amount of break opportunity. Clear example of this was seen with Nadal's clay form. As found in other studies (Schembri& Bedford, 2011), over-round is diminishing in tennis. NO real clear trend by surface exists, with some variability in the 2012 data.

Around 69% of the time the favourite wins, however once this is broken down we find that this rises for Masters and Slams – dramatically for Grass.

The comeback factor showed no surface bias. Finally, we noted that the post adjustment of probability for over round yielded a quite efficient market, especially above the 0.50 mark.

Clearly, consideration of surface is needed for all modelling, and even more important is the consideration of the level of tournament. The large difference in predictability at masters v non masters begs for further investigation and integration into models.

Acknowledgments

We thank tennisinsight.com for the provision of player based data.

Reference

Barnett, T., & Clarke, S. R. (2005). Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*, *16*(2), 113-120.

Barnett, T., O'Shaughnessy, D., & Bedford, A. (2011). Predicting a tennis match in progress for sports multimedia. *OR insight*, *24*(3), 190-204.

Clarke, S. R., &Dyte, D. (2000). Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research*, *7*(6), 585-594.

Del Corral, J., & Prieto-Rodriguez, J. (2010). Are differences in ranks good predictors for Grand Slam tennis matches?. *International Journal of Forecasting*, *26*(3), 551-563.

Koning, R. H. (2011). Home advantage in professional tennis. *Journal of sports sciences*, *29*(1), 19-27.

McHale, I., & Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, *27*(2), 619-630.

Schembri, A., & Bedford, A. (2011). Comparing Market Efficiency with Traditional and Non-Traditional Ratings Systems in ATP Tennis. *New England Symposium on Statistics in Sports.*

PREPROCESSING TENNIS DOUBLES RESULTS TO PRODUCE INDIVIDUAL RATINGS.

Stephen R Clarke

Swinburne University of Technology
sandkclarke@hotmail.com

Abstract

A previous paper investigated using regression analysis and .exponential smoothing to create individual player ratings from doubles tennis results. The exponential smoothing method required both players' ratings to be altered equally depending on whether the pair performed better or worse than predicted. This paper describes a simple algebraic transformation which makes allowance for the relative strength of a player's partners. This results in a single measure of how a player performed against his direct opponent. This can then be used in a ratings program as if they were singles results.

Keywords: tennis, exponential smoothing, player ratings

1. INTRODUCTION

At the second MCS conference I made a case for an adjustive tennis rating system based on margin of victory that rated players from a beginner to the world's Number 1 (Clarke 1994). That paper suggested an exponential smoothing system, similar to the Elo rating system used in chess (Elo 1978). Unfortunately, while some work has been done, little progress has been made in implementing any practical system. Some papers have looked at how the method performs within a restricted range of player abilities. (Bedford and Clarke 2000) compared the performance of an exponential smoothing method with the ATP ranking in predicting the winner of each match in major tournaments. They found the method performed slightly better than the ATP rating for both predicting match winners and tournament seeding. (Clarke 2009; Clarke 2011) used the season's results of a suburban doubles competition to rate all 52 players in the section. An exponential smoothing method adjusts the rating up or down depending on how the result compares with that predicted by the ratings. Here the average of the two players' ratings was compared with that of their opponents', and each player adjusted by the same amount. The final ratings obtained had a high correlation with those obtained using regression analysis to optimise prediction of the set result. This demonstrated the method could be applied successfully to obtain individual ratings from doubles results. This is necessary for a rating system that hopes to rate all standards, as doubles play is the most common format in non-elite tennis.

2 PREPROCESSING

One disadvantage of the system investigated above is that both partners are rewarded/penalised to the same extent for a good/poor team performance. The adjustment to a player's rating depends as much on his partner's performance as on his own. However in many competitions it is possible to transform a player's performance to allow for that of his partner. This paper suggests a method of removing the 'partner effect' from a player's score, to give a proper comparison of how he has gone against his direct opponent.

Consider a simple team competition consisting of three players A, B &C. They play 3 sets, A&B, A&C, and B&C against the similar pairings of the opposition. Usually players measure how many sets or games they are up on the day. However the performance of A over the day includes a contribution from B & C. Thus A could do well because his partners are much better than their respective opponents. But this advantage can be measured in the third set and subtracted from A's results. Similar adjustments for B and C produces a measure of each player's relative strength compared to his direct opponent, his contribution to the team winning margin.

Thus a player's adjusted games up
= players actual games up – (partners winning margin)
= Actual games up – (team winning margin – players actual games up)
= 2*players actual games up – team winning margin.

For example suppose the winning margins are

A&B	+6
A&C	+2
B&C	-2.

Which results in the team winning by 6.

Then A, B &C are respectively +8, +4, 0 up and C could convince himself he has come out even against his opposite number. Note 8+4+0=12, twice the team margin as each set is counted twice. However the adjustment above gives 10, 2, -6 showing that C has clearly been outplayed by his opponent. Note that 10+2-6 = 6, the team winning margin. Note effectively we have solved the simultaneous equations a+b=6, a+c = 2, b+c = -2 to give a=5, b= 1, c= -3. Since each player plays two sets the contributions we calculated are twice these.

The competition considered in (Clarke 2011) is a doubles competition is for teams of 4 men. The team is entered on the card in some order 1, 2, 3 and 4 (usually, but not always, of decreasing ability) and a match consists of each of the 6 possible pairs (1&2, 3&4, 1&3, 2&4, 1&4, 2&3) playing the corresponding pair from the other team in a first to 8 (tiebreaker at 7 all) set. Thus each player plays 3 sets against his direct opponent, each of these including a different one of his teammates and his teammate's direct opponent.

Again, the principle is quite simple. A player's total margin on his 3 sets includes a contribution from each of his other team members. But the total margin in the other 3 sets in which he is not involved includes 2 contributions from each of his team members. We therefore subtract half that margin from the player's margin to obtain his individual performance measure.

Algebraically, we assume each set margin is made up of a contribution from each player's relative standard compared to his direct opponent. Thus if Player 1 is 4 games better than his opponent, and player 2 is 3 games worse the estimated result is4-3 = +1 or 8-7. If these 4 unknowns are a, b, c, & d then player 1's total up on the day (call it player 1s total margin) will be estimated by (a+b)+(a+c)+(a+d) = 3a+ b + c+ d. To get the a we need to remove the b + c+ d. But the other 3 sets in which player 1 did not participate give a measure of (b+c) + (b+d) + (c+d) = 2 (b+c+d). Thus to get 3a, which is player 1's contribution over the 3 sets we take his total margin minus half the total margins in the sets he did not play. Thus

Adjusted margin = (players total margin) – ½(other 3 sets total margin)

This is probably the easiest formula for an individual player to use and understand. For someone doing the calculation for all team members the following is probably better.

Adjusted margin
=(players total margin) – ½(other 3 sets total margin)
= (player's total margin) – ½(team margin – players total margin)
= ½(3*player's margin - team margin)

Basically 3* players margin is 3*(3a+ b + c+ d) = 9a+ 3b+3c+3d and if you take away the team margin = 3(a+ b + c+ d) you are left with 6a which is twice player 1s measure of superiority (ie 3a).

So

Adjusted margin = ½(3*player's margin - team margin)

This is quite easy to implement on a spreadsheet if you already calculate individual players up or down on the day.

Note this is the same as that obtained by fitting the above additive model to the six set results using least squares. ie the estimates given by this simple arithmetic calculation are the same as would be obtained using regression analysis to minimise the errors in predicting the set margin using the algebraic sum of the players rating differentials.

2 APPLICATION

Example: Scores are 82, 48, 82, 68, 68, 68 for a team margin of +2. Players are up 10, 2, 0, -8 respectively (these add to 4, twice the team margin). Applying the above formula we get adjusted figures of 14, 2, -1, -13 (these add up to +2, the team margin).

Note player 1 goes up as his partners are significantly worse that their opponents (they lost 48, 68, 68). Player 4 goes down for the opposite reason (his partners won 82, 82, 68). Player 2 is unchanged as the sets he was not involved in (48, 82, 68) came out level – as a group his partners were the equal of their opponents.

Note this is not saying Player 1 is better or played better than Player 4. It just says that after removing a partner effect, on the day player 1 was 14 games better than his opposition number 1, whereas player 4 was 13 games worse than his direct opponent. For all we know the opposition team number 4 may have been far better than their number 1. It also assumes that players play to the same standard from set to set – I wish that were true. But over a season one would expect those random fluctuations to even out.

As a second example consider the scores 48, 48, 28, 68, 28, 84 for a team margin of -18. Raw scores for players are -16, -2, -6, -12 (total -36) and it looks like all players have failed to hold their own. But adjusted scores of -15, 6, 0, -9 (total -18) show that player 3 has held his own, player 2 is up, and the bad loss is all down to players 1 and 4 being badly beaten by their respective opponents.

Player	Sets played	Games up	Adjusted Games up
EC	33	36	21.5
GM	24	-7	-19
SC	27	44	34
GB	33	43	45
JZ	27	-6	-26.5

Table 1. Individual players season results.

These adjusted figures can be used simply as a better measure of a players performance against his direct opponent, or as input to further analysis. For example they could be used as input to a regression analysis. The first 4 columns of table 1 are from (Clarke 2009) and show the season's totals for the raw figures for each team member for the data as presented in (Clarke 2011). The final 2 columns show the figures obtained by applying the adjustment suggested above. Remembering that on average the adjusted figures will be half the raw figures (since each set is counted once rather than twice), the dominance of GB and lack of same for GM and JZ is highlighted by the adjusted figures.

3 EXPONENTIAL SMOOTHING

Clarke(2011) used a full regression analysis on the complete association set results (276 observations) which allowed for the individual players playing and a home advantage to obtain player ratings. This was used as the 'gold standard' and the ratings obtained compared to those using an exponential smoothing method.

Exponential smoothing operates by adjusting player's ratings depending on a comparison of the predicted and actual set result. As the previous smoothing method used the 276 set results, the adjustment had to be shared equally between the two partners. Nevertheless the correlation between the final season ratings and the 'gold standard' using a smoothing constant of 0.2 was 0.76 when initial ratings were set to zero, and 0.81 when initial ratings were based on position first played.

Here we can use the adjusted figures to smooth each player's ratings directly against his opponent. The data set reduces by a third to 184 (since we are using the days 4 adjusted games up rather than 6 set results). The final ratings give a correlation of 0.83 and 0.84 with the regression ratings depending on whether initial ratings are set to zero or depend on position first played, both slightly higher than those obtained using the set results.

4 CONCLUSION

In many doubles competitions it is possible to make a simple adjustment that removes the partner effect and produces a proper comparison between a player and his direct opponent. This could then be treated in a similar manner to singles results in any rating system. Here we show it gives reasonable results in an exponential smoothing system.

As well as giving better results, this method has the advantage that not all 4 players involved in the match have to be in the rating system. The effect of each player is isolated against his direct opponent, so a player and his direct opponent's rating can be adjusted even if ratings for the other players are not available. This might be important in the implementation of any universal rating scheme, where it would be expected, particularly in lower standard matches, that many players may not have a rating.

References

Elo, A. E. (1978). *The rating of chess players, past and present.* London, Batsford.

Clarke, S. R. (1994). An adjustive rating system for tennis and squash players. *Second Conference on Mathematics and Computers in Sport*. N. de Mestre. Gold Coast, Qld., Bond University: 43-50.

Bedford, A. B. and S. R. Clarke (2000). A comparison of the ATP ratings with a smoothing method for match prediction. *Proceedings of the Fifth Australian conference on Mathematics and Computers in Sport*. G. Cohen and T. Langtry. Sydney, University of Technology Sydney: 43-51.

Clarke, S. R. (2009). Using team doubles competition results to rate non-elite tennis players. *Second International Conference on Mathematics in Sport.* R. Koning and P. Scarf. Groningham, The Netherlands, The Institute of Mathematics and its Applications: 42-46.

Clarke, S. R. (2011). Rating non-elite tennnis players using team doubles competition results. *Journal of the Operational Research Society* 62(7): 1385-1390.

ALTERING THE PROBABILITY OF WINNING A POINT ON SERVE FOR THE MOST AND LEAST IMPORTANT POINT IN TENNIS

Michelle Viney and Anthony Bedford

School of Mathematical and Geospatial Sciences
RMIT University, Melbourne, Australia
Corresponding author: michelleanne.viney@rmit.edu.au

Abstract
Whether the probability of winning a point on serve should be considered constant or variable remains debated. Newton and Aslam (2006) addressed this question by altering the probability on the most and least important point in the game by an arbitrary value of twenty percent. In conclusion, they found that varying probabilities does not dramatically alter the probabilities predicted from a pure iid model. A limitation of Newton and Aslam's research is they selected an arbitrary value of twenty percent without any verification, where this paper will provide analysis on the effects of selecting various values. This research extends on Newton and Aslam's work by evaluating the effectiveness of varying the winning probability of a point on serve. The results indicate that the degree of change impacts the likelihood of winning the set and/or match.

Keywords: Importance, point probability, tennis.

1.INTRODUCTION

Tennis is a popular sport as a spectator, recreational activity and also for sport modellers. Mathematically speaking, the game of tennis is an attractive sport to model as there are only two players to take into consideration. Various models have been developed to forecast the winner of the match. Typically, these models assume the probability of winning a point on serve is independent and identical distributed (iid), where the probability remains constant for the entire match (Barnett and Clarke 2002, Carter and Crews 1974, Fischer 1980 and Schutz 1970). Much debate surrounds the issue of assuming iid (Jackson and Mosurski 1997 and Klaassen and Magnus 2001). Klaassen and Magnus (2001) found that winning the previous point has a positive effect on winning the current point, and those "important" points are more difficult to win for the server than less important points.

Various studies have analysed the effect of altering the probability of winning a point on serve (see Morris 1977, O'Donoghue 2001, Pollard and Noble 2002, 2004 and Viney, Bedford and Kondo, 2013). Overall, the results found that expending more physical and mental effort on important points and relaxing on unimportant points increases the chances of winning the game. Morris (1977) developed an approach to calculate the importance levels at a point, game, set and match level using the Markov Chain model. Importance of a point is defined as the difference in the probability of winning and losing the current point (Morris, 1977). Morris outlined that increasing the probability of winning a point on serve from 0.60 to 0.61 on the important points and decreasing from 0.60 to 0.59 on the unimportant points resulted in increasing the probability of winning the service game by 0.0075.

Extending this concept, Newton and Aslam (2006) applied a Monte Carlo simulator to determine whether increasing or decreasing the probability of winning a point on serve on the most and least important point by twenty percent was more effective than the iid Markov Chain model. Newton and Aslam found that this approach increased the probability of winning greater than the pure iid model (Markov Chain) as the most important point occurs more frequently. The overall conclusion of this research was that varying probabilities does not dramatically alter the probabilities predicted from a pure iid model. While the iid assumption is not perfect, in practise, it appears to perform quite well and the inclusion of non-iid models may introduce unanticipated problems (Newton and Aslam, 2006).

A limitation of Newton and Aslams research is they selected an arbitrary value of twenty percent without any verification. Thus, the aim of this research is to empirically validate Newton and Aslam's findings and to perform an extensive analysis on the effect of altering the probability of winning a point on serve at various intensities.

2. METHODS

The main objective of this paper is to validate and perform an extension analysis on Newton and Aslam's (2006) research. Newton and Aslam developed a tennis Monte Carlo simulator, where a random generating value is produced to determine the winner of each point in the match. Newton and Aslam altered the probability of winning a point on serve using the concept of point importance. How important the current point is in relations to a point, game, set and match level can be calculated using the Markov Chain model, which was developed by Morris (1977). Point importance is the difference between the probability of the server winning and losing the current point (a,b), which is represented as follows:

$$I^{Point}(a,b) = P^{Game}(a+1,b) - P^{Game}(a,b+1) \qquad (1)$$

To alter the probabilities, they adjusted Player A's probability to win a point on serve by an arbitrary value of twenty percent on the most important point in the game, 30-40 and decreased by the same value on the least important point, 40-0. Once the point has concluded the probability returned to the initial value for the next point. It's important to note that the least important point only occurs once in the game, though the most important point can occur an infinite number of times as, for example 30-40 is equivalent to 40-Ad.

To replicate Newton and Aslam's work a Monte Carlo simulator was developed using @Risk, an add-on for Microsoft Excel (Viney, Kondo and Bedford, 2012). To determine whether adjusting the probabilities by twenty percent is the most effective approach, various arbitrary values were implemented. Probabilities were adjusted by quantities of 0.05, 0.10, 0.15, 0.20, 0.25 and 0.30.

To determine the effectiveness of all approaches it was compared against the Markov Chain model, where the probability of winning a point on serve remains constant for the entire set, due to the assumption of independence and identical distribution (iid). The Markov Chain model is typically used to predict outcomes of tennis matches before and during the match. Barnett, Brown & Clarke (2006) applied the properties of the Markov Chain to derive a recursive formula to calculate the probability of winning from any state within a game, set and match.

In terms of a game, the probability of Player A winning the game at point score (a,b) is given by:

$$P_A^{Game}(a,b) = pP_A^{Game}(a+1,b) + (1-p)P_A^{Game}(a,b+1) \qquad (2)$$

with boundary conditions:

$$P_A^{Game}(a,b) = 1 \text{ if } a = 4, b \leq 2$$
$$P_A^{Game}(a,b) = 0 \text{ if } b = 4, a \leq 2$$
$$P_A^{Game}(3,3) = \frac{p^2}{p^2+(1-p)^2},$$

where p is the probability of Player A winning a point on serve which remains constant for the entire match.

In similar fashion, the probability of either player winning a tiebreak set can be calculated using a Markov chain. Let $P_A^{T.Set}(c,d)$ represent the conditional probability of Player A winning a tiebreak set from game score (c,d) when Player A is serving. It is expressed as followed:

$$P_A^{T.Set}(c,d) = P_A^{Game}P_B^{T.Set}(c+1,d) + (1-P_A^{Game})P_B^{T.Set}(c,d+1) \qquad (3)$$

with boundary conditions:

$$P_A^{T.Set}(c,d) = 1 \text{ if } c = 6, 0 \leq d \leq 4, c = 7, d = 5$$
$$P_A^{T.Set}(c,d) = 0 \text{ if } d = 6, 0 \leq c \leq 4, c = 5, d = 7$$
$$P_A^{T.Set}(6,6) = P_A^{Tie-break},$$

where P_A^{Game} represents the probability of Player A winning a game on serve and $P_A^{Tie-break}$ represents the probability of Player A winning a tiebreak game. For a detailed explanation, see Barnett, Brown & Clarke (2006).

3. RESULTS

To determine whether adjusting the probabilities by twenty percent on the most and least important point is the most effective approach, various values were implemented. Probabilities were adjusted by quantities of 0.05, 0.10, 0.15, 0.20, 0.25 and 0.30. Adjusting Player A's probability of winning a point on serve at (30-40) is as follows:

$$\widehat{p_A(2,3)} = p_A(2,3) + \theta \qquad (4)$$

where $\theta = 0.05, 0.10, 0.15, 0.20, 0.25$ and 0.30.

Adjusting Player A's probability of winning a point on serve at (40-0) is as follows:

$$\widehat{p_A(3,0)} = p_A(3,0) - \theta \quad (5)$$

where θ =0.05, 0.10, 0.15, 0.20, 0.25 and 0.30. Traditional point scoring format of 0, 15, 30, 40 and game is represented as 0, 1, 2, 3, and 4, respectively.

For example, when Player A has reached the score-line 40-0 and theta is 0.30, the player will decrease in performance by thirty percent. While at the score-line 30-40 and 40-Ad, Player A will increase in performance by thirty percent. Theta did not exceed thirty percent, as the player should already be playing at a substantial level to have a chance to win the match and increasing more than thirty percent would be deemed an unreachable target.

Table 1 displays a comparison of all approaches in altering the probability of winning a point on serve in a game situation. The theta value selected determines the amount of change in probability. The higher the theta value the larger the difference of change.

Table 1: A comparison between all approaches of the adjusting the probability of winning a point on serve in a game

Score	Markov	Theta Value					
		0.05	0.10	0.15	0.20	0.25	0.30
0-0	0.60	0.60	0.60	0.60	0.60	0.60	0.60
15-0	0.60	0.60	0.60	0.60	0.60	0.60	0.60
30-0	0.60	0.60	0.60	0.60	0.60	0.60	0.60
40-0	0.60	0.55	0.50	0.45	0.40	0.35	0.30
40-15	0.60	0.60	0.60	0.60	0.60	0.60	0.60
40-30	0.60	0.60	0.60	0.60	0.60	0.60	0.60
40-40	0.60	0.60	0.60	0.60	0.60	0.60	0.60
40-Ad	0.60	0.65	0.70	0.75	0.80	0.85	0.90

Analysis was initially performed to outline the effect of altering the probability of winning a point on serve by a particular value. Morris (1977) analysed the effect and gains achieved on increasing effort on different points in a game.

The effect of Player A increasing effort on the most important point in the game at 30-40 in respect to winning a service game, when Player A's probability of winning a point on serve is 0.60, is as follows:

$$\widehat{P_A^{Game}(0,0)} = P_A^{Game}(0,0) + \left(\varepsilon \times N_{(2,3)} \times I_{(2,3)}\right) \quad (6)$$

where ε is the effort contributed where $\varepsilon \in [0,0.30]$ at every 0.01 interval, $N_{(2,3)}$ is the expected number of times (2,3) is played in one game, $N_{(2,3)}$ =0.443 and $I_{(2,3)}$ is the importance of the point in the game at (2,3), $I_{(2,3)} = 0.6923$. Player A's probability of winning a point on serve = 0.60.

The effect of decreasing effort on the least important point in the game at 40-0 in respect to winning a service game, when Player A's probability of winning a point on serve is 0.60, is as follows:

$$\widehat{P_A^{Game}(0,0)} = P_A^{Game}(0,0) - \left(\varepsilon \times N_{(3,0)} \times I_{(3,0)}\right) \quad (7)$$

where ε is the effort contributed where $\varepsilon \in [0,0.30]$ at every 0.01 interval, $N_{(3,0)}$ is the expected number of times (3,0) is played in one game, $N_{(3,0)} = 0.216$, and $I_{(3,0)}$ is the importance of the point in the game at (3,0), $I_{(3,0)} = 0.0492$. Player A's probability of winning a point on serve = 0.60.

For a full explanation of the process and how $N_{(a,b)}$ is calculated refer to Morris (1977). To determine the effect of altering the probability of winning a point on serve, the initial probability was set at 0.60 and ε, the effort contributed, ranged from zero to 0.30, at every 0.01 interval. Figure 1 displays the effect on Player A winning a standard service game when either increasing and/or decreasing their probability at the most and least important point. The combine line in Figure 1 represents the effect of adjusting both the probability on the most and least important point. Player A's starting probability was 0.60 which results in the probability of winning a service game of 0.7357. As represented in Figure 1, as the effort increases on the most important point, the greater the probability of winning a game in comparison to decreasing effort on the least important point. For example, if a player decides to increase his performance by twenty percent on the most important point at 30-40, it results in increasing their probability of winning the game from 0.7357 to 0.7970. Whereas decreasing by the same amount alters the probability of winning the game from 0.7357 to 0.7336. Overall, with every 0.01 incremental change of probability to the most/least important point results in an increased probability of winning a standard service game by 0.003, when the initial probability was 0.60.

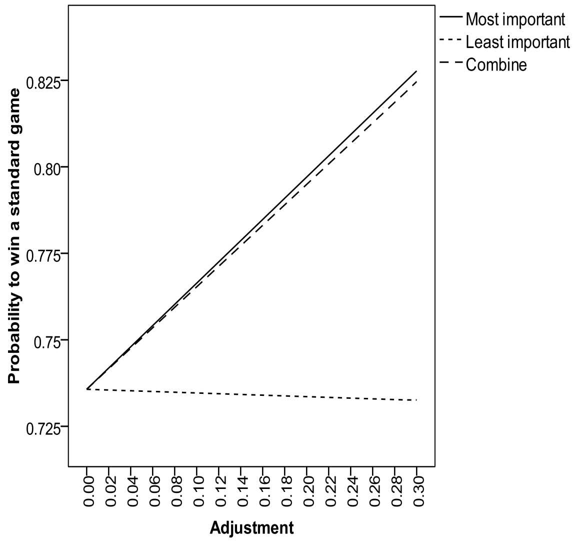


Figure 1: The effect of adjusting the probability of winning a point on serve at the most, least and combining both most and least important point in a standard service game.

Simulation analysis was performed to determine the effect of adjusting the probability of winning a point on serve in respect to winning the first game and set. Ten thousand simulations were performed at the commencement of the set with Player A serving first. Ten thousand simulations were chosen as Viney, Kondo and Bedford (2012) found that simulating greater than ten thousand points, decreased the error rate. Both players' starting probability to win a point on serve was 0.60, where Player A's probability is altered on the most and least important point in the game by 0.05, 0.10, 0.15, 0.20, 0.25 and 0.30. To compare the effects of altering different numerical values, all simulations were linked together, so the same simulated value is applied for accurate analysis. Table 2 displays the difference in adjusting the probability of winning a point on serve in respect to the most and least important point. Player A is assumed to win the first service game and set. Table 2 represents a linear relationship whereas theta increases the probability of Player A winning their service game and set both increase. For example, when adjusting the probabilities by twenty percent, Player A's probability to win their service game and set was 0.804 and 0.602, respectively.

Table 2: Comparing the difference after adjusting the probability to win a point on serve in respect to Player A winning their first service game and the set.

	Markov	0.05	0.10	0.15	0.20	0.25	0.30
Wins 1st service game	0.733	0.749	0.768	0.783	0.804	0.825	0.852
Wins set	0.499	0.518	0.541	0.570	0.602	0.638	0.678

To determine how each approach performs when a player enhances or deteriorates in performance, streaking analysis was carried out. Streaking analysis is a concept applied in the simulator to alter Player A's performance by a particular value. Player A's performance was altered to both increase and decrease at a level of two, four, six, eight and ten percent. This streaking effect can be applied to any phase of the match, but for this research streaking was applied for the entire match.

Player A's increasing streaking effect is represented as follows:

$$P_A = P_A + \sigma \qquad (8)$$

and Player A's decreasing streaking effect is as follows:

$$P_A = P_A - \sigma \qquad (9)$$

where $= 0.02, 0.04, 0.06, 0.08$ and 0.10 .

To analyse the effect of streaking, both player's initial probability of winning a point on serve was 0.60 and only Player A's probability was adjusted. This approach also implements the process of altering the probabilities on the most and least important point by various theta values. Table 3 displays the streaking effect for Player A, when the initial probability of winning a point on serve is 0.60 with a streaking effect of four percent and Theta at 0.05. For example, at the least important point in the game, 40-0, regardless which direction we alter performance, the probability is decreased by five percent due to the point being the least important in the match. Thus, the updated probability to win a point on serve is 0.59 and 0.51 for increasing and decreasing performance at four percent, respectively.

Table 3: An example of streaking analysis with a streaking effect of four percent with theta increasing by 0.05 when the probability of winning a point on serve is 0.60.

Score	Normal	Increase	Decrease
0-0	0.60	0.64	0.56
15-0	0.60	0.64	0.56
30-0	0.60	0.64	0.56
40-0	0.60	0.59	0.51
40-15	0.60	0.64	0.56

Table 4 displays the streaking effect on all approaches with different streaking intensities. It represents a linear relationship where as you increase the streaking value the larger the difference recorded. For example, when altering the probability by twenty percent with a streaking effect of ten percent, the decrease difference is 0.344 as opposed to 0.245 for the increased difference. Comparing between increasing and decreasing performance, the results shows that overall the streaking effect difference is greater when decreasing Player A's performance. It's interesting to note that as we change from a decrease to an increase in performance, the ranking of difference changes. For example, when Player A's performance decreased, Markov and *0.05* recorded the lowest difference, whilst as the performance increased greater differences emerged.

Table 4: Streaking effect on increasing and decreasing performance

Approaches	Streaking effect									
	-0.10	-0.08	-0.06	-0.04	-0.02	0.02	0.04	0.06	0.08	0.10
Markov	0.320	0.257	0.201	0.150	0.071	0.071	0.132	0.191	0.242	0.290
5	0.318	0.257	0.203	0.147	0.069	0.070	0.129	0.190	0.239	0.273
10	0.327	0.262	0.202	0.144	0.071	0.070	0.131	0.183	0.237	0.262
15	0.333	0.264	0.204	0.147	0.070	0.074	0.133	0.180	0.224	0.256
20	0.344	0.275	0.207	0.149	0.067	0.071	0.123	0.171	0.215	0.245
25	0.354	0.275	0.210	0.138	0.062	0.066	0.121	0.164	0.199	0.236
30	0.359	0.278	0.211	0.139	0.058	0.059	0.112	0.149	0.180	0.213

4. DISCUSSION

To obtain a full understanding of how all approaches forecast the outcome of a match, a case study was applied. Ten thousand simulations were performed at the completion of each service game in the match. The case study chosen was when Mikhail Kukushkin was the underdog and defeated Andreas Seppi 6-1, 1-6, 6-4, in the semi-finals at the Kremlin Cup in Moscow. Figure 3 compares all approaches to win the match at all game scores in the match. Overall all approaches follow the same trend with the Markov Chain model for the entire duration of the match. For a deeper analysis on the relationship of all approaches in this case study, we performed analysis one set at a time.

At the commencement of the first set, Kukushkun started as the underdog and won the first set 6-1. Kukushkun only lost three points on serve, whereas Seppi lost eleven. All theta values followed the same trend as the Markov model (Figure 3). The absolute average difference from the Markov model for the entire set was 0.004 to 0.014 for Theta *0.05* to *0.30*, respectively. Kukushkin won both of his first two service games of the match without losing a point. After the completion of Kukushkin's second service game, the Markov model recorded the largest difference of the probability of winning the match at the commencement of the match of a difference of 0.05 and Theta *0.05* recorded the second largest difference at a value of 0.04. Thus in this scenario, no cases accurately reflected Kukushkin's current performance from pre match predictions. At 2-1, Kukushkin broke Seppi's serve and Theta *0.20*, *0.25* and *0.30* recorded the maximum difference from the Markov model at a value of 0.021, 0.030 and 0.031, respectively. At the completion of the set all models increased Kukushkin's probability of winning the match by an average of 0.281, with an average increase difference from the Markov model of 0.009.

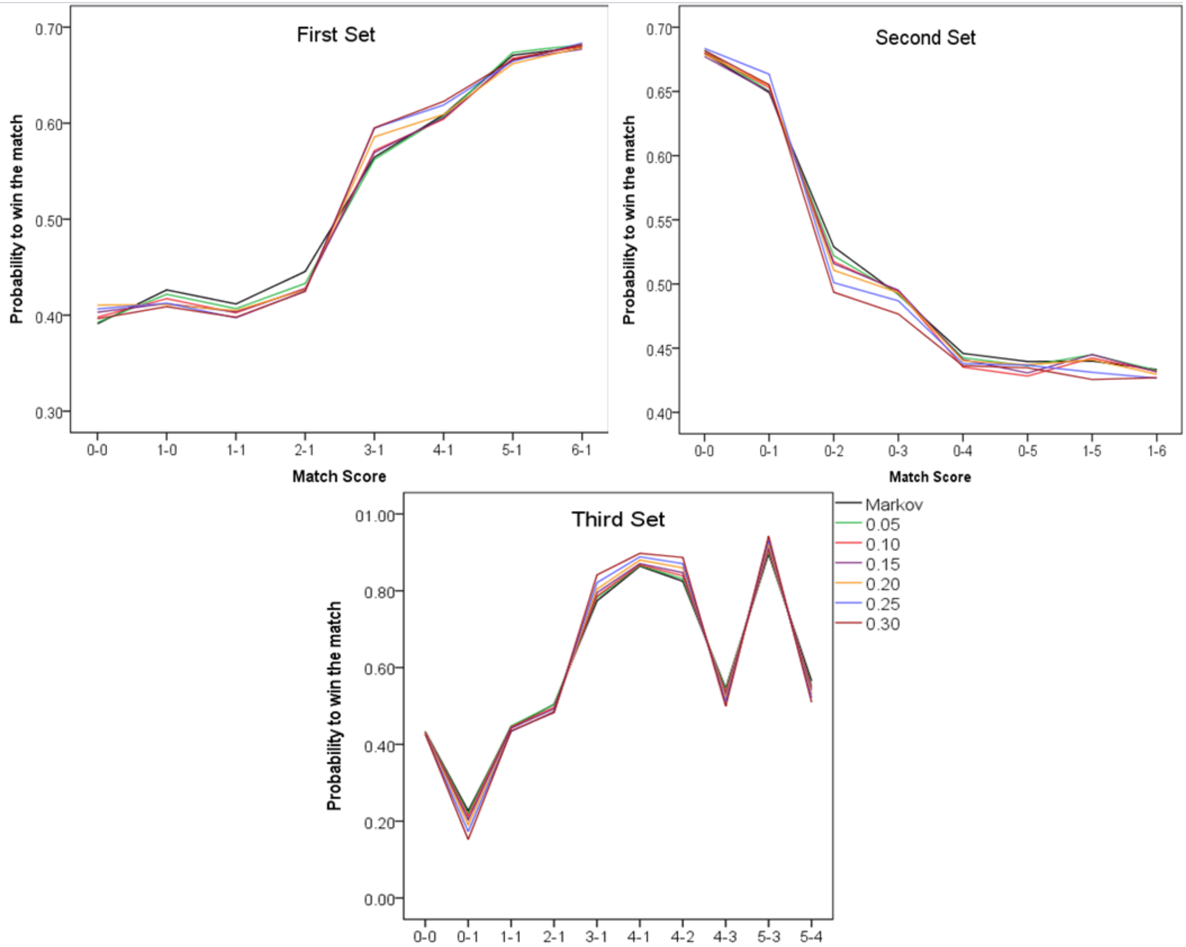


Figure 3: Kukushkin's (Player A) probability to win the match at all game scores in the match.

In respect to the second set, Seppi's form improved and only lost three service points on serve, while Kukushkin lost thirteen points. Seppi won the set 6-1 and broke Kukushkin's serve twice, at 0-1 and 0-3. At the first break of serve, the larger the theta value the larger the difference of change in Kukushkin's probability of winning the match, with a difference of 0.008 to 0.041 from Theta *0.05* to *0.30*. Although, at the second break of serve, all approaches decreased Kukushkin's probability of winning the match by on average 0.05.

In the third and final set there were a total of seven breaks of serves, where Seppi lost four and Kukushkin lost three. In all breaks of service, the larger the theta value the larger the difference in change of Kukushkin's probability of winning the match. For example, Kukushkin lost his first service game at the commencement of the third set, where the Markov and Theta *0.05* alter the probability by 0.207 and 0.213, respectively. While, Theta *0.25* and *0.30* altered the probability by 0.254 and 0.275, respectively.

In this case study all theta approaches altered Kukushkin's probability of winning the match in respect to breaks of service games, though no approach took into consideration how the service games were won. For example, a server winning their service game losing zero points is performing at a greater level than a server winning their service game from deuce. In respect to this case study Kukushkin won two consecutive service games at the eighth and fourth deuce. This indicates that he was struggling to hold serve and consequently lost his next service game to love.

In conclusion, this case study demonstrated that when a break of service occurs, the larger the theta value the larger the difference of change in the probability of winning the match. However, no approach displayed an accurate indication of how the player was performing in the match. For example, if a player was winning their service game by losing zero or five points.

This research has provided valuable insight into the effect of changing the probability of winning a point on serve. Future research should determine the effect of updating the probability of winning a point on serve after every service point. This approach aims to increase the accuracy of forecasting to obtain a precise indication of how players are performing on the day. Although no optimal value can be selected from this research, this research displays a thorough understanding of the effect of altering the probability of winning a point on serve for a range of values.

Reference

Barnett, T.J, Brown, A. & Clarke, S.R. (2006) Developing a model that reflects the outcomes of tennis matches.*Proceedings of the 8th Australasian Conference on Mathematics and Computers in Sport*, 178-188.

Barnett, T. J., & Clarke, S. R. (2002). Using Microsoft Excel to model a tennis match. In *6th Conference on Mathematics and Computers in Sport* (pp. 63-68). Bond University, Queensland, Australia.

Carter Jr, W. H., & Crews, S. L. (1974). An analysis of the game of tennis. *TheAmerican Statistician*, *28*(4), 130-134.

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

Jackson, D., & Mosurski, K. (1997). Heavy defeats in tennis: Psychological momentum or random effect? *Chance*, *10*(2), 27-34.

Klaassen, F. J., & Magnus, J. R. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, *96*(454), 500-509.

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

Newton, P. K., & Aslam, K. (2006). Monte Carlo tennis. *SIAM review*, *48*(4), 722-742.

O'Donoghue, P. G. (2001). The most important points in grand slam singles tennis. *Research quarterly for exercise and sport*, *72*(2), 125-131.

Pollard, G., & Noble, K. (2002). The effect of a variation to the assumption that the probability of winning a point in tennis is constant. In *6th Conference on Mathematics and Computers in Sport* (pp. 227-230). Bond University, Australia.

Pollard, G. H., & Noble, K. (2004). The effect of having correlated point outcomes in tennis. In R. H. Morton, & S. Ganesalingham (Eds.), *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport, Massey University, Palmerston North, New Zealand* (pp. 230-240).

Schutz, R. W. (1970). A mathematical model for evaluating scoring systems with specific reference to tennis. *Research Quarterly. American Association for Health, Physical Education and Recreation*, *41*(4), 552-561.

Viney, M., Bedford, A. & Kondo, E. 2013. "Sensitivity of court-side in tennis?" *Proceedings of the 4th International Conference on Mathematics in Sport,* Leuven, Belgium.

Viney, M., Kondo, E. & Bedford, A. 2012. "In-play modification of service probability in tennis." *Proceedings of the 11th Australasian Conference on Mathematics and Computers in Sport,*Melbourne, Victoria.

PREDICTING THE DRAW IN ENGLISH PREMIER LEAGUE FOOTBALL

Aaron Corris, Anthony Bedford and Elsuida Kondo

School of Mathematical & Geospatial Sciences
RMIT University, Melbourne, Australia

Abstract

This paper determines when the draw should be predicted for English Premier League football. Roughly one in every four matches in the English Premier League ends in a draw, so it would be a great boost to the predictive power of an Elo model to have a methodology of predicting the draw that is reliable. Initial research showed a consistency in draws over time, suggesting they may be predictable. However the banding of probabilities showed that draws were similarly likely across many probability bands, and that only one market-based band had a draw likelihood that exceeded a one in three chance. Three strategies were tested against a control of never picking the draw, with never picking a draw being the clear standout. Optimisation of these strategies resulted in marginally superior predictive power by restricting the draw prediction to tiny ranges that captured unusually high numbers of draws. Ultimately, it was deemed to be of no advantage to predict the draw.

Keywords: football, Elo model, draw, optimisation

1. INTRODUCTION

In the English Premier League, matches have three possible results. These are the home team winning, the away team winning, or a draw. This paper aims to improve the predictive ability of an Elo model by implementing a system which can predict when draws will occur on top of determining whether the home or away team is more likely to win.

If all English Premier League matches from the opening day of the 2002-2003 season to March 2nd 2014 are collated and the number of matches that resulted in draws are counted, it's seen that 1149 of the 4456 matches in that timeframe have ended in draws. This is a considerable number of matches, equating to slightly over one in every four matches. As a result, the maximum percentage of correct predictions for a model that predicts win vs loss is under 75%. To improve this percentage, there needs to be a methodology for reliably predicting draws. The aim of this paper is to determine whether such a methodology exists when it comes to predicting matches using an Elo model.

Before the development of the methodologies for when to predict draws, it is important to look at the statistics of when draws occur. The table below shows the proportions of drawn matches in each season in the dataset.

Season	**Draws**	**Matches**	**%**
203	90	380	23.68%
304	108	380	28.42%
405	110	380	28.95%
506	77	380	20.26%
607	98	380	25.79%
708	100	380	26.32%
809	97	380	25.53%
910	96	380	25.26%
1011	111	380	29.21%
1112	93	380	24.47%
1213	105	380	27.63%
1314	64	276	23.19%
Total	1149	4456	25.79%

Table 1: Draws Split by Season

The season with the most draws is the 2010-2011 season with 111 drawn matches, very closely followed by the 04-05, 03-04 and 12-13 seasons. 2005-2006 clearly has the least draws, with just 77 matches ending level. No other season has had less than 23% of matches end in a draw, with the closest being the in-progress 2013-2014 season. It's clear that this number is fairly consistent from season to season, with no real trend in the data.

Next, it's worth checking whether draws are more likely to occur at a certain time in the season. The following table shows the proportions of draws in each month of competition for the EPL.

Month	Draws	Matches	%
8	95	383	24.80%
9	106	400	26.50%
10	110	427	25.76%
11	132	483	27.33%
12	172	664	25.90%
1	121	465	26.02%
2	109	420	25.95%
3	76	403	18.86%
4	158	518	30.50%
5	70	293	23.89%
Total	1149	4456	25.79%

Table 2: Draws Split By Month

This table suggests that during the majority of the season, the percentage of draws remains quite constant. There are two values that are slightly different from the typical proportion, which are the 18.86% for March and the 30.5% for April. The initial reaction to this was that due to more matches being played in April, fatigue may have been causing matches to be more likely to end in draws, while for March due to comparatively less matches being played teams were better equipped to get a result. However, when looking closer at the number of matches played, March is not too dissimilar to the other lower volume months, and April is far from being the most matches in any month. Thus, in absence of any logical reason for these deviations, it's assumed that these values are merely a statistical quirk rather than a factor that needs to be considered.

2. MODEL AND MARKET BANDING

As it's been established that there are no key factors that need to be controlled for in the draw predictions, the next stage is to compare how the draw is distributed for the Elo model versus the market predictions in the form of betting odds. Firstly, the 3-way result probabilities are viewed to determine whether the market ever predicts a draw. The interesting thing is that in the 4456 matches catalogued in the dataset, on only one occasion did the market predict that the draw was the most likely result. This was the match of Stoke hosting West Brom on 16/03/13, for which bet365 offered odds of $2.80 for a Stoke win, $3.25 for a West Brom win, and $2.75 for a draw. This corresponds to a 35.357% probability of a draw after accounting for the bookmaker's margin. This match ended in a draw as expected, with the final score being 0-0. According to the Elo model this was a 55-45 matchup and according to the 2-way market a 54-46 matchup in favour of Stoke, so it is quite an even match up. However, there are numerous matches which according to these probabilities are more even, so it's quite odd that just this one match has the draw as the favourite. It suggests that the draw may in fact be too unreliable to be predicted with any confidence whatsoever.

The next stage is to compare how the draws are distributed for both the 2-way market model and the Elo model. The following graph shows the frequency of draws for both of these models.

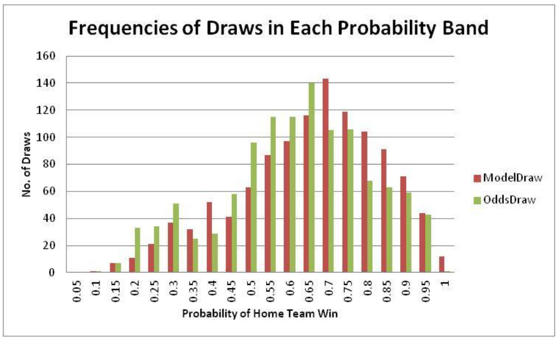


Figure 1: Frequencies of draws in each probability band for 2-way market model (oddsdraw) and Elo model (model draw).

The most draws occurred in the 0.65-0.7 band for the Elo model (143 drawn matches), and the 0.6-0.65 band for the 2-way market model (140 drawn matches). It is interesting that the number of draws seems to be centred roughly on the home ground advantage adjusted probability. As this is just the total number of draws for each band, the proportion of draws needs to be checked to determine whether this is just being skewed by the number of matches in each band. The proportion of draws in each band is shown in the graph below.

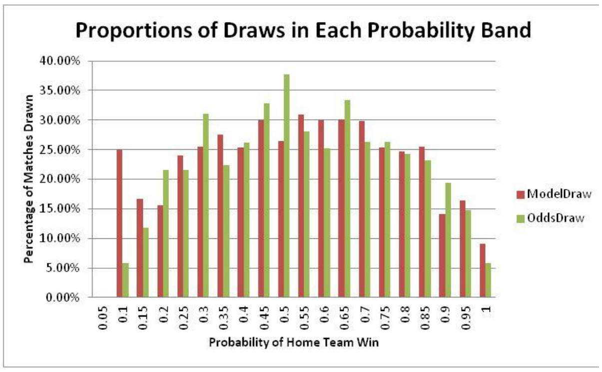


Figure 2: Proportions of draws in each probability band for 2-way market model (oddsdraw) and Elo model (model draw).

It is clear that when correcting for the number of matches there is no longer a point where the draws have a peak. The 2-way market has a slight increase for matches just under 0.5, but this is not particularly significant. The Elo model appears to be over-represented for draws in matches where there is a very strong favourite either home or away, however it's important to note that this is primarily due to a relatively small sample size in these bands. Overall, there's a slight quadratic curve to the percentages of draws, with the edge cases typically being less likely to have a draw, and the central cases having similar proportions of draws. This does not bode well for attempts to predict the draw as there is only a single band that exceeds a one-in-three chance of a draw occurring, which is the 2-way market for matches with home win probabilities between 0.45 and 0.5.

Finally, the draw probability according to the odds is compared to the home versus away split is analysed. The primary purpose of this is to be able to develop a probability of drawing from the Elo model. The following graph is a scatter plot of home win probability minus away win probability versus draw probability.

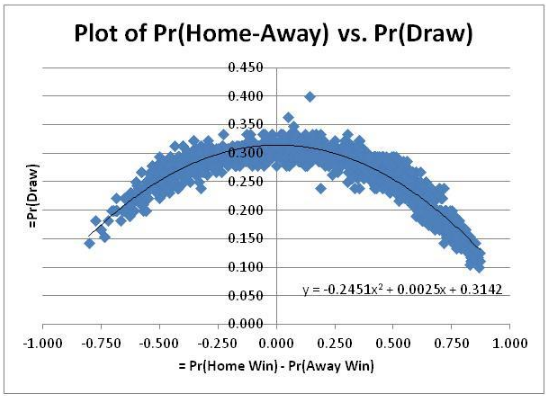


Figure 3: Scatter plot comparing the home-away probability differential to the probability of a draw according to the market

This plot shows a clear quadratic trend for draw probability when compared to the home-away differential. The line of best fit is Pr(Draw) = $-0.245x^2 + 0.0025x + 0.3142$, where x equals the difference between the win probability of the two teams. Using this equation, the probabilities that are generated by the Elo model can be converted to include a draw probability. So, using this equation, a match in which the two teams have an equal probability of winning will have a 0.3142 probability of drawing the match. These probabilities will be used for one of the prediction strategies outlined below.

Initial Draw Prediction Strategies

There are three initial strategies that'll be used to predict the draw. These strategies are as follows;

- pick the draw for the middle 25% of probabilities (matches with win probabilities between 0.375 and 0.625)
- pick the draw for the middle 25 percentiles (matches with win probabilities between the 37.5th and 62.5th percentile, corresponding to matches between 0.577 and 0.699 win probability for the home teams)
- pick the draw for the 25% of matches with the highest draw probability according to the line of best fit to the market odds

These strategies are based upon the fact that roughly 25% of matches end in a draw, so if the draw is perfectly predictable 25% of matches should be predicted as a draw. The results of this initial strategy are shown in the table below.

	Strike Rate
No draw	52.244%
Middle 25% Pr	49.753%
Middle 25th Percentile	48.654%
Upper quartile draws	50.135%

Table of strike rates for initial draw prediction strategies

It is interesting to see that each of these strategies performs worse than never predicting the draw. The best performed of the strategies that picks draws was the upper quartile draws, followed by the matches with draw probabilities between 0.375 and 0.625, and worst performed was the middle 25th percentile. Clearly too many matches are being predicted as draws, so in the next step the range and where the predictions should be centred will be optimised.

3. OPTIMISED DRAW PREDICTIONS

Using the @Risk software, the optimal strike rate for each strategy was search for. This was done by changing the spread of matches that were predicted as draws and in the case of the middle 25% and middle 25th percentile where that spread was centred. 10000 iterations were done, and the results of this optimisation can be seen in the table below.

	Strike Rate	Spread	Centre
No draw	52.289%	0	NA
Middle x% Pr	52.513%	0.00265	0.526
Middle xth Percentile	52.536%	0.00295	0.524
Upper x percentile draws	52.311%	0.00023	NA

Table of optimised draw predictions

Here each of the various draw prediction strategies exceeds the strike rate of never predicting the draw. This appears to suggest that the draw should in fact be predicted sometimes. However, note the extremely small spreads of matches being predicted as draws. For example, the middle x% pr is predicting draws in matches where the home team has a win probability between 0.523 and 0.529. This is a tiny section of matches, a mere 43 (<1%) of the 4456 in the data set. This suggests that predicting the draw for this section of matches isn't actually a good strategy, rather a lucky one that just happens to encapsulate a small cross-section of matches that has an abnormally high proportion of draws.

Based upon these optimisation results, it seems clear that using the Elo model as a basis for draw predictions is not a strategy that is reliably better than simply never picking the draw. Occasionally, one might get lucky and stumble upon a very small pocket of matches which happens to have a greater than typical draw probability, but ultimately attempting to predict the draw in this manner is not beneficial. In other words, the draw is an event that cannot be predicted with any confidence. This validates the market solution of almost never predicting the draw.

4. CONCLUSION

The draw is a frequent event in the English Premier League. It's a fairly constant variable from season to season, and month to month. This consistency suggested that predicting the draw may be possible, as there are no major variations in the probability of the draw. However, further analysis showed that there were no bands of probability where the draw became more likely than the probabilities of one of the two teams to win. Notably only one band exceeded a one in three probability, which was the 0.45-0.5 market band. Three strategies to predict the draw roughly 25% of the time where tested against a strategy of never predicting the draw, and all fell considerably short. The three strategies were optimised for maximum strike rate, but were only marginally better than the no draw strategy due to finding an extremely small band where an abnormally high proportion of draws occurred. Thus the best strategy of predicting the draw is in fact to not predict the draw at all.

UNDERSTANDING THE VELOCITY RATIO OF MALE AND FEMALE OLYMPIC CHAMPIONS IN RUNNING, SPEED SKATING, ROWING AND SWIMMING

Ray Stefani

California State University, Long Beach
Raystefani@aol.com

Abstract

Photos of past Olympic champions in running, speed skating, rowing and swimming show little difference in physique over 80 years. Improvements in performance must therefore be due to better training and better efficiency, the latter including coaching, technique and equipment. Physical laws were applied to the four sports. For running and speed skating, if efficiency is constant, the power-to-weight ratio, P/m, improves velocity. For equally trained men and women, the velocity ratio should equal their relative lean-to-weight. Tested Olympian females were 92% as lean as their male counterparts while Olympic champion women ran 91% as fast as men for 1980-1988. In speed skating, women were 92% as lean, while their Olympic champions skated 92% as fast as men for 1980-1988. If efficiency is constant for rowing, $P / m^{2/3}$ improves the cube of velocity as does $P\ m^{1/3}$ for swimming. For rowing and swimming, assuming equal training and efficiency, the theoretical velocity ratio was calculated using values from tested Olympians as to relative lean-to-weight, relative body mass and relative drag coefficient. For rowing, relative cranking power was also needed. For both sports, the estimated velocity ratio was approximately the $4/9^{th}$ power of the body mass ratio. The estimated swimming velocity ratio of 91% was exactly the velocity ratio of Olympic champions, 1980-1988. The estimated rowing velocity ratio of 90% was exactly the velocity ratio for Olympic champions, 1992-2012. Female champions have improved from being 83% as fast as men 100 years ago in swimming to being 89% as fast in running, 90% as fast in rowing and swimming and 92% as fast in speed skating. In Olympia 2500 years ago, women ran a 500 ft course while men ran 600 ft, making women about 5/6 or 83% as fast. Relative velocity has been remarkably constant over recorded history.

Keywords: Olympics, running, speed skating, rowing, swimming, power, gender differences, velocity, power to weight, Olympic champions

1. INTRODUCTION

Men and women have competed in athletics (track and field) for at least 2,500 years. The ancient Olympics (actually one of the four Pan-Hellenic Games) began at Olympia in 776 BC. That competition was dedicated to the male god Zeus, which meant that only men could compete, under the prevailing religious practices of that day. Unmarried women could and did attend. A few centuries later, a second set of Games were created at Olympia for those unmarried women, called the Heraia Games, dedicated to Zeus' mythological wife, Hera. A group of 16 women was permanently empowered to run the ancient Greek sports program for women. Women competed on different days from the men. At the Olympic Games, the men ran one or more multiple of the **stadia**, a stadium length of 600 Greek feet. Our word "stadium" comes from that measure. At the women's Heraia Games, the stadia was shortened to 500 Greek feet. If women were considered to cover 500 feet in the approximate the same time that men would cover 600 feet, then the relative velocity of women was 500/600 of 83%, some 2,500 years ago.

In 1896, women did not compete in the first modern Games, due this time to the reticence of the Games founder, Baron de Coubertin. It only took four years for women to start competing in a variety of Olympic sports in 1900. Swimming began for women in 1912, athletics in 1928, speed skating in 1960, 1000 m rowing in 1976 and 2000 m rowing in 1988. This paper will explore gender differential behaviour in those sports using starting dates of 1912, 1928, 1960 and 1988 respectively.

Early work at understanding winning performances in general and gender differences in particular was presented in Stefani (2000). The laws of hydrodynamics were used to derive relationships for the power output in rowing, based on elapsed time. Stefani (2002) presented a preliminary derivation of the power output in running, jumping and swimming, along with estimated female/male percent differences in power as developed by Olympic champions.

Stefani (2006) covered a much more sophisticated derivation of power output in running, jumping and speed skating (ground effects events) and in rowing and swimming (hydrodynamic events). The ratio of (women's power) / (men's power) was found employing two methods and the two sets of results were compared. The power ratio based on performances agreed with a power ratio based on physiology, specifically on relative (lean body mass / total body mass), using Olympic data from 1976-2004. It was shown that the power ratio was not as favourable for women in earlier years than the recent ratios consistent with physiology, suggesting unfavourable differences in training and efficiency in the past. Stefani (2007) extended those results by listing the skills required for each event, suggesting which of those improved, and estimating how past champions could have remained competitive in more recent years they had been privy to better efficiency.

A recent paper, Stefani (2014), took a more practical approach for athletes and coaches, using power/weight analysis which included training and efficiency as variables. While an athlete is in training, improvements in ergometer power/weight can be easily measured and then used to directly estimate improvements in performance, without resorting to a time trial. In addition, upper body strength was included (measured by cranking power) which relates to success in throwing and rowing. That paper covered running, jumping, throwing, swimming and rowing.

Stefani (2012) attacks the issue of whether high-tech swim suits may have been a cause of faster swimming; the conclusion is that the swimmer not the suit is responsible for faster times.

This current paper will use a performance measure that is both intuitive and informative: the velocity ratio of women/men for Olympic champions. The laws of physics are used in Section 2 to derive the velocity ratio for the ground reaction events of running and speed skating, due to training, efficiency and physiology. A similar approach is taken in Section 3 for two hydrodynamic events (rowing and swimming). Increases (decreases) in that velocity ratio occur when women improve faster (slower) than men. Section 4 includes salient conclusions.

2. RUNNING AND SPEED SKATING

A more complete explanation of the following physical laws and kinesiology analysis, along with a supporting bibliography, may be found in Stefani (2007, 2014). A bibliography of the references cited in Tables 2-5 may also be found in Stefani (2007, 2014), except for a few additional works cited at the end of this paper.

The power generated by a running or speed skating athlete can be measured on a treadmill or cycling ergometer. Studies show that ergometer power, P, depends on the athlete's lean body mass, LBM, and training (Tr) as given in (1).

$$P = LBM\ Tr \qquad (1)$$

That is, P/LBM is a constant for equally trained athletes of both genders. Enhanced training can improve that ratio for both genders.

A fraction of that generated power, P e, is applied in running and speed skating to the centre of gravity of an athlete with body mass m, where e is the efficiency less than or equal to one. Efficiency depends on some combination of coaching, technique and equipment. That applied power is "absorbed" by that body mass, resulting in a velocity, v in (2), using Newtonian mechanics. The angles in (2) measure the direction of the forward movement of the centre of gravity. These angles differ for running and speed skating,

$$P\ e = m\ v\ f(\text{angles})\ \text{constants} \qquad (2)$$

Side-by-side photos show little physical difference between Olympic champions of the past and present. The increases in velocity must then have followed from some combination of better training, coaching, technique and equipment. If %I/O denotes the percent improvement per four-year Olympiad, the introduction of the rowing ergometer (a training aid) in 1980 increased average %I/O by 508%. Equipment such as the fibreglass pole in 1984 added a one-time 419% to %I/O in the pole vault as did the clap skate in 1998, adding 58% to speed skating %I/O. The Fosbury Flop, a high jumping technique, added an average of 83% when introduced for men in 1968 and for women in 1972.

If both sides of (2) are divided by m, then P/m, the power-to-weight ratio, depends directly on v for fixed e. That is, each 1% increase in P/m while training implies a 1% increase in velocity if efficiency is maintained.

The goal here is to analyse the velocity ratio of women/men, which follows from (1,2), where LTW denotes lean-to-weight given by LBM/m and it is assumed that angles are the same for both genders. Of course constants cancel.

$$v_W/v_M = (LTW_W/LTW_M)\,(Tr_W/Tr_M)\,(e_W/e_M) \quad (3)$$

When men and women are equally trained and efficient, (3) depends only on relative lean-to-weight LTW_W/LTW_M. When the velocity ratio is smaller than given by LTW_W/LTW_M, then training and/or efficiency for women would not be as good as for men. Table 1 contains the average velocity ratio for Olympic champions in running and speed skating. Five periods of Olympic history are used. The first period spans WW1 ending with the second post-war Games when competition had recovered. The second period is similar for WW2. The third period covers the Cold War while the fourth period covers the two boycotted Games followed by recovery in 1988. The fifth period covers the post-1988 anti-doping era. As mentioned earlier, this paper explores gender differential behaviour beginning with 1912 for swimming, 1928 for running, 1960 for speed skating and1988 for 2000m rowing.

Table 2 contains studies from which relative LTW (relative leanness) can be calculated for running and speed skating. For speed skating, women were 92% as lean as men for period four and female champions skated 92% as fast, constant with equal training and efficiency. Women gained from period three to period four (89-92%) and then stayed 92% as fast in the current period.

In running (rounded to two digits) women were 92% as lean in period four while female champions ran 91% as fast. Similarly for period five, women were 91% as lean but champions ran 89% as fast. The actual difference was 1.5% each time. Assuming equal training, women were 1.5% less efficient than men. Why? Female runners are six times as likely to have an ACL tear as men because their pelvis and hips are wider than for men, relative to height, causing relative overstriding and some knee rotation, Williams and Cavanagh (1987), Gilland (2009), Hewitt (2010). The data in Table 1 provides a scaling for that overstriding. Because a female runner has her leg a bit straighter than a man at stride's end, women apparently put 1.5% of the force, intended to move the athlete forward, into the knee and ankle joints, causing the female champion to run 1.5% slower than suggested by relative LTW. The take-away message is that female athletes should strengthen knees and ankles to protect against potential injury.

Women apparently gained in training and efficiency from period two to four (88% to 89% to 91%). Having achieved equality, periods four and five imply a velocity ratio driven by physiology (LTW and overstriding). Why did women lose 2% in period five, returning to a velocity ratio of 89% as in period three? Period four may be the statistical equivalent to the chemical anti-doping passport, wherein a change in blood chemistry suggests doping. We know that steroids were rampant in period four. Anti-doping efforts have ramped up in period five. It may be that women gained more LTW than did men in period four, as women were less lean to begin with and would benefit more with steroid use. The former East Germany was a consistent gold medal winner in women's running events from 1980-1988. Former East German authorities have admitted doping before unification.

3. ROWING AND SWIMMING

The law of hydrodynamics is to rowers and swimmers what Newtonian mechanics is to runners and speed skaters. The kinesiology of a rowing ergometer differs from that of a cycling or treadmill ergometer. That is, P/LBM is not equal for equally trained women and men. An additional cranking effect, Cr, is present. Women are at a disadvantage due to relatively less upper body and shoulder volume. The power registered on a rowing ergometer is given by (4).

$$P = LBM\ Tr\ Cr \quad (4)$$

As in (2), a fraction of that power is applied to a racing shell causing it to move forward with velocity v.

Period	Years	Running		Speed Skating		Rowing		Swimming	
	Period	LTW Ratio N=156	Velocity Ratio N=95	LTW Ratio N=51	Velocity Ratio N=46	Estimate N=1789	Velocity N=49	Estimate N=1815	Velocity N=192
WW1-Recovery	1912-24								83
WW2 - Recovery	1928-52		88						87
Cold War	1956-76		89		89				90
Boycott-Recovery	1980-88	92	91	92	92		90	91	91
Post 1988	1992-14	91	89		92	90	90	90	90

Table 1 Velocity ratios of female/male Olympic champions and estimated ratios based only physiology

Source	Event	Men		Women		LTW_W/LTW_M (sd)
		N	%Fat (sd)	N	% Fat (sd)	
Fleck (1983) US Olympians	Running	24	6.4(1.2)	21	13.7 (3.6)	92.1 (2.5)
Vucetic et al. (2008) Elite Athletes	Running	41	5.8 (2.4)			
Molina (2007) US College Athletes	Running			70	14.2 (1.3)	91.0 (1.8)
Fleck (1983) US Olympians	Speed Skating	31	11.4 (3.2)	20	18.1 (5.0)	92.4 (4.5)
Yoshiga and Higuchi (2003) Elite Athletes	Rowing	120	11.9 (6.2)	71	20.9 (5.2)	89.8 (5.7)
Fleck (1983) US Olympians	Swimming	39	12.4 (3.7)	41	19.5 (2.8)	91.9 (3.5)
Van Erp-Baart et al. (1989) Elite Athletes	Swimming	20	10.7 (3.3)	50	21.4 (5.6)	88.0 (4.4)

Table 2 Relative Lean-to-Weight Ratio LTW (LTW = LBM/m = 100 - %Fat)

Source	Men		Women		Cd_M/Cd_W
	N	Cd(sd)	N	Cd(sd)	
Toussaint (1988)	32	.54(.09)	9	.47 (.07)	1.170
Zampora (2009)	84	.353	66	.318	1.110
Mean					1.123

Table 3 Drag Coefficient Ratio for men/women

Study	Men				Women				Cr_W/Cr_M
	N	m	LBM	Time	N	m	LBM	Time	
Equal m	57	63		436	37	62		477	0.850
Equal LBM	20		52	446	10		51	466	0.930
	77				47				0.890

Table 4 Cranking Ratio Cr for elite rowers (Yoshiga and Higuchi, 2003)

		Men		Women		
Source	Event	N	m (sd)	N	m (sd)	m_W / m_M
Mc Ardle (1981) 1964 & 1968 Olympians	Running and Jumping	354	71.8	181	57.5	1.249
Mc Ardle (1981) 1964 & 1968 Olympians	Swimming	516	73.1	300	58.3	1.254
Pyne (2000) 1988 & 1994 Australian Olympians	Swmming	43	81.8 (7.0)	42	64.8(6.1)	1.262 (.115)
US Olympians 2000	Rowing	19	92.5 (8.1)	19	73.3(6.7)	1.261 (.114)
		932		542		1.256 (.11)

Table 5 Body mass ratio for women/men

The area in contact with the water, approximately the 2/3 power of body mass due to buoyancy, induces drag. The equation of motion is given by (5), where Cd is the drag coefficient.

$$P\ e = v^3\ m^{2/3}\ Cd \text{ constants} \qquad (5)$$

If both sides of (5) are divided by $m^{2/3}$, then $P/m^{2/3}$ becomes the rower's power-to-weight parameter, equal to v^3 if efficiency e remains constant. Rowers can be ranked for placement on a racing shell, based on rowing ergometer $P/m^{2/3}$ as described in Stefani (2000, 2014). Every 1% improvement in that parameter implies a 1/3% increase in velocity, for constant e. The velocity ratio (6) follows from (4,5), where constants cancel.

$$v_W/v_M = [(Tr_W/Tr_M)(e_W/e_M)]^{1/3}$$
$$[(LTW_W/LTW_M)(Cr_W/Cr_M)]^{1/3} (m_W / m_M)^{1/9} \qquad (6)$$

If women are as equally trained and efficient as men, in rowing, then the velocity ratio would be given by the second line of (6). Table 2 (LTW), Table 4(Cr) and Table 5 (body mass) provide the values needed. The values of Cr_W/Cr_M in Table 4 are found by solving (6) using the tabular data. For period five, $(LTW_W/LTW_M)(Cr_W/Cr_M)$ is given by .898 x .890 or .799. That is essentially the same as m_W / m_M which is 1/1.256 or .796. For equal training and efficiency, the estimated velocity ratio of (6) is the same as $(m_W / m_M)^{4/9}$, or 90%, agreeing to two digit accuracy with the observed velocity ratio of 90% for rowing in period five. Women appear to have competed with equal training and efficiency as men over both periods, since the velocity ratio is accurately estimated by physiology only.

For swimming, a treadmill or cycling ergometer can be used to measure power, as given by (1). The fraction of power applied is more complicated than just P e for swimming. A swimmer applying force while immersed in water does so with a propelling efficiency much like that of the propeller on a boat, Stefani (2014), depending on the size of the swimmer. Thus a swimmer's efficiency becomes e = m e_S where e_S depends on coaching, technique and equipment/conditions. The applied power equation is given by (7).

$$P\ m\ e_S = v^3\ m^{2/3}\ Cd \text{ constants} \qquad (7)$$

If the body mass terms are collected on the left side of (7), then a swimmer's power-to weight relationship is P $m^{1/3}$. Every 1% improvement in that parameter implies a 1/3% increase in velocity, for constant e_S.

The velocity ratio follows from (1,7).

$$v_W/v_M = [(Tr_W/Tr_M)(e_W/e_M)]^{1/3}$$
$$[(LTW_W/LTW_M)(Cd_M/Cd_W)]^{1/3}(m_W / m_M)^{4/9} \qquad (8)$$

If women are as equally trained and efficient as men in swimming, then the velocity ratio would be given by the second line of (8). Table 2 (LTW), Table 3(Cd) and Table 5 (body mass) provide the values needed. For period four, $(LTW_W/LTW_M)(Cd_M/Cd_W)$ is .919 x 1.123 or 1.011. For period five, that term is .880 x 1.123 or .996. Both values are close to one. The surprising conclusion is that for equal training and efficiency, the velocity ratio in swimming is closely approximated by $(m_W/m_M)^{4/9}$, the same approximation as for rowing.

Completing the estimated velocities, we obtain 91% for period four and 90% for period five, both of which agree with the actual average velocity ratios in Table 1 for swimming.

It appears that women acquired better training and efficiency from periods one to three, as their speed relative to men increased from 83% to 90%. The relative velocity was nearly the same for the last three periods, agreeing with physiological estimates, indicating equality of training and efficiency.

4. CONCLUSIONS

Side-by-side photographs of Olympic champions taken about 80 years apart show little change in physiology. The large increases in velocity in running, speed skating, rowing and swimming are therefore due to improved training and efficiency (coaching, techniques and equipment). The laws of physics were used to derive the velocity ratio of women/men, in terms of training, efficiency, and physiological variables.

Satisfying the assumption of equal training and efficiency, today's Olympic champions display a velocity ratio for women/men closely approximated by physiology alone. In the hydrodynamic events of rowing and swimming, female elite athletes have 90% the value of men for the $4/9^{th}$ power of their body mass ratio and female Olympic champions are 90% as fast. Female speed skaters are 92% as lean as men and their Olympic champions are 92% as fast. Female runners are 91% as lean but run 89% as fast, losing 1.5% by overstriding, induced by a relatively wider pelvis. If we exclude the velocity ratios for 1980-1988 in running and swimming, a monotonic trend is evident as female athletes gained in training and efficiency, increasing the velocity ratio until differences were due only to physiology. Female swimming champions gained from being 83% as fast in 1912-24 to being 90% as fast in 1956-1976, the same as today (1992-2014). Female rowers have been 90% as fast since 1980-1988. Female speed skaters increased from 89% as fast to 92% as fast in 1980-1988, the same as today. Female runners increased from 88% as fast for 1928-1952 to 89% as fast for 1956-1976, the same as today. There was a bump up in 1980-1988 and then down today for the velocity ratio for female champions in running and swimming, the same time period when East Germany dominated women's running and swimming. Former East German athletes and officials have admitted doping which gave female champions a boost in relative lean body mass, explaining the boost in velocity.

Athletes in training can use a power-to-weight measure to assess progress. For running and speed skating, each 1% by which P/m is increased, velocity increases by 1%, for equal efficiency. For each 1% that $P/m^{2/3}$ (rowing) and $P\ m^{1/3}$ (swimming) increases, velocity increases by 1/3% for equal efficiency. Female runners should strengthen knees and ankles to reduce possible injury.

Over the centuries, it has taken a great deal of effort by female athletes and by fair-minded people of both genders to provide a level playing field for today's female athletes. It is everyone's job to keep it that way.

References

Erp-Baart, et al. (1989) Nationwide survey on nutritional habits in elite athletes, *Int. J. Sports Med*. 10, S3-S10

Galland, M. (2009) Preventing ACL Tears-Why are ACL Tears More Common in Female Athletes, http://www.orthonc.com, May 18, 2009.

Hewitt, T.E. (2010) Why women have increased risk of ACL injury, http://aaos.org/news/aaosnow/nov10/research3.asp.

Malina, R.M (2007) Body Composition of Athletes: Assessment and Estimated Fatness, *Clinics in Sports Medicine*, 26, 37-68.

Stefani, R.T (2000) Power Output of Olympic Rowing Champions, *Olympic Review*, no. XXVI-30, pp.59-63.

Stefani, R.T (2002) The Fundamental Nature of Differential Male/Female World and Olympic Winning Performances, Sports and Rating Systems, *Proceedings of the Sixth Australasian Conference on Mathematics and Computers in Sport*,July 1-3, 2002, pp 32-46.

Stefani, R.T. (2006) The Relative Power Output and the Relative Lean Body Mass of World and Olympic Male and Female Champions with Implications for Gender Equity, *Journal of Sports Sciences*, 24(12), 1329-1339.

Stefani, R.T (2007) The Physics and Evolution of Olympic Winning Performances, Chapter 3, *Statistical Thinking in Sports*, Chapman and Hall/CRC Press.

Stefani, R.T. (2012) Olympic Swimming Gold: The Suit or the Swimmer in the Suit?, *Significance*, 9(2), 13-17.

Stefani, R.T. (2014) The Power-to weight relationships and efficiency improvements of Olympic champions in athletics, swimming and rowing, *International Journal of Sports Science and Coaching*, 9(2).

Vucetic, V. At al. (2008) Morphological Differences of Elite Croation Track and Field Athletes, *Col. Anthropol.* 22 (3) 863-868.

Wlliams K.R. and Cavanagh, P.R. (1987) Biomechanical studies of elite female distance runners, *Int. J. Sports Med.* 8(Suppl. 2) 107-118.

Yoshiga, C.C and Higuchi M. (2003) Rowing performance of female and male rowers, *Scandinavian J. of Medicine, Science and Sport*, 13, 317-331Audas,

Zampora P. et al. (2009) Active and passive drag: the role of trunk incline, *Eur. J. Appl. Physiol*. 106, 195-205.

ANALYSIS OF REFEREE BIAS AND PENALTIES IN NETBALL: A LOOK AT THE ANZ CHAMPIONSHIP

Jason Hunt[a], Anthony Bedford[a] and Nick Sanders[b]

[a]*School of Mathematical & Geospatial Sciences*
RMIT University, Melbourne, AUSTRALIA
[b]*Victorian Institute of Sport*

Abstract
This paper focuses on penalty counts in the ANZ Championship and how they vary according to the teams involved, location and the umpires officiating the matches. The matches were split into four categories for analysis. Of most interest was category 4 – New Zealand side versus an Australian side played in New Zealand (and therefore with umpires from New Zealand). This group produced results that were pointedly different to that of each of the other categories, with the away team conceding on average approximately 12 more penalties than the home side. Category 4 proved to be significantly different to from each of the other three categories, whilst none of the other categories were different from each other. This difference was comprised of approximately 8 contacts and 4 obstructions on average per match. The reasons for this discrepancy are not entirely clear, however it can be linked to crowd factors, umpiring bias, umpiring style differences and team play style differences. It is anticipated this paper may help players, coaches and umpires better understand why this discrepancy might exist and how to go about reducing its effect in the future.

Keywords: Penalties, Umpire bias, ANOVA, Netball

1.INTRODUCTION

Netball is an invasion sport played between two teams, with seven players on court at all times (and up to five players on the bench). The ultimate goal of the game is to score more goals than the opposition by placing to ball into your goal which is a hoop placed at the top of a pole at each end of the court. The ANZ Championship is widely considered to be the highest level of competitive netball in the world, aside from international tests. The competition is composed of five teams from Australia and five teams from New Zealand, who compete in 13 matches each throughout a season and then a finals series to decide the eventual winner. As is the case with every sport, when the rules of the game are infringed upon, the officiating umpires must award free passes against the infringing team. Penalties are the most common occurrence of this in netball, and these actions can be attributed to contact calls and obstruction calls. Contacts can be thought of as any infringement that involves illegal physical contact between players such as attempting to strip an opposition player of possession. Obstructions can be thought of as any infringement that does not involve physical contact, such as defending a player too closely (must be 0.9m or more away from a player whilst defending an opposition player).

The structure of the ANZ Championship means that Australian umpires officiate matches in Australia and New Zealand umpires officiate matches in New Zealand. There is a common perception that umpiring interpretations are different or at least that penalty counts are disparate across the two countries. Possible reasons for why this may include referee bias, crowd influence or differences in play style across countries. There has been a large amount of research in the past surrounding crowd influence and the effect it can have on umpiring decisions in other sports around the world. Watching a match without sound; therefore without the crowd noises and the atmosphere which that creates, has been shown to significantly reduce the number of fouls awarded to the home side in association football (Nevill et al, 2002). The removal of sound also resulted in umpires being less certain of their decisions. This can therefore be linked in closely with home team advantage, which is what clearly creates this atmosphere as the home team will almost always have the larger crowd support. Crowd density is a large factor in football at least, as opposed to crowd size or proximity (Goumas, 2014). The people in the crowd most certainly believe that they can influence the result (Wolfson et al, 2005) so it is very possible that this may extent to influencing the umpiring decisions on the court. The difference in play style between the two countries is much more difficult to discuss or prove – however it is commonly accepted that New Zealand sides traditionally play a defensive zone where as the Australian sides play one on one defence. This may or may not have an effect on the number of penalties that occur.

2.METHODS

Penalty data from the ANZ Championship seasons between 2009 and 2012 (inclusive) was collected from the ANZ Championship website (as supplied by Champion Data). This was split into obstructions and contacts for each team in each match and included locations and scores of matches. This was a total of 260 matches - 65 matches from each season. Finals were not included as they are usually officiated by an umpire from each country. Four of the home and away matches were also excluded from analysis because they were known to either be incomplete or extend into extra time (where it was not possible to obtain a penalty count for the match prior to extra time commencing). The data was then tested to see if the penalty counts have changed dramatically over time prior to any analysis. Following this, the data was split into four categories.

Category	Description
1	Australia vs. Australia matches, played in Australia
2	New Zealand vs. New Zealand matches, played in New Zealand
3	Australia vs. New Zealand matches, played in Australia
4	New Zealand vs. Australia matches, played in New Zealand

Table 1. Venue information and categories

This allows for comparison across the four different types of matches that concern this study and will form the basis of all analyses.
In some cases, thepenalty differential is used as the value of interest, this was defined to be the number of penalties conceded by the away team minus the number of penalties conceded by the home team. Likewise, this was later split to accommodate for contact and obstruction differentials.

3. RESULTS

An ANOVA was used to test whether the mean difference in penalties between away and home teams changed from 2009 to 2012. The resultant p-value (0.895) clearly indicated no significant difference between the years, which meant it was possible to analyse the data as a whole. A look at home and away penalties over time helped to explore this.

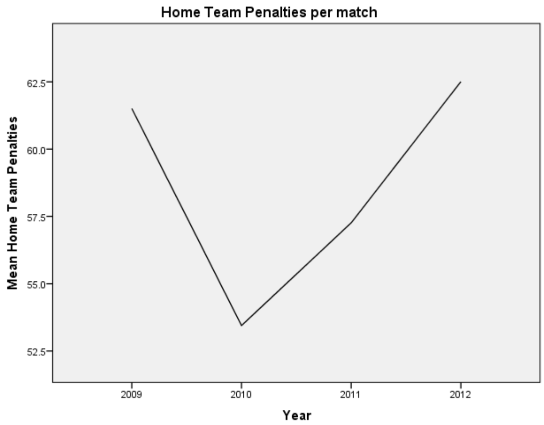


Figure 1. Mean home penalties per match for each year

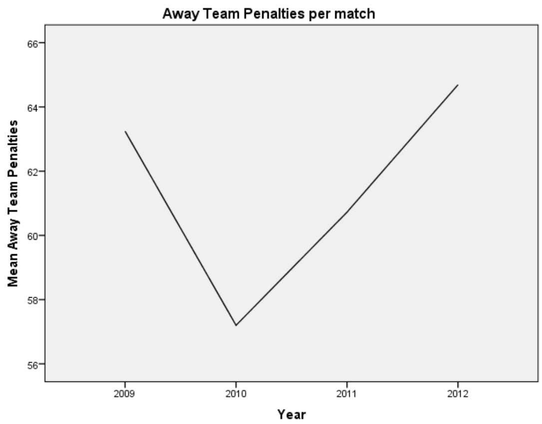


Figure 2. Mean away penalties per match for each year

Even though the numbers of penalties have fluctuated from year to year, the home penalties and away penalties have fluctuated approximately evenly with each other, which explains why the ANOVA test on differentials showed no significant difference.

A preliminary look at penalty counts resulted in Table 2. The first four rows show mean home team penalties for each category, and the second four rows show the mean away team penalties for each category. Generally speaking the penalty means are around the 60 or just under that mark for each side. Underlined is the mean away team penalties in category 4. This figure represents Australian team penalties when playing in New Zealand and it stands out as the discrepancy in these figures with a mean of 68.90. Also of note is that the corresponding home team penalties mean is 56.71, the lowest mean of all of the categories.

	Category		N	Mean	Std. Deviation
Home Team Penalties (Home Team named first, then location of game)	1	Aus/Aus	78	59.13	11.940
	2	NZ/NZ	79	59.75	11.820
	3	Aus/Aus	50	58.00	13.513
	4	NZ/NZ	49	56.71	13.880
	Total		256	58.64	12.585
Away Team Penalties (Away Team named first, then location of game)	1	Aus/Aus	78	60.73	12.184
	2	NZ/NZ	79	58.89	13.186
	3	NZ/Aus	50	59.26	9.583
	4	**Aus/NZ**	**49**	**68.90**	**11.500**
	Total		256	60.33	12.424

Table 2. Mean penalties per side in each category

The analysis then changed toward looking at differentials in penalties rather than penalty counts alone. The differential was calculated by subtracting the home team penalties from the away team penalties. A visual look at this can be seen in Figure 3.

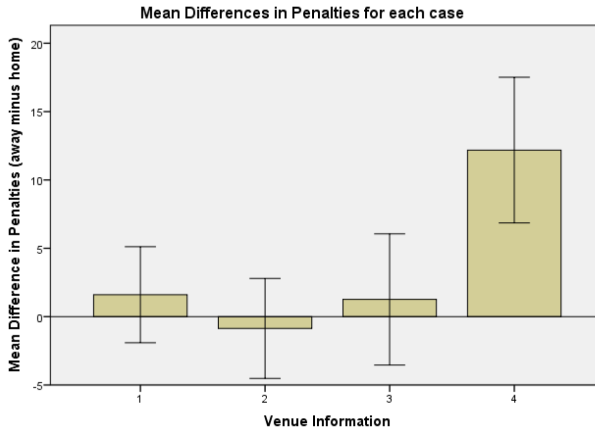


Figure 3. Mean difference in penalty counts with error bars for each category

Again it is category 4 that appears to be the discrepancy in this data, with a much higher mean penalty differential than the other 3 categories. In order to validate this difference, the means were compared for statistically significant differences. This included an ANOVA test (which was significant at alpha level of 0.05) and then post-hoc tests following that.

ANOVA

Difference in Penalties (away minus home)

	SS	DF	MS	F	Sig.
Between Groups	5603.73	3	1867.91	6.72	**<0.00**
Within Groups	70017.12	252	277.85		
Total	75620.84	255			

Table 3. ANOVA test on penalty differential

	(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.
Tukey HSD	1	2	2.463	2.661	0.791
		3	0.343	3.020	0.999
		4	**-10.581**	**3.038**	**0.003**
	2	1	-2.463	2.661	0.791
		3	-2.121	3.012	0.895
		4	**-13.044**	**3.031**	**<0.000**
	3	1	-0.343	3.020	0.999
		2	2.121	3.012	0.895
		4	**-10.924**	**3.351**	**0.007**
	4	**1**	**10.581**	**3.038**	**0.003**
		2	**13.044**	**3.031**	**<0.000**
		3	**10.924**	**3.351**	**0.007**

Table 4. Post-hoc tests on penalty differential

Table 3 indicatedthat at least one pair of match types differ significantly from each other. The post-hoc Tukey tests (in Table 4) show that category 4 is significantly different to each of the other categories, whilst none of the other categories are significantly different. This further suggests that category 4 is the discrepancy in this data set.

When the data was split into obstructions and contacts, it gave a slightly take on the discrepancies.

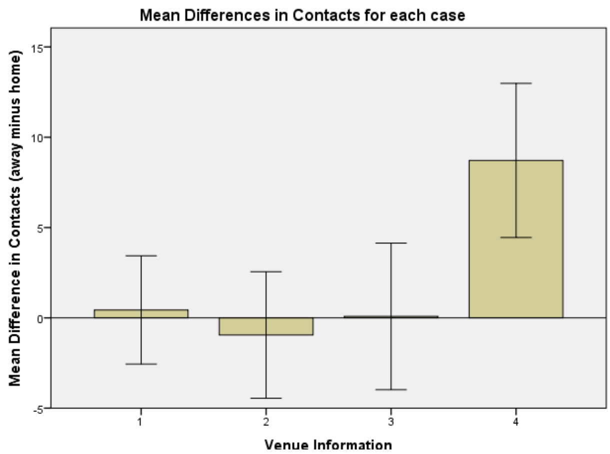


Figure 4. Mean difference in contact counts with error bars for each category

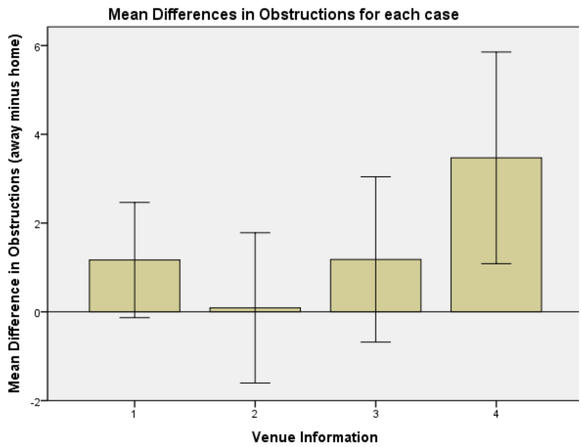


Figure 5. Mean difference in obstruction counts with error bars for each category

These results indicate that more of the overall difference in penalties can be attributed to contacts – approximately twice as much of a difference comes from contacts compared to obstructions.

By comparing the number of obstructions and contacts Australian teams are penalised with in Australia (against sides from New Zealand) versus the same opposition in New Zealand some more information comes to light.

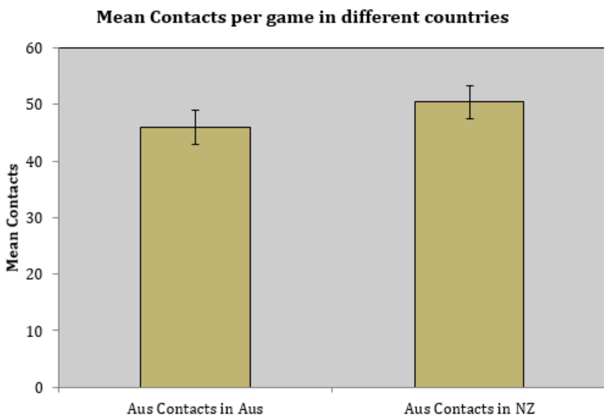


Figure 6. Mean contact counts with error bars for Australian sides against NZ sides

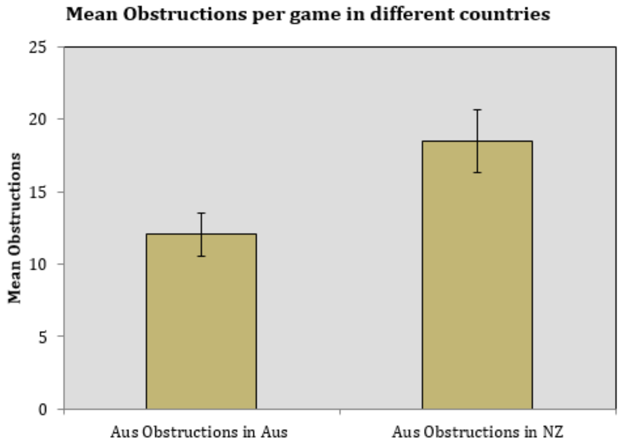


Figure 7. Mean obstruction counts with error bars for Australian sides against NZ

Even though the difference in contact numbers is larger than that of obstructions, the difference in obstruction may be more significant given that there are much less obstruction calls in a match. Australia a penalised with approximately 10% more contacts when playing in New Zealand, as shown in Figure 6. However they concede approximately 50% more obstructions under the same constraints (Figure 7).

4. DISCUSSION

Clearly throughout the analyses it was category 4 that was consistently the discrepancy in the data set. This indicated matches played in New Zealand between one side from New Zealand and one side from Australia. Although the number of home team penalties remained relatively consistent for each category, the away team penalty mean for category 4 was more than 8 penalties higher than any of the other three categories. This is curious particularly because category 3 does not share this property despite it also being an Australia versus New Zealand match up. An average differential of 12 penalties (Figure 3)against the away side is a very large number given an average count is around the 59 mark for any given match.

As expected, the ANOVA and post-hoc tests identified that category 4 was the noticeable discrepancy in the data set, and that it was in fact different to each of the other three categories significantly. In addition, none of the other categories proved to be statistically different from one another. Given that it has been proven than home teams can be inadvertently advantaged by a referee due to the home crowd, we could reasonably expect to see this evident in each category. However it is categories 1 and 3 that best display this behaviour at a weak level. Interestingly category 2 has a mean differential of penalties in the negatives – that is to say that home teams are on average receiving more penalties than the away teams. This only makes the result from category 4 more glaring.

By looking at contacts and obstructions separately, we may be able to better understand how and/or it is that category 4 is so unusual. Both contacts and obstructions contribute to the overall differential in category 4; approximately 8 penalties come from contacts and 8 from obstructions. Again the other three categories are approximately equal for bother contacts and penalties and appear to differ at least somewhat from category 4 (especially in the case of contacts).However this difference could potentially be somewhat misleading. It may be more telling to look at these numbers in the context of how many contacts and obstruction calls there actually are throughout a match (Figures 6 and 7). There are considerably more contacts in a match than there are obstructions generally speaking. This means that the increase in contact counts is less of a leap so to speak. However given the relatively small number of obstructions paid in matches, what may seem like a small increase in face value is actually almost a 50% increase in obstructions that Australian sides concede in New Zealand against the same opposition. Meanwhile the contacts number increases by approximately 10%. It is possible to look at these changes from both perspectives, but whichever way it is observed, category 4 is always the odd one out.

When trying to put a finger on why this is the case, it possible to go down several routes. The most obvious explanation is perhaps the most controversial, in that umpires from New Zealand are being biased against Australian sides (or *for* sides from New Zealand, depending on your perspective). This is a common complaint from some members of the netballing community in Australia. This could potentially be in part attributed to crowd factors, however that would not adequately explain why this issue is only prevalent for that particular category of match. Even arguing that crowds in New Zealand are more 'hostile' towards away teams could not explain this discrepancy since away sides from New Zealand do not experience this differential at all – in fact slightly the opposite.

It's also difficult to attribute this result entirely to a difference in playing style from country to country. Category 3 shows no sign of the extraordinary results that category 4 contain – and both of these scenarios have Australian sides playing against sides from New Zealand.

Another popular view is that the umpiring styles from country to country differ; but again this fails to fully explain why categories 3 and 4 are different. This has also been a popular belief held in the netballing community. One possible way of explaining this might be a combination of the above views. Potentially the Australian game style coupled with the New Zealand umpiring style produces a higher number of penalties. Conversely, the New Zealand game style may not overly infringe on the Australian umpiring style and therefore not draw the extra ire of the umpires. Were this to be the case, the onus would then be on the Australian sides to adjust better to the different umpiring style when playing matches in New Zealand.

It is likely that the discrepancy can be somewhat attributed to a combination of many of these factors, but it is unclear at this stage as to how much each factor plays a part.

5. CONCLUSION

We have shown through looking at matches over the course of four years in ANZ Championship netball that there is some inconsistency in penalty counts – specifically involving New Zealand sides and Australian sides facing off in New Zealand. For matches played in New Zealand, Australian sides average approximately 12 more penalties than their opposition. This difference is made up of approximately 8 contacts and 4 obstructions per match. In contrast, the other categories of matches have an average differential of no more than 3 penalties. Possible explanations for this discrepancy include different umpiring techniques between countries, different playing styles between countries crowd factors and umpiring bias. Further research could attempt to better understand the reasons for this discrepancy and work towards to resolving it. Further investigation into whether the quality of the sides involved has any bearing on these results could also make the problem clearer.

References

Goumas, C. (2014). Home advantage and referee bias in European football. *European Journal of Sport Science, 14*(SUPPL.1), S243-S249.

Nevill, A. M., Balmer, N. J., & Mark Williams, A. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise, 3*(4), 261-272.

Wolfson, S., Wakelin, D., & Lewis, M. (2005). Football supporters' perceptions of their role in the home advantage. *Journal of Sports Sciences, 23*(4), 365-374.

A TEN TEAM AFL FINALS SYSTEM

Anthony Bedford and Steve Gabriel

School of Mathematical & Geospatial Sciences
RMIT University, Melbourne, Australia.

Abstract

In this paper, we present a final ten system for the AFL that utilises a world cup style system. We manage to keep the finals system in a four week window, meeting existing constraints. We evaluate the system via simulation and determine the system's likelihoods, fairness, timely completion and balance.

Keywords: simulation, finals, scheduling

1. INTRODUCTION

The aim of this paper is to create a practical and commercially attractive AFL Final 10 system, for potential use in an 18 to 20 team AFL.
Currently, the AFL consists of 18 teams, with 8 teams contesting the four week finals month concluding with the AFL Grand Final in late September.
The growth of the AFL in recent decades, and some issues with previous finals systems, has meant that various finals systems have been used.

As mentioned in our other paper in these proceedings, prior to 1972, the most common finals system was a four team, three weeks structure. There were twelve teams in the VFL from 1944 until 1986 inclusive.
From 1972 to 1990 inclusive a final five was used, expanding to a final six for three years to 1993. The Macintyre Final 8 was adopted in 1994 and used up to and including 1999. This matched 1st and 8th against each other in Week 1, and 2nd against 7th etc. The two lowest ranked losers would be eliminated in the first week, meaning that individual matches' results did not have predetermined consequences.
Since the year 2000, a new final 8 system replaced the Macintyre Final 8. This is still in use in 2014. The current system will not be explained in this paper, however the probabilities of teams winning the premiership will be referred to, as calculated by Lowe and Clarke (2000).

2. METHODOLOGY

The Final 10 system devised in this paper increases the number of matches in the finals series to 11, from the current nine-match structure. An extra two matches are played in the second week. Recent history of AFL/VFL finals is shown below, in terms of number of finals teams, number of finals, and (line) the percentage of teams contesting the finals.

As in Figure 1, teams are given home game and/or bye privileges based on season finishing positions.

The system brings in a new element to AFL, four of the teams contest a mini-finals within the finals in the first two weeks. The four teams finishing from 3rd to 6th play 'within their group' over the first half (ie two weeks) of the finals, with the top two rewarded with proceeding further. This is similar to a four team World Cup group structure, where each four team group plays round-robin, after which the top two proceed to the round of sixteen.

Byes would be given to the fi teams at the end of the regular season. The top team is allocated as A1, and the team finishing second is denoted A2. These teams are in Group A.

There are three main differences between the World Cup groups process, and the system proposed here:

1. Due to the limitation of a four-week finals system, only four matches rather than six are possible for the group round-robin. That is, each team would play two of the other teams with Group B, rather than all three.
2. In the World Cup, points are allocated 3/1/0 for Win/Draw/Lose respectively, while in

the proposed AFL Group B, points would be 4/2/0 in keeping with the main season.

3. In the World Cup, teams tied on points are separated on Goal Difference. In the AFL system, the percentage system could also be carried through from the main season. Only the scores from the finals would be used to calculate the Group B percentages.

Weeks 3 and 4 would remain as two Preliminary Finals (in effect these are semi-finals) followed by the Grand Final.

Victorian based clubs, or two interstate clubs contesting the flag. Of course if one team is a Melbourne team, and the other an 'interstate' team then this is a home-away match with the appropriate home v away probabilities.

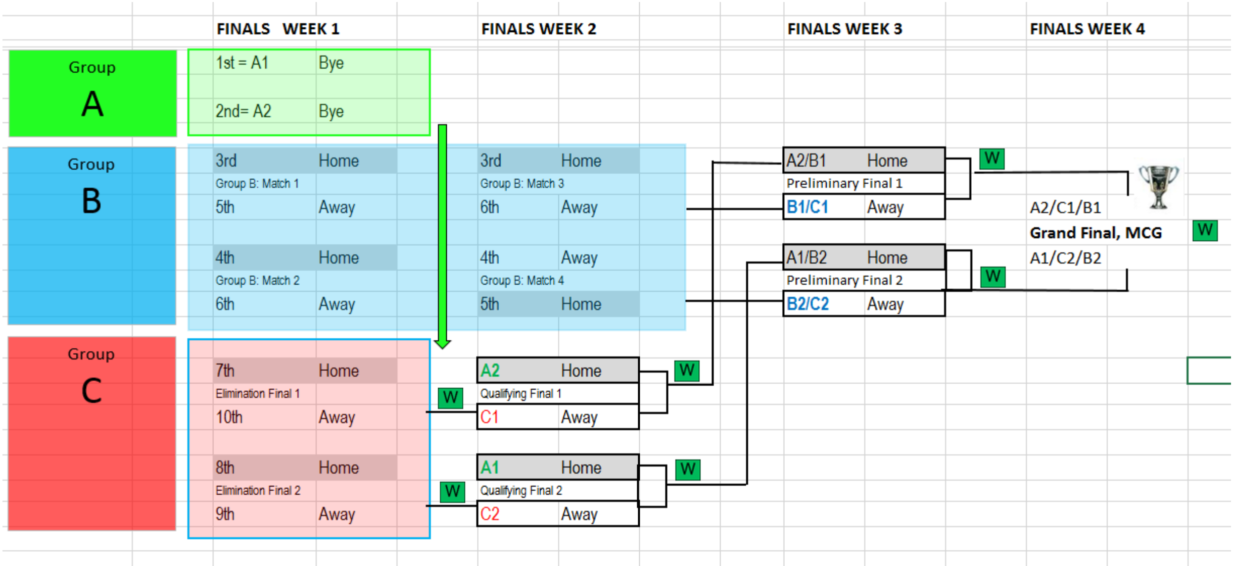


Figure 1: Devised Final 10 system

The AFL currently gives the higher ranked team the benefit of a 'home' final in all finals matches except for the Grand Final, which is played at the MCG. The new system would continue this pattern.

Analysis of the 2013 season reveals a 57.8% win rate for the home team where they have a home-away advantage. This was determined by allocating matches as either home - away (advantage to one team) or neutral (equal chance to both teams) by making the following assumptions:

1. A home advantage is gained only when the home team plays in their home city/state and doesn't need to take a flight and the opposing side is required to fly interstate.
2. All matches not fitting the rules in 1 above are deemed neutral.
3. The Grand Final is played at the MCG: this will be modeled as a neutral match for two

So for the purposes of modeling fairness of the Finals 10 system, the following grounds are pooled together as being effectively located together, with home teams as shown in Figure 2. The remaining boutique grounds used in 2013 were deemed to be neutral, as both teams needed to travel from their home states, and the stadiums were used infrequently. For the finals system the boutique grounds aren't used and so don't come into play. A simple comparison is shown:

	Week1	Week2	Week3	Week4
Current 8	4	2	2	1
Proposed	4	4	2	1

Table 1: Comparison of games played

Code	Team	Location 1			Location 2	Location 3			Location 4	Location 5	Location 6
G10_CAR	Carlton	MCG	ES	SS							
G10_COL	Collingwood	MCG	ES	SS							
G10_ESS	Essendon	MCG	ES	SS							
G10_GEE	Geelong	MCG	ES	SS							
G10_HAW	Hawthorn	MCG	ES	SS							
G10_MEL	Melbourne	MCG	ES	SS							
G10_NME	Nth Melbourne	MCG	ES	SS							
G10_RIC	Richmond	MCG	ES	SS							
G10_STK	StKilda	MCG	ES	SS							
G10_WBD	Western Bulldogs	MCG	ES	SS							
G21_ADE	Adelaide				AS						
G21_PAD	Port Adelaide				AS						
G22_GWS	GWS					SCG	ANZ	SK			
G22_SYD	Sydney					SCG	ANZ	SK			
G23_FRE	Fremantle								PS		
G23_WCE	West Coast								PS		
G24_BRI	Brisbane									G	
G25_GCS	Gold Coast										MS

Abbreviations: ES: Etihad Stadium (Docklands), SS: Simonds Stadium (Geelong), AS: Aami Stadium (Adelaide)
SK: Skoda Stadium (Sydney), PS: Pattersons Stadium (Perth), G: Gabba (Brisbane), MS: Metricon Stadium (Carrera)

Figure 2: Pooled Grounds

Group B consists of the next four teams as ranked at the end of the regular season. They play amongst themselves in the first two weeks, with the top two going through to the third week. This would be a new feature of VFL/AFL finals, a mini-finals within the finals much like a World Cup group where two teams out of four succeed in going through to the 'elimination phase'.

Here, however, each team would play two rather than three of the other teams. These four are awarded 'home' and 'away' fixtures in weeks one and two according to their season ending positions – 3rd is given two home matches which are against 5th and then 6th. The team finishing 6th has no home matches in this phase, and the 4th and 5th teams are allocated one home and one away in the two week round-robin as shown above.

Other combinations of fixtures in this phase would have been possible. In the structure selected, the match-ups of 3rd v 4th and 5th v 6th are avoided.

After the end of Finals Week 2, the top two ranked teams from the group are decided, and allocated as B1 (top) and B2 (second). How are the two Group B winners decided and ranked? Firstly on wins/points whereby 4 points are given for each win. Of course any team achieving 8 points/ 2 wins will always proceed to Week 3, while any teams not winning either game cannot proceed. Percentage from within the finals after the four matches is the next separator. If teams are still equal, then end of season points (and then percentage) would be used. It would be very unlikely for regular season data to be required.

3.RESULTS

The simulation procedure was as follows:

Stage 1.

Teams were allocated a number and a team code which includes two digit number for home game location. For example G10 is for Melbourne/Vic based clubs, and G25 for Gold Coast. The first block of inputs (uniform dist) provides a random end of season ladder, which of course includes random final 10 teams to participate in the 10 team finals.

Team No	Team Code	Team name
1	G10_CAR	Carlton
2	G10_COL	Collingwood
3	G10_ESS	Essendon
4	G10_GEE	Geelong
5	G10_HAW	Hawthorn
6	G10_MEL	Melbourne
7	G10_NME	Nth Melbourne
8	G10_RIC	Richmond
9	G10_STK	StKilda
10	G10_WBD	Western Bulldogs
11	G21_ADE	Adelaide
12	G21_PAD	Port Adelaide
13	G22_GWS	GWS
14	G22_SYD	Sydney
15	G23_FRE	Fremantle
16	G23_WCE	West Coast
17	G24_BRI	Brisbane
18	G25_GCS	Gold Coast

Top 10			
Position	Team Code	Position	Team No
1	G22_SYD	1	14
2	G25_GCS	2	18
3	G21_ADE	3	11
4	G22_GWS	4	13
5	G10_WBD	5	10
6	G23_WCE	6	16
7	G10_MEL	7	6
8	G10_GEE	8	4
9	G10_CAR	9	1
10	G10_RIC	10	8

Stage 2: Finals Week 1

Week 1 of the finals (four matches) are played out with the competing teams allocated either 50%-50% or 57.8%-42.2% depending on whether the matches are neutral or home-away (advantage) using the rules discussed previously.

In Figure 3, Western Bulldogs and GWS win their Group B games. In the elimination finals Richmond and Geelong both survive, while Melbourne and Carlton are eliminated.

						Input x 4	
				Neutral or H-A?	Pr (win)	Rand()	Winner
				1=Neutral; 2=Home_Away			
Group B Match 1	Home	3rd	G21_ADE	G21	0.578		
				2		0.579674	G10_WBD
	Away	5th	G10_WBD	G10	0.422		
Group B Match 2	Home	4th	G22_GWS	G22	0.578		
				2		0.34302	G22_GWS
	Away	6th	G23_WCE	G23	0.422		
Elimination Final 1	Home	7th	G10_MEL	G10	0.5		
				1		0.637153	G10_RIC
	Away	10th	G10_RIC	G10	0.5		
Elimination Final 2	Home	8th	G10_GEE	G10	0.5		
				1		0.43154	G10_GEE
	Away	9th	G10_CAR	G10	0.5		

Figure 3: Stage 2

Stage 3: Finals Week 2

The second week includes another four matches – the deciding third and fourth Group B games, and two qualifying finals which bring together the top two teams (fresh from a bye in Week 1) and the surviving Group C teams.

After appropriate teams are allocated into these finals, the simulation structure is the same as for the first week.

The two losing teams from the qualifying matches are eliminated. In this example, as shown below, Richmond and Geelong will be out, while Gold Coast and Sydney will proceed further.

						Input x 4		
				Neutral or H-A?	Pr (win)	Rand()	Winner	
			Team Cod	1=Neutral; 2=Home_Away			Team Code	Team No.
Group B Match 3	Home	3rd	G21_ADE	G21	0.578			
				2		0.5	G21_ADE	11
	Away	6th	G23_WCE	G23	0.422			
Group B Match 4	Home	5th	G10_WBD	G10	0.578			
				2		0.5	G10_WBD	10
	Away	4th	G22_GWS	G22	0.422			
Qualifying Final 1	Home	2nd	G25_GCS	G25	0.578			
				2		0.5	G25_GCS	18
	Away	7th/10th	G10_RIC	G10	0.422			
Qualifying Final 2	Home	1st	G22_SYD	G22	0.578			
				2		0.5	G22_SYD	14
	Away	8th/9th	G10_GEE	G10	0.422			

Figure 4:Stage 3

In matches 3 and 4 of Group B, Adelaide and Western Bulldogs won. Therefore after all four Group B matches, the points ladder for Group B is:

Teams	Total Points
G21_ADE	4
G22_GWS	4
G10_WBD	8
G23_WCE	0

There are no more matches for these teams in the Group B two week 'mini-series', so two must proceed and two will be eliminated. Here, Western Bulldogs (WBD) has 8 points after winning both of its matches. It will proceed into the preliminary finals week. West Coast (WCE) lost both its matches so will not proceed.

Adelaide and GWS both won one of their matches and are equal on four points. In reality their 'in finals' percentages would decide which proceeds to Week 3. In the simulation, actual scores were not generated, although this could be added relatively easily. Instead, four more inputs are used to decide percentage ranking for the Group B teams. From the two teams in dispute, the one with the higher percentage rank proceed, and the other is eliminated:

All other matches in the finals except for the Grand Final would need to be decided on the match day, so the contingency of extra time would be retained from the current AFL finals system. Currently the AFL's policy for the Grand Final is the drawn matches will be completely replayed on the following weekend – a crowd and revenue bonanza.

Stage 4: Finals week 3, Preliminary Finals

The benefit (where team pairings aren't neutral) of home finals in the preliminary finals goes to the teams from the higher group. The top two teams A1 and A2 then will always play home PFs when they make it to Week 3, while Group B teams will get home finals only where they are competing against Group C teams.

In this example, it is assumed that Gold Coast versus Western Bulldogs is played at Carrara. As the capacity is only 25,000 there, perhaps the match would be played in Brisbane's Gabba (approx. 42,000 capacity) instead however, or even the MCG.

This page decides which two teams 'proceed' from Group B, and designates B1 and B2.

			Week 1	Week 2						
	Teams	Number Codes	Match points (4/0)	Match points (4/0)	Total Points	Percentage Rank	Aggregate for Order	Group B outcome(Rank)	B1/B2/OUT	Team Code
3rd	G21_ADE	11	0	4	4	3	4.33	3	OUT	G21_ADE
4th	G22_GWS	13	4	0	4	2	4.50	2	B2	G22_GWS
5th	G10_WBD	10	4	4	8	1	9.00	1	B1	G10_WBD
6th	G23_WCE	16	0	0	0	4	0.25	4	OUT	G23_WCE

Figure 4: Progression from Group B

						Input x 2		2 x Output
			Team code	Neutral or H-A?	Pr (win)		Winner	
				1=Neutral; 2=Home_Away			Team Code	Team No.
Preliminary Final 1	Home	A2/B1	G25_GCS	G25	0.578			
	B1 plays Home only if opponent is from C group.			2		0.014235	G25_GCS	18
	Away	B1/C1	G10_WBD	G10	0.422			
Preliminary Final 2	Home	A1/B2	G22_SYD	G22	0.5			
	B2 plays Home only if opponent is from C group.			1		0.267459	G22_SYD	14
	Away	B2/C2	G22_GWS	G22	0.5			

Figure 5: Stage 4

However all ‘Home’ teams are assumed to be given the benefit of home city finals up to Week 3, which is relatively commercially practical for all teams *except* GCS. For GWS playing a home final at ANZ Stadium or the SCG, this would be treated as a home – away game except against Sydney.

As shown in Figure 5, Gold Coast and Sydney both won their matches and are rewarded by proceeding to the Grand Final.

Stage 5: Grand Final

			Team code	Neutral or H-A?	Pr (wi	Input		1 x Output	1 x Output
				1=Neutral; 2=Home_Away			Winning Team Code	Winning Team No.	Position in 10
	Home/Away/Neutral	A2/B1/C1	G25_GCS	G25	0.5				
Grand Final	Match has H-A probabilities if one team is VIC, the other non-VIC.			1		0.887503	G22_SYD	14	2
MCG, Melbourne	Away/Home/Neutral	A1/B2/C2	G22_SYD	G22	0.5				

Figure 6: Stage 5

In this example, the match between GCS and SYD is designated as neutral, both teams playing away at the MCG. As shown on the right, the winning team Sydney’s starting position in the finals was 2nd.

3. DISCUSSION

Fairness is the main criterion that needs to be tested. The finals system structure should reward teams for finishing higher on the ladder by the structure of the match pairings and the rules to decide matches locations and therefore the possibility of home team advantage. At the same time, to keep interest in the series, the higher finishing teams should not be virtual certainties to make the Grand Final. In addition, the probabilities ideally wouldn’t change from one finishing position to the next in too large a step.

Previous work by Lowe and Clarke (2000) showed that for the current AFL finals system the teams have the following probabilities of winning the premiership, assuming all games are 50-50 in the first table, and in the home-away advantage model in the second table.

<u>Current AFL 8-team finals system</u>

Equal Probability Model	
Team	Pr (Win) %
1	18.75
2	18.75
3	18.75
4	18.75
5	6.25
6	6.25
7	6.25
8	6.25

Home ground advantage	
Team	Pr (Win) % *
1	20.7
2	20.7
3	18.5
4	18.5
5	5.20
6	5.20
7	3.90
8	3.90

* median of groups

The top four teams in the current system are given considerably higher chance of winning the GF.

In the home ground advantage model, the gap between 4th and 5th chances of winning the premiership is 13.3 percentage points.

The proposed Top 10 Finals system resulted in the following probabilities, applying models based on the models of Lowe and Clarke (2000).

The equal probability model assumes 50%-50% chances in all matches for the competing teams;

	Pre-finals position	Matches to win	Pr (premiership)
Group A	1st	3	0.5 ^ 3 = 0.125
	2nd	3	0.5 ^ 3 = 0.125
Group B	3rd	3 or 4	0.5 ^ 3 = 0.125
	4th	3 or 4	0.5 ^ 3 = 0.125
	5th	3 or 4	0.5 ^ 3 = 0.125
	6th	3 or 4	0.5 ^ 3 = 0.125
Group C	7th	4	0.5^4 = 0.0625
	8th	4	0.5^4 = 0.0625
	9th	4	0.5^4 = 0.0625
	10th	4	0.5^4 = 0.0625
			Total = 1.000

The Group B teams have a 50% chance of making it through the two-week ‘group stage’ as there is no home-away advantage and half of the four teams will proceed.

A simulation of 50,000 iterations was run to determine probabilities with the home-away advantage where one team travels while the other doesn't. The procedure was detailed previously. Groupings of teams to factor in home-away advantage are different to Lowe and Clarke.

The results for all teams combined were:

Ladder	Pr(Win)	diff
1	15.6%	
2	15.3%	0.3%
3	13.5%	1.8%
4	12.5%	1.0%
5	12.3%	0.2%
6	10.8%	1.5%
7	5.5%	5.3%
8	5.4%	0.1%
9	4.5%	0.9%
10	4.6%	-0.1%
	100.0%	

Dividing the teams into three groups based on location:

Type 1 teams (Victorian based): CAR, COL, ESS, GEE, HAW, MEL, NME, RIC, STK, WBD

Type 2 teams (located in two-team cities): WCE, FRE, ADE, PAD, SYD, GWS

Type 3 teams (single-team locations): GCS, BRI

Type 1 teams

Ladder	Pr(Win)
1	15.2%
2	15.1%
3	13.5%
4	12.5%
5	12.3%
6	10.9%
7	5.6%
8	5.5%
9	4.7%
10	4.8%
	100.0%

Type 2 teams

Ladder	Pr(Win)
1	16.1%
2	15.6%
3	13.4%
4	12.6%
5	12.1%
6	10.7%
7	5.5%
8	5.3%
9	4.1%
10	4.4%
	100.0%

Type 3 teams

Ladder	Pr(Win)
1	16.3%
2	15.5%
3	13.9%
4	12.5%
5	12.8%
6	10.6%
7	5.1%
8	4.7%
9	4.4%
10	4.2%
	100.0%

For type one and type two teams, the 10th team has a slightly higher chance than 9th of winning the premiership using the simulation results as shown above. More iterations would be required to confirm this however.

4. CONCLUSIONS

A relatively fair ten-team AFL finals system within four weeks has advantages over the current finals system. The main ones are the extra two matches in the second week, and the much better deal for the fifth and sixth in particular, which have significantly higher probabilities of winning the premiership, and are guaranteed at least two weeks in the finals (compared to one currently).

Also, the Final 10 system devised rewards the top two with a week off in Week 1, whereas currently they have little advantage over third and fourth. The 10 team system is less predictable than the current 8 team system, while also being fairer when comparing teams relative starting positions.

Also, of course, 9th and 10th would have some chance in this system, compared to missing out completely in any Final 8 system. While they would have a low chance of winning the main prize, it's not unlikely they could have an impact beyond Week 1, especially for a fast finishing team that perhaps had early season injuries.

The simulation as run could be improved in several ways, as already mentioned. Accuracy could be improved, and realism could be improved in Group B percentages for example. Also, more work could look at changing the Group B match-ups to see how that would affect probabilities.

References

Lowe, J. & Clarke, S.R.(2000). An Analysis of the AFL Final Eight System. *The 5th Australasian Conference on Mathematics and Computers in Sport.*

AFL Record Season Guide 2013, AFL Media

2013 AFL Record Grand Final edition, Slattery Media.

APPLIED STATISTICS FOR SPORTS PSYCHOLOGY STUDENTS

"IF YOU BUILD IT THEY WILL COME - OUR SPORTS STATISTICS FIELD OF DREAMS"

Steven J Bowe[1], Denny Meyer[1], Paul Bracewell[2], and Karl Jackson [3]

[1] Swinburne University of Technology, Melbourne.
[2]CricHQ, Wellington, New Zealand.
[3] Champion Data, Melbourne.

Contact author: Steven J Bowe, e-mail: sbowe@swin.edu.au

Abstract

Sports Statistics texts tend to focus on mostly introductory statistics and probability. Those textbooks available generally are very USA sports focused, despite the increasing availability of large sports data sets from other world sports. An invitation to design and teach sports statistics to a group of sports psychology students, with a good background in data analysis methodologies, provided an opportunity to explore a range sporting data sets, including: cricket, basketball, tennis, Australian Rules Football, soccer and rugby union.Also, to introduce statistical methods and toolsused mostly in manufacturing (quality control)and economics (forecasting) that arebeing applied to monitor and predict the performance of elite athletes and teams. This unique 12 week unit integrates real examples and real data. The unit is delivered face-to-face through a mixture of lectures (including guest speakers i.e. Champion Data Statistician) and computer lab tutorials. Assessments involved online computer lab quizzes, journal articles, group projects with a presentation and an examination related to all topics. Evaluation of student feedback at the end of this unit was very positive and encouraging, especially from those with a strong interest in sports performance.

Key Words: sports data, monitoring performance, forecasting, rankings and ratings.