ARTICLE POSTPRINT

Using performance data to identify styles of play in netball: An alternative to performance indicators

Hayden Croft[a] and Bobby Wilcox[b] and Peter Lamb[c]

[a]Institute of Sport and Adventure, Otago Polytechnic, Dunedin, New Zealand; [b]New Zealand Netball, Auckland, New Zealand; [c]School of Physical Education, Sport and Exercise Sciences, University of Otago, Dunedin, New Zealand

ARTICLE HISTORY
Compiled December 18, 2017

ABSTRACT
The advent of sports technology has led to large, high-dimensional, performance data sets, which pose decision-making challenges for coaches and performance analysts. If large data sets are managed poorly inaccurate and biased decision-making may actually be enabled. This paper outlines a process for capturing, organising, and analysing a large performance data set in professional netball. 250 ANZ Championship matches, from the 2012–2015 seasons, where analysed. Self-organising maps and a *k*-means clustering algorithm were used to describe seven games-styles, which were used in a case study to devise a strategy for an upcoming opponent. The team implemented a centre-pass (CP) defence strategy based on the opponent's previous successful and unsuccessful performances. This strategy involved allowing the oppositions Wing-attack to receive the CP while allowing their Goal-attack to take the second pass. The strategy was monitored live by the coaches on a tablet computer via a custom-built dashboard, which tracks each component of the strategy. The process provides an alternative to use of conventional performance indicators and demonstrates a method for handling large high-dimensional performance data sets. Further work is needed to identify an ecologically valid method for variable selection.

KEYWORDS
performance analysis; netball; self-organising maps; performance indicators; performance data; dimensionality reduction

1. Introduction

Technology and the data it can generate poses a challenge for performance analysts and coaches in sport. Performance analysis software, such as Sportscode^{TM} (Hudl, Lincoln, Nebraska), allows coaches to sort through and analyse game events. Commercial entities, such as Opta^{TM} (London, United Kingdom) and Prozone^{TM} (Stats

CONTACT H. Croft. Email: hayden.croft@op.ac.nz

LLC, Chicago, Illinois), collect performance data describing various actions performed, where those actions took place on the pitch, who was involved, and other descriptive information, such as the outcome of the action (O'Donoghue & Holmes, 2015). The number of data points describing players' positions and performance data alone do not necessarily pose a problem for the analyst, with modern database filtering and sorting techniques along with improved computer processor speeds and memory. The increase in dimensionality, or the number of variables, describing performance, however, drastically increases the complexity in sorting through data sets to find important information (Bellman, 2013).

According to M. Hughes and Bartlett (2002), most coaches often review match statistics to reinforce their opinions, rather than to inform, on events they remember from certain matches. When deciding on strategies for upcoming matches, coaches are likely constrained by their experiences. More generally, during decision-making, individuals often identify solutions based on familiar situations with a known solution from their past (Nash & Collins, 2006). There is now an opportunity for coaches to improve decision-making with performance data if it aligns with their way of thinking and the constraints they face. Given the increased availability of high-dimensional performance data, as well as coaches' propensities to utilise new information to confirm their bias's, finding a way of capturing, organising and presenting this information back to coaches in an effective way remains a challenge for analysts. There are many ways matches can be coded to represent individual and team performance (M. Hughes & Bartlett, 2015). Often actions are coded with other corresponding information, such as players involved, location on the court, or time of the match. Frequency tables and charts are, in most cases, limited to two variables, which requires a large number of charts to compare all variable combinations. More importantly, these two-way comparisons mask higher-dimensional relationships among the match variables and is a downfall of using performance indicators.

Many sports are well suited to collecting performance data. This paper will focus on netball, which is a team, court sport, with many similarities to basketball. Some of the main differences between netball and basketball include the following: the player in possession of the ball cannot move, there is no backboard behind the hoop and there are restrictions on where players can move and shoot the ball. Netball is played in mostly Commonwealth countries, with the majority of participants being female. As with other court sports, research has focussed primarily on identifying key performance indicators biomechanical assessment of technique (Delextrat & Goss-Sampson, 2010; O'Donoghue, Mayes, Edwards, & Garland, 2008). Normative data describing performance indicators are informative; however, describing performance with summary statistics covers up the interaction between the teams over the course of a match. That norms for certain performance indicators have been established does not necessarily imply that they are desired in specific matches, which may explain the hesitancy of coaches to use normative data to inform their game strategies (Nash & Collins, 2006).

Self-organising maps (SOMs) present an opportunity to characterise the high-dimensional interaction between sports teams (see Croft, Lamb, & Middlemas, 2015; Lamb & Croft, 2016, for applications in rugby union). SOMs are a type of neural network useful for clustering and visualising high-dimensional information on a low-dimensional output map (Kohonen, 2013). SOMs enable match performance of one team to be compared, in the context of all matches in the dataset, to the performance of the opponent. Importantly, because of the neighbourhood function and competitive learning strategy, the original topology of the input distribution is preserved (Vesanto & Alhoniemi, 2000) – unlike approaches based on means and statistical benchmark

Netball Performance Variables	
Shooting Accuracy (%)	Centre Pass Reception (Goal Defence, 1st)
Possession Conversion (Centre Pass to Circle, Total)	Centre Pass Reception (Goal Defence, 2nd)
Possession Conversion (Centre Pass to Score, Total)	Centre Pass Accuracy (Successful)
Possession Conversion (Turnover to Circle, Total)	Centre Pass Accuracy (Attempts)
Possession Conversion (Turnover to Score, Total)	Centre Pass Accuracy (%)
Offensive Rebounds	Feeding into the Circle (Goal Attack, Successful)
Defensive Rebounds	Feeding into the Circle (Goal Attack, Attempts)
Total Losses	Feeding into the Circle (Wing Attack, Successful)
Feeding Accuracy (%)	Feeding into the Circle (Wing Attack, Attempts)
Centre Pass Reception (Goal Shoot, 2nd)	Feeding into the Circle (Centre, Successful)
Centre Pass Reception (Goal Attack, 1st)	Feeding into the Circle (Centre, Attempts)
Centre Pass Reception (Goal Attack, 2nd)	Penalties Conceded (Centre)
Centre Pass Reception (Goal Attack, Total)	Penalties Conceded (Wing Defence)
Centre Pass Reception (Wing Attack, 1st)	Penalties Conceded (Goal Defence, In Circle)
Centre Pass Reception (Wing Attack, 2nd)	Penalties Conceded (Goal Defence, Out of Circle)
Centre Pass Reception (Wing Attack, Total)	Penalties Conceded (Goal Keep, In Circle)
Centre Pass Reception (Centre, 2nd)	Penalties Conceded (Goal Keep, Out of Circle)
Centre Pass Reception (Wing Defence, 1st)	Gains (Intercepts/Tips)
Centre Pass Reception (Wing Defence, 2nd)	Gains (Opposition Error)
Centre Pass Reception (Wing Defence, Total)	Gains (Rebounds)

Table 1.: Forty variables selected for analysis with the SOM algorithm.

values.

This case-study reports the workflow and coordination between the sports scientist and the coach in developing and executing tactical plans in professional netball. We also demonstrate the role of SOMs in reducing match statistics down to game styles and feeding back information to the coach, during the match, to enable tactical adjustments based on the game style coupling of the two teams.

2. Methods

2.1. *Data collection*

Notational data were manually coded by the second author using spreadsheets and video footage from each match in the ANZ Championship from the 2012 to 2015 seasons. In total, 250 matches were notated and 124 match variables defined for each team across all matches in the competition. Examples of these variables include:

- playing positions for first and second centre pass receptions (frequency);
- centre pass to score attempts and completions (frequency and percentage);
- feeding into the shooting circle accuracy and success (frequency and percentage);
- shooting attempts and success (frequency and percentage);
- turnovers gained and conceded (frequency).

Based on the recommendations of Vincent, Stergiou, and Katz (2009) a database was created to allow an organised data structure that could be used to select and analyse any combination of matches from the four seasons collected. Data representing 40 match performance variables were selected from the original set of 124. Two coaches and the performance analyst, from the case study team, selected the 40 variables that best represented performance in their opinions as seen in Table 1. This was both a

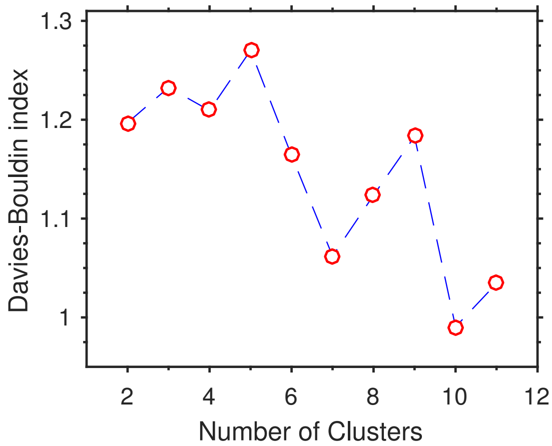

Figure 1.: The Davies-Bouldin Index for clusters 2, . . ., 11.

strength and a weakness of the process as it introduced biases, however created better ecological validity for the team featured below. Each match consisted of two team performances, resulting in a [500 40] input data matrix for SOM training.

2.2. *Self-organising map procedures*

The basic architecture of SOMs consist of an output layer of nodes connected to an input layer of nodes. The input nodes are represented by input vectors, which, in this case, represent a set of match performance data. Therefore, input vector, x_i, represents the ith match performance in the input matrix. Each node on the output map has an associated weight vector with the same dimensionality, $d = 40$, as the input. Data were normalised linearly to a range of [0, 1]. The SOM was batch-trained with 12 rough-training and 36 fine-tuning steps, resulting in a 14 row by 8 column output map (see Lamb, Bartlett, Lindinger, & Kennedy, 2014, for details). Training parameters were guided by minimising quantisation and topographical errors (Kaski & Lagus, 1996). All SOM procedures were performed in MATLAB (R2016a, The MathWorks, Inc., Natick, USA); procedures used in this analysis incorporated functions in the SOM Toolbox (Alhoniemi, Himberg, Parviainen, & Vesanto, 2012).

2.3. *Game style clustering*

The nodes on the map tend to cluster as a combined result of a) the number of map nodes being less than that of the input nodes and b) the competitive learning strategy and neighbourhood function, which are key features of the SOM algorithm. Furthermore, because of these features in the algorithm, similar map regions tend to represent similar input data. Therefore, we partitioned the nodes into clusters, which correspond to different game styles. The k-means clustering algorithm was run several times for different number of clusters, from 2 to $\sqrt{m} = 11$, where m is the number of map nodes (Vesanto & Alhoniemi, 2000). We decided that seven clusters balanced the trade-off between minimising the Davies-Bouldin Index (Davies & Bouldin, 1979) and maintaining sufficient input node representation within each cluster to analyse specific team performances (see Figure 1). The interpretation of the clusters is explained in the next section.

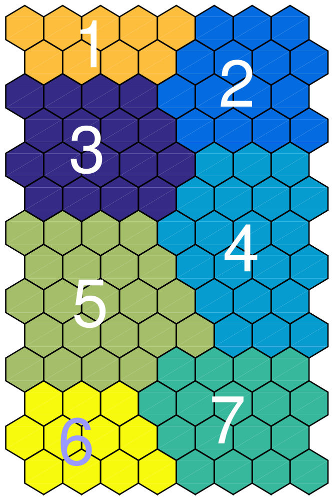

Figure 2.: Cluster partitioning for 7 clusters using k-means algorithm. Numbers indicate cluster membership.

The output map in Figure 2 is shown as a hexagonal lattice of nodes, with the clusters identified and numbered accordingly. Any given input vector can be visualised on the output map by identifying its best-matching node in the output, which is defined as the node whose weight vector has the shortest Euclidean distance to the respective input vector. We highlight these best-matching units using hit histograms, for which the size of the marked node indicates the number of inputs that node best-matches (Figure 3).

3. Map interpretation

3.1. *Seven netball game styles*

The game styles, or clusters, were interpreted by viewing the individual components (i.e. variables) and the distribution of their values across the output map (three example components are shown in Figure 4). Clusters were characterised by an expert international performance analyst and the lead author, as described below.

Game style 1: Safety first

Wing Attack (WA) and Goal Attack (GA) play a dominant role; both are highly involved in Centre Pass (CP), while Centre (C) and GA are dominant feeders. This game style involves very accurate feeding, shooting and offensive rebounds, and low loss of possession rate. CP and Turnover (T/O) conversion rates are very high. Similar to game style 3 (see below), however, this style is not as effective defensively evidenced by low intercept rate, and a high in-circle penalty count.

Game style 2: A strong attacking style

Similar to game style 1, WA and GA play a dominant role. Matches represented by this style are consistent with low loss of possession rates leading to high conversion rates with good shooting accuracy and a high frequency of offensive rebounds. This style is less effective defensively than other styles mainly due to a low frequency defensive

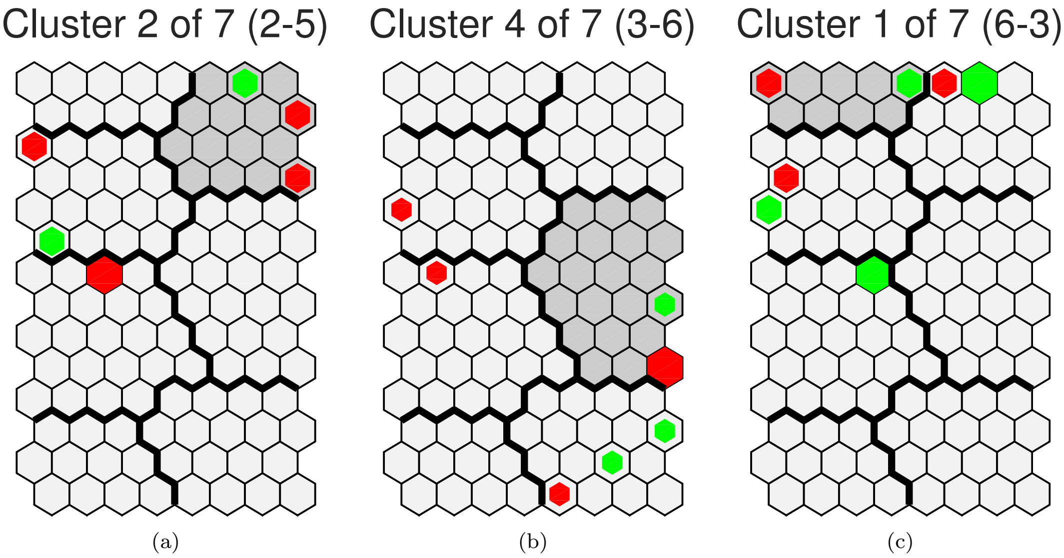

Figure 3.: Best-matching nodes for the performance of “Team A’s” opponent for games in which Team A played the specified cluster: a) Cluster 2, b) Cluster 4 and c) Cluster 1. Red nodes indicate wins by Team A and green indicates wins by the opponent. Size of the highlighted node reflects the relative number of matches represented. For example, in b) the large red hexagon represents three matches and all other coloured hexagons represent one match.

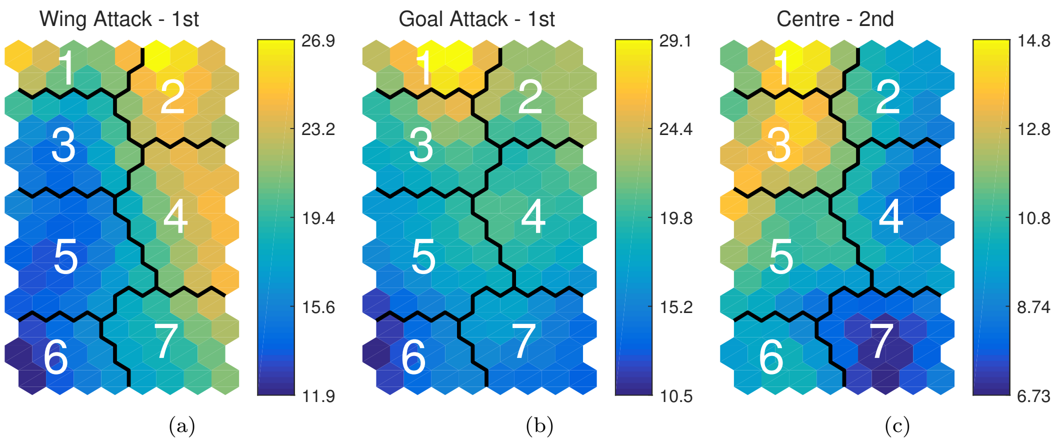

Figure 4.: Component planes visualisation for three example variables related to centre passing: a) centre pass receptions for Wing Attack on the first pass, b) centre pass to Goal Attack on the first pass and, c) centre pass to Centre on the second pass.

rebounds and few errors by the opposition. WA and GA are both highly involved in the CP, with WA being the dominant circle feeder.

Game style 3: Another 'safety first' style of playing

This game style is characterised by a high frequency of back passes off CP, with frequent involvement of C and Wing Defence (WD). Feeding is very accurate, leading to a high CP to score conversion; there is also a moderate T/O to score conversion and a low loss of possession rate. Shooting accuracy and offensive rebounding success is high indicating very effective play once the ball enters the circle. Defensively, this style is not as effective as other styles, show by a low intercept rate and low defensive rebounding success. Despite this style's relatively poor defensive performance, this is generally a successful style owed to its offensive effectiveness.

Game style 4: Reasonably balanced style

In this style, the WA is dominant on the centre pass, and is a main feeder. Matches represented by this style display effective defensive tactics with high intercept rates and a high frequency of possession gains through opposition errors. Some risk is taken with moderate loss in rates in possession. This style, however, is not as effective as other styles in the shooting circle, with a low frequency of offensive rebounds and moderate shooting accuracy.

Game style 5: A low scoring-low loss rate style

This game style can be characterised by low involvement on the CP from WA, GA and Goal Shoot (GS), and only moderate involvement from Wing Defence (WD) – so no clear pattern of where the CP is going. Low CP involvement seems to indicate that this is simply a low scoring style. A low loss rate overall combined with high offensive rebounds results in moderate CP and T/O conversion and shooting accuracy is quite variable.

Game style 6: A high-risk game style

This game style is associated with a very dominant defence. On attack WD and Goal Defence (GD) receive many CPs; however, feeding is inaccurate. Shooting accuracy is low and loss of possession rates are high, indicating very low CP and T/O to score conversion rates. Defensively, this style excels in intercepting the ball, while conceding very few penalties in the circle.

Game style 7: GA plays a second shooter role

In game style 7, the GA has a low involvement in the CP, and feeds the ball into the circle very infrequently. WA does most of the feeding into the circle with C not highly involved in feeding. Defensively, this style is associated with moderate loss rates, low shooting accuracy and low offensive rebounding success, leading to low CP and T/O to score conversion.

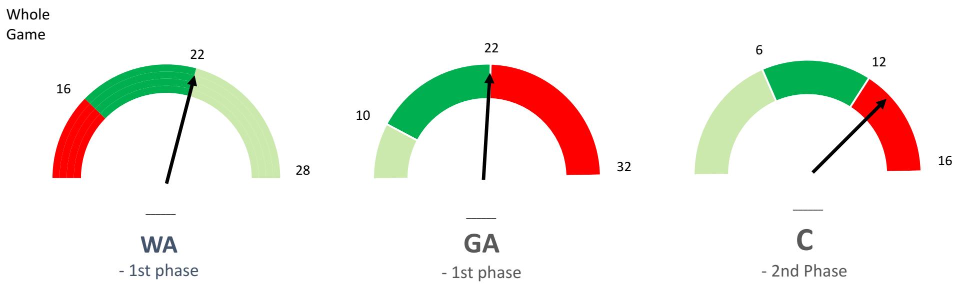

Figure 5.: Dashboard used to track three example variables, with respect to game style driven game strategy, live during a match.

3.2. *Case study – centre pass strategy*

Visualising wins and losses for an example team ("Team A"), for each of the seven game styles and the performance of their opponent (Figure 3) for finding game styles that were more or less successful, based on their win-loss record. Figure 3 shows that when "Team A" played a game represented by cluster 1 their record was 6–3, compared to their less successful styles in clusters 2 and 4, resulting in records of 2–5 and 3–6, respectively. These three game styles were chosen for further analysis, in an effort to devise a strategy related to defeating "Team A". Further inspection of the coloured cells (i.e. hexagons), which represent the opponent's performance, may provide further insight into the coupling between teams, although this case study is focussed on the preparation for an upcoming match against "Team A".

To develop a strategy, the 40 component planes diagrams (of which, three are shown in Figure 4), were examined by the analyst to identify which variables best distinguished these game styles. Three variables showed a visually apparent difference in value between the identified clusters:

- Wing Attack – 1st phase reception (WA1)
- Goal Attack – 1st phase reception (GA1) and
- Centre – 2nd phase reception (C2)

These three variables were chosen to distinguish between the successful performance associated with cluster 1 for Team A and the less successful performance associated with clusters 2 and 4 (Figure 4).

As a result, a strategy was devised by the coach to prevent the GA1 and C2, by instructing two defenders to "mark" each of those players at the appropriate time. WA1 was allowed to occur by not "marking" this player. To track progress during the match a dashboard was created as seen in Figure 5 to indicate whether the team had successfully (dark green) or unsuccessfully (red) implemented the strategy. Figure 5 indicates the execution of this strategy was "border line" which interestingly coincided with a 1-goal win to the team.

3.3. *Summary of the Process*

This paper has outlined a process to handle large multi-dimensional data sets, making them interpretable, unbiased and useful to the coaches during a game. There were multiple stages involved, which included:

(1) Capturing and organising data;
(2) Organising games topographically with a SOM algorithm;
(3) Grouping these games with k-means clustering;
(4) Describing each of seven groups using the 40 variable maps;
(5) Plotting wins and losses for a team for each of the seven groups;
(6) Using the less successful game style to devise a strategy to defeat the opponent;
(7) Using a simple dashboard to measure the effectiveness of the strategy.

4. Discussion

As discussed earlier coaches have a propensity to use information to confirm their current beliefs (Mercier & Sperber, 2011). This phenomenon, known as confirmation bias, leads to decision-making based on samples of the evidence, potentially taken out of context, which unfairly represent the situation. This means that the recent growth in data availability may be to the detriment of coach decision-making as easier access to information will allow greater confirmation bias. There are many reasons that coaches might look to reinforce their biases; however, it is fair to consider that pressures associated with job retention, defending their decisions, ideas, theories and philosophies (Cassidy, Jones, & Potrac, 2009) would all contribute. This paper presents a process to help reduce biases and improve decision-making.

The detailed descriptions above, of the seven netball game styles, is a useful resource for coaches and players who are interested in better understanding professional netball in Australasia. Additionally, the use of SOMs within a wider process provides an alternative approach to the use of performance indicators in netball, as seen in other research (O'Donoghue et al., 2008). Game styles require further research as to identify how each interact with and affect others during a game, as previously explored in rugby union (Lamb & Croft, 2016) providing counter strategies for teams playing specific styles.

Recent research (A. Hughes, Barnes, Churchill, & Stone, 2017) still reports the predictors of success based on one or two PIs, yet research into SOMs (Lamb & Croft, 2016) have shown, that for teams that measure well against performance indicators, some counter-strategies are less successful. The case study in this paper showed a strategy to defeat a specific opponent based on their past performances. The complex interaction between three variables (GA1, WA1 and C2) showed that a successful performance requires a focus on more complex strategies that are not evident when traditional performance indicators are used. This could explain why many studies using performance indicators often only identify a small number of significant results (A. Hughes et al., 2017; Jones, Mellalieu, & James, 2004; Kraak & Welman, 2014). The interaction between certain performance indicators needs to be investigated further, as linear, inverse and mixed multi-dimensional relationships may provide more than one strategy for winning.

5. Conclusion

This paper builds on earlier work (Croft et al., 2015; Lamb & Croft, 2016) by not only applying SOMs to a different sport, but also introducing a further development around coach focused variable selection and introducing a live dashboard to track a strategic finding from the SOM analysis. This process is one that other sports could adopt and provides an alternative to the traditional performance indicator approach first described by M. Hughes and Bartlett (2002).

Secondly, this paper concludes that further work is required to develop a method for choosing which variables to use in the SOM analysis. Although there are several techniques that can be implemented for dimension reduction including, principle component analysis, linear discriminant analysis and canonical correlation analysis (Shaw & Jebara, 2009), these approaches do not provide strong ecological validity and do not necessarily align with the areas of the game that coaches can either control or are interested in. The problem with letting a coach select the dimensions or variables that are analysed, despite being more valid, is that is preserves their biases and includes them in the analysis process.

References

Alhoniemi, E., Himberg, J., Parviainen, J., & Vesanto, J. (2012). *SOM Toolbox 2.1 for MATLAB.* Retrieved from `https://github.com/ilarinieminen/SOM-Toolbox`

Bellman, R. E. (2013). *Dynamic programming.* Mineola, USA: Dover Publications, Inc.

Cassidy, T., Jones, R., & Potrac, P. (2009). *Understanding sports coaching: The social, cultural and pedagogical foundations of coaching practice* (2nd ed.). Abingdon, UK: Routledge.

Croft, H., Lamb, P., & Middlemas, S. (2015). The application of self-organising maps to performance analysis data in rugby union. *International Journal of Performance Analysis in Sport*, *15*, 1037–1046.

Davies, D., & Bouldin, D. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*(2), 224–227.

Delextrat, A., & Goss-Sampson, M. (2010). Kinematic analysis of netball goal shooting: A comparison of junior and senior players. *Journal of Sports Sciences*, *28*(12), 1299–1307.

Hughes, A., Barnes, A., Churchill, S., & Stone, J. (2017). The use of performance indicators that discriminate winning and losing in elite men's and women's rugby union. *International Journal of Performance Analysis in Sport*, *17*(4), 534–544.

Hughes, M., & Bartlett, R. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences*, *20*(10), 739–754.

Hughes, M., & Bartlett, R. (2015). The use of performance indicators in performance analysis. In M. Hughes & I. Franks (Eds.), *Essentials of performance analysis in sport* (2nd ed., pp. 89–112). Abingdon, UK: Routledge.

Jones, N., Mellalieu, S., & James, N. (2004). Team performance indicators as a function of winning and losing in rugby union. *International Journal of Performance Analysis in Sport*, *4*(1), 61–71.

Kaski, S., & Lagus, K. (1996). Comparing self-organizing maps. In *International Conference on Artificial Neural Networks* (pp. 809–814). Berlin: Springer.

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52–65.

Kraak, W. J., & Welman, K. E. (2014). Ruck-play as performance indicator during the 2010 Six Nations Championship. *International Journal of Sports Science & Coaching*, *9*(3), 525–537.

Lamb, P., Bartlett, R., Lindinger, S., & Kennedy, G. (2014). Multi-dimensional coordination in cross-country skiing analyzed using self-organizing maps. *Human Movement Science*, *33*,

54–69.
Lamb, P., & Croft, H. (2016). Visualizing rugby game styles using self-organising maps. *IEEE Computer Graphics & Applications*, *36*(6), 11–15.
Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–111.
Nash, C., & Collins, D. (2006). Tacit knowledge in expert coaching: Science or art? *Quest*, *58*, 465–477.
O'Donoghue, P., & Holmes, L. (2015). *Data analysis in sport (Routledge studies in sports performance analysis).* Abingdon, UK: Routledge.
O'Donoghue, P., Mayes, A., Edwards, K., & Garland, J. (2008). Performance norms for British National Super League Netball. *International Journal of Sports Science & Coaching*, *3*(4), 501–511.
Shaw, B., & Jebara, T. (2009). Structure preserving embedding. In *Proceedings of the 26th Annual International Conference on Machine Learning.* Montréal, CAN.
Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, *11*(3), 586–600.
Vincent, J., Stergiou, P., & Katz, L. (2009). The role of databases in sport science: Current practice and future potential. *International Journal of Computer Science in Sport*, *8*(2), 50–66.