Using contextual player movement and spatial control to analyse player passing trends in football

Bart Spencer[1*], Matthew Hawkey[12], Sam Robertson[1]

[1] Institute for Health & Sport (IHeS), Victoria University, Australia
[2] Professional Referees Organization (PRO), USA
[*] Corresponding Author: bartholomew.spencer@vu.edu.au

Abstract

This study explored the fitting of player motion models in football using contextual player movement behaviours. Measurement of a team's spatial control using player motion models fit on average player displacements underestimates a player's ability to perform high effort displacements during moments of importance (such as winning the ball). We address this by producing *commitment*-based motion models that model the effects of player momentum on their likelihood to attempt to win the ball (via interceptions). Motion models were fit on approximately 46000 samples of player movement relative to the ball across 264 matches from the 2018 MLS season. This approach to motion modelling incorporates contextual movement behaviour whilst remaining low-dimension, hence is intuitive and interpretable to non-technical staff. Resultant models are used to estimate a player's spatial influence over their surroundings, given their momentum and proximity to the ball. Player influence was used to calculate the attacking team's spatial control which serves as a proxy for passing risk – that is, a pass is deemed high risk if passed to a region of low control. It was found that attacking players are the senders and receivers of high risk passes due to their positioning in areas of opposition control. Furthermore, this analysis was able to differentiate the team's star player (as measured via goals and xG throughout the season), who was the receiver of the highest risk passes. This analysis has applications in player profiling, tactics, and recruitment.

Keywords

Spatiotemporal, Player motion, Passing networks, Soccer, Performance analysis, MLS

1.0 Introduction

The concept of spatial control has been a recurrent theme in football literature. Utilisation of space has been linked to successful outcomes (Rein, et al., 2017), with high performing players shown to generate valuable space (Fernandez & Bornn, 2018). Furthermore, space creation has been used as a method for evaluating the quality of passes (Horton, et al., 2015). Outside of football, applications of spatial control include valuation of court locations in basketball (Cervone, et al., 2016) and evaluation of decision-making in Australian football (Spencer, et al., 2019a).

Early approaches to the measurement of space control produced individual dominant regions (DR) – Voronoi-like bounded regions of space assigned to the player that could reach them before any other (Gudmundsson & Horton, 2017). Recent applications of dominant regions have incorporated the effects of player motion via physics-based modelling (Taki & Hasegawa, 2000; Fujimaru & Sugihara, 2005) or via modelling of observed displacements to produce realistic representations of player movement, given the effects of player momentum (Horton, et al., 2015; Brefeld, et al., 2018).

This discrete representation of space fails to model the contests that develop by the time the ball is moved within a player's DR. A continuous representation of space such as in Fernandez and Bornn (2018) or Spencer et al. (2019a) is more logical. These approaches consider the varying spatial influence of individual players to quantify the attacking team's spatial (or pitch) control. Measuring space in this manner can be achieved by using player *motion models*. Motion models measure the effects of a player's prior motion (e.g., velocity and orientation) on future displacements. In short, they answer the following – what are the effects of player momentum on their future movements?

Recent approaches to modelling player motion have involved sampling relative displacements over different time intervals, thereby learning the bounds (Horton, et al., 2015) or distribution (Brefeld, et al., 2018) of possible displacements. The bounding method exemplified by Horton et al. (2015) assumes equal likelihood of displacements which is unrealistic (Brefeld et al., 2018; Spencer, et al., 2019b). Brefeld et al. (2018) addressed this by fitting the distribution of player displacements using Kernel Density estimation (KDE). A limitation of the existing approaches is a lack of consideration for movement context. Whilst momentum may permit a future displacement, the likelihood of the player performing this action is dependent on contextual features such as the ball, field position and tactical behaviour. For example, a player may have a higher likelihood of performing a high effort displacement when attempting to win the ball.

The objective of this study was to develop player motion models that are representative of movement behaviour relative to possession outcomes (i.e., winning the ball). This approach has previously been exemplified on the movements of Australian footballers, where it was found that contextual motion models produced different models to those fit on average player displacements (Spencer, et al., 2019b). Resultant models are used to compute the spatial control of teams on a continuous scale that considers the proximity of players from the ball (such as in Fernandez & Bornn, 2018), while also considering their momentum. Applications of these measures are exemplified in the analysis of passing.

2.0 Methods

2.1 Data & pre-processing

Optical tracking datasets were collected by Metrica Sports for select matches from the 2018 Major League Soccer (MLS) season. A total of 264 matches were used to produce player motion models, while passing applications were exemplified on a single, randomly chosen match, with players deidentified. Player performance data from this season (e.g., goals, shots, xG) was provided by OptaPro.

Player positional information (x, y, t) and match events (e.g., pass, intercept) were recorded at a 25 Hz using optical tracking systems. Match events and player positions were consolidated to infer ball position. Player tracking time series were smoothed using a 5 Hz moving average filter before being down-sampled to 5 Hz to reduce error.

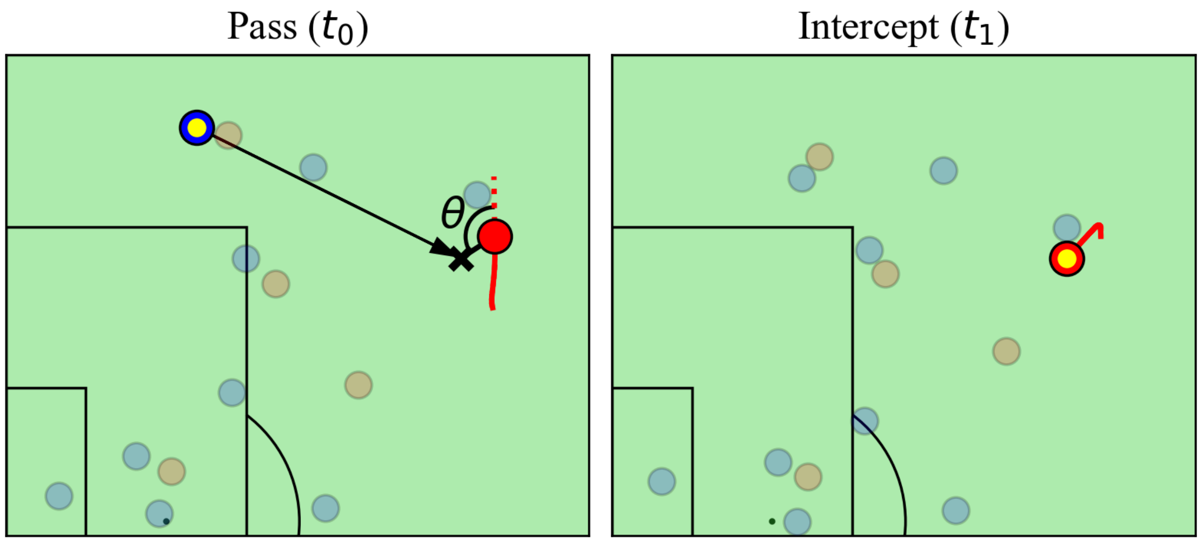

Figure 1. The sampling process for commitment-based motion models. The player with possession is denoted with a yellow point. (a) At t_0, the blue player begins a pass which results in an intercept (b) at t_1 by the red player. At t_0, the relative position of the intercept (X) is recorded. This is dependent on player orientation and the angle of reorientation required (θ) to win the ball. The red player's orientation and movement over the preceding second are denoted by dotted and solid red lines, respectively.

2.1 Player motion

The commitment-based motion models from Spencer et al., (2019b) are used to incorporate contextual player movement information into the development of motion models in football. These model the effect of momentum on players' likelihood of specific movement behaviours (such as intercepting a pass). We define a match event (e.g., interception), from which movement behaviour is sampled. Hence, these models pose the following – given the present position of the ball, what is the likelihood that a player will reposition to attempt to win the ball, were it kicked to a specified location? This behaviour is learned through observation of displacements during intercepts.

The sampling process is visualised in Fig. 1. At the time of the pass (t_0), the relative location of the forthcoming intercept, the interceptor's velocity, and the time between the pass and the intercept are recorded. This becomes an observed player displacement (i.e., the player moving at the sampled velocity was able to reposition to the relative location within the ball's travel time). From these observations, referred to as the *intercept* dataset, a four-dimensional distribution of player displacements (towards intercepts) is produced (Fig. 2a).

While the output is similar to traditional motion models (e.g., Brefeld, et al., 2018; Horton, et al., 2015), it is important to consider those instances whereby the ball was not intercepted in order to determine how frequently players reposition to perform interceptions. Random positions and times across successful passes were sampled and treated as potential interceptions that were not achieved. These points form the *non-intercept* dataset (Fig. 2b). This process allows us to learn player movement behaviour when attempting to win the ball. Visually, the *intercept* and *non-intercept* datasets have cleared differences (Fig. 2). For example, the density of the *non-intercept* dataset is low at the player's location (0, 0), reflecting a high likelihood of interception when the ball passes directly by an opponent.

The two distributions are combined to produce commitment models using the following function:

$$p_i(x) = \frac{w f_{int}(x)}{w f_{int}(x) + (1-w) f_{non-int}(x)} \tag{1}$$

where $p_i(x)$ is the probability of displacement to location x by player i and f_{int} and $f_{non\text{-}int}$ are the distributions for the *intercept* and *non-intercept* datasets. Given the sample size inequality between the datasets, a weighting coefficient, w, is used to weight the distributions, hence countering their normalisation. Resultant models are continuous in four dimensions (x-, y-, velocity, time). Results of this process (in two-dimensions) are visualised in Fig. 2.

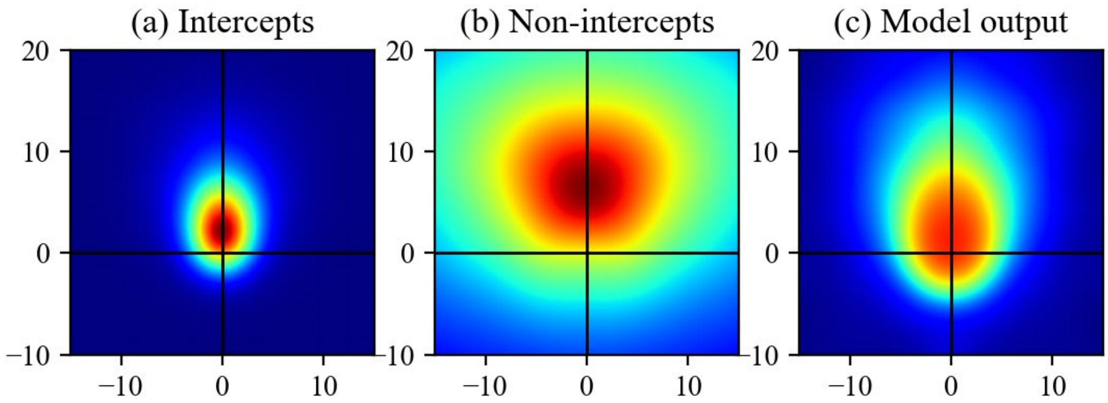

Figure 2. Distributions of relative pass locations for $t = 2$ seconds. (a) Distributions of relative locations for passes that were intercepted. The interceptor is positioned at the (0, 0) point of each plot at the beginning of the pass. (b) Relative pass locations that were not intercepted by the focus player. (c) The combined distribution (see Eq. 1) produces the likelihood of interception in the player's vicinity.

2.2 Spatial control

The commitment-based motion models described in 2.2 are used to measure players' spatial influence. If a player is likely to commit to an event at location X (i.e., high probability of intercepting the ball at X based on their positioning and momentum), they are considering to be applying pressure to (or influencing) X. Derivation of these models with respect to the ball produces measures of influence that consider the ball's distance. Hence, if a team has a spatial advantage at X, that advantage will be retained if the ball was passed directly to X (a property not present in Voronoi tessellations). The *spatial control* (SC) of team i at location X is measured as follows:

$$SC(X) = \frac{\sum_i p_i(X)}{\sum_i p_i(X) + \sum_j p_j(X)} \quad (2)$$

Eq. 2 produces SC in the range [0, 1] where SC(X) > 0.5 signifies a spatial advantage to the attacking team, i, over their opponent, j, at X.

2.3 Analysis

2.3.1 Passing outcomes

The SC of passes was extracted for passes from a single match. SC at the receiver was recorded at the time of the pass. Passes were classed as successful or unsuccessful (where unsuccessful passes were those resulting in interception). To assess the explanatory power of SC, differences in SC between the two classes were analysed. Theoretically, unsuccessful passes should more frequently be to regions of opposition control (SC < 0.5) than successful passes.

2.3.2 Passing networks

Passing interactions between teammates have been used to understand within-team relationships and their relationship to success (e.g., Goncalves, et al., 2017). However, there has been little consideration of the quality of passes between teammates. Understanding the SC of passes between teammates provides insights into tactical behaviour and more advanced player profiling. For example, if passes from one player to another are frequently high risk (i.e., low SC of the receiver), this may be indicative of player decision-making or of the perceived strength of the receiver (hence an effort from defenders to apply pressure, or a tendency for teammates to execute riskier passes to valuable targets). Passing interactions between a single team were analysed based on passing frequency and the mean SC of these passes.

3.0 Results

3.1 Player motion models

Motion models were fit on approximately 46000 samples of intercept behaviour. The sample size of the *intercept* and *non-intercept* classes was 6922 and 35959 respectively. w was set to 0.15 based on class samples. Commitment-based models are presented in Fig. 3 and a comparison between models using commitment- and displacement-based methods is presented in Fig. 4.

Player's spatial influence widens as time increases (Fig. 3). When moving at high speeds, greater influence is noted at further distances along a player's trajectory, while influence behind the player decreases. This is in line with findings from previous studies (e.g., Taki & Hasegawa, 2000; Horton, et al., 2015; Brefeld, et al., 2018).

Displacement-based models are narrower than commitment-based models (Fig. 4). Furthermore, a greater portion of the area under displacement-based models is in front of the player, with minimal displacements observed behind the player.

3.2 Passing outcomes

A visualisation of SC is presented in Fig. 5. In this example, the attacking team possesses the ball close to the middle of the field. SC is moderate within the region surrounding the ball, low towards the attacking goal, and high towards the opposite boundary.

The distribution of SC for successful and unsuccessful passes is presented in Fig. 6. Successful passes are negatively skewed and bimodal with peaks at approximate SC of 0.58 and 0.95. The former is a pass to a teammate behind pressured by an opponent, with a slight spatial advantage to the teammate, and the latter is to a teammate who is under minimal pressure. Conversely, unsuccessful passes are generally intercepted in areas of opposition control (SC < 0.5). The mean SC of these classes are 0.69 and 0.39 respectively. Furthermore, 17.3% of the distribution of the successful class is between SC of 0 and 0.5 ($0 \leq SC \leq 0.5$) compared to 75.9% for the unsuccessful class. These results were found to be significantly different using Dunn's pairwise test ($p < 0.001$).

3.3 Passing networks

The passing relationships between pairs are visualised in Fig. 7 and Table 1. The mean receiving SC was lowest for the attacking players (*H*, *I*, *J*, *K*). Furthermore, these four players were amongst the five players with the lowest mean passing SC. These results are logical as SC is expected to be lower in opposition territory where attackers more frequently position. The pass completion rates of players throughout the season had a moderate positive correlation with the mean SC of their passes (Spearman's rank correlation coefficient = 0.33).

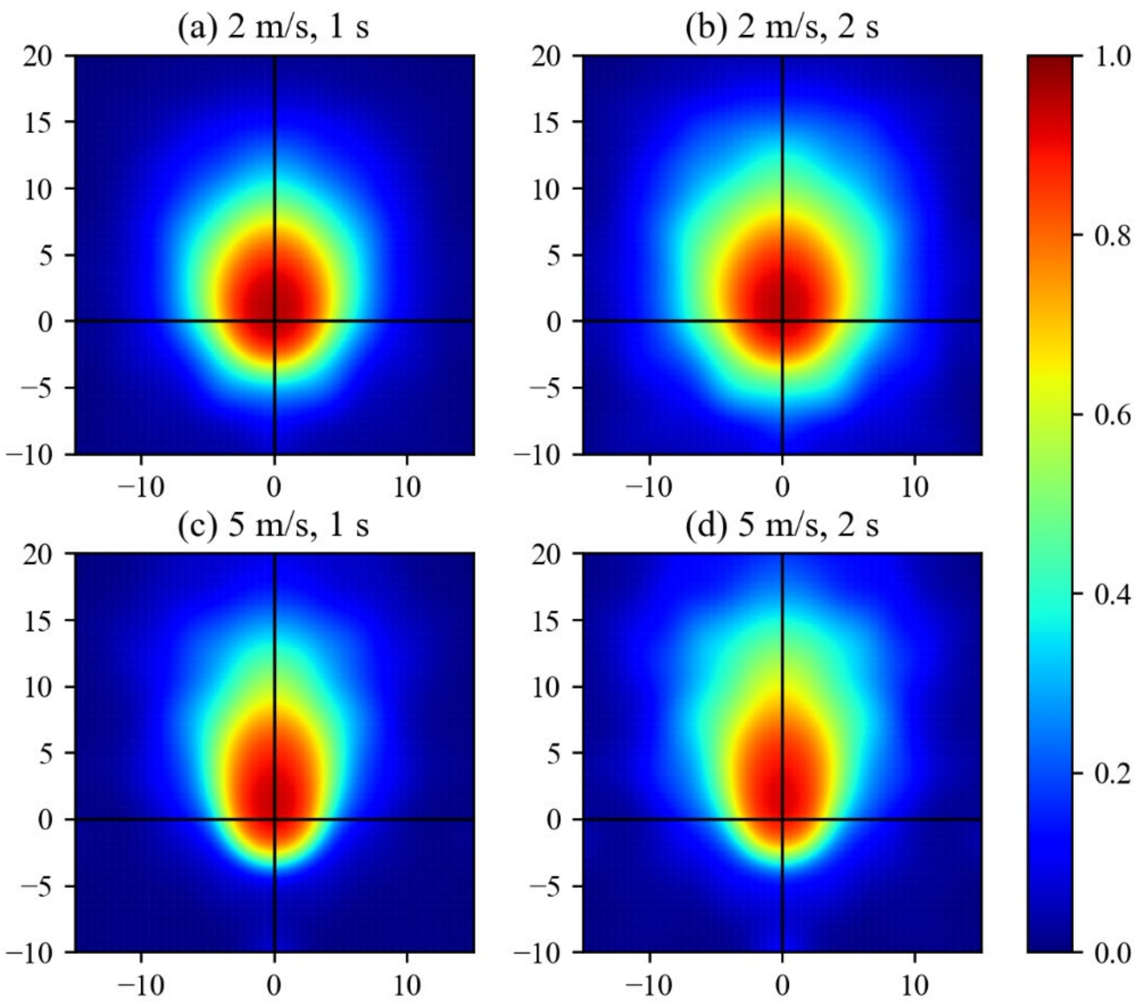

Figure 3. Player commit-models for a player moving at (a) 2 m/s with 1 s to reposition, (b) 2 m/s with 2 s to reposition, (c) 5 m/s with 1 s to reposition, and (d) 5 m/s with 2 s to reposition.

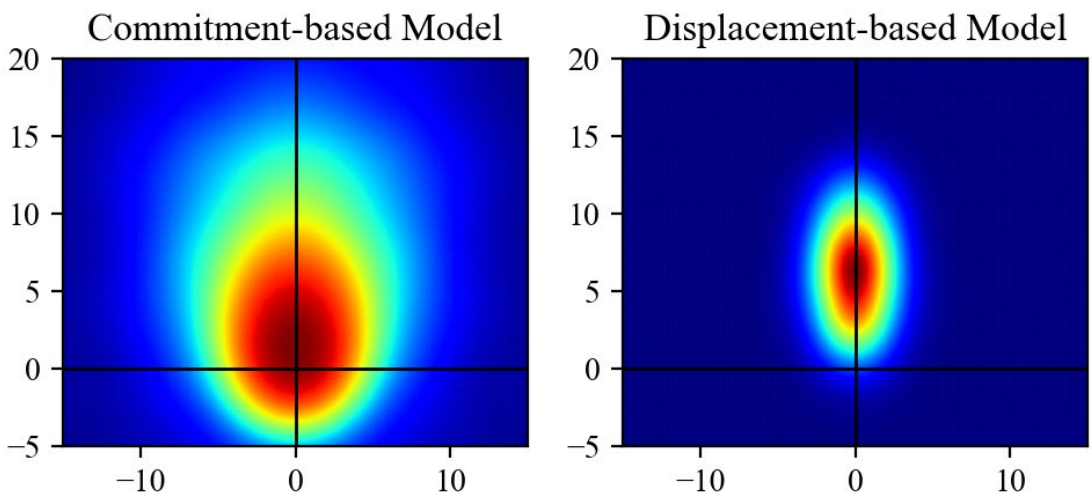

Figure 4. Comparison of player motion models derived from (a) commitment-based methods, and (b) displacement-based methods.

The player with the lowest receiving and passing SC relationships is *J*. This player recorded the highest goals, shots and xG (OptaPro) for this team across the 2018 season. This may indicate a tendency for teammates to perform riskier passes to player *J*, *J*'s perceived threat by opponents (hence, an effort to limit their space) or *J*'s positioning in higher value space, which is theorised to have higher opposition occupancy (Fernandez & Bornn, 2018).

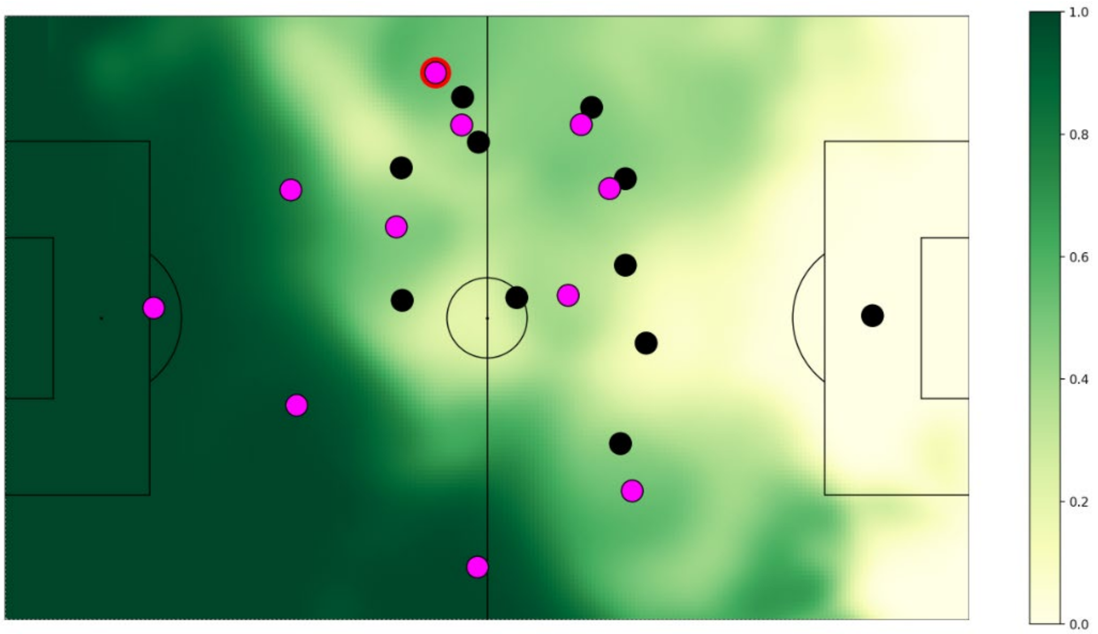

Figure 5. Spatial control for the attacking team (magenta) relative to the defending team (black). The player with possession is circled in red.

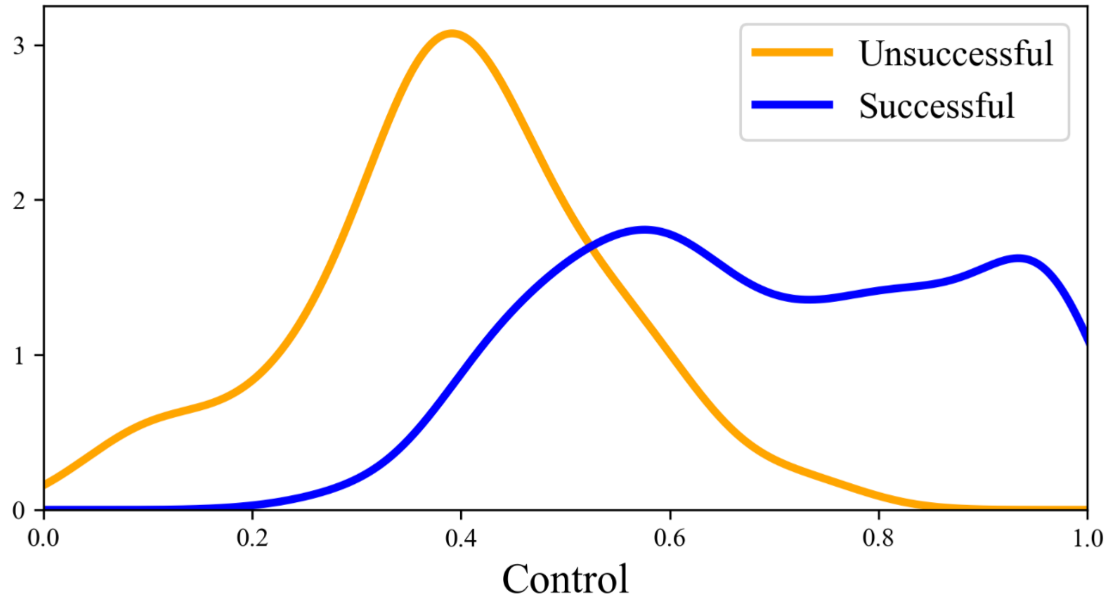

Figure 6. The distribution of spatial control for successful (blue) and unsuccessful (orange) passes.

4.0 Practical applications

4.1 Discussion

This study presented a new method for modelling player motion in football. This approach considers the context of player displacements. Hence, measurements of spatial control derived from these models produce results more reflective of player movement behaviour, in comparison to models containing greater assumptions (e.g., Brefeld, et al., 2018). A further advantage of these models is their continuity in four-dimensions. Representing data on a continuous scale removes bias from arbitrability bounding metrics (e.g., training load, Carey, et al., 2018). Previous implementations using empirical player displacements have represented velocity as a categorical variable (e.g., Horton, et al., 2015; Brefeld, et al., 2018).

A consideration of these models is their computational complexity. Their dimensionality increases complexity over low dimension models (Brefeld, et al., 2018) or influence models that represent motion using Gaussians (e.g., Fernandez & Bornn, 2018). The latter can be run in real time (Fernandez & Bornn, 2018), hence is preferable for applications that can be actioned during matches. For post-match analytics, commitment-based models provide more realistic measurements of space while retaining an interpretable output. Whilst higher-dimension deep learning models have been applied in football (Fernandez, et al., 2019; Le, et al., 2017), motion models produce an output that is both intuitive and interpretable by non-technical staff (e.g., Fig. 3).

Commitment-based models rely on a weighting coefficient to produce realistic probabilities of player commitment. However, regardless if w is set incorrectly, resultant commitment probabilities (p_i) can be used because their outputs are relative (that is, $p_i = 0.8$ signifies higher commitment likelihood than $p_i = 0.6$, however may not equal an 80% chance of commitment). In football, successful passes could theoretically contain multiple missed interceptions along their path. Hence, automatically calculating w presents difficulties. In this study, w was estimated based the frequency of successful and unsuccessful passes.

The attacking team's SC over a region was measured as its proportion of ownership of said region relative to the occupancy of all players (Eq. 2). It should be noted that this implementation of SC implies a linear relationship between winning and control. It is possible that beyond a certain spatial advantage, additional control is unnecessary. In Fernandez and Bornn (2018), pitch control was measured via the logistic function of the influence differential between teams. Future work should explore variations in SC computation and their correlation to possession outcomes.

4.2 Applications & future work

Measuring a team's spatial control has applications in performance analysis and may inform tactics, list management and recruitment. A primary application of the motion models in this study was the analysis of passes, where the risk of a pass was estimated via the attacking team's SC. Using this approach, the risk of passes between player pairs was measured for a single match. Player relationships were quantified via the mean SC of passes between each pair.

The primary findings of this analysis were that the attacking players were receivers of riskier passes than defensive players, and that the team's highest performer (as quantified via goals, shots and xG throughout the season) was both the passer and receiver of the riskiest passes between pairs (player *J*). This player is one of the analysed team's designated players (star players not counted in a team's salary cap), hence the ability to differentiate this player from their peers is notable.

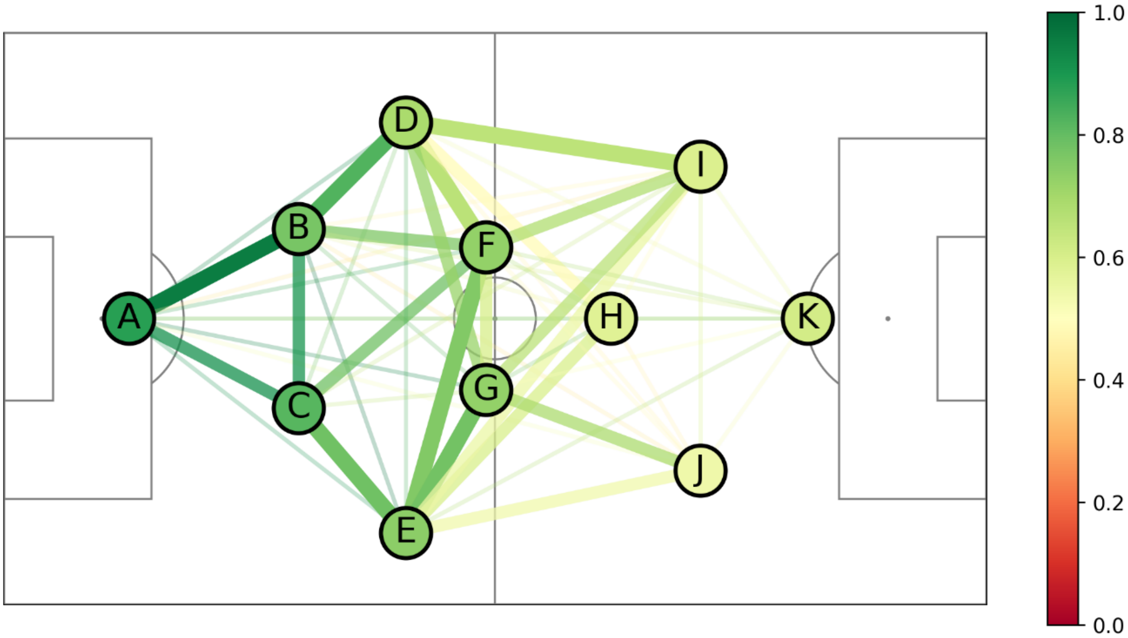

Figure 7. Player passing network diagram. Player (x, y) location corresponds to playing position. The width of the link between pairs corresponds to frequency of passing, and the colour corresponds to the mean SC of passes. Player positions are coloured according to their mean passing and receiving SC.

Table 1. Average pass SC between player pairs from a single match. Passes between defensive positions (top left) are of generally higher SC than those between forward positions (bottom right). Blank cells indicate no passes between the pair.

		Receiver											
		A	B	C	D	E	F	G	H	I	J	K	Mean
Passer	A		0.98	0.92	0.91	0.97	0.84		0.52	0.44	0.54	0.84	0.77
	B	0.97		0.95	0.80		0.80		0.45	0.47	0.44	0.59	0.68
	C	0.97	0.89		0.87	0.82	0.80	0.61		0.63			0.80
	D	0.99	0.90				0.67	0.63	0.38	0.63	0.50	0.60	0.66
	E		0.99	0.82	0.90		0.64	0.75	0.57	0.44	0.51	0.63	0.69
	F	0.97	0.82	0.81	0.66	0.86		0.60	0.70	0.61	0.52	0.43	0.70
	G	0.99	0.87	0.87	0.86	0.85	0.68			0.74	0.49		0.79
	H		0.71		0.59	0.66	0.72	0.84		0.53	0.49	0.65	0.65
	I			0.61	0.74	0.68	0.76	0.60	0.49		0.61	0.52	0.63
	J				0.41	0.70		0.80	0.45	0.52		0.51	0.56
	K		0.76		0.39	0.83	0.87	0.56	0.53	0.56	0.45		0.62
	Mean	0.98	0.87	0.83	0.71	0.80	0.75	0.67	0.51	0.56	0.51	0.60	

Longitudinal analysis of passing behaviour is required to identify the cause of *J*'s passing behaviours. Identifying if this trend is a coaching directive by the team or their opponent would provide insights that could be exploited within the league. Should certain teams tend to overcommit on high performing players, this could be exploited by leveraging alternative attacking channels. The optimal distribution of player performance in football should be researched to understand the effects of high performers. In Australian football, it was found that there was a negatively skewed distribution of player performances in losses compared to a more even distribution during wins (Robertson, et al., 2016).

These analyses could be used to develop player decision-making profiles. Initial results revealed differences between the passing and receiving risk of players in similar positions (e.g., *J* and *K*). It is likely that individuals' approach to their position differs between players, hence measuring these behaviours is valuable from a coaching and recruitment perspective. An extension of this would be defining types of passing with consideration of SC. The spatial components of passes between players have been clustered in Australian football to define passing categories based on their distance, risk and change in a field equity metric (Spencer, et al., 2019b). Similar analysis could be conducted in football, with results informing training prescription based on match demands. A multivariate understanding of passes is important in training prescription (Browne, et al., 2019).

Finally, the measure of risk (SC) presented in this study models player behaviour relative to ball movements. Using similar methodology, it is possible to measure different types of pressure such as pressing or tackling threats from opponents.

5.0 Conclusions

A variation of player motion models incorporating contextual movement behaviour was exemplified in this study. Resultant commitment-based motion models quantified a higher likelihood of displacements requiring reorientation, when the displacement is to intercept a pass, compared to traditional displacement-based motion models. As such, when analysing player movements relative to possession outcomes, commitment-based models are more representative of player movement behaviours. The spatial occupancy of the attacking team was measured to quantify the risk of passes between players. It was found that attacking players are the receivers of higher risk passes which is logical due to their positioning in high value space.

6.0 References

Brefeld, U., Lasek, J., & Mair, S. (2018). Probabilistic movement models and zones of control. *Machine Learning*, 108, 127-147. doi: 10.1007/s10994-018-5725-1

Browne, P., Sweeting, A., Davids, K., & Robertson, S. (2019). Prevalence of interactions and influence of performance constraints on kick outcomes across Australian football tiers: Implications for representative practice designs. *Human Movement Science*, 66, 621-630.

Carey, D., Crossley, K., Whiteley, R., Mosler, A., Ong, K., Crow, J., & Morris, M. (2018). Modelling training loads and injuries: The dangers of discretization. *Medicine and Science in Sports and Exercise*, 50(11), 2267-2276.

Cervone, D., Bornn, L., & Goldsberry, K. "NBA court realty," in *Proceedings of the 10th Sloan Sports Analytics Conference*, (Boston, MA).

Fernandez, J., & Bornn, L. (2018). "Wide open spaces: a statistical technique for measuring space creation in professional soccer," in *Proceedings of the 12th Sloan Sports Analytics Conference*, (Boston, MA).

Fernandez, J., Cervone, D., & Bornn, L. (2019). "Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer," in *Proceedings of the 13th Sloan Sports Analytics Conference*, (Boston, MA).

Fujimaru, A., & Sugihara, K. (2005). Geometric analysis and quantitative evaluation of sport teamwork. *Systems and Computers in Japan,* 36(6), 49-58. doi: 10.1002/scj.20254

Goncalves, B., Coutinho, D., Santos, S., Lago-Penas, C., Jimenez, S., & Sampaio, J. (2017). Exploring team passing networks and player movement dynamics in youth association football. PLoS One 12(1), e0171156.

Gudmundsson, J., & Horton, M. (2017). Spatiotemporal analysis of team sports. *ACM Computing Surveys*, 50(2), 22.

Horton, M., Gudmundsson, J., Chawla, S., & Estephan, J. (2015). "Automated classification of passing in football," in Pacific-Asia Conference on Knowledge Discovery and Data Mining, eds T. Cao, et al. (Cham: Springer), 319-330. doi: 10.1007/978-3-319-18032-8_25

Le, H., Carr, P., Yue, Y., & Lucey, P. (2017). "Data-driven ghosting using deep imitation learning," in *Proceedings of the 11th Sloan Sports Analytics Conference*, (Boston, MA).

Rein, R., Raabe, D., & Memmert, D. (2017). Which pass is better? Novel approaches to assess passing effectiveness in elite soccer. *Human Movement Science*, 55, 172-181.

Robertson, S., Gupta, R., & McIntosh, S. (2016). A method to assess the influence of individual player performance distribution on match outcome in team sports. *Journal of Sports Sciences*, 34(19), 1893-1900.

Spencer, B., Jackson, K., Bedin, T., & Robertson, S. (2019a). Modelling the quality of player passing decisions in Australian Rules football relative to risk, reward and commitment. *Frontiers in Psychology,* 10, 1777.

Spencer, B., Jackson, K., & Robertson, S. (2019b). Fitting motion models to contextual player behaviour. *arXiv preprint arXiv:1907.10762*.

Taki, T., & Hasegawa, J. (2000). "Visualization of dominant region in team games and its application to teamwork analysis," in *Proceedings Computer Graphics International 2000*, (Geneva: IEEE).