Predicting match outcome in professional Dutch football using tactical performance metrics computed from position tracking data

Floris Goes, Matthias Kempe and Koen Lemmink

Predicting match outcome in professional Dutch soccer using tactical performance metrics computed from position tracking data

Floris Goes[1*], Matthias Kempe[1] and Koen Lemmink[1]

[1] University of Groningen, University Medical Center Groningen (UMCG), Department of Human Movement Sciences, Groningen, The Netherlands

f.r.goes@umcg.nl

Abstract

Quality as well as quantity of tracking data have rapidly increased over the recent years, and multiple leagues have programs for league-wide collection of tracking data. Tracking data enables in-depth performance analysis, especially with regard to tactics. This already resulted in the development of several Key Performance Indicators (KPI's) related to scoring opportunities, outplaying defenders, numerical balance and territorial advantage. Although some of these KPI's have gained popularity in the analytics community, little research has been conducted to support the link with performance. Therefore, we aim to study the relationship between match outcome and tactical KPI's derived from tracking data. Our dataset contains tracking data of all players and the ball, and match outcome, for 118 Dutch premier league matches. Using tracking data, we identified 72.989 passes. For every pass-reception window we computed KPI's related to numerical superiority, outplayed defenders, territorial gains and scoring opportunities using position data. This individual data was then aggregated over a full match. We then split the dataset in a train and test set, and predicted match outcome using different combinations of features in a logistic regression model. KPI's related to a combination of off-the-ball features seemed to be the best predictor of match outcome (accuracy of 64.0% and a log loss of 0.67), followed by KPI's related to the creation of scoring opportunities (accuracy of 58% and a log loss of 0.69). This indicates that although most (commercially) available KPI's are based on ball-events, the most important information seems to be in off-the-ball activity. We have demonstrated that tactical KPI's computed from tracking data are relatively good predictors of match outcome. As off-the-ball activity seems to be the main predictor of match outcome, tracking data seems to provide much more insight than notational analysis.

---

[*] Presenting & Corresponding Author

1   Introduction

Soccer is one of the most popular global sports, and match performance analysis has been the subject of intensive research over several decades[1]. Soccer, nowadays, is a multi-billion industry that embraces mathematical ideas as teams are constantly searching for ways to improve their odds at winning, while spectators are trying to predict the outcome of a game to win money on the gambling market[2]. As a result, analyzing tactics and match performance in soccer is of particular interest to a broad and varied audience.

Traditionally, tactical analysis has been conducted based on observational assessment by experts or by means of notational assessment on-ball events like passes, dribbles, and tackles[3]. Despite the limitations of notational data, the focus on ball-events like passes in itself is understandable. A pass is the most frequent ball-event in a match and passing is a – or even the – key aspect of tactics in soccer. However, notational data only provides discrete low-level data, and thus only tells us what happens with the ball. Therefore it has limited practical value[4]. Teams might even dominate typical summary statistics like possessions, shots and number of passes, but still fail to score[5]. Nevertheless, notational analysis is still frequently used for tactical analysis by broadcasters, teams and scientists[3,6]. One could argue however that it would be much more interesting to look at what is going on with the 21 players not carrying the ball during a ball-event like passing. Yet achieving this requires not only notational data, but also position tracking data.

As opposed to notational event data, automatically generated position tracking data provides the opportunity to derive high-level continuous data off all players and the ball at the same time[7]. As a result of technological innovation and the league-wide implementation of position tracking systems in for example the German Bundesliga and the Dutch Eredivisie, the quantity and quality of available tracking data rapidly increased over the recent years[3,6]. Despite this increasing availability, the potential of position tracking data to analyze tactical performance has not been harnessed as tracking data is mostly used by analysts to monitor physical performance[8]. However, this data allows us to automatically study the complex interactions of all players on the field during every pass, and can therefore be regarded as a potential game changer for tactical analysis in soccer[3].

The limited practical use of position tracking data for tactical analysis might be explained by two reasons. First of all, most scientific work on tactical analysis using position tracking data – although of great scientific importance – has relatively little practical implications. Only a minority of the work investigated a link between the features they used for tactical analysis and actual match performance, and most of them did not find a clear relationship. In order to derive practical meaning from these types of analyses, we therefore propose it is critical to study the link between tactical features and match performance. Secondly, one could argue that as position tracking data is characterized by a much higher complexity and volume in comparison to notational event data, it challenges the typical data management and data analytics methods[9] commonly employed in sports science, and can therefore be considered big data. As a result, we propose that unlocking the potential of this data for tactical analysis requires the implementation of skills and techniques from other domains than sports science.

In conclusion, one could argue position tracking data harnesses the potential to provide in-depth insights in the complex tactics of soccer, and these insights can theoretically be used in the analysis and maybe even the prediction of performance. However, in order to achieve this and derive practical meaning from tactical analysis, the link between tactical features derived from position tracking data and actual match performance first has to be established. With the current paper, we therefore aim to study the relationship between tactical features derived from position tracking data and match outcome. To achieve this, we will use a match outcome prediction model based on tactical key performance indicators (KPI's). The results of our study could allow analysts to derive more practical meaning from tactical analysis using position tracking data, and scientists could use these KPI's to study the relationship between their tactical features and match success.

2   Quantifying Tactical Behavior

Tactics, often referred to in research as tactical behavior, can be defined as the management of space and time by a group of cooperating individuals, in interaction with the opponent while constantly adapting to the conditions of play, in order to achieve a common goal[3,10]. This common goal is related to ball-possession status, as teams have different tactical objectives when attacking and defending[11]. When in possession of the ball, teams aim to move the ball in the direction of the opponents goal, increase the effective play area through depth and width mobility, create numerical superiority in key offensive areas of the field, destabilize the defense, and ultimately create scoring opportunities[12]. On the other hand, when defending, teams aim to keep the opponent away from the goal, keep the effective play area small, move in unity to prevent destabilization, and keep numerical superiority close to their own goal[12]. These common goals are widely considered the general principles of play in soccer[11,12]. Achieving these goals can be seen as successful tactical behavior, and a relationship between tactical behavior and match outcome is widely assumed.

In order to study the relationship between tactical behavior and match outcome, one first has to quantify successful tactical behavior. As we are mainly concerned with offensive tactical behavior, we focused on tactical features related to the offensive principles of play. For this purpose, we first need to discuss how existing tactical features (either commercially available or derived from scientific research) can be related to the offensive principles of play.

First of all, moving the ball towards the opponent's goal and subsequently creating scoring opportunities (*zone principle*) can be assumed to have the most direct relationship with scoring goals in comparison to the other principles of play. Existing features like *expected goals* (*xG*)[13] (Optasports, London, United Kingdom), and Link's *dangerousity*[14] feature directly quantify this tactical principle. Both features are computed using distance and angle between the goal and the ball carrier, and award higher values for locations closer to the goal. As *xG* is typically computed using only notational event data, it is relatively inaccurate and does not take the pressure of defenders or any other of-the-ball activity into account. Therefore, it provides low-level information. *Dangerousity* is computed in a somewhat similar fashion, yet it is computed based on position tracking data and takes defensive pressure as well as of the ball activity into account as moderating factors. Therefore, *dangerousity* could be regarded as a high-level expected goals model.

Secondly, gaining numerical superiority (*balance principle*) is often believed to be of key importance for creating high probability scoring opportunities, as it will contribute to space creation and destabilization of the defense[15]. Numerical superiority and outplaying defenders can be analyzed from an on-the-ball as well as an off-the-ball perspective: teams can try to outplay defenders through passing and dribbling, and they can position their off-the-ball players in key areas of the field. Existing features like *Packing-Rate*[16] and *Impect*[16] (Impect GmbH, Cologne, Germany) have gained popularity in especially the German Bundesliga. They quantify the number of outplayed (*packed*) opponents or defenders through passing, and can easily be derived from position tracking data. Off-the-ball superiority has gained considerably less attention in the literature, but can also be directly derived from the position tracking data.

Finally, it is often believed in soccer that keeping the effective play area large when in possession of the ball (*space mobility principle*) is another prerequisite for space creation and destabilization of the defense. The effective play area of the attacking team can be defined as the attacking team's surface, and can be derived from position tracking data using the *Convex Hull* method[17]. One can compute the *Convex Hull* for every timeframe in ball possession and take the average as an indicator of space mobility.

3   Feature Engineering

3.1   Features Related to Zone

To quantify tactical performance with regard to the *zone*, *balance* and *space mobility* principles, we constructed separate features for every principle of attacking play. For the current study we adapted features currently available in science and practice. As in most cases limited technical details underlying a certain feature are publically available, and in order to solve some feature-specific limitations, we choose to construct our own adaptation of these features rather than exactly replicate existing features. All feature construction was conducted in Python 3.6 using the NumPy, Pandas and SciPy libraries.

To quantify tactical performance on the *zone* principle, we constructed a low-level and high-level *zone* feature, partly adapted from the work by Link[14]. First, we determined the low-level *zone* value based on the position of the ball-carrier relative to the goal in every pass and reception (**Error! Reference source not found.**). *Zone* values could range from 0 (furthest from the goal) to 1 (closest to the goal). The high-level feature was then computed by adding on-ball-pressure to the model. Pressure on the ball was computed using the model proposed in Andrienko et al[18]. This model computes a pressure value *PR* (0-100%) based on the distance off all defensive players to the ball carrier, and the angle of all defensive players towards the threat direction (in this case the direction from the ball carrier to the goal). In this model, 0 represents no pressure at all, while values of 100% represent high pressure from the defenders close to the ball-carrier. As we assume high pressure increases the difficulty of creating a scoring opportunity, the zone value Z is penalized by PR as shown in Eq. 1.

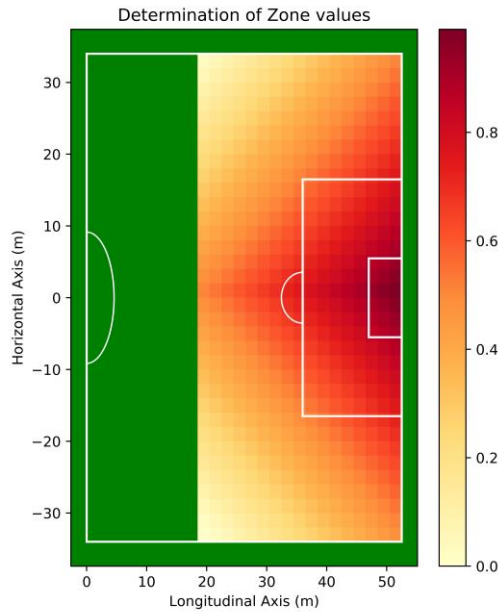$$Z = Z \ * (1 - PR) \tag{1}$$

Figure 1 - Visual representation of zone values computed for every pass and reception. Color bar represents the zone values (range 0.0 - 1.0)

Both the low-level and high-level *zone* were computed for every successful pass and reception and then aggregated over the full match. This resulted in mean and total low- and high-level zone values for passers and receivers on a team.

3.2   Features Related to Balance

To quantify tactical performance on the *balance* principle, we constructed two passing features and three off-the-ball balance features. Our passing features follow the description of the *Packing-Rate[16]* and *Impect[16]*. We computed the number of outplayed opponents based on the longitudinal coordinates of the pass, reception and all the opposing players, and we computed the number of outplayed defenders based on the longitudinal coordinates of the pass, reception and the last 6 players on the field plus the goalkeeper (**Error! Reference source not found.**). Note that the number of outplayed opponents can also be negative in the case of a backwards pass. Furthermore, note that we only looked at the X-coordinates   to   determine   what   defenders   are   outplayed.
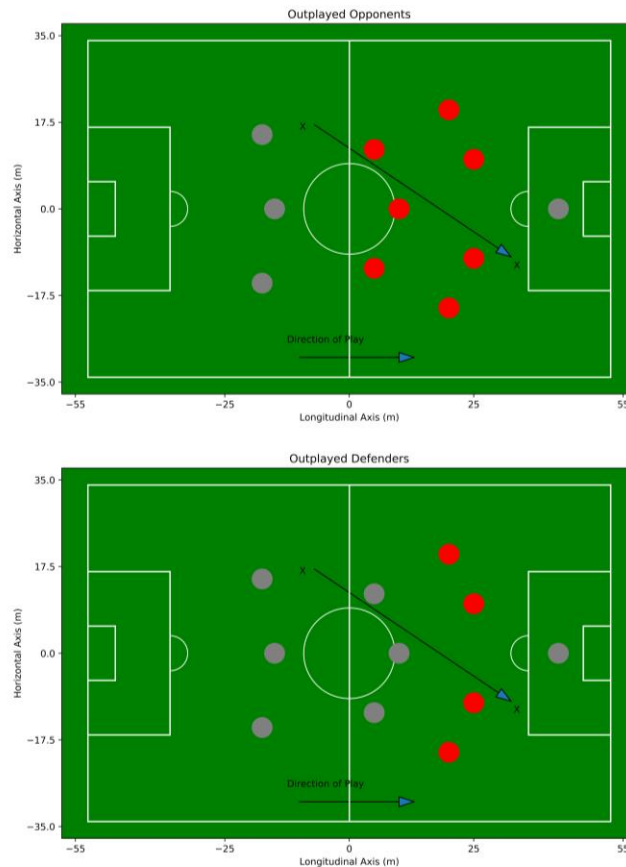

Figure 2 - Visual representation of outplayed opponents (top) and outplayed defenders (bottom) as a result of pass. Outplayed opposing players are shown in red, other opposing players are shown in grey. Note that in our approach the number of outplayed defenders is based on the last 6 outfield players and the goalkeeper.

As off-the-ball balance features we computed numerical superiority scores for the attacking team on the opposing half, in the final 3rd and in the score-box. To do so, we assessed numerical balance (by counting players of both teams) in a certain area (i.e. final 3rd or score-box) during every pass-reception window, and awarded points for every window in which the attacking team had numerical superiority in that area (+1 player = 1 point, +2 players = equal 2 points, etc.).

3.3   Features Related to Space Mobility

Finally, as a quantification of the *space mobility* principle, we computed the average attacking team's surface area for every attack during a game, over the duration of the complete possession. The attacking team surface area (S_A) on every timestamp t in the game was computed as the *Convex Hull* of an array P_t containing the positions of all n outfield players (the goalkeeper was excluded), using the QHull implementation in the SciPy library (eq. 2 & 3).

$$P_t = [[X_i^t + Y_i^t], [X_{i+1}^t + Y_{i+1}^t], [\ldots, \ldots], [X_n^t, Y_n^t]] \tag{2}$$

$$S_A = \text{ConvexHull} \parallel P_t \parallel \tag{3}$$

4   Modelling Match Performance in Soccer

To evaluate tactical performance of a team in relation to the different principles, and analyze the relationship between tactical performance and match outcome, we collected and processed position tracking data on both teams for matches played during 4 consecutive Dutch Eredivisie seasons. Players were tracked with a semi-automatic optical tracking system (SportVU; STATS LLC, Chigago, IL) that captures the X and Y coordinates of all players and the ball at 10 Hz. Our dataset contained 118 matches in which 26 unique teams played each other. As we were only concerned with the differences between winning and losing teams, we excluded matches that ended in a draw. This resulted in a final dataset that consists of 25 teams that played in 89 matches that resulted in a win or a loss and contained 98.718 pass attempts of which 60.524 passes were successful.

The data of every single match were first pre-processed with ImoClient software (Inmotio Object Tracking B.V., The Netherlands). Pre-processing consisted of filtering the data with a weighted Gaussian algorithm (85% sensitivity) and automatic detection of ball possessions and ball events based on the tracking data. Both the tracking data and the ball event data were then imported as individual data frames in Python 3.6 and automatically processed on a match-by-match basis. We then computed the low-level and high-level zone feature for every pass and reception, the number of outplayed opponents and outplayed defenders for every pass, the numerical superiority in 3 areas for every pass-reception window, and the team surface area of all outfield players for every timeframe the team was in possession of the ball. All features were computed according to the methods as described in section 3.

Table 1 - Descriptive statistics (mean ± std.) of winning and losing teams on the various principles of play. ‡ (p < .05) and ⁂ (p < .01) denote significant differences between winning and losing teams.

	Wins (N = 89)	Losses (N = 89)	Mean Diff.	Effect Size (Cohen's d)
Zone Principle				
Low-level zone passer (Mean)	0.031 ± 0.013	0.028 ± 0.012	+10.7%	0.24
Low-level zone receiver (Mean)	0.040 ± 0.014	0.037 ± 0.014	+8.1%	0.24
High-level zone passer (Mean)	0.022 ± 0.010	0.020 ± 0.010	+10%	0.21
High-level zone receiver (Mean)	0.032 ± 0.012	0.028 ± 0.011	+14.3%	0.28‡
Low-level zone passer (Total)	10.62 ± 5.40	9.55 ± 4.54	+11.2%	0.21
Low-level zone receiver (Total)	13.54 ± 6.21	12.36 ± 5.26	+9.5%	0.20
High-level zone passer (Total)	7.11 ± 3.70	6.51 ± 3.57	+9.2%	0.16
High-level zone receiver (Total)	10.10 ± 4.52	9.14 ± 4.01	+10.5%	0.22
Balance Principle				
Outplayed defenders (Mean)	0.23 ± 0.10	0.21 ± 0.09	+9.5%	0.19
Outplayed opponents (Mean)	0.39 ± 0.17	0.39 ± 0.16	+2.6%	0.13
Outplayed defenders (Total)	71.01 ± 29.69	67.88 ± 30.57	+4.7%	0.11
Outplayed opponents (Total)	119.69 ± 49.46	121.91 ± 50.88	-1.8%	-0.04
Half Superiority (Total)	2.82 ± 7.67	1.87 ± 5.78	+50.8%	0.14
Final 3rd Superiority (Total)	3.11 ± 3.52	2.22 ± 3.04	+40.0%	0.27⁂
Score Box Superiority (Total)	0.84 ± 1.51	0.76 ± 3.39	+10.5%	0.03‡
Space Mobility Principle				
Team Surface Area (mean)	979.76 ± 99.12	966.41 ± 96.70	+1.4%	0.14

To compare performance between winning and losing teams, we aggregated all feature scores into mean (values per pass), and total (sum over a full match) scores. We then took the means and standard deviations of all winning and losing teams for a between-group comparison (Table 1). As most features scores were not-normally distributed, and variances were heterogenic, we conducted Kruskal-Wallis tests to statistically compare both groups. We found that winning teams had a significantly increased mean high-level zone score for pass receivers ($H(176) = 4.16$, $p < 0.05$), and a significantly increased superiority score in the final 3rd ($H(176) = 6.90$, $p < 0.01$) and score box ($H(176) = 5.09$, $p < 0.05$) compared to losing teams.

As a next step, we predicted match outcome based on several combinations of performance features. To do so we first split the data set in a training set that contained 80% of the data, and a test set that contained 20% of the data, stratified on match outcome. Furthermore, we scaled our features to the same scale using a robust scaling algorithm. We then fitted a 5-fold cross-validated Logistic Regression model to our training dataset and predicted winning and losing probability for both teams in every match.

First, we fitted the model using only the features that had shown (significant) power to discriminate between winning and losing teams (Table 1), as we expected this model to perform the best. Based on the mean high-level zone receiver score (β_1), the total final 3rd superiority score (β_2), and the total score box superiority score (β_3), we were able to predict binary match outcome with an accuracy of 64% and a log loss of 0.67, based on the following regression equation (4):

$$\text{Outcome} = -0.0167 + 0.136\,\beta_1 + 0.130\,\beta_2 - 0.0162\,\beta_3 \tag{4}$$

Then, we fitted models for all three discussed principles of play, to see what principle has the strongest relation with success. In cases where we had both mean and total values for a variable, we opted for the mean as this consistently proved to be a better discriminator. For performance on the *zone* principle, we fitted a model using the mean low-level zone for passers (β_4) and receivers (β_5), and the mean high-level zone for passers (β_6) and receivers (β_7). Based only on zone features, we were able to predict binary match outcome with an accuracy of 58% and a log loss of 0.69, using the following regression equation (5):

$$\text{Outcome} = -0.7e^{-6} + 0.00028\,\beta_4 + 0.00035\,\beta_5 + 0.00014\,\beta_6 + 0.00054\,\beta_7 \tag{5}$$

For performance on the *balance* principle, we fitted a model using the mean outplayed defenders (β_8) and opponents (β_9), and the total half superiority (β_{10}), final 3rd superiority (β_{11}), and score-box superiority scores (β_{12}). Based only on balance features, we were able to predict binary match outcome with an accuracy of 58% and a log loss of 0.70, using the following regression equation (6):

$$\text{Outcome} = 0.018 + 0.97\,\beta_8 - 0.65\,\beta_9 - 0.06\,\beta_{10} + 0.38\,\beta_{11} - 0.04\,\beta_{12} \tag{6}$$

Finally, for performance on the *space mobility* principle, we fitted a model using the mean team surface area per attack (β_{13}). Based only on a space mobility feature, we were able to predict binary match outcome with an accuracy of 64% and a log loss of 0.69, using the following regression equation (7):

$$\text{Outcome} = 0.003 + 0.06\,\beta_{13} \tag{7}$$

5 Discussion

The aim of this study was to analyze the relationship between tactical features derived from position tracking data and match outcome. To achieve this we constructed features that quantify performance on three main principles of attacking play in soccer[11,12], and studied the relationship between performance on these principles and binary match outcome (win or lose). Our results indicate differences between winning and losing teams are relatively small, but especially features that are either directly related to off-the-ball activity (numerical superiority) or at least incorporate off-the-ball activity (high-level zone for receivers) are able to discriminate between winning and losing teams and predict match outcome with fair accuracy. Based on these results we were able to confirm the relationship between tactical performance on the *zone* and *balance* principles, but not on the *space mobility* principle. Furthermore, our results indicate some of the features that have gained considerable popularity within the analytics community over the recent years seem to have limited practical value.

To study tactical performance on the *zone* principle, we constructed low-level and high-level zone features for both the passer and receiver in every pass. Our low-level feature has some resemblance with the popular expected goals (*xG*) feature[13], and – while we derived it directly from the tracking data – could be approximated with notational analysis. Our high-level feature accounts for defensive pressure and therefore requires position tracking data of all players on the field. Both the high-level and low-level features showed some discriminative power between winning and losing teams, with low to medium effect sizes, yet only the mean high-level zone for receivers was significantly increased in winning teams in comparison to losing teams. Based on these results we conclude winning teams more often seem to bring the ball into a position from which scoring opportunities can be created. Both high- and low-level features seem capable of capturing this principle, yet high-level features seem to have more discriminative power. As Optasport's *xG* is typically only computed for actual shots, and we

computed *zone* values for every pass and reception, one has to be cautious in generalizing our results to interpret actual *xG* values.

To assess performance on the *balance* principle, we used both on-the-ball and off-the-ball features. Our on the-ball-features are focused on outplayed opponents and defenders, and resemble the popular *Packing-Rate*[16] and *Impect*[16]. Although these features have gained considerable popularity in especially the German Bundesliga over the recent years[19], and multiple claims have been made about a possible link with match outcome, our research does not support such a relationship. Whereas winning teams did show a slightly higher mean number of outplayed defenders per pass, there was no difference in the mean and total number of outplayed opponents between winning and losing teams, and adding these features to the prediction model decreased prediction accuracy. Off-the-ball features on the other hand seemed to be a strong discriminator between winning and losing teams, as winning teams had significantly increased superiority scores in the final 3rd and the score box. Interestingly, the effect for score-box superiority was only small, but still significant, and leaving this feature of the prediction model harmed the accuracy of the prediction. The lack of a relationship between outplayed opponents/defenders and match outcome might be explained by methodologic limitations. One could for example argue that one should not only look at how many players were passed in the longitudinal direction but also in the lateral direction, and that in some areas of the field passing backwards can be more effective. However, to closely resemble existing approaches we choose not alter the approach for the current study.

Finally, performance on the *space mobility* principle did not seem to have a clear relationship with match outcome, despite the fact that space mobility is assumed to be a key aspect of offensive performance[11]. The absence of a clear effect might be explained by the fact that we used the team's surface area to assess space mobility. One could argue that although the team's surface is a valid feature to describe the effective area of play, space mobility actually refers to attackers dynamically creating depth by moving away from the ball at the right moment. It is questionable whether this dynamic effect is captured by a collective variable that is aggregated over all timeframes in possession of the ball.

Although capturing performance in easily interpretable KPI's is popular within the analytics community as well as the media, the reality of soccer seems much more complex. One likely explanation for the absence of a strong relationship between most popular KPI's and match outcome might be the fact that these KPI's are typically related to frequent events like passing, that are then aggregated over the full match. As there is a large match-to-match variability and actual tactics depend heavily on the interaction with the opponent; features like the *Packing-Rate* might be more dependent on the playing style of both teams than the actual match outcome. Soccer is a low-scoring game, and one could argue that in order to accurately predict match outcome, one should capture the rare events that lead to offensive success. One such an example is our proposed superiority score. Although highly discriminative between winning and losing, achieving final 3rd superiority also proved to be a rare event. The average superiority score of 3.11 in winning teams indicates these teams only achieve a +1 numerical superiority in the final 3rd on 3 occasions during a match, and these occasions seem to have a big importance for match outcome.

6   Conclusion

With this study, we have shown that although soccer is a complex game that is often considered highly unpredictable, the outcome of a match can be modelled with a fair accuracy. However, despite popular belief, soccer is not really a numbers game that can be analyzed based on simple KPI's of frequent events aggregated over the full course of a match. Discriminating between winning and losing teams and understanding tactical performance requires advanced features that can only be derived from position tracking data and heavily focus on off-the-ball rather than on-the-ball performance.

Disclosure Statement

The authors of this paper reported no conflicts of interest

Acknowledgements

References

1. Dubitzky, W., Lopes, P., Davis, J. & Berrar, D. The Open International Soccer Database for machine learning. *Mach. Learn.* **108,** 9–28 (2019).
2. Constantinou, A. C., Fenton, N. E. & Neil, M. Profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using Bayesian networks. *Knowledge-Based Syst.* **50,** 60–86 (2013).
3. Rein, R. & Memmert, D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *Springerplus* **5,** 1410 (2016).
4. Vilar, L., Araujo, D., Davids, K. & Bar-Yam, Y. Science of winning soccer: Emergent pattern-forming dynamics in association football. *J. Syst. Sci. Complex.* **26,** 73–84 (2013).
5. Brooks, J., Kerr, M. & Guttag, J. Using machine learning to draw inferences from pass location data in soccer. *Stat. Anal. Data Min.* **9,** 338–349 (2016).
6. Brink, M. S. & Lemmink, K. A. P. M. Performance analysis in elite football: all in the game? *Sci. Med. Footb.* **2,** 253–254 (2018).
7. Memmert, D., Lemmink, K. A. P. M. & Sampaio, J. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sport. Med.* **47,** 1–10 (2017).
8. Sarmento, H. *et al.* Match analysis in football: a systematic review. *J. Sports Sci.* **32,** 1831–1843 (2014).
9. Gandomi, A. & Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manage.* **35,** 137–144 (2015).
10. Gréhaigne, J.-F., Godbout, P. & Bouthier, D. The Foundations of Tactics and Strategy in Team sports. *J. Teach. Phys. Educ.* **18,** 159–174 (1999).
11. Clemente, F. M., Martins, F. M. L., Mendes, R. S. & Figueiredo, A. J. A systemic overview of football game: The principles behind the game. *J. Hum. Sport Exerc.* **9,** 656–667 (2014).
12. Costa, I. T., Garganta, J., Greco, P. J. & Mesquita, I. Tactical Principles of Soccer: concepts and application Tactical Principles of Soccer. *Rev. Mot.* **15,** 657–668 (2009).
13. Expected Goals (xG). Available at: https://www.optasports.com/services/analytics/advanced-metrics/. (Accessed: 3rd April 2019)
14. Link, D., Lang, S. & Seidenschwarz, P. Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLoS One* **11,** e0168768 (2016).
15. Rein, R., Raabe, D. & Memmert, D. "Which pass is better?" Novel approaches to assess passing effectiveness in elite soccer. *Hum. Mov. Sci.* **55,** 172–181 (2017).
16. Impect. Available at: https://www.impect.com/en/#idea. (Accessed: 3rd April 2019)
17. Moura, F. A., Barreto Martins, L. E., Anido, R. D. O., Leite De Barros, R. M. & Cunha, S. A. Quantitative analysis of Brazilian football players' organisation on the pitch. *Sport. Biomech.* **11,** 85–96 (2012).
18. Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., von Landesberger, T. &

Weber, H. Visual analysis of pressure in football. *Data Min. Knowl. Discov.* **31,** 1793–1839 (2017).

19.   Packing Rate – the Art of Outplaying - RWTH AACHEN UNIVERSITY - English. Available at: http://www.rwth-aachen.de/cms/root/Die-RWTH/Aktuell/Pressemitteilungen/Juni/~ljyt/Packing-Rate-die-Kunst-des-Ueberspiele/?lidx=1. (Accessed: 3rd April 2019)