Journal of Sports Sciences

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rjsp20

Expected Pass Turnovers (xPT) - a model to analyse turnovers from passing events in football

Andrew J Peters, Nimai Parmar, Michael Davies, Matt Reeves, Mladen Sormaz & Nic James

JOURNAL OF SPORTS SCIENCES
https://doi.org/10.1080/02640414.2024.2379697

Routledge
Taylor & Francis Group

SPORTS PERFORMANCE

OPEN ACCESS Check for updates

Expected Pass Turnovers (xPT) - a model to analyse turnovers from passing events in football

Andrew J Peters [a,b], Nimai Parmar [a], Michael Davies[a,b], Matt Reeves[c], Mladen Sormaz[b] and Nic James [a]

[a]Faculty of Science & Technology, Middlesex University, London, UK; [b]Data Analytics Department, Leicester City Football Club, Leicester, UK; [c]Sports Science & Medical Department, Leicester City Football Club, Leicester, UK

ABSTRACT
The aim of this study was to create a novel metric, Expected Pass Turnovers (xPT), that could evaluate possession retention from player-passing events in football. Event and positional data were analysed from all 380 matches in the 2020/21 English Premier League season, which encompassed 256,433 passes in the final dataset. A logistic mixed-effects model was implemented to attribute the probability of each pass getting turned over. The use of positional data enabled the identification of a) opposition players present in radii surrounding the ball carrier and b) availability of teammates with respect to the ball carrier. The addition of these positional features improved the accuracy (+6.1 AUC Score) of the model. xPT serves as a practitioner Key Performance Indicator, as analysts can identify players that lose possession more often or not than expected, given the situational context of each pass, from game to game. Future work may include modelling the turnover probability of dribble and carry actions, as this would lead to a more comprehensive understanding of turnover events in football.

ARTICLE HISTORY
Received 24 April 2023
Accepted 5 July 2024

KEYWORDS
Expected pass turnovers; logistic regression; football metric

Background

Key Performance Indicators (KPIs) are used to evaluate the performance of individuals and collectively at the team level (Hughes et al., 2012). A tactical performance is deemed to be of high quality if pre-defined KPIs are achieved, which generally leads to individual and team success (Wright et al., 2014). In football, KPIs can be categorised based on their data source, and can be extracted from event data, positional data or even a combination of both. Event data is comprised of individual events that occur during a football match, such as passes, pressures and tackles. A pressure event occurs when player(s) of the defending team attempt to deny the space for the in-possession ball carrier. Basic measures of these events include counts, and their calculated proportions (such as the proportion of pressure actions in the opposition half), which can be assessed to evaluate a team's performance (Lago-Penas & Dellal, 2010). The frequency of events may, however, be skewed due to the variation in possession duration for different teams. Hence, defensive actions (tackles, interceptions & foul events) can be aggregated and normalised based on the number of passes the opposition makes, to create the team pressure proxy: passes per defensive action (PPDA; Trainor, 2014). The number of opposition passes is used in this calculation, as this has a significant correlation with possession time (Collet, 2013) and hence is a suitable proxy to measure opposition possession. The measurement of passes can be favoured over measuring the exact time of possession due to a) the available nature of passes a team makes within a dataset and b) it can be difficult to distinguish when a particular team has possession, such as when the ball is in contest (e.g., patterns of play originating from long balls and set pieces). Overall, the limitation when using single events to evaluate team or player performance, is due to their isolated nature as a football match encompasses on average 1682 (±101) events (Pappalardo, Cintia, Rossi, et al., 2019) in complex possession sequences.

The complexity of KPIs increased when separate event types were aggregated and analysed in the context of possession strings, which enabled football events to be assessed in a broader context (McHale & Scarf, 2007; Pappalardo, Cintia, Ferragina, et al., 2019; Rudd, 2011). Hence, each event a player partakes in can be objectively valued by such models, referred to as possession value models. For example, the expected threat (xT; Singh, 2019) combines shooting events with ball carry and pass events in a Markov chain model. The aim of this possession value model was to attribute value not just to the player shooting, but also to those players involved earlier in the possession. Hence, this model can objectively and quantitatively measure the expected threat value of each player involved in the possession string, which can be used as a player evaluation tool among practitioners. Although this model had statistical robustness (Van Roy et al., 2020), only event-level information was included. Hence, the lack of positional contextual features, such as the location of teammates, means that there is potential to improve the model accuracy of such frameworks.

Further possession value models attempted to estimate a player's contribution to improving or reducing their team's probability of scoring (Decroos et al., 2019; Introducing on-ball value obv, 2022). "Valuing Actions by Estimating Probabilities" (VAEP) uses a gradient boosting

CONTACT Andrew J Peters andrewpeters1994@gmail.com Faculty of Science & Technology, Middlesex University, London NW4 4BT, UK
Supplemental data for this article can be accessed online https://doi.org/10.1080/02640414.2024.2379697

machine learning approach to estimate the probability of scoring and conceding a goal. However, VAEP tends to assign high values to players who frequently score goals and is hence biased towards attacking players (Van Roy et al., 2020). The possession value model devised by data provider, *Statsbomb*, called On-Ball Value (OBV) is another tool that enables thorough player analysis, as they break down a player's action contributions into separate player value categories: pass OBV, dribble & carry OBV, goalkeeper OBV, shot OBV, defensive action OBV. However, the output from these machine learning methods tends to be abstract values that can be difficult to interpret, potentially leading to slow practitioner uptake. Also, the utility of the OBV method is reduced, since the exact details of the commercial model are not provided. Academic studies presenting full model workings would therefore advance models in this field.

Other possession value models focused specifically on passing events (Brooks et al., 2016; Power et al., 2017; Goes et al., 2022). One method (Brooks et al., 2016) applied a supervised machine learning algorithm to derive a value from the relationship between pass locations and shot opportunities. However, this model used event-level features alone, which provides limited contextual information given that event data fails to capture the spatial and temporal interactions between players (Garganta, 2009). However, other methods have combined positional with event data to evaluate pass risk (Power et al., 2017). This method utilised tracking data to craft positional features, such as the speed of the intended receiver and nearest defender. The model accuracy is improved following the introduction of these novel tracking features, as indicated by a reduction in log loss and RMSE scores. A further method that evaluated pass risk assessment (Goes et al., 2022), demonstrated that there is a large variance in passing decision-making between players of different positional roles.

Although many KPIs evaluate attacking events in football, research is sparse for models evaluating possession breaking down, or turnover events (Forcher et al., 2022). However, at the team-level, understanding the ability to retain possession is a key characteristic of playing style (Hewitt et al., 2016). For example, teams that engage in a more established offensive playing style will have more possession and will hence have more players with good pass retention than teams that adopt a defense-to-attack transition strategy (Hewitt et al., 2016). A player's ability to retain possession has previously been analysed by considering passing events, as they represent more than 80% of the events that occur in a football match (Cintia et al., 2015). Hence, a player's pass completion rate, which is a fraction of the total number of successful passes and the total number of all passes a player makes, is used as a proxy for possession retention. However, players who fulfil different tactical roles may have inflated or deflated pass completion numbers. For example, under a Work Domain Analysis (WDA) framework (Vicente, 1999), central defensive midfielders may serve the functional purpose of connecting defence and attacking players (Berber et al., 2020). Thus, to achieve this, one player may engage in the object-related process of risk-taking (Berber et al., 2020) which may involve playing line-breaking passes with a high likelihood of interception. However, another player may simply protect and hold the ball with a lateral or backwards pass. Hence, it may not be logical to use pass completion to rank a player's passing ability.

Overall, there is evidence that there is greater accuracy for KPIs where positional data is combined with event-level information (Anzer & Bauer, 2021; Power et al., 2017). Hence, the aim of this research was to develop a KPI that leverages positional as well as event data, to better evaluate pass completion on the individual level. The proposed KPI, Expected Pass Turnovers (xPT), will aim to contextualise each pass in terms of situation and difficulty by using positional features, unlike simple pass completion metrics and other pass models (Brooks et al., 2016). The novelty within xPT lies within the method that creates it. Whereas previous methods have considered all passes existing within a homogenous group (Power et al., 2017), the underlying data structure is in fact different. For example, variation in pass completion exists at the player-level, as some players attempt riskier passes, at the positional level (Goes et al., 2022), as players in advanced positions have to keep the ball when subject to higher defensive pressure, and at the match-level, as different weather conditions may present challenges to performing a suitable passing game. Hence, by including player, position and match as independent random effects in a logistic mixed model approach, xPT considers this substructure for every pass made.

Methods

Data sample

This study was conducted with event and positional data from the 2020/21 Premier League Season, using data provided by *Statsbomb* (Statsbomb 360 - see how we're changing the game, 2022). This positional data only included the positional information of players within the television broadcast camera's range (Figure 1).

To assess the limitations of the dataset, all passing events were analysed to determine the number of players per frame Figure 2(a) and pass length Figure 2(b). As the average player count within each frame remained large (13.64), coupled with a small average pass distance (16.34 metres), it may be concluded that the dataset contains the necessary information to capably model passing performance. In addition, as the aim of football is to progress the ball to the opposition goal, it could be concluded that the important positional context exists around the ball carrier, which was captured by the dataset.

Initial data wrangling

The 2020/21 Premier League Season dataset was filtered to contain exclusively pass events. As football pitches vary slightly in length and width, the positional co-ordinates were converted to metres. This was achieved by scaling the co-ordinates according to the pitch dimensions.

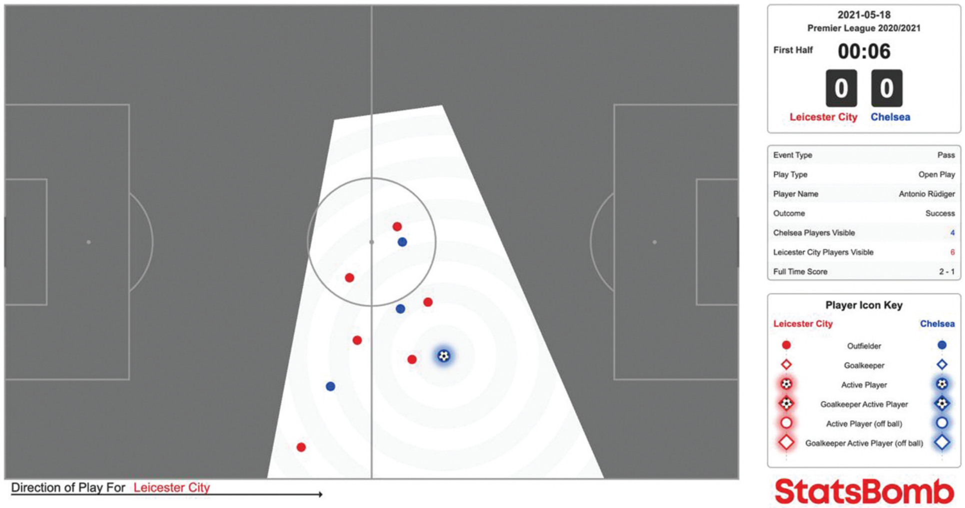

Figure 1. Limitation of the statsbomb 360 dataset (no positional information obtained from grey area). Frame taken during a successful passing action.

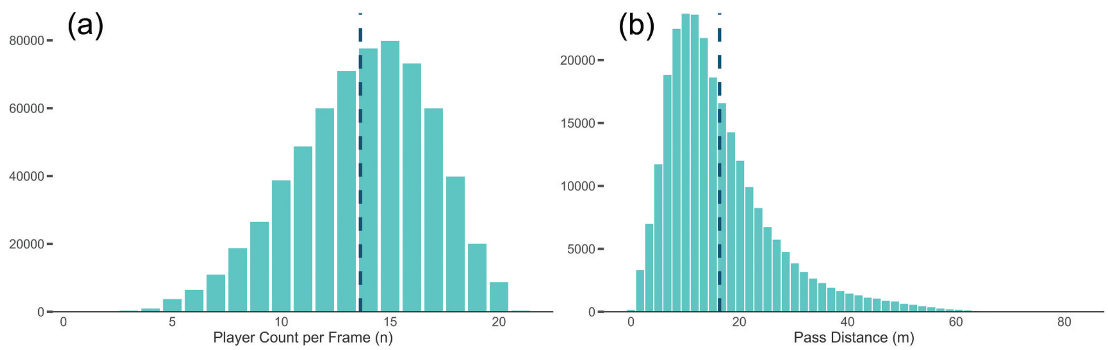

Figure 2. Distribution of a) number of players in frame (n), and b) pass length (m), for passes captured by the initial dataset.

Exclusion criteria for the dataset were passes whose outcome was not characterised as successful, unsuccessful or out of play; the initial pass from set piece situations (such as goal kick, corner kick, free kick and kick off); passes defined as a cross; passes played by goalkeepers (all of these events were coded by the data provider). The initial pass from a set piece was excluded because it is subject to no pressure and is not representative of a general free-flowing match situation. Although the initial pass from set piece situations were excluded, the remaining passes that occurred within the same pattern of play, as determined by the data provider, were retained for analysis. Crosses were excluded as they are passes to a high threat area that provide a direct goal-scoring opportunity (Yamada & Hayashi, 2015), rather than directly to a player. Hence, crosses have a lower likelihood of completion and were judged to be unrepresentative of a general pass sample. Goalkeeper passes were also removed as the aim of this KPI was to assess the performance of outfield players only. All eligible passes were labelled as a turnover event if the pass outcome was out (ball goes out of bounds) or unsuccessful (ball does not reach a teammate and is still in play; label = 1), with all successful pass outcomes (ball does reach a teammate and is still in play) labelled as a non-turnover (label = 0). These labels served as the response variable in the statistical model which employed a logistic mixed model approach. The final dataset contained 256,433 passes from all 380 matches in the 2020/21 English Premier League season.

Statistical analysis

The positional data enabled the identification of a) opposition players near the ball carrier and b) unmarked teammates in specified directions with respect to the ball carrier. In order to classify players' proximity to the ball, three radii of increasing length were drawn around the ball carrier and the number of opposition players positioned within each radii calculated. These radii were named "pressure" radii, as the number of opposition players present in each represented the intensity of pressure that the ball carrier was subject to Figure 3(a). In addition, a teammate was considered available or unmarked if they had no opponent player within a radius of 2 metres. Teammates were then categorised based on four locations with respect to the ball carrier, which was previously implemented by Van Roy et al. (2021): left (−100 < degrees < −45), right (45 < degrees < 100), in front (−45 < degrees < 45) and behind (the remaining angle; Figure 3(b)). This resulted in four independent binary features for each pass (left option, front

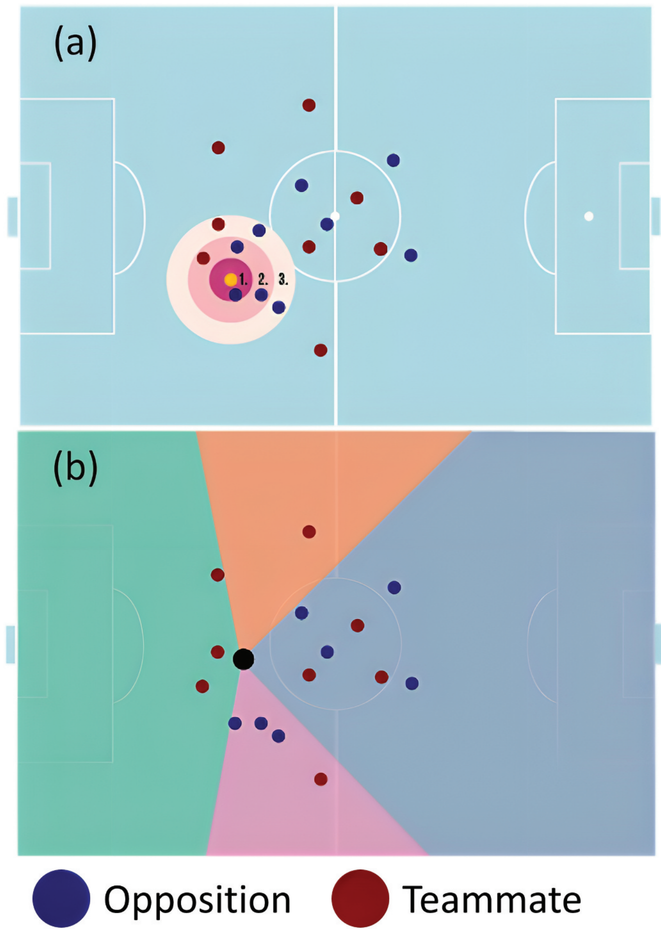

Figure 3. Pitch displaying a) pressure radii of different sizes surrounding ball carrier, and b) compartmentalisation of pitch into four different segments with respect to the ball carrier.

option, right option & back option), confirming if at least one teammate was unmarked within each location.

The event data enabled the selection of fixed effects features: ball location (x & y co-ordinate), ball progression (distance ball moved & progression towards goal), ball speed, phase of play (play pattern name), pass angle and pass type. Passes from open play were categorised differently as one-touch passes from an interception or from a ball recovery (Table 1).

In order to account for the non-independence of each observation, random effects were included in the model using the random intercept method. Different player abilities, playing positions and match conditions lead to variation in pass completion performance and hence the variables "player id",

Table 1. Description of features in the model.

Feature	Definition	Type	Unit
Ball Carrier X Co-ordinate	The x co-ordinate of the pass origin.	Event	Metres (m)
Ball Carrier Y Co-ordinate	The y co-ordinate of the pass origin	Event	Metres (m)
Ball Movement Speed	The speed of the pass, taken as the distance travelled divided by the time taken to get to it's destination.	Event	Metres per second (m/s)
Closest Defender	The distance in metres the closest opposition player is to the ball.	Event	Metres (m)
Distance Ball Moved	The distance in metres that the ball moved.	Event	Metres (m)
Pass Angle	The angle in radians of the pass, calculated clockwise from 0 representing straight ahead, to π.	Event	Radians (rad)
Pass Type Name	Passes were labelled to be from one touch pass from an interception or from a loose ball recovery. The remaining passes were labelled as normal.	Event	Name
Percentage Increaes in Distance Towards Goal (Ball Progression)	The percentage increase in distance the ball moved towards the centre of the opposition goal.	Event	Percentage (%)
Play Pattern Name	The phase of play relevant to the pass event including: throw-in, kick-off, goal kick, free kick, counter attack, corner kick. All remaining play patterns were labelled as regualr play.	Event	Name
Three Pressure Radii	The number of opposition players independently present within three radii of varying length surrounding the ball carrier.	Positional	Count (n)
Unmarked Teammates	Teammates that had no opponent player within 2 metres, across four different directions (options), were considered as unmarked.	Positional	Binary (1 = marked, 0 = unmarked)

Table 2. Position group classification of centre midfield players.

Formation	Initial Position	Labelled Playing Position
4-4-2	Centre Midfield	Defensive Midfield
4-4-1-1	Centre Midfield	Defensive Midfield
4-2-2-2	Centre Midfield	Defensive Midfield
4-2-3-1	Centre Midfield	Defensive Midfield
4-3-2-1	Centre Midfield	Defensive Midfield
3-4-3	Centre Midfield	Defensive Midfield
3-4-2-1	Centre Midfield	Defensive Midfield
3-4-1-2	Centre Midfield	Defensive Midfield
4-1-2-1-2	Centre Midfield	Attacking Midfield
4-3-3	Centre Midfield	Attacking Midfield
4-5-1	Centre Midfield	Attacking Midfield
4-1-4-1	Centre Midfield	Attacking Midfield
3-5-2	Centre Midfield	Attacking Midfield
3-5-1-1	Centre Midfield	Attacking Midfield

"playing position" & "match id", respectively, were modelled as random effects in a cross-classified multilevel design. Playing positions were categorised into seven outfield groups: centre back, full back, wing back, defensive midfield, attacking midfield, forward & winger. As some playing positions may have overlapping roles and hence can be difficult to differentiate, practitioner analysts were consulted to identify what formations characterised these different positional groups from game to game (Table 2).

As many features were used, the Variance Influence Factor (VIF) was calculated to assess multi-collinearity. A generalised VIF (GVIF) was used as some features had greater than 1 degree of freedom (e.g., the different play patterns). In order to make the GVIF comparable across the different dimensions of the dataset, the GVIF^(1/(2*Df)) was inspected. None of these values exceeded 1.5, suggesting that considerable multi-collinearity was not present in the model (Johnston et al., 2018).

The model accuracy was determined by predicting pass turnovers which was learned from the training data (80% of initial dataset) and applied to the test data (remaining 20%). The motivation for using this 80/20 split was driven by the Pareto principle (Pareto, 1896), which states that 80% of the effect is driven by 20% of the causes. A decision boundary was applied, whereby prediction probabilities less than 0.5 were considered non-turnover events and those greater than 0.5 were considered a turnover event.

Applications

The model was applied to a single, randomised match in the 2020/21 English Premier League to demonstrate the model's applications. During the game, the xPT values for all outfield players, including substitutes, for all passes was calculated. These values were averaged per player, and multiplied by 100 to determine the expected turnovers per 100 passes of each player.

As some players repeatedly pass the ball backwards to teammates, whereas others perform more dangerous passes into advanced areas, it was important to consider xPT in the context of a ball progression or threat metric. The model was applied to all the centre backs that completed at least 25% of the minutes across the 2020/21 English Premier League Season. As different position groups were used as a random effect in the model, to enable a fair comparison, the centre back group was only used to demonstrate this proof of concept. Again, the xPT values were averaged per player, and multiplied by 100, to determine the expected turnovers per 100 passes of each player. The trade-off between pass OBV (Introducing on-ball value obv, 2022) and xPT performances were analysed to

Figure 4. Log of Odds Ratios (circles) for all features in the model.

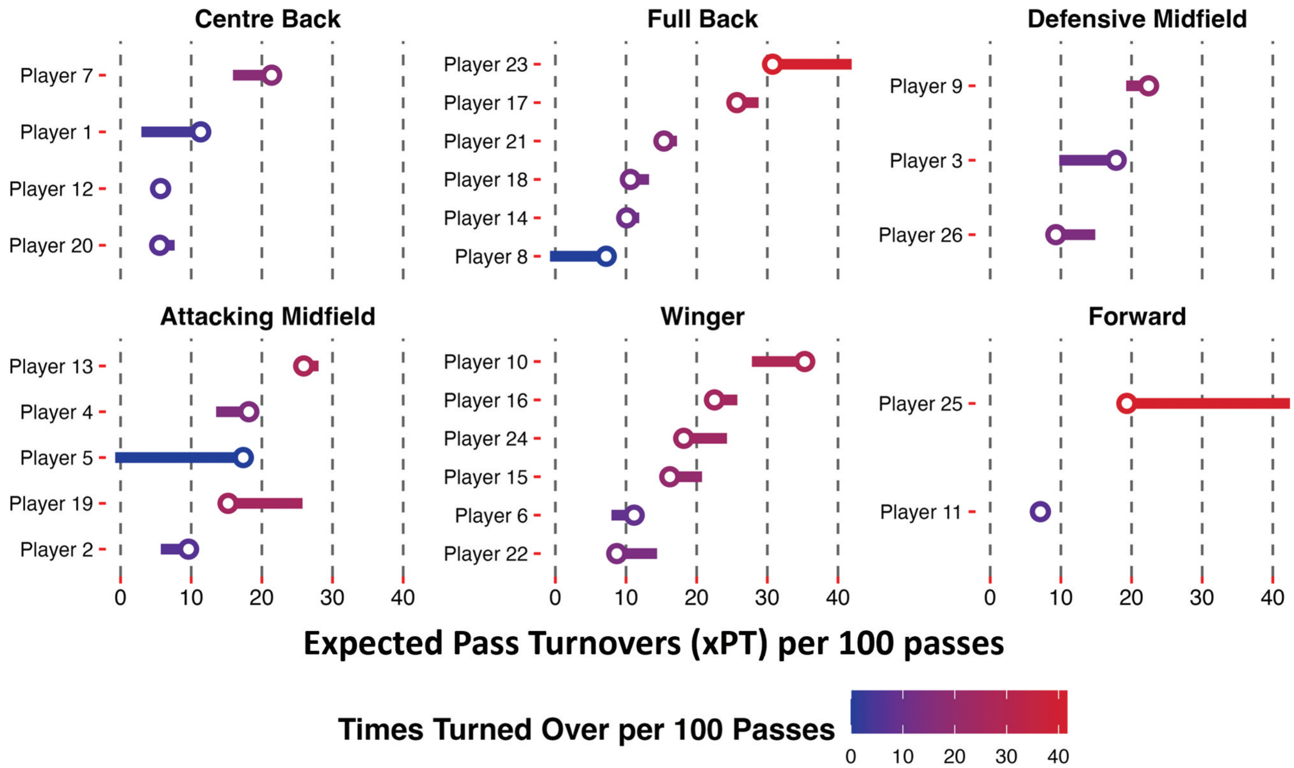

Figure 5. Normalised expected pass turnover (xPT) values obtained from a randomly selected match. Circles display xPT value, and the line ends display the actual turnovers suffered.

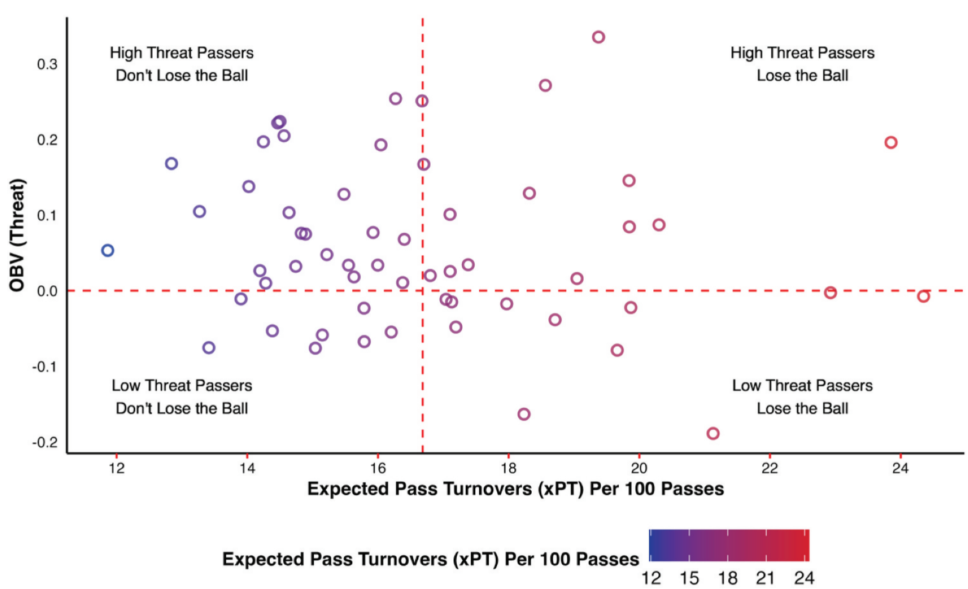

Figure 6. Scatter plot of player OBV (threat) vs. Expected pass turnovers (xPT) per 100 passes for centre backs.

consider those players who improve their team's goalscoring probability, yet frequently retain the ball more than expected.

Results

Model evaluation

All features had a significant association with a turnover event following a Wald Test and were plotted to inspect their respective odds ratios (Figure 4). The speed at which the ball moved was the greatest predictor of a turnover event (Log OR 3.76), with a counter-attacking situation being the biggest predictor of a non-turnover event (or completing a pass) as indicated by its negative log odds ratio value (Log OR −1.47). With respect to the positional features, the availability of teammates to the left (Log OR −0.49) and right (Log OR −0.43) of the ball carrier had the largest impact in non-turnover prediction. Overall, the player random effect had the greatest variation in determining the pass turnover probability ($\sigma 2 = 0.08$), whereas the playing position ($\sigma 2 = 0.03$) and match ($\sigma 2 = 0.02$) had a lower degree of variation.

Following training and test subset partitioning, the model achieved an accuracy of 0.85. The model was also subject to a 5 K-Fold cross-validation procedure which matched these model values (0.85 accuracy for all folds), suggesting the model's high accuracy and that the model was not prone to overfitting. A further method to determine model accuracy is the comparison of the true positive rate with the false-positive rate. The area under the receiver operator curve (ROC AUC) was calculated at 80.1%, demonstrating that it was an excellent predictive model (Mandrekar, 2010).

The model was also evaluated with and without the positional features: pressure radii and availability of unmarked players. In addition, the model was further evaluated with and without the random effects: "player id", "playing position" and "match id". Overall, the model inclusive of positional-fixed effects and random effects was more accurate (Log Loss 7.65; AIC 187,516; AUC 80.1), compared with the model without random effects (Log Loss 7.82; AIC 190,243; AUC 78.8) & the model without both random effects & positional features (Log Loss 8.32; AIC 203,520; AUC 72.7). A significant ($p < 0.001$) Chi-square likelihood ratio test suggests an improvement in model accuracy when utilising the mixed model method, compared to the other naive models.

Applications

The average xPT value per 100 passes of individual players from a randomised match in the 2020/21 English Premier League was plotted (Figure 5). The players' names were removed to preserve anonymity. In the below example, player 25 was expected to get turned over from ≈20/100 or 20% of their passes (indicated by the circle). It was also possible to determine how frequently these players actually got turned over from their passes. Thus, player 25 got turned over for ≈40% of their passes (indicated by end of line). This demonstrates that player 25 lost the ball 20% more than expectation, given the passing scenarios that they were exposed to. Hence, it is possible to evaluate the pass completion performance of players relative to expectation at the match-level, which can be utilised by practitioner performance analysts to evaluate a player's passing performance.

The relationship between OBV (Introducing on-ball value obv, 2022) and xPT was analysed and plotted (Figure 6) for the centre back group. The players' names were removed to preserve anonymity. Players existing in the top left of the figure represent centre backs who play passes that enhance their team's chances of scoring, yet frequently retain possession.

Discussion

The aim of this study was to create a KPI that evaluates the likelihood of players losing possession of the ball through passing actions, using a combination of event and positional data. Passing events were considered as they're the most frequent event in football, representing more than 80% of all events (Cintia et al., 2015). As greater possession implies greater team success (Hook & Hughes, 2001; Jones et al., 2004), evaluating pass completion may lead to critical insight into positive team performance. This is what xPT achieves, as a player's xPT performance can be interpreted after a match and compared with expectation. This is more informative than interpreting a player's pass completion rate, which is a fraction of the total number of successful passes and total number of all passes a player makes. This is because different players occupy different pitch regions and play in different match conditions, which is accounted for in our model with the inclusion of "player id", "positional group" and "match id" random effects respectively.

Although it is possible to evaluate passing retention through other methods such as a team passing networks (Ievoli et al., 2021), this approach may present many flaws. For example, although it is easy to identify strong passing links between players, many of these networks do not consider changes in formation or personnel that may occur throughout a game, and therefore the network will be skewed if players occupy different positions. In addition, they only portray information about successful passes and not those that are turned over. In contrast, xPT evaluates passing retention as it considers turnover events, yet also retains the spatial information of where each pass occurred despite alterations in players and personnel. In addition, the xPT metric advances on previous passing metrics that use event-data alone (Brooks et al., 2016), as it also combines positional features. The AUC score increased by 6.1 following the addition of positional features, indicating an improvement in model accuracy. This demonstrates the importance of using positional data to develop KPIs in football, as it can capture player positioning across the pitch, providing greater situational context which cannot be achieved when using event data alone. Furthermore, the model accuracy is improved with the addition of the random effects "player id", "positional group" & "match id". Whereas previous methods have considered all passes existing within a homogenous group (Power et al., 2017), it was important to consider the substructure of the pass dataset in a mixed model approach.

The accuracy of our model is in line with those previously published. Spearman et al. (2017) achieved an accuracy of 80.5% after mathematically modelling the probability of passing success. A further method that assessed the risk-reward of pass decisions (Goes et al., 2022), achieved a prediction accuracy of 84.5% with their test-set. In the current study, an average prediction accuracy of 85% was achieved following a 5-fold cross-validation procedure, indicating that our model outperformed others in the field.

A further utility of xPT is that it can be used in combination with outputs from possession value models. Whereas these models focus on how a pass action can enhance a team's probability of scoring (xT; McHale & Scarf, 2007; Pappalardo, Cintia, Ferragina, et al., 2019; Rudd, 2011; Singh, 2019), xPT can objectively quantify the pass risk, or how likely the pass is getting turned over. Hence, it is possible to identify players who play threatening passes, yet retain possession of the ball through their passing, as xPT reveals an additional layer of information with respect to a player's passing habits.

Limitations and future work

Overall, this study was conducted on the event and positional data from the 2020/21 Premier League season. This was an unprecedented season, owing to disruptions caused by the COVID-19 pandemic and most notably, a lack of spectators at matches. To ensure reproducibility, a similar study should be conducted on a dataset originating from a non-disrupted season. In addition, although important player positional information exists near the ball-carrier (particularly for pressure), the *Statsbomb 360* (https://statsbomb.com/360-data/) dataset does not include positional data for players outside of the television broadcast camera's range. Future studies should consider using full tracking data, to analyse positional patterns of players outside the broadcast range. Additional future modelling work may consider turnover events other than passes such as from dribbles. These models could be combined, giving a more comprehensive understanding of turnover events in football.

Conclusion

In this paper, the Expected Pass Turnover (xPT) KPI was developed. Unlike previous KPIs that modelled goalscoring probabilities (VAEP; Decroos et al., 2019) and attacking progression (xT; Singh, 2019), this model considers the probability of passes getting turned over. The model features derive from a combination of event and positional sources, increasing the model accuracy (+6.1 AUC Score). Using this KPI, analysts can identify players that lose possession more often or not than expected from passing actions. This can be used at the practitioner level to evaluate performance and meet pre-defined objectives. Going forward, this model can be applied to data from different seasons and leagues, to examine the variability of turnovers in football. Future studies may include the use of tracking data, which would grant positional information of every player present on the pitch rather than those players present within the broadcast camera frame. The accuracy of the model may improve further with the inclusion of more sophisticated model features such as pitch control (Spearman et al., 2017) to better evaluate the space afforded to each player.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported that there is no funding associated with the work featured in this article.

ORCID

Andrew J Peters http://orcid.org/0000-0002-2586-1523
Nimai Parmar http://orcid.org/0000-0001-5540-123X
Nic James http://orcid.org/0000-0002-4181-9501

Data availability statement

The data that support the findings of this study are available from *Statsbomb*. Restrictions apply to the availability of this data, which were used under licence for this study.

References

Anzer, G., & Bauer, P. (2021). A goal scoring probability model for shots based on synchronized positional and event data in football (soccer). *Frontiers in Sports and Active Living*, *3*. https://doi.org/10.3389/fspor.2021.624475

Berber, E., McLean, S., Beanland, V., Read, G. J. M., & Salmon, P. M. (2020). Defining the attributes for specific playing positions in football match-play: A complex systems approach. *Journal of Sports Sciences*, *38* (11–12), 1248–1258. https://doi.org/10.1080/02640414.2020.1768636

Brooks, J., Kerr, M., & Guttag, J. (2016). *Developing a data-driven player ranking in soccer using predictive model weights*. https://doi.org/10.1145/2939672.2939695

Cintia, P., Rinzivillo, S., & Pappalardo, L. (2015). Superoxide dismutase activity, hydrogen peroxide steady-state concentration, and bactericidal and phagocytic activities against Moraxella bovis, in neutrophils isolated from copper-deficient bovines. *Biological Trace Element Research*, *171*(1), 94–100. https://doi.org/10.1007/s12011-015-0505-2

Collet, C. (2013). The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. *Journal of Sports Sciences*, *31*(2), 123–136. https://doi.org/10.1080/02640414.2012.727455

Decroos, T., Bransen, L., van Haaren, J., & Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax, NS, Canada (pp. 1851–1861).

Forcher, L., Altmann, S., Forcher, L., Jekauc, D., & Kempe, M. (2022). The use of player tracking data to analyze defensive play in professional soccer - A scoping review. *International Journal of Sports Science & Coaching*, *17* (6), 1567–1592. https://doi.org/10.1177/17479541221075734

Garganta, J. (2009). Trends of tactical performance analysis in team sports: Bridging the gap between research, training and competition. *Revista Portuguesa de Ciências do Desporto*, *9*(1), 81–89. https://doi.org/10.5628/rpcd.09.01.81

Goes, F., Schwarz, E., Elferink-Gemser, M., Lemmink, K., & Brink, M. (2022). A risk-reward assessment of passing decisions: Comparison between positional roles using tracking data from professional men's soccer. *Science and Medicine in Football*, *6*(3), 372–380. https://doi.org/10.1080/24733938.2021.1944660

Hewitt, A., Greenham, G., & Norton, K. (2016). Game style in soccer: What is it and can we quantify it? *International Journal of Performance Analysis in Sport*, *16*(1), 355–372. https://doi.org/10.1080/24748668.2016.11868892

Hook, C., & Hughes, M. (2001). Patterns of play leading to shots in Euro 2000. *Pass. Com*, *1*, 295–302.

Hughes, M., Caudrelier, T., James, N., Redwood-Brown, A., Donnelly, I., Kirkbride, A., & Duschesne, C. (2012). Moneyball and soccer – an analysis of the key performance indicators of elite male soccer players by position. *Journal of Human Sport & Exercise*, *7*(2), 402–412. https://doi.org/10.4100/jhse.2012.72.06

Ievoli, R., Palazzo, L., & Ragozini, G. (2021). On the use of passing network indicators to predict football outcomes. *Knowledge-Based Systems*, *222*, 106997. https://www.sciencedirect.com/science/article/pii/S0950705121002604

Introducing on-ball value (obv). (2022). Retrieved April 13, 2022, from. https://statsbomb.com/2021/09/introducing-on-ball-value-obv/

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity*, *52*(4), 1957–1976. https://doi.org/10.1007/s11135-017-0584-6

Jones, N., James, P. D., & Mellalieu, S. (2004). Possession as a performance indicator in soccer. *International Journal of Performance Analysis in Sport*, *4*(1), 98–102. https://doi.org/10.1080/24748668.2004.11868295

Lago-Penas, C., & Dellal, A. (2010). Ball possession strategies in elite soccer according to the evolution of the match-score: The influence of situational variables. *Journal of Human Kinetics*, *25*(2010), 93–100. https://doi.org/10.2478/v10078-010-0036-z

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

McHale, I., & Scarf, P. (2007). Modelling soccer matches using bivariate discrete distributions with general dependence structure. *Statistica Neerlandica*, *61*(4), 432–445. https://doi.org/10.1111/j.1467-9574.2007.00368.x

Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2019). Playerank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach. *ACM Transactions on Intelligent Systems and Technology*, *10*(5), 1–27. https://doi.org/10.1145/3343172

Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D., & Giannotti, F. (2019). A public data set of spatio-temporal match events in soccer competitions. *Scientific Data*, *6*(1), 236. https://doi.org/10.1038/s41597-019-0247-7

Pareto, V. (1896). *Cours d'conomie Politique Profess a l'Universit de Lausanne. I.*

Power, P., Ruiz, H., Wei, X., & Lucey, P. (2017). Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, United States (pp. 1605–1613).

Rudd, S. (2011). A framework for tactical analysis and individual offensive production assessment in soccer using Markov chains. *New England Symposium on Statistics in Sports*, Harvard University, Cambridge, MA, United States.

Singh, K. (2019). *Introducing expected threat (xt)*. Retrieved April 21, 2022, from. https://karun.in/blog/expectedthreat.html

Spearman, W., Basye, A., Dick, G., Hotovy, R., & Pop, P. (2017). Physics-based modeling of pass probabilities in soccer. *MIT Sloan Sports Analytics Conference*, Boston, MA, United States.

Statsbomb 360 - see how we're changing the game. (2022). Retrieved April 21, 2022, from. https://statsbomb.com/360-data/

Trainor, C. (2014). *Defensive metrics: Measuring the intensity of a high press*. Retrieved February 20, 2022, from. http://statsbomb.com/2014/07/defensive-metrics-measuring-the-intensity-of-a-high-press/

Van Roy, M., Robberechts, P., & Davis, J. (2021). *Optimally disrupting opponent build-ups*. Paper Track.

Van Roy, M., Robberechts, P., Decroos, T., & Davis, J. (2020). *Valuing on-the-ball actions in soccer: A critical comparison of xt and vaep*.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive and healthy computer-based work*. CRC Press.

Wright, C., Carling, C., & Collins, D. (2014). The wider context of performance analysis and it application in the football coaching process. *International Journal of Performance Analysis in Sport*, *14*(3), 709–733. https://doi.org/10.1080/24748668.2014.11868753

Yamada, H., & Hayashi, Y. (2015). Characteristics of goal-scoring crosses in international soccer tournaments. *Football Science*, *12*, 24–32. https://api.semanticscholar.org/CorpusID:202723043