Routine Inspection: A Playbook for Corner Kicks

Laurie Shaw[1](✉) and Sudarshan Gopaladesikan[2]

[1] Harvard University, Cambridge, USA
laurie.shaw@cfa.harvard.edu
[2] Benfica SL, Lisbon, Portugal
sgopaladesikan@slbenfica.pt

Abstract. We present a set of tools for identifying and studying the offensive and defensive strategies used by football teams in corner kick situations: their corner playbooks. Drawing from methods in topic modelling, our tools classify corners based on the runs made by the attacking players, enabling us to identify the distinct corner routines used by individual teams and search tracking data to find corners that exhibit specific features of interest. We use a supervised machine learning approach to identify whether individual defenders are marking man-to-man or zonally and study the positioning of zonal defenders over many matches. We demonstrate how our methods can be used for opposition analysis by highlighting the offensive and defensive corner strategies used by teams in our data over the course of a season.

1 Introduction

A cross from the winger is blocked by a defender, who concedes a corner. The crowd roars, anticipating a goal-scoring opportunity. The statistics, however, predict a different outcome. According to data collected by OPTA, in the 2019–20 season of the Portuguese Primeira Liga, just 45 goals have been scored from 3082 corners – a 1.5% conversion rate. In the German Bundesliga the conversion rate was 1.9%, and in Spain's La Liga it was only 0.8%. Similarly, Power et al. (2018) found that only 2.1% of corners resulted in a goal over three seasons of the English Premier League. Given that corner kicks are one of the few attacking situations that occur in nearly every match – and that they can easily be rehearsed in training – the low average conversion rate is somewhat surprising. Indeed, some teams have been able to achieve a much higher success rate [1].

Previous work has focused on studying general measures of corner strategy, such as delivery type (inswinging, outswinging or straight), the ball delivery zone (e.g. near post, far post, center), the number of attacking players involved and whether they were 'static' or 'dynamic' [1–5]. Casal et al. (2015) analyzed the factors that lead to a shot or a goal, finding that match time, the number of intervening attackers, and whether the attack was 'dynamic' were the most significant variables. Power et al. (2018) found that a goal was more likely to be

U. Brefeld et al. (Eds.): MLSA 2020, CCIS 1324, pp. 3–16, 2020.
https://doi.org/10.1007/978-3-030-64912-8_1

scored on the second ball (i.e. after a touch from a teammate) than directly from the corner kick. From a defensive perspective, they found that hybrid systems concede the most dangerous shots relative to pure zonal or man-to-man systems.

With the advent of player tracking data, it is now possible to perform a detailed analysis of the synchronised runs made by the attacking players – the rehearsed *routines* that define the core of offensive corner tactics. Tools developed for this purpose would enable analysts to identify and study the corner routines used frequently by an opponent over a large number of matches. Furthermore, while previous studies of corner defence have focused on general classifications of the defensive system (zonal, man marking or hybrid), it is quite rare for teams to employ a purely zonal or man-marking system: most use some form of a hybrid system [1]. The challenge is to identify which defenders are man-marking, which defenders have been assigned to zones and where those zones are located.

This paper aims to meet these challenges. Using a combination of statistical and machine learning techniques and a large sample of tracking data, we have developed tools to classify corner routines based on the runs made by the attacking players, enabling us to identify the distinct corner routines employed by teams in our data. We have trained a supervised learning algorithm to identify whether *individual defenders* have been assigned a man-marking or zonal role and studied the positioning of the zonal players of teams in our sample.

The paper is organised as follows. In Sect. 2 we describe our data and the selection of our corner sample. In Sect. 3 we describe our technique for classifying corner routines based on the runs made by individual players. In Sect. 4 we describe our methodology for identifying the roles of the defending players during corner kicks, making use of a unique training set provided by the analysts at SL Benfica. In Sect. 5 we demonstrate how our methods can be used for opposition analysis, before concluding in Sect. 6.

2 Data

Our analysis uses tracking and event data for 234 matches from a single season of an elite professional league. The tracking data for each match consists of the positions of all 22 players and the ball, sampled at a frequency 25 Hz. Individual player identities are tagged in the data, enabling tracking of each player over time. The event data consists of a log of each on-ball event that took place during a match (e.g. passes, shots, tackles and interceptions), including the identity of the players involved and the time and location on the pitch at which the event occurred.

We use the event data to provide an initial estimate for the time at which each corner kick was taken during a match. To identify the exact frame we use a combination of factors, including the acceleration of the ball and a *ball in play* flag included in the tracking data. After removing short corners (which are not included in this analysis) and a small number of corners for which ball tracking errors were identified, we were left with a sample of 1723 corner kicks. Finally, to aid comparisons, we reflected the positions of players and the ball so that corners always appear to be taken from the left side of the pitch.

3 Classifying Corner Routines

Figure 1 illustrates two examples of corner routines in our sample. The offensive strategies are clearly very different: the starting positions of the players, their trajectories and the delivery target of the ball. One of the main goals of this work is to develop tools to search tracking data to identify the different corner routines used by teams over many matches. We achieve this by developing a classification system for describing the main features of a routine based on the runs made by the individual players in the attacking team. This system enables us to find corners that have similar features, or to quickly search the data for corners that exhibit a particular feature of interest.

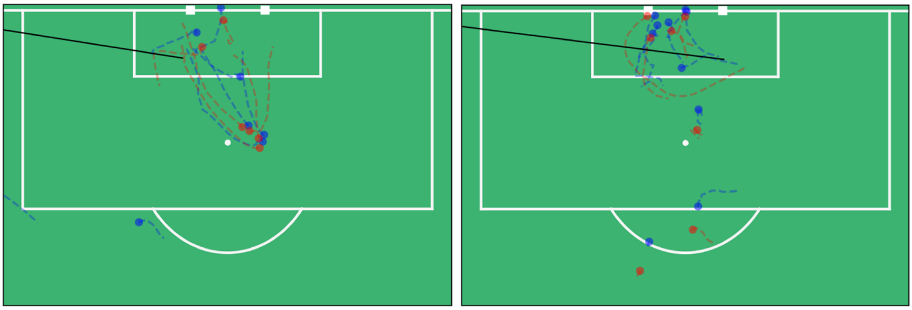

Fig. 1. Two corners in our data depicting very different offensive strategies. The red (blue) markers indicate the positions of the attacking (defending) players two seconds before the corner is taken. The dashed lines indicate each player's trajectory, from two seconds before the corner was taken until approximately two seconds after. The solid black line indicates the path of the ball over the same time period. (Color figure online)

Our methodology has two key steps:

1. a data-defined zonal system for classifying player runs based on their start and target locations; and
2. a topic model for identifying player runs that frequently co-occur in corner routines.

We now summarize each of these steps in more detail. Note that the trajectory of the ball does not feature in our system. This is because the ball does not always reach the intended target, either because it was intercepted or because the cross was not sufficiently accurate. We recognize that the ball can also be delivered using different trajectory types: curving away from goal, curving towards goal, floated or flat. We intend to incorporate this information in a follow-up study evaluating the effectiveness of different corner routines.

3.1 Classifying Player Runs

The basic building blocks of a corner routine are the individual runs made by the players. We define a run entirely in terms of the initial and target location; we do not attempt to model a player's trajectory between those locations. Initial positions are measured exactly two seconds before the corner is taken – this corresponds to the moment at which the average speed of players typically starts to increase as they begin their runs. The target locations are defined as being either the positions of the players exactly one second after the first on-ball event following the corner, or two seconds after the corner is taken, whichever occurs first[1]. It is impossible to know the true intended target location of a player, we simply assume that attacking players always reach their target.

We allocate players to pairs of zones based on their initial and target locations. These zones are defined using the distribution of the initial and target positions of all the attacking players in our sample of corners. The process starts with the distribution of the target positions. The upper panel of Fig. 2 indicates the target positions of nearly 15000 attacking players measured over the 1723 corners in our sample (only players in the attacking quarter of the field are plotted). The cloud of points in the top-left corner corresponds to the positions of the corner takers shortly after each corner is taken.

Target zones are defined by fitting a 15-component Gaussian Mixture Model (GMM) using the expectation-maximisation algorithm [6,7]. We find that 15 components (that is, 15 bivariate normal distributions) are sufficient and that adding further components does not result in a significant improvement in the log-likelihood. The lower-left panel of Fig. 2 shows each of the 15 components in the GMM. The seven components of the model located in the penalty area are indicated by blue ellipses and labelled a to g. We henceforth refer to these as the *active zones*. Individual points belonging to the same active zone are coloured accordingly. Players with a target position near one of these seven active zones are assumed to be directly involved in the corner routine: these are referred to as *active players*. Players that do not end near an active zone are ignored in our analysis.

The upper-right panel of Fig. 2 shows the initial positions of attacking players, two seconds before the corner is taken. Active players are coloured blue and form two groups: those that start inside the six yard box, and the players that are initially clustered around, and depicted slightly to the right of, the penalty spot. Points coloured black are players that were not actively involved in the corner (including the corner taker, who is no longer involved after taking the corner). To define the initial zones of active players we fit a 6 component GMM model (with some outlier removal) to their initial positions. The six components of our fit are labelled 1 to 6 in the lower-right panel of Fig. 2. The fit was not improved significantly by adding additional components to the model.

Allocating players to initial and target zones enables a simple encoding of player runs. Active players are allocated to an initial zone (1–6) and a target

[1] Measuring positions one second after the first ball event helps to identify the target position of players aiming to reach a flick-on.

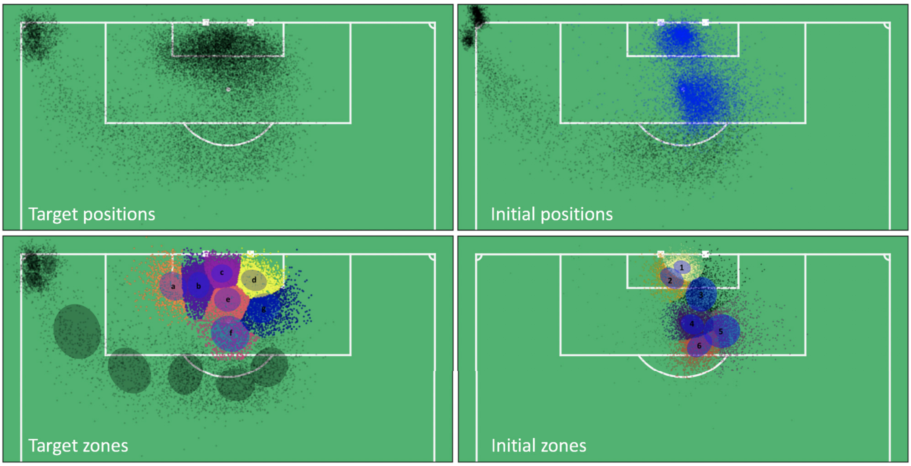

Fig. 2. (*upper left*) The target positions of all 15000 attacking players in our sample of corners. (*lower left*) The results of the 15-component GMM fit to the target positions. The seven 'active zones' in the penalty area are represented by blue ellipses and labelled *a-g*. Individual points are coloured according to the active zone to which they belong. (*upper right*) The initial positions of all 15000 attacking players. Players coloured blue are tagged as 'active'. (*lower panel*) The results of a 6-component GMM fit to the initial positions of the active players.

zone (a-g) based on their initial and target positions. For example, in the left panel of Fig. 1 the four attacking players that start their runs next to the penalty spot are initially in zone *4* and move into target zones *b*, *c* and *d*. Their runs are therefore encoded as $\{4b,4b,4c,4d\}$.

3.2 Topic Modelling

The runs made by the attacking players are coordinated and synchronised: some players will attempt to draw away defenders, while others will attempt to intercept the ball. The second step of our method is to identify the types of runs that are frequently combined in corner routines. To achieve this, we draw inspiration from topic modelling by making the analogy between runs and words, combinations of runs and topics, and corner kicks and documents.

We use non-negative matrix factorization (NMF) to represent the corners in our data in terms of a basis set of run combinations. NMF decomposes an initial matrix, called the *term matrix*, into two lower-rank, non-negative matrices, W and H [8,9]. Our term matrix has the following dimensions: 42 rows by 1723 columns. The rows represent all 42 combinations of the 6 initial and 7 target zones and each column corresponds to a corner in our data set. Each element in the term matrix is given by the sum of the probabilities that each active player in the corner made a run between a specific pair of initial and target zones. For each player, this is calculated by multiplying the GMM posterior probability that the player started in the initial zone by the probability that he ended in the target zone. W represents the run combinations (or topics) that frequently co-occur in the data; H tells you how to construct each corner in the data from the run combinations.

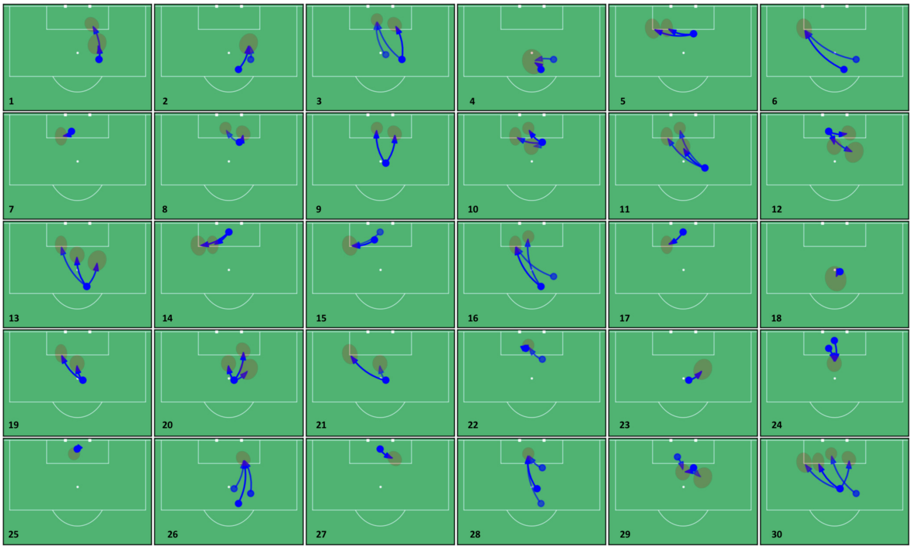

Fig. 3. The thirty features, or frequently co-occurring runs, identified by our topic model.

We find that the corners in our data can be accurately reconstructed from a set of 30 run combinations (henceforth referred to as *features*). These features are shown in Fig. 3. In some cases, a feature consists of just a single run – this is because the same run may occur in many different corner routines. We can represent every corner in our data in terms of a sub-sample of these features and group together corners that exhibit similar features to identify distinct routines. For example, the corner depicted in the left panel of Fig. 1 exhibits features 9 and 19, which describe the runs from the penalty spot towards the near post, goal center and far post. The corner depicted in the right panel of Fig. 1 strongly exhibits feature 12, which describes runs from the near post towards the far post. Both corners also exhibit feature 25: an attacking player standing close to where the goalkeeper would be located.

We can also use the feature representation to rapidly search large numbers of matches to find corners that exhibit a certain feature of interest, or a combination of features. Figure 4 shows other corners in our sample that strongly exhibit feature 12. Three teams in particular make frequent use of this corner routine.

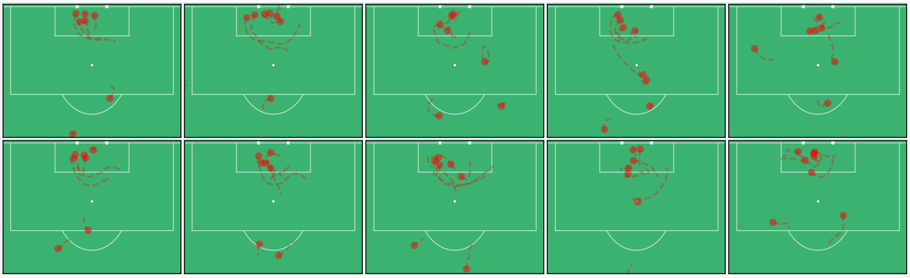

Fig. 4. Ten corners in our sample that strongly exhibit feature 12 – runs from the near post round to the far post. Only attacking team players are shown. Large dots indicate the initial positions of each player and dashed lines show their trajectories.

4 Identifying Defensive Roles

The defining feature of defensive strategy in corner kick situations is the use of man- or zonal-marking systems. Man-to-man marking requires a player to closely track a specific opponent, while zonal marking requires a player to defend a spatial region (we consider defending a goal post as zonal marking). Few teams use an exclusively zonal marking system and it is rare for a team to have no zonal marking players whatsoever. For example, Power et al. (2018) find that 80% of the teams in their data employed hybrid systems (i.e., a mixture of man-to-man and zonal marking).

For this reason we work on the level of individual players, using supervised machine learning to classify the role of each defender in the penalty area. We use the popular XGBoost implementation of gradient boosted decision trees [10] to calculate the probability that each defender in the penalty area is marking man-to-man. Gradient boosted decision trees have been shown to be a powerful tool for solving classification problems in an efficient manner [11].

The key distinction between man-to-man and zonal marking is that, in the former, a player is marking a moving target rather than a static region. Treating attackers near the goalkeeper as a special case and emphasizing the locomotive reaction of the defenders, we selected the metrics listed below as predictive variables. These variables were vetted by video analysts at SL Benfica[2]:

1. initial position (x coordinate)*
2. initial position (y coordinate)*
3. distance between start and target positions*
4. initial proximity to goalkeeper*
5. average distance travelled by attacking players in the same initial zone*
6. average distance travelled by other defenders in the same initial zone*
7. initial zone
8. target zone

[2] Our methodology does not identify the specific opponent a defender is man-marking.

9. # of attacking players in the same initial zone
10. # of other defenders in the same initial zone

To provide data for training and testing, analysts at SL Benfica watched a sample of 500 movies of corners selected from our data set and manually identified the jersey numbers of the man-to-man and zonal defenders. In total, 3907 defenders were evaluated: 55% were tagged as marking man-to-man and the remainder tagged as zonal; the two classes are therefore well-balanced. Figure 5 shows the initial positions of zonal (upper-left) and man-marking players (upper-right) in the analysts' sample. It is clearly evident that player positions are a strong discriminator.

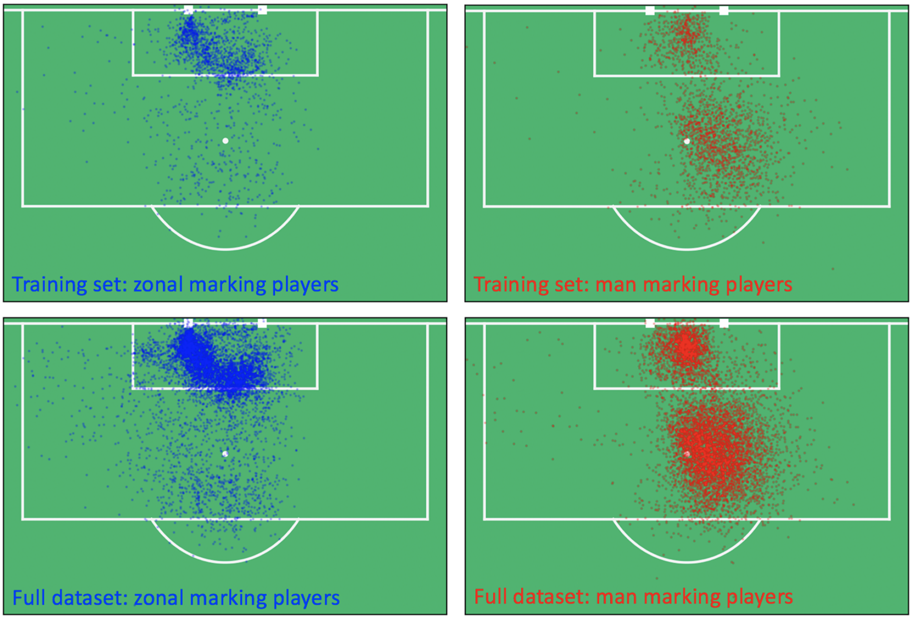

Fig. 5. (*upper*) The initial positions of zonal (left) and man-marking players (right) based on the manual classifications made by Benfica's analysts. (*lower*) Results for the full sample, based on the classifications predicted by the trained XGboost model.

We use 66% of the analysts' annotated sample as the training set, with the remaining 33% reserved as a test set to evaluate the performance of the model. We obtain a classification accuracy of 83.5%; training instead with 10-fold cross validation results in a classification accuracy of $83.4\% \pm 2.1\%$. The most predictive features, selected via their F-score, are marked with an asterisk in the list above. We apply the decision tree to our remaining sample of 1223 corners to predict the roles of each defending player. The initial positions of players classified as zonal (lower-left) and man-marking (lower-right) are shown in Fig. 5.

There is clearly a strong resemblance to the analyst-annotated sample. Note that the spatial distribution of man-marking defenders also resembles the initial positions of the active attacking players shown in Fig. 2.

5 Applications: Opposition Analysis

Anticipating how an opponent will play in different phases of a game is a crucial aspect of pre-match preparation [12,13]. However, evaluating an opponent using video is a time-consuming process and so club analysts are often limited to studying approximately five matches to inform their reports for the coaching staff. Such a small sample of matches provides only a limited insight into the range of set-piece strategies that might be utilized by the opponent.

Our methodology enables us to rapidly identify the key features of the corner strategies used by teams over a large sample of their matches. In this section we demonstrate its application to opposition analysis, highlighting the corner routines used frequently by teams in our sample over the course of a season and analysing the systems they employed to defend corners.

5.1 Offensive Strategy

Figure 6 shows four corner routines used regularly by teams A (red), B (blue), C (green) and D (black) in our data set. Each panel shows a specific example of a routine that the team used multiple times throughout the season. Distinct routines were identified by clustering corners based on their feature activation (columns of the H matrix). The circles indicate the starting position of each player and the dashed lines indicate their runs in the few seconds that follow the corner kick. The black dotted line indicates the trajectory of the ball. All corners are translated so that the ball is delivered from the left corner of the pitch.

It is clear that our methodology has identified a set of distinct routines for each team. A popular strategy used by almost every team in our sample is the *jellyfish*. In this strategy, three or four players start in a cluster outside the six-yard box before making gradually diverging runs towards the box (see rows 3 and 4 in Fig. 6). Closer inspection reveals that teams employ different versions of this strategy, varying the position of the initial cluster and the length of run made by each player. One example is the *love train* – in which players start in a line, rather than a cluster – as popularised by the England team at the 2018 World Cup. Team B regularly employed the *love train*, with the line starting near the penalty spot (third row). They also used a variation of the *jellyfish* in which the players started in the far corner of the penalty area (fourth row). Team A employed a variation in which a player makes a run around and behind the initial cluster, aiming for the far post (fourth row) and an unusual routine in which four players start at the far edge of the six-yard box before running horizontally towards the ball (second row).

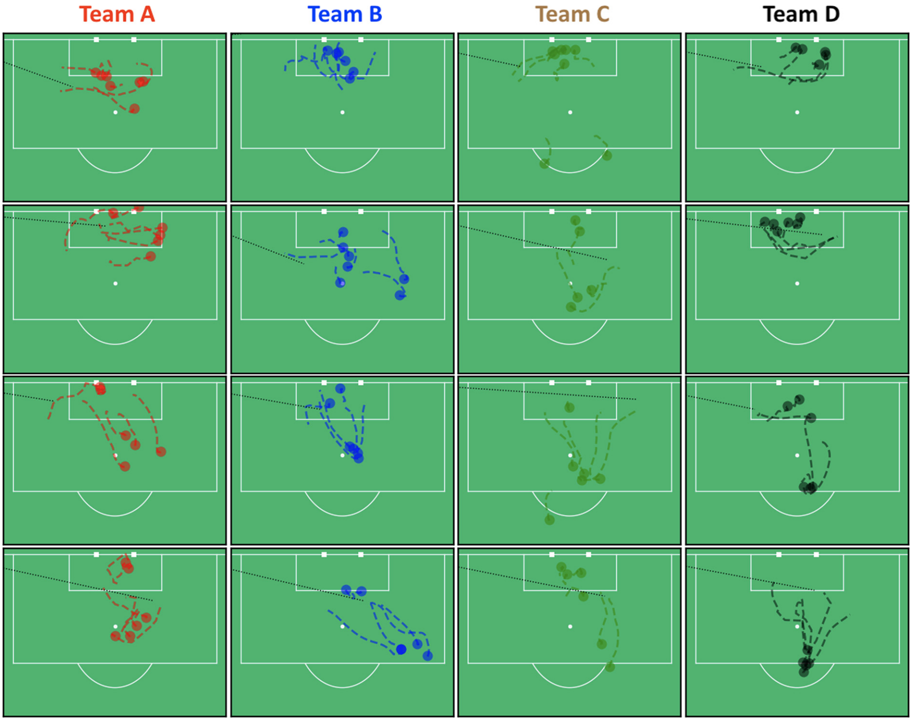

Fig. 6. Examples of popular corner routines employed by four teams in our dataset. (Color figure online)

Another class of routines is the six-yard box *overload*, in which four or five attacking players initially crowd the six-yard box. Figure 6 shows some distinct variations: the second row of Team D shows a routine in which two players positioned in front of the near post circle out of the six-yard box and round towards the far post to intercept a deep delivery. The first row shows the reverse of this: two players at the far post run around the box to intercept a near post delivery. Team B also regularly employed variants of the *overload* (e.g. first row).

The teams in our data did not alternate randomly from one routine to another as they took corners. Rather, they would typically use a routine regularly over a series of consecutive matches and then discard it, perhaps reintroducing the routine later in the season. For example, Team C attempted the corner depicted in the fourth row five times over three consecutive games, discarded it for six games and then used it three times in one game. This emphasizes the need to scan over a large number of matches to fully scout the range of offensive strategies that might be employed by an opponent in their next match.

5.2 Defence

The objective of offensive corner routines is to disrupt the defensive structure of their opponents and create sufficient space for the target player to have a clear shot on goal. The fundamental question for the attacking team is therefore: *what is our opponent's strategy for defending corner kicks?* Fig. 7 shows the total number of players allocated to zonal or man-marking roles for four teams in our sample during their first 30 matches of the season (the teams depicted are not the same set of teams discussed in the previous section). Red indicates the number of players in zonal marking roles, blue indicates the number of players in man-marking roles. The large dots connected by lines show the median number of players in either role over all the corners defended in each match; the small pale dots show the numbers in either role for the individual corners in each match.

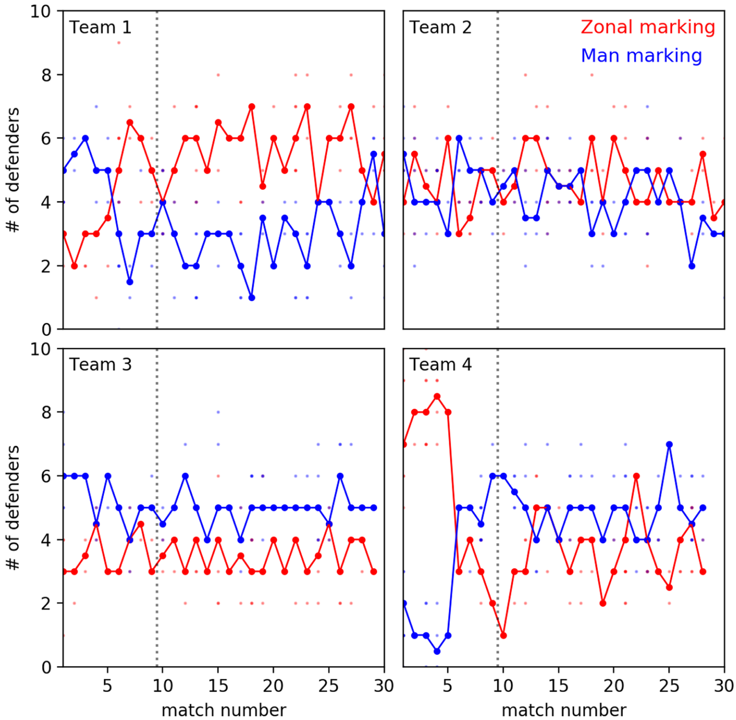

Fig. 7. The number of man-marking and zonally marking defenders over the first 30 matches for four teams in our dataset. Red (blue) lines indicate the median number of zonal (man-marking) defenders over the corners in each match. The small points indicate the numbers for the individual corners. Matches to the left of the dashed lines were included in the analyst-annotated sample. (Color figure online)

Team 1 defended with a predominantly zonal system, supplemented by two or three man-marking players. Team 2 adopted an almost perfectly hybrid system with similar numbers of man-marking and zonal defenders, while Team 3 consistently defended with more players in man-marking roles. Team 4 changed their defensive system during the season. In their first five matches they used a predominantly zonal system, with eight outfield players assigned to zones and typically just one player man-marking. In game six they made a wholesale change to their defensive strategy, significantly increasing the number of man-marking defenders and halving the number of players assigned to zonal roles. The motivations behind the change seem evident: in the first five matches they conceded 2 goals from 32 corners, including 1 from 2 during game five. Following the change, the rate at which they conceded goals from corners halved to 4 goals in 120 corners.

Having identified the players allocated to a zonal role, we can investigate where the zones were located. Figure 8 presents heat-maps for the spatial distribution of the zonal defenders for Teams 1–4 over the 30 matches. A darker shade of blue indicates a higher occupation rate for a given location.

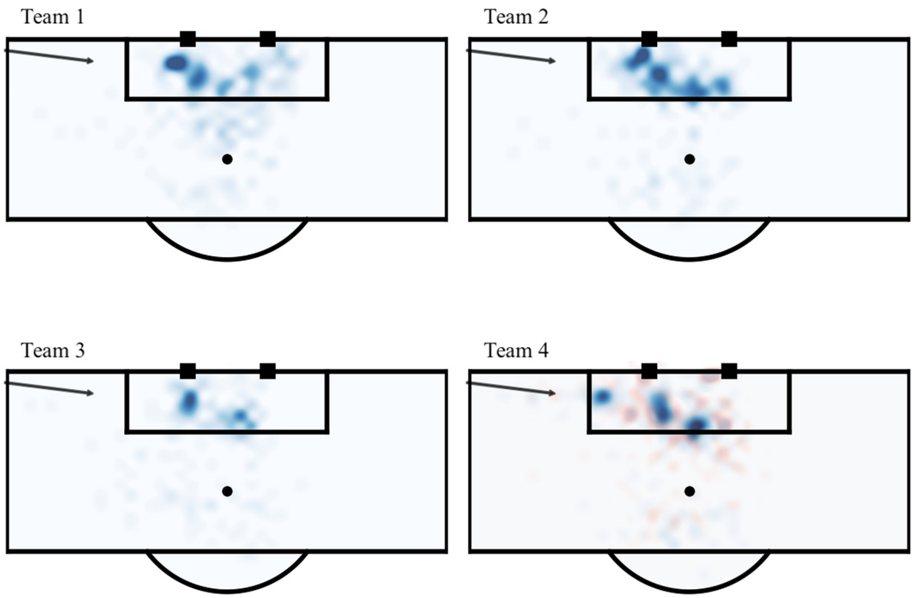

Fig. 8. The spatial distribution of zonal marking defenders for Teams 1–4 in Fig. 7 over their 30 matches. The red shading for Team 4 indicates the distribution of zonal defenders over their first 5 matches, the blue shading for their remaining 25 matches. The arrow indicates the direction from which the corner is taken. (Color figure online)

Team 1 used a predominantly zonal system and this is reflected in spatial distribution of their zonal players. There are four clearly distinct peaks in the distribution forming a circle around the goalmouth, the most prominent being in front of the near post. There is a fainter fifth peak next to the far post, suggesting a zonal defender was sometimes positioned there, and a cloud between the six-yard box and the penalty spot where the remaining zonal players were positioned (although not necessarily in the same place in every corner). A similar pattern is observed for Team 2, albeit with no defender at the far post and fewer zonal players outside the 6-yard box. Team 3 is a predominantly man-marking team, and there were only two areas routinely defended by a zonal player: next to the front post and at the edge of the six-yard box.

Team 4 is an interesting case. The red regions indicate the distribution of zonal players in their first five games, during which they employed a predominantly zonal marking system. There is insufficient data to identify exactly where each player was stationed, but they were clearly distributed broadly throughout the 6-yard box (and players guarding either post are also visible). After the tactical change in game 6, Team 4 typically retained at least three zonal players, with one player being positioned significantly in advance of the near post.

6 Conclusion

Using a season's worth of player tracking and event data, we have conducted an in-depth analysis of the offensive and defensive strategies employed by teams in corner kick situations. By studying and classifying the runs made by the attacking team, we have created a system for describing corners that enables us to characterise a corner routine in terms of the run combinations made by the attacking players. This allows us to search a large sample of corners to find certain characteristics (e.g. a particular run, or combinations of runs) or group similar corner routines.

We have also presented a supervised learning model for classifying the role of each defending player in corner situations. Using a sample of 500 corners manually annotated by club analysts, we trained the XGBoost algorithm to predict whether each defending player was instructed to man-mark or zonally mark, obtaining a cross-validated classification accuracy of $83.4 \pm 2.1\%$. The main predictors of player role are their initial positions, distance travelled during the corner, proximity to the goalkeeper, and the average distance travelled by nearby teammates and opponents.

We have demonstrated how these tools can be applied to provide unprecedented insights into the strategies used by teams in corner kick situations, identifying the distinct corner routines employed by four teams over the course of a season. Our methodology for identifying the defensive roles of individual players will also help analysts and coaches to find vulnerabilities in the defensive systems of their opponents and exploit them.

A natural next question to ask is: *which attacking routines are most effective against a certain defensive set-up?* We have refrained from providing an empirical answer to that question in this paper because of the limited size and scope of our data set. In a follow-up paper we will make use of a significantly larger sample of tracking data to empirically investigate the most effective strategies for boosting the quality and quantity of chances created in corner kick situations.

Acknowledgments. We acknowledge Devin Pleuler at Toronto FC for his advice and insights, and Tiago Maia and Jan Schimpchen from SL Benfica for helping to produce the training data for our defensive role classification model.

References

1. Power, P., Hobbs, J., Ruiz, H., Wei, X., Lucey, P.: Mythbusting set-pieces in soccer. In: MIT Sloan Sports Analytics Conference (2018)
2. Beare, H., Stone, J.A.: Analysis of attacking corner kick strategies in the fa women's super league 2017/2018. Int. J. Perform. Anal. Sport **19**(6), 893–903 (2019)
3. Casal, C.A., Maneiro, R., Ardá, T., Losada, J.L., Rial, A.: Analysis of corner kick success in elite football. Int. J. Perform. Anal. Sport **15**(2), 430–451 (2015)
4. Pulling, C.: Long corner kicks in the English premier league: deliveries into the goal area and critical area. Kinesiol. Int. J. Fundam. Appl. Kinesiol. **47**(2), 193–201 (2015)
5. Pulling, C., Newton, J.: Defending corner kicks in the English premier league: near-post guard systems. Int. J. Perform. Anal. Sport **17**(3), 283–292 (2017)
6. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc.: Ser. B (Methodol.) **39**(1), 1–22 (1977)
7. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. SSS. Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7
8. Kuang, D., Choo, J., Park, H.: Nonnegative matrix factorization for interactive topic modeling and document clustering. In: Celebi, M.E. (ed.) Partitional Clustering Algorithms, pp. 215–243. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-09259-1_7
9. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing Systems, pp. 556–562 (2001)
10. Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
11. Sagi, O., Rokach, L.: Ensemble learning: a survey. Wiley Interdisc. Rev. Data Mining Knowl. Discov. **8**(4), e1249 (2018)
12. Carling, C., Williams, A.M., Reilly, T.: Handbook of Soccer Matchanalysis: A Systematic Approach to Improving Performance. Psychology Press, London (2005)
13. Sarmento, H., Marcelino, R., Anguera, M.T., CampaniÇo, J., Matos, N., LeitÃo, J.C.: Match analysis in football: a systematic review. J. Sports Sci. **32**(20), 1831–1843 (2014)