Modeling Defensive Dynamics in Football: A Hidden Markov Model-Based Approach for Man-Marking and Zonal Defending Corner Analysis

Sean Groom[1,2] , **Dan Morris**[3] , **Liam Anderson**[1] and **Shuo Wang**[1]
[1]The University of Birmingham
[2]Nottingham Forest FC
[3]Work completed while at Nottingham Forest FC
s.wang.2@bham.ac.uk

Abstract

Elite football clubs increasingly rely on data science to enhance decision-making in player analysis, recruitment, and tactical understanding. Despite this, there is a significant bias toward offensive play in existing research. Quantifying defensive performance, particularly in open play of invasive team sports, poses substantial challenges due to its complexity. However, corner kicks present a more controlled environment that facilitates detailed defensive analysis because they involve preplanned strategies implemented by coaches. This paper outlines a Hidden Markov Model (HMM) framework tailored for analyzing corner kick scenarios. The model differentiates between zonal and man-marking defenders while estimating specific man-marking assignments and zone locations. Unlike existing supervised methods for analyzing defensive behaviour during corner kicks, this unsupervised approach eliminates the need for a labelled marking assignment dataset, making it less cumbersome to implement. It also provides coaches with more actionable insights by spatially detailing teams' man-marking and zonal behaviours across the pitch, which is useful for tactical planning.

1 Introduction

Machine learning has become increasingly important in sports analytics, enhancing team tactics and player recruitment across various sports, including football [Gudmundsson and Horton, 2017]. In football, data providers collect extensive datasets, such as event data that includes labelled actions (e.g., shots, passes) with their spatial pitch coordinates, and tracking data from stadium-based multi-camera systems detailing player positions and velocities [Memmert and Rein, 2018]. These datasets support the use of AI and machine learning to provide coaches with sophisticated tools for detailed performance analysis and tactical planning, however developing such tools to evaluate the defensive performance of outfield players remains challenging [Forcher *et al.*, 2022]. Offensive play in soccer is often evaluated using clear metrics like goals and assists, but defensive metrics are less straightforward. Measures such as tackles and interceptions don't fully capture a defender's effectiveness, as a skilled defender can significantly impact a game by subtly influencing attackers into taking less threatening actions. Defending players aim to minimize the likelihood of conceding opportunities while maximizing the chances of regaining possession. The positioning of players off the ball, both individually and as a team, is a key part of defense. However, disentangling the influence of each player and accurately measuring individual defensive performance poses significant challenges. To this end, set pieces such as corner kicks offer a more regular environment within which models can be more easily developed as they offer teams the opportunity to execute pre-planned movements. Corner kicks occur on average 10 times [Shaw and Gopaladesikan, 2020] in each game, with teams scoring from 2.1% of all corners [Power *et al.*, 2018], whilst this is relatively few compared to set plays in other sports such as American football, it is still sufficiently important for numerous clubs to employ dedicated set-piece coaches. Coaches may instruct players to either man-mark—following an opponent closely to reduce their probability of successfully interacting with the ball—or to zonally mark, which involves guarding a designated static area of the pitch. Typically, teams use a hybrid defense combining both strategies, with the distribution between man-marking and zonal marking reflecting individual coaching preferences.

The aim of this paper is to extend the HMM framework first used in basketball analytics [Franks *et al.*, 2015] [Keshri *et al.*, 2019] for use in football corner kick analysis, this is accomplished by incorporating a new style of defending, zonal-marking, into the model. This extension allows the HMM to be used in football, which is a more challenging setting due to the larger pitch size and increased number of players with more unique roles. The paper is organized as follows: related work is described in Section 2, the methodology is explored in Section 3, model results are explored in Section 4, the direction of future research is proposed in Section 5 and conclusions are finalised in Section 6.

2 Related Work

Corner kicks have been extensively studied in both sports science [Pulling and Newton, 2017] [De Baranda and Lopez-Riquelme, 2012] and, increasingly, through the lens of machine learning. Power et al. [Power *et al.*, 2018] first used neural networks to classify defensive strategies employed during corners, identifying whether teams used man-marking, zonal, or hybrid defenses. They found teams conceded most goals from corners as a result of the second interaction with the ball after the corner is taken rather than directly from the corner, as well as that teams who placed men guarding both the corner posts conceded the most chances. Shaw et al. [Shaw and Gopaladesikan, 2020] clustered attacking runs based on their start and end locations, alongside creating a labelled dataset to train a gradient boosted model to detect whether a player was zonally marking or not using hand crafted features derived from tracking data. These approaches both cannot determine who a defender is marking, so to bridge this gap Bauer et al. [Bauer *et al.*, 2022] developed an approach that uses a convolutional neural network with a long short-term memory neural network to assign defenders various roles, including zonal defending and man-marking roles, alongside others such as positioning for potential counter attacks. This approach improved upon the accuracy of Shaw et al. when predicting which players are zonally defending whilst additionally providing information about who is marking who at each time step.

Hidden Markov models have been effectively utilized in basketball analytics to generate man-marking assignments during open play, however they do not account for zonal defending. The HMM used by Franks et al. [Franks *et al.*, 2015] treats each defender j independently, modelling their position with an emission distribution that at time t, given they are marking attacker k, is normally distributed with mean

$$\mu_{tk} = \gamma_o O_{tk} + \gamma_b B_t + \gamma_h H, \tag{1}$$

where O_{tk} is the attackers position and B_t and H is the position of the ball and hoop respectively. The weights $\Gamma = [\gamma_o, \gamma_b, \gamma_h]$ are constrained such that $\Gamma \mathbf{1} = 1$. In this model they make the simplifying assumption that the probability of a defender transitioning to mark any other attacker at each time step is homogeneous. Keshri et al. [Keshri *et al.*, 2019] advanced this model by introducing an energy-based bond breaking model for transitions between marking assignments. This enhancement produces more realistic transition probabilities by considering the man-marking choices of other teammates. Alongside this, Keshri et al. also allow the emission distribution that describes the defender's positioning to vary its parameters across the court depending upon where the attacker being marked is located. They fit individual emission distributions for each attacking player, which significantly improves the accuracy of modelling the defender's position. This is particularly evident in the case of players known for their long-range shooting, such as Stephen Curry, who is more closely guarded across the court than others.

Wang et al. [Wang *et al.*, 2024], in collaboration with Liverpool Football Club, developed a set of models named TacticAI using graph neural networks to predict receivers and shots, as well as to generate suggestions on how to improve the outcomes of corner kicks. By representing player tracking data as fully connected graphs, the team utilized clustering techniques within TacticAI's latent space to identify similar corner kick scenarios. This method and Shaw et al.'s approach both provide the ability to analyse similar historical corner situations. However, both approaches share a common limitation: they do not incorporate man-marking assignments, making it challenging to assess the defensive effectiveness of individual players during these sequences.

The objective of this research is to adapt a HMM framework for analyzing football corner kicks, extending existing models to incorporate zonal defending while still generating man-marking assignments. This approach eliminates the need for generating a labelled dataset, a significant hurdle in the deployment of Bauer et al.'s model, providing coaches and analysts with an automated method for analyzing corner kick situations, enhancing tactical decision-making and strategic planning.

3 Methodology

In this study, all results shown are obtained by utilising the entire 2022-23 English Premier League season's tracking data. Corner kick sequences are extracted from tracking data, which are aligned with corresponding event data. Short corner deliveries, which involve a quick pass to a teammate close to the corner rather than a cross into the penalty area, are then removed using the event data. To simplify the analysis, goalkeepers are not considered in man-marking assignments, and sequences where a player has been sent off are excluded to maintain consistency in team zone locations. To standardize the data, all sequences are adjusted—translated and reflected—so that they consistently appear as if taken from the top-right corner of the pitch. The tracking data used begins one second prior to the execution of the corner kick. Each sequence is then trimmed, either ending two seconds after the corner is taken or at the occurrence of the second event following the corner kick, whichever comes first. A total of 3145 corner sequences are collected, consisting of 1818 inswinging deliveries and 1327 outswinging. Per team, the average number of inswinging corner deliveries collected was 90.9 ± 20.7 whilst the average number of outswinging deliveries collected was 66.4 ± 17.8

A HMM [Bishop, 2006] is utilized to infer defensive assignments, specifically to identify which defender is marking which attacker, and who is adopting a zonal marking strategy. At each time step t, a defender j, numbered from 1 to 10, may either mark one of the ten outfield attackers k, or be in a zonal marking state. Our HMM employs N hidden states, where $N = K + 1$, with K representing the number of outfield attackers, to accommodate all man-marking states plus one additional state for zonal defense. The set of hidden states $Q = \{q_1, q_2, \ldots, q_N\}$ is such that a defender in state q_k is man-marking attacker k, and state q_N indicates zonal marking. For each defending team, zones $z \in \{1, \ldots, 10\}$ are estimated. The zones are then assigned to each of the defender's possible hidden states in such a way that only one defender in an individual corner sequence can occupy each zone that is estimated for their team. This reflects the instructions a coach would give to a defender, assigning each zonal defender to a specific area.

In this model, the task of assigning specific zones on the pitch to each defender is handled using the Kuhn-Munkres algorithm [Kuhn, 1955], [Munkres, 1957], also known as the

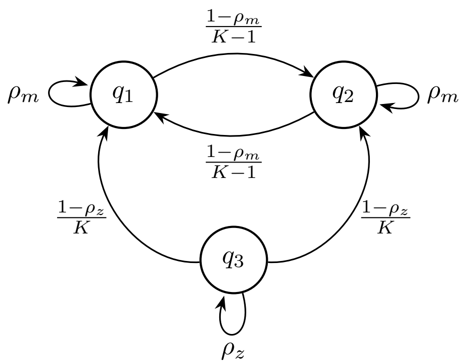


Figure 1: A simplified Markov chain diagram with three states illustrating transition probabilities. This represents the Markov chain for an individual defender in the case where there are only two outfield players ($K = 2$) that the defender can mark. States q_1 and q_2 are man-marking states and q_3 is a zonal state. ρ_m represents the probability of man-marking the same opponent in the next frame of tracking data, while ρ_z is the probability of continuing to zonally mark their assigned zone.

Hungarian algorithm. This algorithm is utilized to efficiently solve assignment problems where the goal is to minimize the total cost of assigning tasks to agents. In the context of our analysis, the "tasks" are zones on the pitch that defenders are responsible for during a corner kick. Each zone, z, is defined by a bivariate Gaussian distribution. Given a set of zonal distributions, the distance between each defender's initial location at the start of the corner sequence and the zone's means are calculated. The Kuhn-Munkres algorithm is then applied to determine the optimal assignment of defenders to these zones such that the total distance between the zone means and their assigned defender's starting positions is minimised.

Figure 1 illustrates a simplified example of the transition model used in our analysis, specifically for a scenario with only two outfield players ($K = 2$). This model is adapted from the approach used by Franks et al. and includes similar assumptions about the probabilities of marking transitions between frames of tracking data. In this model, each man-marking defender can either continue to man-mark the same attacker or switch to mark the other attacker. The probability of continuing to man-mark the same attacker in the next frame is constant for all players, denoted by ρ_m. Conversely, the probability of switching to mark any other attacker is equally likely across the other players. The probability that a player continues to mark a specific zone is constant and represented by ρ_z. In this model, a player cannot transfer from one zonal state to another but can transfer from their assigned zonal state to any other man-marking state with probability $1 - \rho_z/K$. As a result of this transition model the number of man-marking defenders tends to increase towards the end of a sequence as players transition out of their zonal states and the defensive structure of the team becomes more disorganised after the corner is taken.

Regarding the emission distribution, a similar approach is taken to that of Keshri et al. however, due to the relatively limited amount of tracking data captured during corner kicks when compared to open play, it is not practical to estimate unique emission distribution parameters for each outfield player. Instead, in this paper we estimate two unique emission distributions for each defending team, one for inswinging deliveries and another for outswinging deliveries, and then following a similar approach to Keshri et al. the pitch is separated into a grid and model parameters are estimated for each bin.

Adapting this further for use in football, separate emission distributions are used depending on whether a defender is occupying a man marking or zonal hidden state. In Frank's and Keshri's models, the defender j's position at time t, given they are marking attacker k, is normally distributed with mean given in Equation 1. However, for this corner model we will only use

$$\mu_{tk} = \gamma_o O_{tk} + \gamma_g G, \tag{2}$$

where G is the centre of the defending goal, because the quality of the ball location tracking data during corners is poor. For man marking states the pitch is separated into 3m by 3m bins, labelled l, and unique parameters $\Gamma^{(l)} = [\gamma_o, \gamma_g]_l$ and $\sigma^{2(l)}$ are estimated for each bin while constrained such that $\Gamma^{(l)}\mathbf{1} = 1$. This constraint indicates that the parameter γ_o quantifies the average proximity of a defender to the attacker they are marking, measured along the line from that attacker to the centre of the defender's goal. While Premier League pitches differ in size, the dimensions of the penalty areas are standard. Bins are therefore defined with a fixed size in order for bins surrounding the penalty box to be consistent. The total number of bins is calculated by dividing the league maximum pitch length and width by the length and width of a bin, respectively, and then rounding up to the nearest whole number. These quotients are then multiplied together to give us the total bin count for the entire pitch. Given a defender is marking the attacker k, who's position O_{tk} is located within bin l, the defender j's position is modelled as a normal distribution with mean $\mu_{tk}^{(l)} = Z_{tk}\Gamma^{(l)}$, and covariance $\sigma^{2(l)}$, where $Z_{tk} = [O_{tk}, G]^\top$. For each zone, z, we use a bivariate Gaussian emission distribution with mean μ_z and covariance σ_z^2.

3.1 Model Training

Model parameters, $\lambda = \{\Gamma, \sigma^2, \mu_z, \sigma_z^2, \rho_m, \rho_z\}$, are estimated using an expectation maximisation approach as outlined in algorithm 1. We denote the set of P observation sequences as $\mathbf{O} = [\mathbf{O^{(1)}}, ..., \mathbf{O^{(P)}}]$ where $\mathbf{O}^{(p)} = [\mathbf{O}_1^{(p)}, ..., \mathbf{O}_{T_p}^{(p)}]$. Here, $\mathbf{O}^{(p)}$ represents the sequence of tracking data for an individual corner kick sequence p, which consists of T_p frames. First, K zone locations are randomly initialized by sampling from a uniform distribution along the six-yard line in the y-direction and extending five meters on either side of the six-yard line in the x-direction. Each zone is initialised with a covariance $\sigma_z^2 = 2I$, where I represents the identity matrix.. The parameters $\Gamma^{(l)}$ are initialized as $[0.8, 0.2]$ for each l. Both ρ_m and ρ_z are set to 0.95 initially. All initial state probabilities are uniformly set to $1/K$, ensuring that each state has an equal probability at the start.

Treating each defender independently, expected state occupancy and transition counts are calculated using the Forward-Backward algorithm [Rabiner, 1989]. The observation likelihood of an individual defender in the sequence p, $L_{pj} = P(\mathbf{O}^{(p)}|\lambda, j)$, is calculated using the forward-backward algorithm. The complete observation log likelihood is therefore

$$\sum_{p,j,n} \log \alpha_{j,T_p}(n),$$

where $\alpha_{j,T_p}(n)$ is the forward probability at time T_p for defender j whilst they occupy state n in corner sequence p. For each team K zonal locations are estimated, each zone z, is described by a normal distribution with a mean that is estimated via

$$\hat{\mu_z} = \frac{\sum_{p=0}^{P} \frac{1}{L_{pj}} D_{0j} E_{0jN}^{(p)}}{\sum_{p=0}^{P} \frac{1}{L_{pj}} E_{0jN}^{(p)}},$$

where D_{0j} is the position of defender j who is assigned to the zone at time $t = 0$. $E_{0jN}^{(p)}$ is the expected state occupancy of defender j occupying their assigned zonal state at time $t = 0$ in corner sequence p. The zonal distribution has covariance which is estimated using

$$\hat{\sigma_z^2} = \frac{\sum_{p=0}^{P} \frac{1}{L_{pj}} \sum_{t=0}^{T_p} E_{tjN}^{(p)} (D_{tj} - \hat{\mu_z})(D_{tj} - \hat{\mu_z})^\top}{\sum_{p=0}^{P} \frac{1}{L_{pj}} \sum_{t=0}^{T_p} E_{tjN}^{(p)}}.$$

Transition probabilities ρ_m and ρ_z are each estimated using

$$\hat{\rho_m} = \frac{\sum_{p,j} \sum_{n=0}^{n=K} \sum_{t=0}^{T_p-1} \xi_{tjnn}^{(p)}}{\sum_{p,j} \sum_{n=0}^{n=K} \sum_{t=0}^{T_p-1} E_{tjn}^{(p)}}$$

and

$$\hat{\rho_z} = \frac{\sum_{p} \sum_{t=0}^{T_p-1} \xi_{tjNN}^{(p)}}{\sum_{p} \sum_{t=0}^{T_p-1} E_{tjN}^{(p)}}$$

respectively, where ξ_{tjnn} is the expected state transition count, the probability of defender j being in state n at time t, and state n, at time $t+1$. Initial state probabilities for each defender are updated using the expected state occupancy values $E_{0jn}^{(p)}$.

To estimate $\Gamma^{(l)}$ and $\sigma^{2(l)}$ for each bin l, we define the design matrix $\mathbf{X}^{(l)}$ as follows:

$$\mathbf{X}^{(l)} = \bigoplus_{\substack{k \in \text{Attackers} \\ \text{dist}(k,l) \leq N_H}} [O_{tk}, G],$$

where $\bigoplus$ denotes the concatenation operation, O_{tk} represents the position of attacker k that the defender is marking at time t, and G is the goal position. The matrix includes entries for attacker k who is positioned in bin l or in any neighbouring bin within N_H hops of bin l. In this analysis, we set the parameter N_H equal to one so only immediate neighbours are included. Any player whose position falls outside the defined bins is included in the nearest possible bin. With our modified design matrix, in the M-step we update $\Gamma^{(l)}$ and $\sigma^{2(l)}$ as outlined by Franks et al.,

Algorithm 1 Algorithm for HMM Training

Initialisation:
Initialise zonal distributions
Initialise all bin's $\Gamma^{(l)}$
Initialise transition probabilities ρ_m and ρ_z
Initialise initial state distribution
while not converged **do**
  Expectation Step (E-step):
  for each corner sequence O_p in $O = [O_1, \ldots, O_P]$ **do**
    Assign defending players to zones using the Hungarian algorithm.
    for each outfield defending player **do**
      Compute forward and backward probabilities using current HMM parameters
      Compute sequence observation likelihood L_{pj}
      Compute expected state occupancy and expected state transition counts
      Compute hidden state sequence using the Viterbi algorithm
    end for
  end for
  Maximization Step (M-step):
  Estimate new zone positions and covariances
  Update transition probabilities
  Update initial state distribution
  for each man-marking bin l **do**
    Estimate $\Gamma^{(l)}$ and $\sigma^{2(l)}$
  end for
  Check for convergence criteria
end while

$$(\hat{\Gamma}^{(l)}, \hat{\sigma}^{2(l)}) \leftarrow \underset{\Gamma^{(l)}, \sigma^{2(l)}}{\arg\max} \sum_{t,j,k} \frac{E_{tjk}}{\sigma^{2(l)}} (D_{tj} - \Gamma^{(l)} X_{tk}^{(l)})^2,$$

where E_{tjk} is the expected state occupancy for defender j occupying man-marking hidden state k at time t.

4 Results

All results presented are obtained by training five separate HMMs for each team and delivery type and then selecting the model which yields the best observation likelihood after convergence. Figure 2 presents the league average values of γ_o, calculated solely from inswinging corner deliveries. This calculation includes each bin l where the average count of entries in $\mathbf{X}^{(l)}$ across all teams exceeds 1,000. The figure illustrates the man-marking behaviour typical of an average Premier League team. As expected, players are marked more closely near the goal, with a slight bias toward closer marking at the near post compared to the far post. Additionally, the figure highlights that defenders tend to mark potential short corner recipients more tightly than other attackers positioned outside the box. It also reflects the rule that defenders must maintain a distance of at least 10 yards from the corner spot until the ball is kicked.

Figure 3 displays the inswinging γ_o parameters for Manchester United when compared to the league average, chosen for their consistent defensive system under a single man-

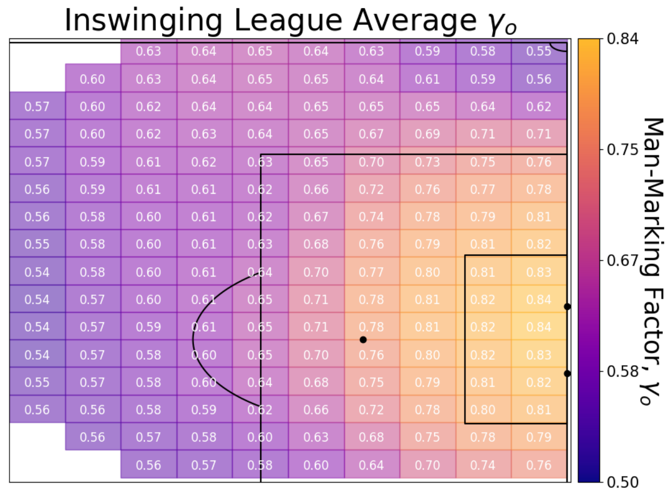


Figure 2: This figure displays the league average values of γ_o, which quantifies the average proximity of a defender to an attacker, for each bin. The bins included are those where the average count of entries in the design matrix $\mathbf{X}^{(l)}$ for each team exceeds 1,000. Solid black lines in the figure denote the pitch markings, including the touchlines and the penalty box. Solid black dots are used to indicate the penalty spot and the goal posts.

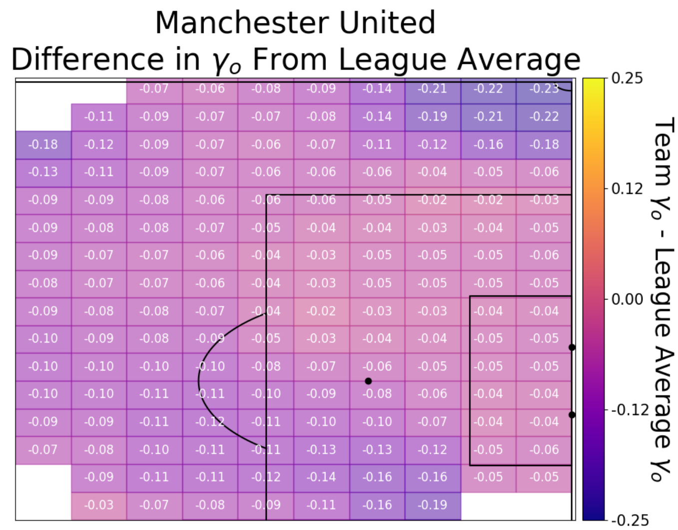


Figure 3: This figure displays the difference in γ_o parameters for Manchester United when compared to the league average for each bin when defending inswinging corners. The parameter γ_o represents how closely defenders man-mark attackers located in each bin. This figure includes each bin where the count of entries in the design matrix $\mathbf{X}^{(l)}$ exceeds 1,000. Solid black lines in the figure denote the pitch markings, including the touchlines and the penalty box. Solid black dots are used to indicate the penalty spot and the goal posts. Estimated using 105 inswinging corner sequences.

ager throughout the 2022-23 season. The figure shows that Manchester United's man-marking was generally less tight compared to the league average, notably when defending against potential short corners and deep players who are located on the far side of the box. In comparison, Figure 4 shows the γ_o parameters for Manchester City as they defend inswinging corners, when compared to the league average it can be seen that Manchester City typically man-mark more closely across the pitch, except for attackers far from the corner spot who are near the edge of the box. Figure 5 shows an example frame of tracking data with man-marking assignments displayed. The particular frame is the first in a sequence of Manchester United defending an inswinging corner delivery. If a defender is not assigned a man then they are zonally defending. Figure 6 illustrates the estimated inswinging and outswinging defensive zones for Manchester United. Each zonal state is represented by a cross marking the mean position and an ellipse that encloses a 66% confidence region around the mean. The colour of each zonal state indicates the average duration, as a fraction of corner sequence length that a defender remains in that zonal state before transitioning to a man-marking state. This duration reflects the extent to which each zonal state corresponds to a true defensive zone for the team. Longer durations in a zonal state suggest a true zone location, indicating that defenders predominantly maintain their positions. Conversely, shorter durations typically involve defenders who are initially in zonal states but soon switch to man-marking as the play unfolds. The variations between the two figures highlight the team's strategic adjustments in response to different types of corner deliveries. Figure 7 shows the estimated defensive zones for Manchester City. Given Manchester City's possession based style of play it is not surprising to note that they faced both the fewest inswinging and outswing corners of all teams in the Premier League with 50 and 32 corners respectively. Despite this, the zones estimated for Manchester City seem reasonable to the eye.

5 Discussion and Future Work

Unlike supervised methods such as Bauer et al., which may generalize effectively to new teams, the adaptability of this unsupervised method to unfamiliar team behaviours without sufficient tracking data remains uncertain. Whilst HMMs typically need less training data when compared to deep learning approaches, this limitation may make it particularly challenging to analyze opponents in non-league competitions or newly promoted teams at the start of a season when the amount of tracking data available is limited. To better understand these constraints, a sensitivity analysis is recommended. This analysis would examine the variance in model parameters when trained on limited data, providing insights into the model's robustness.

As an unsupervised method, this approach lacks a gold-standard set of labels for validating the generated man-marking assignments, making it difficult to measure the accuracy of these assignments quantitatively. The effectiveness of model improvements can thus be hard to verify, with visual inspections only offering limited assurance. Manually recording some man-marking assignments could allow for quantitative assessments of different modelling choices, such as the bond-breaking transition model used by Keshri et al., and their impact on performance. Despite these limitations, this model's insights have been found to be beneficial by do-

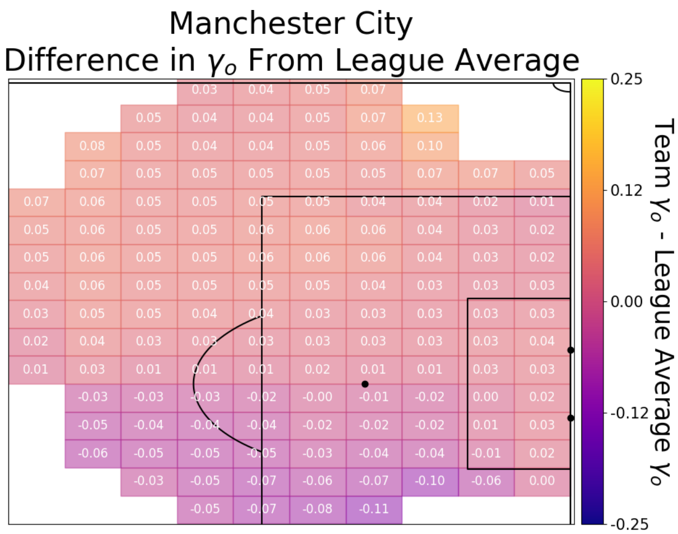


Figure 4: This figure displays the difference in γ_o parameters for Manchester City when compared to the league average for each bin when defending inswinging corners. The parameter γ_o represents how closely defenders man-mark attackers located in each bin. This figure includes each bin where the count of entries in the design matrix $\mathbf{X}^{(l)}$ exceeds 1,000. Solid black lines in the figure denote the pitch markings, including the touchlines and the penalty box. Solid black dots are used to indicate the penalty spot and the goal posts. Estimated using 50 inswinging corner sequences.

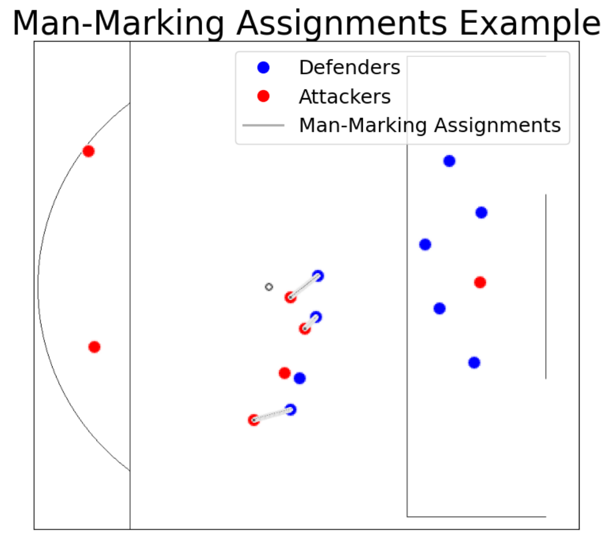


Figure 5: This figure illustrates man-marking assignments for Manchester United, generated by the HMM, as they defend against an inswinging corner kick. This visualization shows the initial frame at the beginning of a corner sequence, excluding goalkeepers. Each defender without a man-marking assignment is zonally marking. As the sequence unfolds and attackers start to make runs, an increasing amount of defenders transition out of their zonal state and begin to man-mark.

main experts at Nottingham Forest FC.

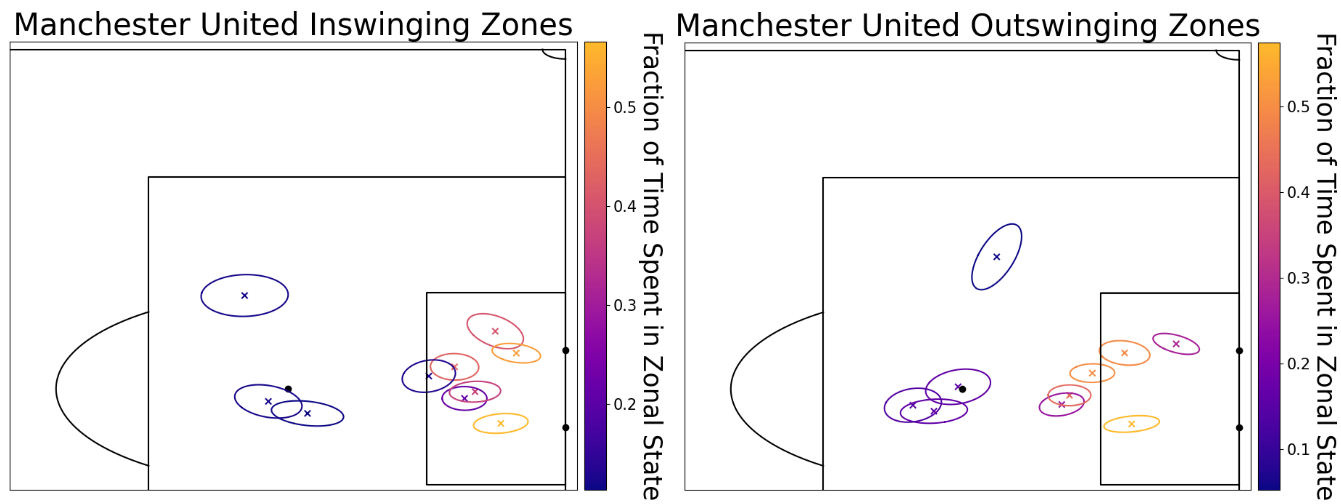
Further exploration could also consider the influence of choices like bin size, the number of neighbouring bins included in parameter estimation, and the possible transitions in a player's Markov chain. An interesting option would be to allow each zone to have a unique transition probability ρ_z. These investigations would help refine the model's precision and applicability in corner kick analysis. The accuracy of man-marking assignments can potentially be enhanced by collaborating directly with end user analysts to determine the optimal number of zones for each team. This process could be facilitated by providing visual aids, such as generating figures similar to Figures 6 and 7 or each team. Alternatively, analysts could be given the opportunity to recommend the number of zonal states based on their expert judgment, independent of the model's outputs. Then after applying the Hungarian algorithm with less than K zones, man-marking assignments could be generated using the Viterbi algorithm, only allowing the players assigned to the reduced set of zones to enter the zonal states throughout an individual corner sequence.

Previous studies have demonstrated the effectiveness of Graph Neural Networks (GNNs) in evaluating defensive performance during both open play [Stöckl *et al.*, 2021] and set pieces [Wang *et al.*, 2024] by estimating reception and shot probabilities for each player. However, the lack of man-marking assignments has made it difficult to accurately credit defenders for their actions. These GNN analyses have also employed fully connected graphs, with the HMM outputs described in this paper, it is now feasible to explore a variety of graph representations using man-marking assignments or expected state occupancy values to determine whether a pair of nodes should be connected with an edge.

Future research will focus on examining how different graph representations affect model performance in reception and shot prediction tasks. Additionally, we plan to assess the impact of incorporating HMM outputs as node and edge features to enhance model accuracy. Utilizing GNN models in conjunction with the detailed man-marking and zonal information provided by our HMM could reveal new insights into the effectiveness of various defensive strategies during corner kicks. Looking ahead, combining HMM and GNN approaches could establish a new framework for quantifying outfield defensive performance in a novel way.

6 Conclusion

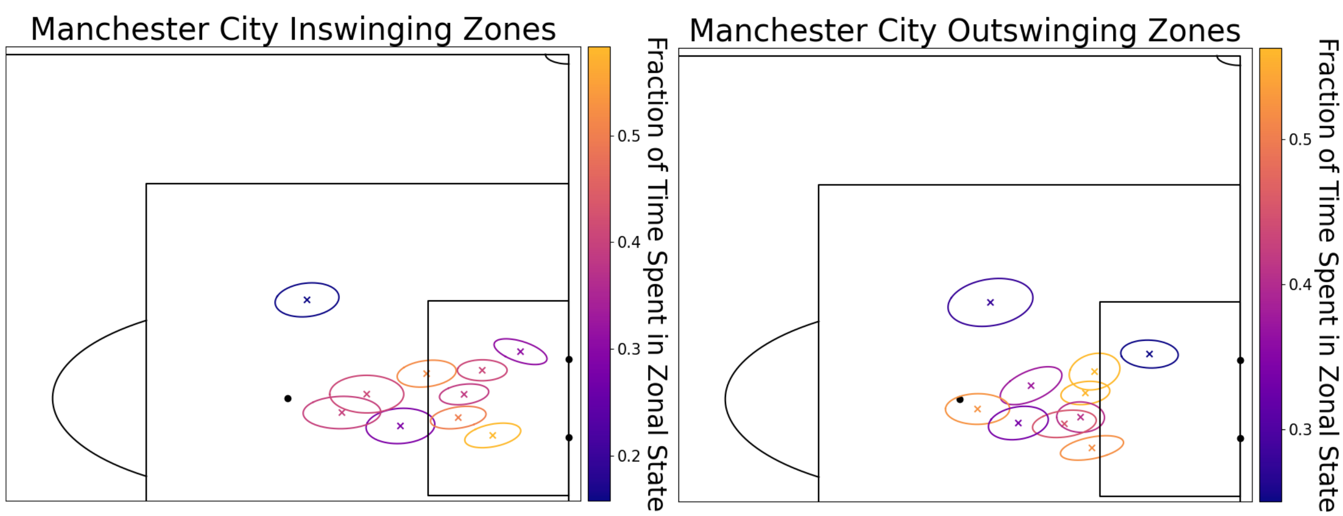
This study introduces a HMM framework for analyzing defensive strategies in football, which extends an unsupervised method previously utilized in basketball analytics to provide man-marking assignments. By applying this technique, it is possible to quantify differences in marking behaviour across the pitch and to identify zonal defending. This methodology offers potential benefits for coaching staff in terms of automated reporting and tactical planning, facilitating a more nuanced understanding of defensive dynamics. Furthermore, this approach may serve as a foundation for developing more precise metrics of defensive performance, contributing to the broader field of sports analytics.



(a) Estimated zonal states for Manchester United defending inswinging corners. Estimated using 105 inswinging corner sequences.

(b) Estimated zonal states for Manchester United defending outswinging corners. Estimated using 66 outswinging corner sequences.

Figure 6: This figure illustrates the estimated zonal states for Manchester United when defending against inswinging and outswinging corner deliveries. Each sub-figure presents 10 zonal emission distributions, marked by a cross indicating the mean position and encircled by a 66% confidence interval. The colour of each zone denotes the average fraction of the corner sequence duration that a defender spends in that zonal state before transitioning to a man-marking state. Solid black lines in the figures denote the pitch markings, including the touchlines and the penalty box. Solid black dots are used to indicate the penalty spot and the goal posts.



(a) Estimated zonal states for Manchester City defending inswinging corners. Estimated using 50 inswinging corner sequences.

(b) Estimated zonal states for Manchester City defending outswinging corners. Estimated using 32 outswinging corner sequences.

Figure 7: This figure illustrates the estimated zonal states for Manchester City when defending against inswinging and outswinging corner deliveries. Each sub-figure presents 10 zonal emission distributions, marked by a cross indicating the mean position and encircled by a 66% confidence interval. The colour of each zone denotes the average fraction of the corner sequence duration that a defender spends in that zonal state before transitioning to a man-marking state. Solid black lines in the figures denote the pitch markings, including the touchlines and the penalty box. Solid black dots are used to indicate the penalty spot and the goal posts.

References

[Bauer *et al.*, 2022] Pascal Bauer, Gabriel Anzer, and Joshua Wyatt Smith. Individual role classification for players defending corners in football (soccer). *Journal of Quantitative Analysis in Sports*, 18(2):147–160, 2022.

[Bishop, 2006] Christopher M. Bishop. *Pattern recognition and machine learning*. New York : Springer, [2006] ©2006, 2006. Textbook for graduates.;Includes bibliographical references (pages 711-728) and index.

[De Baranda and Lopez-Riquelme, 2012] Pilar Sainz De Baranda and David Lopez-Riquelme. Analysis of corner kicks in relation to match status in the 2006 world cup. *European Journal of Sport Science*, 12(2):121–129, 2012.

[Forcher *et al.*, 2022] Leander Forcher, Stefan Altmann, Leon Forcher, Darko Jekauc, and Matthias Kempe. 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, 2022.

[Franks *et al.*, 2015] A. Franks, A. Miller, L. Bornn, and K. Goldsberry. Characterizing the spatial structure of defensive skill in professional basketball. *Annals of Applied Statistics*, 9(1):94–121, 2015. Franks, Alexander Miller, Andrew Bornn, Luke Goldsberry, Kirk Franks, Alexander/0000-0002-9329-206X.

[Gudmundsson and Horton, 2017] J. Gudmundsson and M. Horton. Spatio-temporal analysis of team sports. *ACM Computing Surveys*, 50(2), 2017. Gudmundsson, Joachim Horton, Michael 1557-7341.

[Keshri *et al.*, 2019] Suraj Keshri, Min-hwan Oh, Sheng Zhang, and Garud Iyengar. Automatic event detection in basketball using hmm with energy based defensive assignment. *Journal of Quantitative Analysis in Sports*, 15(2):141–153, 2019.

[Kuhn, 1955] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

[Memmert and Rein, 2018] D. Memmert and R. Rein. Match analysis, big data and tactics: Current trends in elite soccer. *Deutsche Zeitschrift f¸r Sportmedizin*, 69(3):65–72, 2018.

[Munkres, 1957] James Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.

[Power *et al.*, 2018] Paul Power, Jennifer Hobbs, Héctor Ruiz, Xinyu Wei, and Patrick Lucey. Mythbusting setpieces in soccer. In *MIT Sloan Sports Analytics Conference*, 2018.

[Pulling and Newton, 2017] Craig Pulling and Jay Newton. Defending corner kicks in the english premier league: near-post guard systems. *International Journal of Performance Analysis in Sport*, 17(3):283–292, 2017.

[Rabiner, 1989] L. R. Rabiner. A tutorial on hidden markov-models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. Rabiner, lr 1558-2256.

[Shaw and Gopaladesikan, 2020] Laurie Shaw and Sudarshan Gopaladesikan. Routine inspection: A playbook for corner kicks. In *Machine Learning and Data Mining for Sports Analytics: 7th International Workshop, MLSA 2020, Co-located with ECML/PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings 7*, pages 3–16. Springer, 2020.

[Stöckl *et al.*, 2021] Michael Stöckl, Thomas Seidl, Daniel Marley, and Paul Power. *Making Offensive Play Predictable -Using a Graph Convolutional Network to Understand Defensive Performance in Soccer*. MIT Sloan Sports Analytics Conference, 2021.

[Wang *et al.*, 2024] Zhe Wang, Petar Veličković, Daniel Hennes, Nenad Tomašev, Laurel Prince, Michael Kaisers, Yoram Bachrach, Romuald Elie, Li Kevin Wenliang, Federico Piccinini, William Spearman, Ian Graham, Jerome Connor, Yi Yang, Adrià Recasens, Mina Khan, Nathalie Beauguerlange, Pablo Sprechmann, Pol Moreno, Nicolas Heess, Michael Bowling, Demis Hassabis, and Karl Tuyls. Tacticai: an ai assistant for football tactics. *Nature Communications*, 15(1):1906, 2024.