*Journal of the Royal Statistical Society Series C: Applied Statistics*, 2023, **72**, 1095–1126
https://doi.org/10.1093/jrsssc/qlad085
Advance access publication 29 August 2023
**Discussion Paper**

# Flexible marked spatio-temporal point processes with applications to event sequences from association football

**Santhosh Narayanan[1], Ioannis Kosmidis[2] and Petros Dellaportas[3,4]**

[1]The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, United Kingdom
[2]Department of Statistics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom
[3]Department of Statistical Science, University College, Gower Street, London WC1E 6BT, United Kingdom
[4]Department of Statistics, Athens University of Economics and Business, 76 Patission Str., Athens 10434, Greece

*Address for correspondence:* Santhosh Narayanan, The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK. Email: snarayanan@turing.ac.uk

[Read before The Royal Statistical Society at an online meeting organized by the RSS Statistics in Sport Section on Tuesday, 22 November 2022, Dr B Hale in the Chair]

## Abstract

We develop a new family of marked point processes by focusing the characteristic properties of marked Hawkes processes exclusively on the space of marks, allowing a separate model specification for the occurrence times. We develop a Bayesian framework for their inference and prediction that can naturally accommodate covariate information to drive cross-excitations, offering broad flexibility for real-world applications. The framework is applied to in-game event sequences from association football, resulting in inferences about previously unquantified characteristics of game dynamics, extraction of event-specific team abilities and predictions for event occurrences, such as goals or fouls in a specified interval of time.

**Keywords:** Bayesian inference, branching structure, Hamiltonian Monte Carlo, team abilities

## 1 Introduction

Football is one of the most popular team sports and is an example of an invasive sport, where two opposing teams compete for the possession of the ball with the dual objective of attacking to score a goal and defending against attacks from the opposition. Most analyses in football is typically done manually by studying video footage or using simple frequency analysis of match events. Hence, there is huge scope to improve the efficiency of the data-analytic methods as well as the quality of performance evaluation. However, the analysis of football data is mathematically and statistically challenging due to the continuous interaction between players within and across the two teams. As an introduction, we describe the event data from football and survey the existing work in this area before arguing how marked point processes are well suited to developing a modelling foundation to achieve our goal of describing the game dynamics.

### 1.1 Football event data

Over the last decade, the availability of spatio-temporal data from team sports has inspired research into the application of statistical methods for team and player performance evaluation. A comprehensive survey of the recent research efforts into the spatio-temporal analysis of team sports is provided in Gudmundsson and Horton (2017). There are two primary types of spatio-

**Received:** March 8, 2021. **Revised:** January 4, 2022. **Accepted:** May 12, 2022

temporal data collected from team sports. Movement data consists of samples of time-stamped locations in the plane tracking the movement of all players and the ball during the game. Player movement is captured using fixed cameras in optical tracking systems that process the images to obtain the trajectories. Event data streams, on the other hand, record the sequence of events that occur during the game and are collected manually by trained analysts who watch video feeds of the games through special annotation software. As our work is motivated by the availability of event data from football, we focus on reviewing research that uses event data streams. Event data is less dense than movement data but richer in the sense that they contain more information about what is happening in the game. Events broadly fall into two categories: player events such as passes and shots; and stoppage events such as fouls, end of the game, etc. Every event is annotated with a time-stamp, its location, its type (pass, foul, etc.), the players involved, and team information.

A popular research topic based on event data is the network analysis of player interaction. Models for player interaction can quantify a team's playing style as well as the importance of an individual player within the team. Players are identified as nodes of a network and are connected using directed edges whose weights are proportional to the number of successful passes between the two players. Passing networks were first applied to team sports in Passos et al. (2011) to study a team's collective behaviour in water polo. Grund (2012) studied the degree centrality of passing networks in football, which quantifies the importance of nodes in the network based on the number of edges. They showed that teams that rely heavily on key players performed relatively worse. Duch et al. (2010) used flow centrality to assess player performance by capturing the fraction of times that a player intervenes in those paths that result in a shot on goal. They also take into account defensive behaviour by letting each player start a number of paths proportional to the number of balls they recover. Clemente et al. (2015) studied the density and heterogeneity of passing networks and showed that high heterogeneity leads to the formation of sub-communities, meaning there is a low level of cooperation between the players of a team. Pena and Touchette (2012) looked at other centrality measures such as closeness and eigenvector centrality as well as clustering in football passing networks.

Another use of event data is in the identification of frequently occurring sequences of passes between a small group of players within the same team. In Borrie et al. (2002), passes are identified by the zones in the pitch they start and end in and frequently occurring sequences are detected by also taking into account the time intervals between passes. Wang et al. (2015) proposed an unsupervised approach to automatically detect tactical patterns in football. They present the Team Tactic Topic model based on Latent Dirichlet Allocation to identify tactics from pass sequences. Interesting visualisations are provided for the most successful tactics as well as how a team's tactical patterns evolve over a season. Van Haaren et al. (2016) also look at the automatic discovery of patterns in attacking strategy. They use a data-driven approach to determine a number of spatial features about the areas occupied during a continuous possession phase of a team. The features are then used to cluster similar phases together to identify frequently occurring event sequences within the cluster. Decroos et al. (2017) partition the game using overlapping intervals to create subsequences of events to use as a feature to predict a goal event in the near future. They compute the similarity between subsequences using Dynamic Time Warping, a distance measure for time-dependent sequences.

Extracting game states from event sequences to quantify the value of player actions or to make predictions of the game outcome is another interesting area of research. Routley and Schulte (2015) used Markov decision processes for valuing player actions in Ice Hockey. Game states are derived from contextual features like game score and time remaining along with the recent history of events. The associated reward for an action in the Markov decision process gives the value of the player action. A similar approach based on game states is taken in Decroos et al. (2018) to value player actions in football. They train a classification model to calculate the probability a game state will lead to a goal in the near future, where each game state is described using over 150 features. The value of a player's action is then calculated by the shift in the predicted goal probability before and after the action. Other approaches for predicting goal probabilities based on a current game state are by Mackay (2017) and Robberechts et al. (2019). Approaches based on game states involve significant effort into feature engineering, and with the use of learning algorithms like gradient boosting that limit parameter interpretations, the methods provide, typically, little insight into the dynamics of the game.

The major focus of existing methods in team sport analysis appears to be tailored towards individual player performance evaluation or identifying specific patterns in team play. Most approaches take the route of summarising the event data into compact representations like networks and game states. In this paper, we take a more holistic approach to studying football as a dynamic system and model the entire sequence of events within a game. Such a model, which captures all event interactions, is attractive for predicting the occurrence of the rare goal-scored events that determine the outcome of the game.

## 1.2 Point processes

Phenomena that are observed as a sequence of events happening over time can be represented using point processes. While point processes can describe a random collection of points in any general space, we limit ourselves to the case in which the points denote events that occur along a time axis. Such point processes, having a natural order in which the points occur, are suitable for a wide range of real-world applications and are well studied in probability theory.

As in Daley and Vere-Jones (2003, Section 6.4), processes in which points are identified only by the occurrence times are referred to as univariate point processes. Multivariate point processes, on the other hand, are those in which the realisation of a discrete random variable, say $m$, with a finite number of categories is recorded along with the occurrence times. Marked point processes are processes where $m$ is allowed to be a continuous random variable. An example application of a marked point process with continuous marks is in seismology, where the magnitude of an earthquake is recorded in addition to the time of occurrence. In this paper, we model event sequences observed in football using marked point processes with discrete marks used to denote the event type.

When event sequence data are analysed using point process models, an important distinction is between empirical models and mechanistic models as noted by Diggle (2013). Empirical models have the solitary aim of describing the patterns in the observed data, while mechanistic models go beyond that and attempt to capture the underlying process that generated the data. Mechanistic models for marked point processes are typically specified using a joint conditional intensity for the occurrence times and the marks and in general are not flexible enough to be applied to complex real-world phenomena. The joint modelling of the components of the process can also be challenging and it is common to make strong restrictive assumptions like separability (González et al., 2016) to simplify the model. In this paper, we present a flexible mechanistic modelling framework for marked point processes that are suitable for a wide range of applications without the need for assumptions like separability.

We produce a family of marked point processes that generalises the classical Hawkes process, a mathematical model for self-exciting processes proposed in Hawkes (1971) that can be used to model a sequence of arrivals of some type over time, for example, earthquakes in Ogata (1998). Each arrival excites the process in the sense that the chance of a subsequent arrival increases for a period of time after the initial arrival and the excitation from previous arrivals add up. Marked Hawkes processes are typically specified using a joint conditional intensity function for the occurrence times and the marks (see, for example, Rasmussen, 2013, expression 2.2), and captures the magnitudes of all cross-excitations between the various event types as well as the rate at which these excitations decay over time. Excitation leads to the clustering of events in time as the process is driven by an intensity that increases with every arrival for a short period of time. However, in applications like the event sequences observed in football, the events tend not to cluster in time and the marked Hawkes process model is not suitable. The joint modelling of the times and the marks have to be decoupled to restrict the excitation property of the process exclusively to the dimension of the marks.

Similar to the decomposition of a multivariate distribution function that motivated the partial likelihood in Cox (1975), we factorise the joint conditional distribution for a marked point process into probability density functions for each event time conditioned on past occurrences times and marks, and probability mass functions for the event marks conditioned on the time of occurrence and the filtration of the process. Therefore, an alternative approach to specifying a marked point process model is to specify the conditional distribution functions for the times and the marks separately. We derive the conditional distribution function for the marks from a marked Hawkes process, which gives us then the freedom to specify the conditional distribution for the times separately. In this way, we are able to construct marked point process models that retain the characteristic properties of Hawkes processes, such as excitation for the marks, while avoiding the strong clustering of event times.

We develop a framework for Bayesian inference of such flexible marked point processes, which is realised through the Stan (Stan Development Team, 2020) software for statistical modelling. Stan implements a variant of the Hamiltonian Monte Carlo algorithm, originally proposed by Duane et al. (1987), to generate samples from the posterior distribution of the parameters. The Bayesian models we consider are compared using the out-of-sample log predictive density.

We define marked point processes for the modelling of touch-ball events in football, which along with time and event type information also carry location information. As we illustrate, the family of marked point processes can be readily enriched to handle all times, event types and locations. We are also able to incorporate team information in a direct way that captures the relative abilities of the teams for each event type. We develop a method based on association rules (Agrawal et al., 1993) to reduce the complexity introduced by the model extensions we introduce. The rule-based approach identifies significant event interactions within sequences by placing thresholds on particular measures of significance. We then evaluate the accuracy of the excitation-based models by comparing them against two baseline models and confirm the superior performance of the models with excitation effects.

We provide a detailed parameter description showing how the model parameters can be used to gain valuable insights into football. The excitation component of the proposed model captures both the magnitudes and the durations of all pairwise event interactions across different locations. From the conversion rate parameters, we are able to confirm the well-known home advantage effect, and quantify the relative performance of each team when playing games at their home venue compared to away. The conversion rate parameters are also driven by team information, via the team ability parameters which, for example, can capture the relative ability of a team to convert one successful pass to another and retain possession of the ball. We also discuss how the team ability parameters can be used to obtain rankings for the teams by event type, which can be used as predictors of team performance. The team ability parameters also capture some interesting differences in the playing styles of the teams that are not immediately apparent just by looking at the event data. In this way, the model along with its parameters can be used to develop a deeper understanding of the game-play by the coaching staff and inform strategic decision-making. The proposed model can also be used to simulate the sequence of events in a game to obtain real-time predictions of event probabilities. The simulator results in predictions that can enhance, among others, the viewing experience of televised games. Finally, like Hawkes Processes, the proposed model also allows the recovery of the hidden branching structure of the process that quantifies the relative contributions of the background process and previous occurrences to the triggering of a new event.

The developments in this paper can be readily applied to many other team sports like rugby, hockey, basketball, etc. As none of the methods have been tailored specifically to football or even sports for that matter, they can also be applied to a wide range of applications that generate event data streams.

## 2 Data

### 2.1 Description and descriptives

The data that motivated this work was provided by Stratagem Technologies Ltd, and consists of all touch-ball events from all English Premier League games in the 2013/2014 season. A touch-ball event is an event where a player has acted on the ball by touching it with some part of their body. We identified mistakes in the original data, with the most critical issues relating to impossible sequences of consecutive events (e.g. a dribble a few seconds after a goal). Such data issues have been addressed in a systematic way, using the data-cleaning workflow in the publicly available PhD thesis by Narayanan (see, Narayanan, 2021, Section 5.3). In total, the data consists of over half a million touch-ball events recorded over the season along with other attributes. A snapshot of the data is provided in Table 1. The league is contested by a total of 20 teams and follows a round-robin tournament schedule, where each team plays every other team at their home and away venues, which results in a total of 380 games over the season.

Each game comprises two halves that are separated by an interruption of approximately 15 min. In what follows, we refer to each uninterrupted game half as a game period. For each touch-ball event, we have records of the event type, time-stamp, $(x, y)$ co-ordinates of its location in the playing field, team and unique player identifiers, game period, and if the touch-ball event is a Pass, the

**Table 1.** Events and their attributes from the first 20 s of the game between Southampton and West Ham United on 15 September 2013

| Second | Minute | Team_id | Player_id | Type | $x$ | $y$ | Outcome | End_$x$ | End_$y$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 14 | 29,544 | Pass | 50.1 | 48.8 | Successful | 51.1 | 48.2 |
| 2 | 0 | 14 | 21,683 | Pass | 51.1 | 48.2 | Successful | 39.2 | 47.8 |
| 4 | 0 | 14 | 71,714 | Pass | 39.2 | 47.8 | Successful | 29.5 | 77.6 |
| 6 | 0 | 14 | 118,244 | Pass | 30.8 | 79.6 | Unsuccessful | 33.5 | 79.7 |
| 12 | 0 | 20 | 12,533 | BallRecovery | 34.9 | 89.9 | | | |
| 13 | 0 | 20 | 12,533 | Pass | 35.9 | 88.3 | Successful | 37.3 | 76.1 |
| 15 | 0 | 20 | 8,247 | Pass | 34.9 | 77.0 | Unsuccessful | 44.9 | 85.9 |
| 16 | 0 | 14 | 71,714 | Interception | 53.2 | 16.7 | | | |
| 18 | 0 | 14 | 69,375 | Pass | 43.1 | 23.1 | Unsuccessful | 70.9 | 9.7 |

*Note*. For each event, we have records of the event time-stamp, team and player ids, event type, $(x, y)$ co-ordinates of its location in the playing field, and if the event type is a Pass, the event outcome (successful/unsuccessful) and the end $(x, y)$ co-ordinates.

**Table 2.** Frequencies of the 22 distinct types of touch-ball events in the data

| Event type | Frequency | Event type | Frequency |
|---|---|---|---|
| Pass | 376,924 | SavedShot | 4,971 |
| BallRecovery | 36,908 | Save | 4,910 |
| Clearance | 25,462 | CornerAwarded | 4,100 |
| Tackle | 14,581 | MissedShots | 4,076 |
| TakeOn | 13,607 | OffsidePass | 1,582 |
| BallTouch | 13,517 | Claim | 1,181 |
| Aerial | 12,871 | Goal | 1,052 |
| Interception | 10,422 | Punch | 380 |
| Dispossessed | 8,897 | ShotOnPost | 187 |
| Foul | 8,238 | Smother | 122 |
| KeeperPickup | 5,208 | CrossNotClaimed | 81 |

event outcome (successful/unsuccessful) and the end $(x, y)$ co-ordinates. Table 2 gives the frequency of each of the 22 distinct touch-ball event types recorded in the data.

Figure 1 shows the trajectory of the ball during an attacking move that led to a goal in the 18th minute of the game between Arsenal and Norwich City on 19 October 2013. The goal was scored by Jack Wilshere for Arsenal and was voted as the best goal in the English Premier League for the 2013/2014 season.

Latent game characteristics, such as the home advantage, and differences in playing styles and formations between the teams are also reflected in the touch-ball events. For example, Figure 2 compares the concentration of ball-touches for Arsenal and Chelsea between their home and away games in the 2013/2014 season. The playing field is plotted so that the team is always attacking to the right. It is clear that when the teams play at home the density of events is higher towards the opponent's goal. In fact, the point process modelling framework developed in this paper allows us to quantify home advantage by, for example, learning the relative ability of each team to retain possession when playing at home compared to away (see Section 6.6).

## 2.2 Data preparation

We combine the types and outcomes of touch-ball events into in-play and terminal composite events. The in-play composite events are Win, Dribble, Successful Pass (Pass_S), Unsuccessful

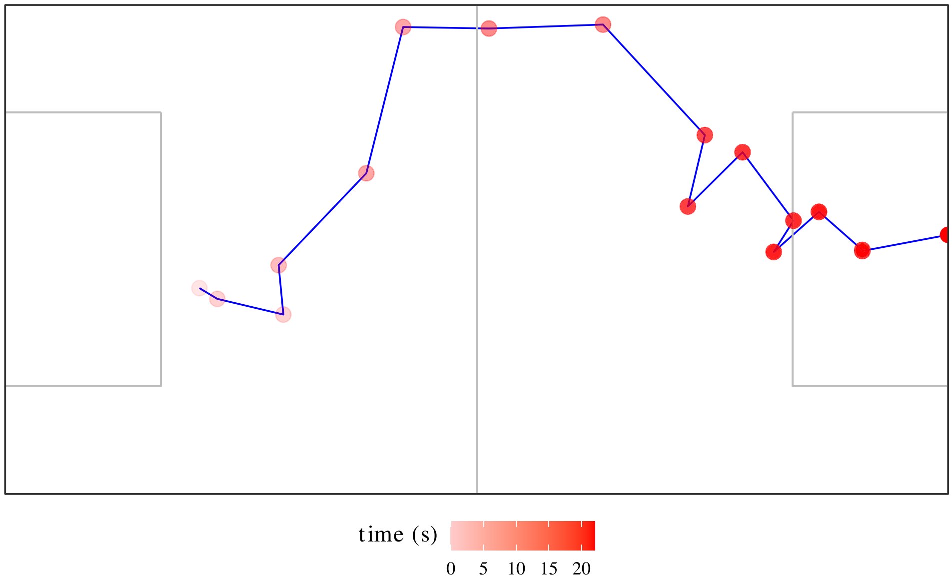



**Figure 1.** Tracing the locations of the sequence of events that led to the goal scored by Jack Wilshere for Arsenal against Norwich City (voted the best goal of the 2013/2014 season).

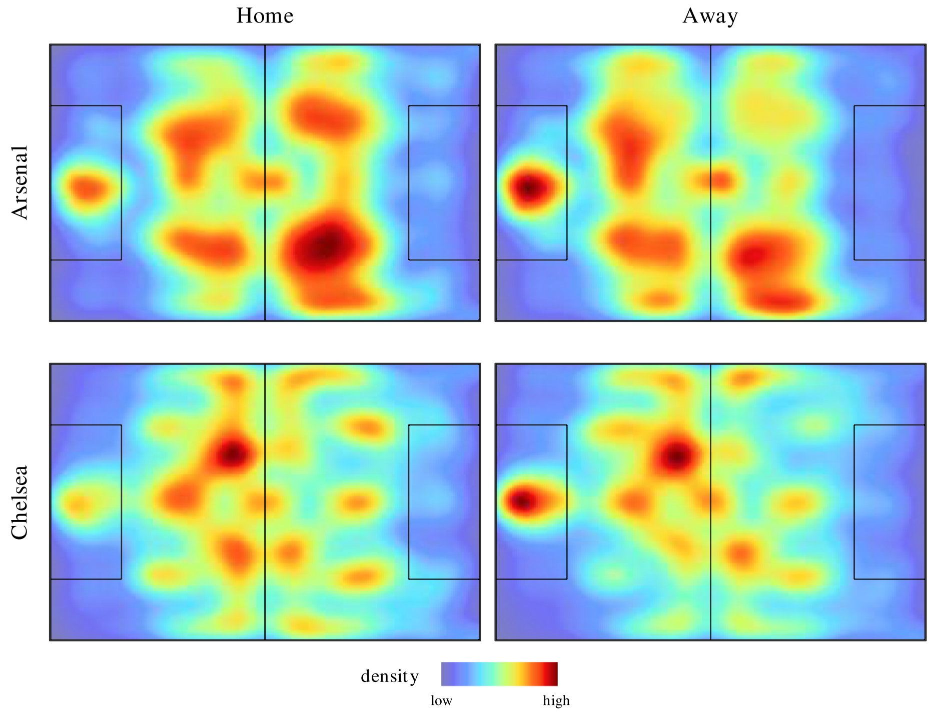



**Figure 2.** Heat maps showing the density of ball-touches for Arsenal and Chelsea in their home and away games in the 2013/2014 season. In all heat maps, the team is attacking to the right.

Pass (Pass_U), Shot, Keeper, Save, Clear, and Lose. Win denotes a player regaining possession of the ball from the opponent. Dribble is taking the ball forward with repeated slight touches. Passes are deemed to be successful when the ball is received by a team-mate and unsuccessful otherwise. Shots include all attempts on the opponents' goal, including those missing the target. The Keeper

**Table 3.** Composite event types along with their labels and observed frequencies in the data

| *M* | Mark label | Count | *M* | Mark label | Count |
|---|---|---|---|---|---|
| 1 | Home_Win | 10,864 | 16 | Away_Win | 10,829 |
| 2 | Home_Dribble | 3,432 | 17 | Away_Dribble | 3,123 |
| 3 | Home_Pass_S | 152,140 | 18 | Away_Pass_S | 140,975 |
| 4 | Home_Pass_U | 42,344 | 19 | Away_Pass_U | 41,462 |
| 5 | Home_Shot | 5,127 | 20 | Away_Shot | 4,107 |
| 6 | Home_Keeper | 3,273 | 21 | Away_Keeper | 3,555 |
| 7 | Home_Save | 2,208 | 22 | Away_Save | 2,702 |
| 8 | Home_Clear | 11,780 | 23 | Away_Clear | 14,059 |
| 9 | Home_Lose | 16,534 | 24 | Away_Lose | 16,515 |
| 10 | Home_Goal | 597 | 25 | Away_Goal | 455 |
| 11 | Home_Foul | 4,229 | 26 | Away_Foul | 4,009 |
| 12 | Home_Out_Throw | 8,982 | 27 | Away_Out_Throw | 8,396 |
| 13 | Home_Out_GK | 3,084 | 28 | Away_Out_GK | 3,697 |
| 14 | Home_Out_Corner | 2,321 | 29 | Away_Out_Corner | 1,779 |
| 15 | Home_Pass_O | 814 | 30 | Away_Pass_O | 768 |

event denotes the goalkeeper taking possession of the ball into their hands by picking it up or claiming a cross. The Keeper event is unlike any other in-play event, as the goalkeeper is allowed to hold the ball without being challenged for a period of time while waiting for opponents to clear the goal area. As a result, there is often a delay before the next event even though the ball is technically in-play. Saves are events where the goalkeeper prevents a shot from crossing the goal line. Clear events are those where a player moves the ball away from their goal area to safety while the Lose event is when a player loses possession of the ball. The terminal composite events are those which result in the ball going out-of-play and are Goal, Foul, Out_Throw, Out_GK, Out_Corner, and Offside Pass (Pass_O). The terminal events interrupt the game resulting in a delay before play resumes. Each composite event is tracked for both the home and away teams. For this reason, we append 'Home' or 'Away' as a prefix to the event label to distinguish between the events of the two teams playing the game. This results in $M = 30$ distinct composite events, whose labels and observed frequencies are given in Table 3.

For each touch-ball event, the data also contains the associated $(x, y)$ co-ordinates on the playing field. We partition the playing field into 3 zones of equal area. The zones and their corresponding labels are shown in Figure 3. Zone 1 is the region where the home team defends their goal, zone 3 is the region where the home team attacks, and zone 2 is the midfield region. It is natural to expect that the control a team has on the game depends on the zone the ball is at. For example, the home team is expected to retain possession of the ball more successfully in zone 1 as compared to zone 3.

Table 4 shows a snapshot of the data after its preparation, including a unique identifier for each game period in the data.

## 3 Marked point processes

### 3.1 Conditional intensity function

Sequences of events over time are conveniently represented as realisations of a point process. Oftentimes, the events can carry additional information, which are assumed to be realisations of random variables, referred to as marks. The collection of the times $\{t_i\}$ at which the events occur and the marks $\{m_i\}$ is a marked point process, whose ground process, is the process for $\{t_i\}$ only.

A marked point process is typically specified through its joint conditional intensity function

$$\lambda^*(t, m) = \lambda_g^*(t) f^*(m \mid t), \tag{1}$$

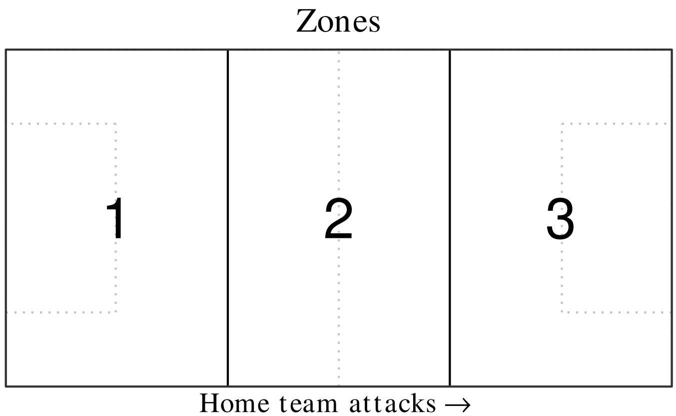



**Figure 3.** Mapping from event location in ($x$, $y$) co-ordinates to zones.

**Table 4.** Snapshot of the final data prepared for modelling where each event, indexed by $i = 1, \ldots, n$, consists of the following components, the time of occurrence $t_i$, the zone $z_i$, and the mark $m_i$

| $i$ | id | Period | Team_id | Time ($t_i$) | Zone ($z_i$) | Mark ($m_i$) |
|---|---|---|---|---|---|---|
| 1 | 101 | 1 | 1 | 0 | 2 | 18 |
| 2 | 101 | 1 | 1 | 1 | 2 | 19 |
| 3 | 101 | 1 | 2 | 3 | 1 | 8 |
| 4 | 101 | 1 | 1 | 6 | 3 | 16 |
| 5 | 101 | 1 | 1 | 8 | 3 | 18 |
| 6 | 101 | 1 | 1 | 15 | 2 | 18 |
| 7 | 101 | 1 | 1 | 16 | 1 | 19 |
| 8 | 101 | 1 | 2 | 19 | 1 | 12 |

*Note*. The home and away team information for each game is assumed to be known and the first event ($t_1$, $z_1$, $m_1$) in each game period is considered to be deterministic and therefore, not modelled.

where $\lambda_g^*(t)$ is the conditional intensity of the ground process and $f^*(m \mid t)$ is the conditional probability density or mass function of the mark $m$ at time $t$. Both $\lambda_g^*(t)$ and $f^*(m \mid t)$ in equation (1) are understood as being conditional on $\mathcal{F}_{t^-}$, which is the filtration of the marked point process up to but not including $t$.

## 3.2 Marked Hawkes processes

Marked Hawkes processes are point processes whose defining characteristic is that they self-excite, meaning that each arrival increases the rate of future arrivals for a period of time. More formally, consider a realisation of a marked point process, consisting of event times $\{t_i\}$ with $t_i \in \mathbb{R}^+$ and $t_i > t_{i-1}$, and marks $m_i \in \{1, \ldots, M\}$ ($i = 1, \ldots, n$), where $M$ is the number of discrete marks. The marked Hawkes process is most intuitively specified using its mark dependent conditional intensity function $\lambda^*(t, m)$, which for an exponentially decaying intensity is (Rasmussen, 2013, expression 2.2)

$$\lambda^*(t, m) = \mu\delta_m + \sum_{t_j<t} \epsilon\beta e^{-\beta(t-t_j)} \gamma_{m_j \to m}. \tag{2}$$

In equation (2), the parameter $\mu > 0$ is a constant background intensity and $\delta_m \in (0, 1)$ is the background mark probability for mark $m$ with $\sum_{m=1}^{M} \delta_m = 1$. The parameter $\epsilon \in (0, 1)$ is the excitation factor, $\beta > 0$ is the exponential decay rate, and $\gamma_{m_j \to m} \in (0, 1)$ is the probability the excitation from an event of mark $m_j$ triggers an event of mark $m$, with $\sum_{m=1}^{M} \gamma_{m_j \to m} = 1$ for any $m_j \in \{1, \ldots, M\}$.

## 3.3 Limitations of the marked Hawkes process model

The specification in equation (2) describes a marked Hawkes process that is linear in the sense that the excitations from different arrivals add up, not only increasing the probability of triggering an event of a particular type, but also concentrating the occurrence times for a certain period of time. For this reason, marked Hawkes processes have proven useful in a wide range of applications, where events tend to cluster in time, such as the modelling of earthquakes (Ogata, 1998), gang violence (Mohler et al., 2011), and financial market events (Bowsher, 2007).

However, in applications like the modelling of event sequences in football, each event triggers other events of a particular type with high probability, while it is not necessarily true that the occurrence times cluster.

As an illustration that the observed events in football do not exhibit clustering, consider only the collection of the times $\{t_i\}$ at which the events occur. A succinct method to investigate the aggregation of the points is using the non-parametric Ripley's $K$-function summary (Ripley, 1977), which is the reduced second-moment measure. An estimator of the $K$-function in the one-dimensional case is derived in Diggle (1985) as

$$\hat{K}(t) = \frac{T}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} w_{ij} \mathbb{1}(|t_i - t_j| \leq t), \tag{3}$$

where $(0, T)$ is the time interval over which the $n$ points are observed, $\mathbb{1}(.)$ is the indicator function, and $w_{ij}$ is an edge correction taking values $w_{ij} = 1$ if $|t_i - t_j| \leq \min(t_i, T - t_i)$ and $w_{ij} = 2$, otherwise.

For a homogeneous Poisson process, $K(t) = 2t$. If $K(t) > 2t$, the process is said to be *over-dispersed* relative to the Poisson and exhibits some degree of clustering. If $K(t) < 2t$, the process is *under-dispersed* relative to the Poisson and tends towards regular occurrences. Figure 4 shows the $K$ function estimate, $\hat{K}(t) - 2t$ for $t \in \{1, 2, \ldots, 100\}$, of the observed event times from the first game of the season between Aston Villa and Arsenal ($n = 1{,}279$). We also compare the estimates from those observed times with the estimates of the events simulated from several one-dimensional Hawkes processes with a conditional intensity of the form, $\lambda^*(t) = \mu + \sum_{t_j < t} \epsilon \beta e^{-\beta(t - t_j)}$. Hawkes I (green) is the fitted Hawkes process with parameters $(\mu, \epsilon, \beta) = (0.4183, 0.0035, 0.0004)$ estimated from the observed times using maximum likelihood. Note that the estimated $\epsilon$ is very close to 0, indicating no clustering. A Hawkes process with $\epsilon = 0$ is the trivial case with no excitation that reduces to a Poisson process (orange) with an estimated rate $\mu = 0.4189$. Hawkes II (pink) has parameters $(\mu, \epsilon, \beta) = (0.2594, 0.4, 0.01)$ and Hawkes III (purple) has parameters $(\mu, \epsilon, \beta) = (0.1068, 0.8, 0.01)$. Hawkes II and Hawkes III are examples of processes with moderate and severe clustering, respectively, whose $\mu$ parameters were estimated from the observed times using maximum likelihood after fixing $\epsilon, \beta$. The box plots of the estimates for the Hawkes and Poisson processes were computed using 100 independent simulations of each process over the same time interval as the observed times.

The $\hat{K}(t) - 2t$ values for the Hawkes II and Hawkes III processes quickly get above 0 and increase with $t$ demonstrating the behaviour of processes with different degrees of clustering. On the other hand, the $\hat{K}(t) - 2t$ values for the fitted Hawkes process (Hawkes I) and the Poisson process concentrate around 0, being indicative of the expected behaviour of processes where points do not cluster. The $\hat{K}(t) - 2t$ values from the observed times range from $-1.4$ to $-0.9$ indicating slight under-dispersion relative to the Poisson process. In other words, the observed times in football exhibit no clustering and in fact show evidence of being more regular than the Poisson process.

Another method to investigate the aggregation of points is by looking at the distribution of the inter-arrival times. Figure 5 shows the empirical distribution function of the first 10,000 inter-arrival times from the first seven games of the league season. We also plot the cumulative distribution functions of a homogeneous Poisson process and a Hawkes process, whose parameters are

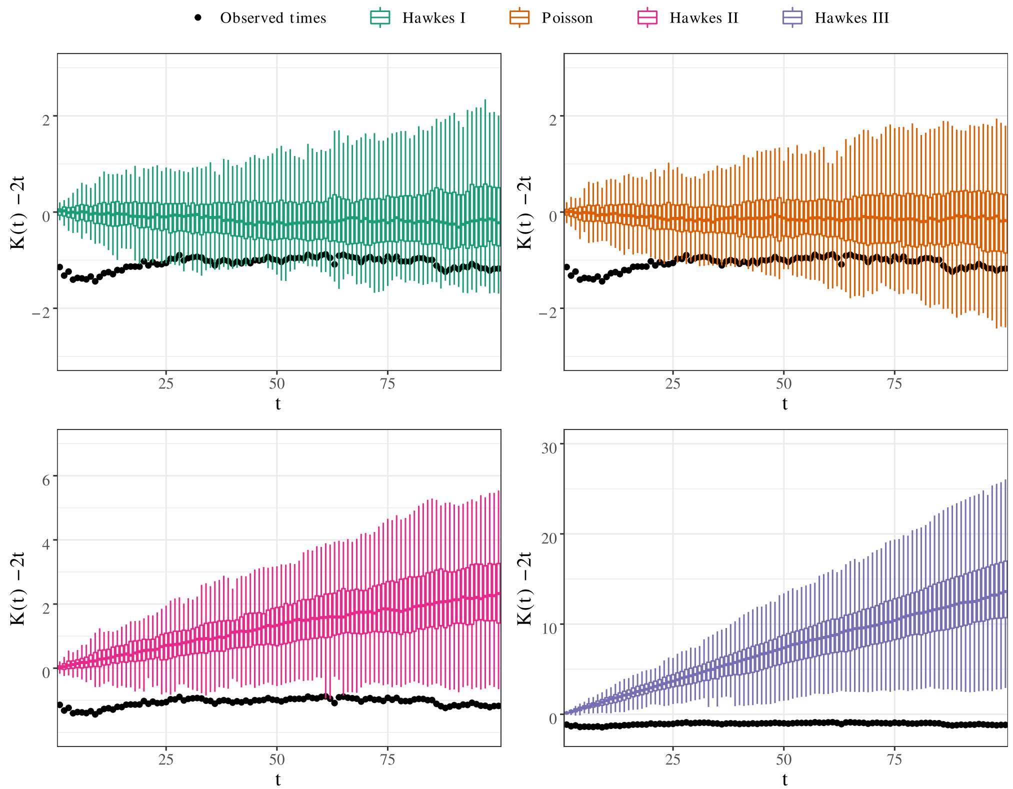



**Figure 4.** The $K$ function estimate $\hat{K}(t) - 2t$ of the observed event times (black points) from the first game of the season between Aston Villa and Arsenal. Hawkes I (top left) is a Hawkes process with parameters $(\mu, \epsilon, \beta) = (0.4183, 0.0035, 0.0004)$ estimated from the observed times using maximum likelihood. A Hawkes process with $\epsilon = 0$ is the trivial case with no excitation that reduces to a Poisson process (top right) with an estimated rate $\mu = 0.4189$. Hawkes II (bottom left) has parameters $(\mu, \epsilon, \beta) = (0.2594, 0.4, 0.01)$ and Hawkes III (bottom right) has parameters $(\mu, \epsilon, \beta) = (0.1068, 0.8, 0.01)$. Hawkes II and Hawkes III are examples of processes with moderate and severe clustering, respectively, whose $\mu$ parameters were estimated from the observed times using maximum likelihood after fixing $\epsilon, \beta$. The box plots of the estimates for the Hawkes and Poisson processes were computed using 100 independent simulations of each process over the same time interval as the observed times.

estimated from the aforementioned 10,000 events using maximum likelihood. The fitted Hawkes process is far from the empirical distribution function, and almost identical to the fitted Poisson process confirming that the arrival times in football do not cluster. This is further evidence that Hawkes processes are not appropriate for modelling events such as those observed in football, which tend not to cluster in time. Figure 5 also includes the cumulative distribution function of a fitted Gamma process, which, as is apparent provides an excellent fit to the observed inter-arrival times.

# 4 Specification of flexible marked point processes

## 4.1 Decoupling the modelling of times and marks

According to the decomposition of a multivariate distribution function in Cox (1975, expression 2), the likelihood of a marked point process observed in $(0, T)$ can always be factorised as

$$\mathcal{L}(\mathcal{F}_{t_n} \mid \zeta, \boldsymbol{\theta}) = \prod_{i=1}^{n} \{g(t_i \mid \mathcal{F}_{t_{i-1}}; \zeta) f(m_i \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\theta})\} \{1 - G(T \mid \mathcal{F}_{t_n}; \zeta)\}, \tag{4}$$

where $g$, $G$, and $f$ are the conditional density and distribution function for the times, and the probability mass function for the marks, respectively, and $\zeta, \boldsymbol{\theta}$ are unknown parameter vectors that may or may not share components. The last term in equation (4) accounts for the fact that the

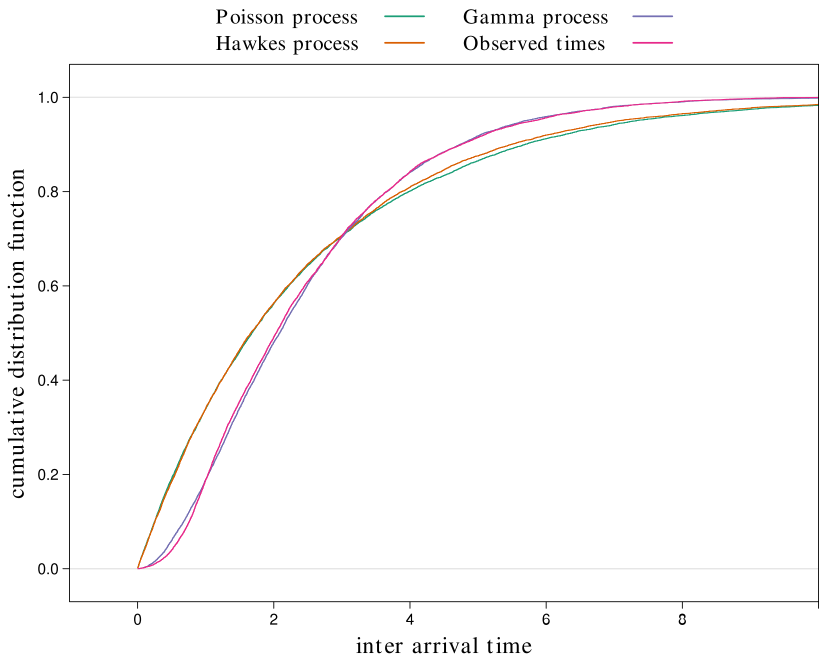



**Figure 5.** Comparing the cumulative distribution functions (CDFs) of the inter-arrival times of events simulated from a Poisson process (green), a Hawkes process (orange), a Gamma process (purple), and observed event times in football (pink). Empirical CDFs were computed using 10,000 inter-arrival times in each case.

unobserved occurrence time $t_{n+1}$ must be after the end of the observation interval $(0, T)$. Therefore, an alternative approach to specifying a marked point process is to specify the functions $g(\cdot \mid \mathcal{F}_{t_{i-1}}; \zeta)$ and $f(\cdot \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\theta})$, separately, and combine them as in equation (4). The key insight in the current work is to derive the specification for the marks $f(\cdot \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\theta})$ from the joint conditional intensity function of a marked Hawkes process model, and then specify a probability density function for the times $g(\cdot \mid \mathcal{F}_{t_{i-1}}; \zeta)$ best suited to our application. In this way, we can restrict the characteristic excitation property of marked Hawkes processes exclusively to the modelling of the marks, and have the freedom to specify a different model for the occurrence times.

By the definition of the conditional intensity function for a marked point process in equation (1), $f(m_i \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\theta}) = \lambda^*(t_i, m_i) / \sum_{m=1}^{M} \lambda^*(t_i, m)$. Plugging in $\lambda^*(t, m)$ from equation (2) in the latter expression, gives

$$f(m_i \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\theta}) = \frac{\delta_{m_i} + \sum_{t_j < t_i} \alpha^* \mathrm{e}^{-\beta(t_i - t_j)} \gamma_{m_j \to m_i}}{1 + \sum_{t_j < t_i} \alpha^* \mathrm{e}^{-\beta(t_i - t_j)}}, \tag{5}$$

where $\alpha^* = \frac{\epsilon\beta}{\mu}$. Expression (5) makes it immediately apparent that the parameters $\mu$ and $\epsilon$ of the marked Hawkes process as specified by equation (2) are not always identifiable for general specifications of $g(\cdot \mid \mathcal{F}_{t_{i-1}}; \zeta)$ in equation (4). Apart from a mathematical fact, this is also rather intuitive, because $\mu$ and $\epsilon$ in equation (2) characterise the evolution of the Hawkes process in the time dimension and the sequence of marks is not sufficient to identify them.

The specification of the marked point process likelihood is complete once a probability density function $g(\cdot \mid \mathcal{F}_{t_{i-1}}; \zeta)$ for the event times is specified.

We should highlight here that the marked point processes from the factorisation in equation (4) are generally different to the ones that result by assuming the separability of the conditional intensity functions (see, for example, González et al., 2016, Section 6.5). A separable conditional intensity function has the form

$$\lambda^*(t, m) = \lambda_g^*(t) f^*(m)$$

and implies that the conditional distribution of the mark does not depend on the occurrence time $t$. Separability is a convenient assumption because it allows for the sequence of marks to be modelled separately from the sequence of times. In contrast, the factorisation in equation (4) allows the conditional distribution of the mark to depend on the time of occurrence as well as the history, still allowing for estimating $\boldsymbol{\theta}$ separately from $\boldsymbol{\zeta}$, if $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ do not share components.

The proposed marked point process model also allows the recovery of the hidden branching structure of the process, a key feature of Hawkes Processes (Hawkes & Oakes, 1974). In Section 6.8, we calculate the branching structure probabilities and quantify the relative contributions of the background process and previous occurrences to the triggering of a new event.

## 4.2 Parameter interpretation

In equation (5), the mark probability of each event in the sequence is determined by a combined additive effect from a background component and all previous occurrences. The first term $\delta_{m_i}$ in the numerator is the mark probability associated with the background component, while each term $\alpha^* e^{-\beta(t_i - t_j)} \gamma_{m_j \to m_i}$ is the contribution from the excitation caused by a previous occurrence in the sequence.

The background mark probability $\delta_m \in (0, 1)$ is the probability an event has a mark $m$ if the event is triggered solely by the background component. The excitation factor $\alpha^* \geq 0$ is a scaling factor applied to the contributions from the previous occurrences to the event mark probability. Large values of $\alpha^*$ indicate a stronger dependence of the process on its history because the contributions from previous occurrences are weighted higher relative to the background component. The decay rate $\beta > 0$ is the exponential rate at which the excitations from previous occurrences decay over time. The parameter $\gamma_{m_j \to m_i} \in (0, 1)$ is the probability the excitation from an event of mark $m_j$ triggers an event of mark $m_i$. In other words, $\gamma_{m_j \to m_i}$ can be viewed as the conversion rate for the transition from an event with mark $m_j$ to an event with mark $m_i$.

In summary, as in marked Hawkes processes, the specification for the marks in equation (5) captures not only all cross-excitations between the various marks but also the rate at which these excitations decay over time.

## 4.3 Covariate-driven cross-excitation

The conditional distribution of marks with probability mass function (5), allows to drive the cross-excitation of the marks using covariates. The conversion rates $\gamma_{m_j \to m}$ can be linked to a covariate vector $\boldsymbol{x} = (x_1, \ldots, x_p)^\top$ observed at the current time through the baseline-category logit specification (see, for example, Agresti, 2007, Section 6.1)

$$\log\left(\frac{\gamma_{m_j \to m}}{\gamma_{m_j \to M}}\right) = \phi_{m_j \to m} + \boldsymbol{\omega}_m^\top \boldsymbol{x} \qquad (m = 1, \ldots, M - 1), \tag{6}$$

where $\boldsymbol{\omega}_m$ is an unknown $p$-vector of regression parameters. Then, keeping all covariates apart from $x_t$ fixed, $\omega_{mt}$ is the log of the ratio of odds for category $m$ vs. the baseline category $M$ at $x_t + 1$ to that at $x_t$ $(t = 1, \ldots, p)$. Also, by setting all covariates $x_t$ equal to 0, $\phi_{m_j \to m_i}$ is the log of the ratio of odds for category $m$ vs. the baseline category $M$. The covariate vector $\boldsymbol{x}$ can include a combination of process-specific covariates that are time-invariant, and event-specific covariates. For example, in Section 5.2, we use equation (6) to parameterise the marked point process in terms of the relative abilities of teams for each event type, and produce team rankings per event type.

## 4.4 Spatio-temporal marked point processes

We can readily extend the factorisation of the likelihood in equation (4) to include conditional densities for the event locations, when the latter are observed. We can write

$$\mathcal{L}(\mathcal{F}_{t_n} \mid \boldsymbol{\psi}) = \prod_{i=1}^{n} \left\{ g(t_i \mid \mathcal{F}_{t_{i-1}}; \boldsymbol{\zeta}) h(z_i \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\eta}) f(m_i \mid t_i, z_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\theta}) \right\} \left\{ 1 - G(T \mid \mathcal{F}_{t_n}; \zeta) \right\}, \tag{7}$$

where $\{z_i\}$ is the collection of random variables corresponding to the spatial component of the process, which is characterised by the conditional probability mass or density functions $h(\cdot \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\eta})$ with $\boldsymbol{\eta}$ being a parameter vector that may or may not share parameters with $\boldsymbol{\zeta}$ and $\boldsymbol{\theta}$, and $\boldsymbol{\psi} = (\boldsymbol{\zeta}^\top, \boldsymbol{\eta}^\top, \boldsymbol{\theta}^\top)^\top$. The filtration $\mathcal{F}_{t_i}$ now includes all times, marks and locations up to time $t_i$.

If the process ends at the last occurrence time $t_n$, then the last term $1 - G(T \mid \mathcal{F}_{t_n}; \boldsymbol{\zeta})$ in equations (4) and (7) is not part of the likelihood (see, for example Lindqvist, 2006, Section 4.2). This is the case in the modelling of touch-ball event sequences in football we consider here, where the process ends with or immediately after the last event observed in each half of the game.

## 5 Bayesian modelling of in-game event sequences

### 5.1 Preamble

The framework for specifying flexible marked point processes of Section 4 is rather attractive for the modelling of in-game event sequences in football and other team sports. Firstly, cross-excitation of in-game events is a natural assumption because any event in an event sequence is likely to be triggered by one or more of the previous events. For example, following a corner kick, the next event is with high probability one among a shot on goal, a defensive clearance or a claim by the keeper. Such effects can be naturally and readily captured by the parameters of the conditional mark distribution in equation (5), which involves not only the magnitudes of all cross-excitations between the various event types but also the rate at which these excitations decay over time. Secondly, the preliminary analyses in Section 3.3 provides strong evidence that occurrence times do not necessarily cluster, as off-the-shelf marked Hawkes processes imply. Hence, the freedom to use a more flexible conditional distribution for the occurrence times, such as a Gamma process, is a rather attractive prospect. Furthermore, as discussed in Section 4.3, team information can be incorporated into the model in a direct way as covariate information to drive the cross-excitation based on the relative abilities of the teams.

Overall, as we demonstrate later, the framework of Section 4 can be used to provide valuable explanatory tools into the underlying dynamics of the game for the coaching staff and inform strategic decision-making. It can also produce predictions of events, such as goals, in a specified time horizon, and of game outcomes that can enhance, among other things, the viewing experience in televised games.

### 5.2 Excitation-based models

Assume that the touch-ball events in $S$ game periods are $S$ realisations of independent spatio-temporal point processes, with the $s$th realisation involving $n_s$ events. Denote by $t_{si}$, $m_{si}$, and $z_{si}$ the occurrence time, mark, and location of the $i$th event in the $s$th realisation, respectively. Each of the $S$ independent spatio-temporal marked point processes have a likelihood as in equation (7) after dropping the last term, and with conditional probability mass function for the marks as in equation (5). The product of the $S$ likelihoods is the overall likelihood based on the $S$ game periods.

The probability density functions for the occurrence times within each period is set to

$$
\begin{gathered}
g(t_{si} \mid \mathcal{F}_{st_{si-1}}; \boldsymbol{\zeta}) = p(t_{si} - t_{si-1} \mid m_{si-1}, \boldsymbol{a}, \boldsymbol{b}) \\
t_{si} - t_{si-1} \mid m_{si-1}, \boldsymbol{a}, \boldsymbol{b} \sim \text{Gamma}(a_{m_{si-1}}, b_{m_{si-1}}),
\end{gathered} \tag{8}
$$

where $\mathcal{F}_{st_{si}}$ denotes the filtration of the $s$th process up to time $t_{si}$.

By this specification, the time to the next event is modelled using a gamma distribution with shape and rate parameters that are specific to the mark of the last observed event. In this way, we wish to capture the differences in the expected time to the next event across the different event types. For example, we expect a shorter time to the next event after an in-play event like a Pass, compared to that of an out-of-play event like a Foul. Even within the group of out-of-play events, we expect a shorter delay following a Throw-in as compared to a Goal event.

For a discrete set of locations $\{1, \ldots, Z\}$, the conditional probability mass function for the current location is set to

$$h(z_{si} \mid t_{si}, \mathcal{F}_{st_{si-1}}; \boldsymbol{\eta}) = \eta_{(z_{si-1}, m_{si-1}) \to z_{si}}, \tag{9}$$

where $\eta_{(z_{si-1}, m_{si-1}) \to z_{si}}$ is the probability of transitioning into location $z_{si}$ given the location $z_{si-1}$ and the mark $m_{si-1}$ of the last observed event. Expression (9) models the sequence of locations as a discrete first-order Markov chain (see, for example, Norris, 1997) with a transition probability matrix $\boldsymbol{\eta}$. The current state of the Markov chain is determined by the combination of the location and the mark of the last observed event, and the probability of transitioning into the next location depends only on the current state. The state space of the Markov chain is given by the Cartesian product $\{1, \ldots, Z\} \times \{1, \ldots, M\}$.

We consider four alternative parameterisations for the conditional probability mass function for the marks. The S$\beta$ (scalar $\beta$) spatio-temporal marked point process results from equations (8), (9), and a conditional mark distribution of the form (5), that is

$$f(m_{si} \mid t_{si}, \mathcal{F}_{st_{si-1}}; \boldsymbol{\theta}) = \frac{\delta_{m_{si}} + \sum_{t_{sj} < t_{si}} e^{\alpha - \beta(t_{si} - t_{sj})} \gamma_{m_{sj} \to m_{si}}}{1 + \sum_{t_{sj} < t_{si}} e^{\alpha - \beta(t_{si} - t_{sj})}}, \tag{10}$$

where $\alpha = \log(\alpha^*)$. The V$\beta$ (vector $\beta$) model results from equations (8), (9), and

$$f(m_{si} \mid t_i, \mathcal{F}_{st_{si-1}}; \boldsymbol{\theta}) = \frac{\delta_{m_{si}} + \sum_{t_{sj} < t_{si}} e^{\alpha - \beta_{m_{sj}}(t_{si} - t_{sj})} \gamma_{m_{sj} \to m_{si}}}{1 + \sum_{t_{sj} < t_{si}} e^{\alpha - \beta_{m_{sj}}(t_{si} - t_{sj})}}, \tag{11}$$

where $\beta_m$ is the exponential decay rate of the excitation caused by an event of mark $m$. V$\beta$ allows the decay rates to depend on the mark of the event causing the excitation. Hence, S$\beta$ is formally nested in V$\beta$ and results when $\beta = \beta_1 = \cdots = \beta_M$. The M$\beta$ (matrix $\beta$) process results from equations (8), (9), and

$$f(m_{si} \mid t_{si}, z_{si}, \mathcal{F}_{st_{si-1}}; \boldsymbol{\theta}) = \frac{\delta_{m_{si}|z_{si}} + \sum_{t_{sj} < t_{si}} e^{\alpha - \beta_{m_{sj} \to m_{si}|z_{si}}(t_{si} - t_{sj})} \gamma_{m_{sj} \to m_{si}|z_{si}}}{\sum_{m=1}^{M} \left[ \delta_{m|z_{si}} + \sum_{t_{sj} < t_{si}} e^{\alpha - \beta_{m_{sj} \to m|z_{si}}(t_{si} - t_{sj})} \gamma_{m_{sj} \to m|z_{si}} \right]}, \tag{12}$$

where $\beta_{m \to m'|z}$ is the decay rate of the excitation caused by an event of mark $m$ on an event of mark $m'$ at location $z$. Specification (12) allows the decay rates to vary both with the pair of marks involved in the excitation and across locations, and allows the background mark probabilities $\boldsymbol{\delta}$ and event conversion rates $\boldsymbol{\gamma}$ to vary across location. The M$\beta$ model can be used to account for scenarios like those where a Corner event excites a Pass event in the short term and a Shot event in the longer term ($\beta_{\text{Corner} \to \text{Pass}|3} > \beta_{\text{Corner} \to \text{Shot}|3}$). It also allows us to capture effects such as how a team is more likely to make more passes and retains possession of the ball in the defensive zone, while attempting more shots on goal in the attacking zone ($\gamma_{\text{Pass} \to \text{Pass}|1} > \gamma_{\text{Pass} \to \text{Pass}|3}$ and $\gamma_{\text{Pass} \to \text{Shot}|3} > \gamma_{\text{Pass} \to \text{Shot}|2}$). The final model we consider is the M$\beta$A (matrix $\beta$ with abilities) where the baseline-category logits of the conversion rates in equation (6) are driven by team information as

$$\log\left(\frac{\gamma_{m_{sj} \to m|z}(c)}{\gamma_{m_{sj} \to M|z}(c)}\right) = \phi_{m_{sj} \to m|z} + \omega_{cm} \quad (m = 1, \ldots, M-1; c = 1, \ldots, C). \tag{13}$$

In the above expression, $\phi_{m \to m'|z}$ is a location-dependent baseline conversion, and $c$ is the team in possession of the ball attempting the event conversion. The parameter $\omega_{cm}$, then, reflects the ability of a team to complete a conversion to an event of mark $m$.

## 5.3 Prior distributions

The shape and rate parameters of the Gamma distributions for the inter-arrival times in equation (8) are assigned independent exponential priors with rates $a'$ and $b'$, respectively. The probability mass function for the locations specified in equation (9) models the locations as a multinomial distribution given the current state of the Markov chain. The conjugate prior for the multinomial distribution is the Dirichlet distribution (see, for example, Gelman et al., 2013, Section 3.4) and therefore we assign a Dirichlet prior on the multinomial probabilities $\boldsymbol{\eta}$ with a common concentration rate parameter $v$. The background mark probability vector $\boldsymbol{\delta}$ in the S$\beta$, V$\beta$, M$\beta$, and M$\beta$A models is also assigned a Dirichlet prior with concentration hyper-parameter $\delta'$. The location-specific mark probability vectors $(\delta_{1|1}, \ldots, \delta_{M|1})^\top, \ldots, (\delta_{1|Z}, \ldots, \delta_{M|Z})^\top$ in the M$\beta$ and M$\beta$A models are assigned independent Dirichlet priors with concentration hyper-parameter $\delta''$. The excitation factor α is assigned a normal prior with mean 0 and standard deviation $\sigma_\alpha$. The decay rate parameter $\beta$ in the S$\beta$ model, the parameters $\beta_1, \ldots, \beta_M$ in the V$\beta$ model, and their location-specific counterparts in the M$\beta$ and M$\beta$A models are assigned independent exponential priors with a common rate $\beta'$. The parameters $\phi_{m\to m'|z}$ and $\omega_{cm}$ $(m, m' = 1, \ldots, M; z = 1, \ldots, Z; c = 1, \ldots, C)$ in the M$\beta$A model are assigned independent Normal priors with mean 0 and standard error $\sigma_\gamma$. The conversion rate parameters $(\gamma_{m\to 1}, \ldots, \gamma_{m\to M})^\top$ in the S$\beta$ and V$\beta$ models, and their location-specific counterparts in the M$\beta$ model are assigned independent Dirichlet priors with a common concentration rate parameter $\gamma'$.

## 5.4 Posterior distributions

The time and location conditional distributions corresponding to equations (8) and (9) share no parameters with each other, and no parameters with any of the conditional mark distributions for the S$\beta$, V$\beta$, M$\beta$, and M$\beta$A models. Furthermore, the likelihood in equation (7) can be factorised into a term depending only on the time parameters $\boldsymbol{\zeta}$, a term depending only on the location parameters $\boldsymbol{\eta}$ and a term depending only on the mark parameters $\boldsymbol{\theta}$. Given that the priors for $\boldsymbol{\zeta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\theta}$ are also independent, the derivation of, or sampling from, the posterior distributions can be performed separately for each of those parameters.

The priors for the location parameters $\boldsymbol{\eta}$ are conjugate, so the posterior for $\boldsymbol{\eta}$ is readily obtained. If $\boldsymbol{y} = \{y_{i\to j}\}$, for $j \in \{1, \ldots, Z\}$, are the observed counts of transitions originating from the state $i$ where $i \in \{1, \ldots, Z\} \times \{1, \ldots, M\}$, then the posterior distribution of each row of the transition matrix $\boldsymbol{\eta}_i$ is a Dirichlet distribution with concentration parameters $(y_{i1} + v, \ldots, y_{iZ} + v)$.

Posterior sampling for the parameter vectors $\boldsymbol{a}, \boldsymbol{b}$ in equation (8) of the conditional distributions for the times, and the parameters $\boldsymbol{\theta}$ of the conditional distributions for the marks in each of the S$\beta$, V$\beta$, M$\beta$, and M$\beta$A models is carried out using the variant of the Hamiltonian Monte Carlo procedure (Duane et al., 1987) that is implemented in Stan (Stan Development Team, 2020).

We have also implemented posterior sampling using a Metropolis-within-Gibbs procedure, which, though, proved to mix poorly in artificial data sets for the S$\beta$, V$\beta$, M$\beta$, and M$\beta$A, rendering it computationally infeasible. As in the case of the Hawkes process, the poor mixing stems from the presence of strong correlations between the model parameters as well as the flatness of the likelihood function (Veen & Schoenberg, 2008). Stan, on the other hand, implements the No-U-Turn Sampler (Hoffman & Gelman, 2014) that automatically calibrates tuning parameters in a warm-up phase and can efficiently sample from complex posterior distributions.

## 5.5 Model complexity

The conditional mark distribution in the M$\beta$ and M$\beta$A models involves a large number of parameters. There are $M^2 Z$ decay rate parameters and $M(M-1)Z$ baseline conversion rate parameters which makes posterior sampling a computationally challenging task. We have developed a screening procedure based on the association rule learning (see, for example, Agrawal et al., 1993) that operates on the data involved in the likelihood and eliminates parameters prior to posterior sampling.

The screening procedure retains only those parameters that capture the most significant event interactions and depends on two constants that need to be chosen. The first is the window size $W$ for the number of transient events, and any event is allowed to be triggered by only one of the $W$ events leading up to it. The other is the number of event pairs $N$ considered in each of the three zones and sets a threshold on the number of significant event interactions that are identified. Full details on the association rule-based screening procedure are given in Section S2 of the online supplementary material.

### 5.6 Model evaluation

Let $\boldsymbol{X}^{(\text{train})}$ be the set of the training data on which the likelihood is based on, consisting of $n^{(\text{train})}$ events, and let $\boldsymbol{\psi}^{(1)}, \ldots, \boldsymbol{\psi}^{(R)}$ be $R$ samples from the posterior distribution. Denote by $\boldsymbol{X}^{(\text{test})}$ the set of held-out test data, consisting of $n^{(\text{test})}$ events.

One method to evaluate the predictive accuracy of each model is to use the log point-wise predictive density (Vehtari et al., 2017) computed on the test data, using the posterior samples

$$\widehat{lpd} = \sum_{(t,z,m)\in \boldsymbol{X}^{(\text{test})}} \log\left(\frac{1}{R}\sum_{r=1}^{R} L(t, z, m \mid \mathcal{F}_{t^-}, \boldsymbol{\zeta}^{(r)}, \boldsymbol{\eta}^{(r)}, \boldsymbol{\theta}^{(r)})\right), \tag{14}$$

where $L(t, z, m \mid \mathcal{F}_{t^-}, \boldsymbol{\zeta}^{(r)}, \boldsymbol{\eta}^{(r)}, \boldsymbol{\theta}^{(r)})$ is the likelihood of $(t, z, m)$ given the filtration $\mathcal{F}_{t^-}$ at the posterior sample $\boldsymbol{\zeta}^{(r)}, \boldsymbol{\eta}^{(r)}, \boldsymbol{\theta}^{(r)}$. Large values of $\widehat{lpd}$ indicate better predictive accuracy.

Apart from S$\beta$, V$\beta$, M$\beta$, and M$\beta$A, we also evaluate the predictive accuracy of two simpler baseline models that do not include Hawkes-like excitation effects. The first baseline model termed FOMC model is based on the factorisation of the likelihood of marked spatio-temporal processes in expression (7) with models for the times and locations as in equations (8) and (9), respectively, but with the conditional probability mass function for the marks being a first-order Markov chain

$$f(m_i \mid t_i, z_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\theta}) = \theta_{(z_i, m_{i-1})\to m_i}. \tag{15}$$

In this specification, $\theta_{(z,m)\to m'}$ is the probability of the event mark $m'$ given that last observed event has location $z$ and mark $m$. The second baseline model, termed MSTHP, is the marked spatio-temporal homogeneous Poisson process (Daley & Vere-Jones, 2003, Section 7.3), which has likelihood

$$\mathcal{L}^{(P)}(\boldsymbol{q} \mid \boldsymbol{\rho}) = \prod_{m=1}^{M}\prod_{z=1}^{Z} \rho_{mz}^{q_{mz}} \exp\{-T\rho_{mz}\}, \tag{16}$$

where $\rho_{mz}$ is the Poisson rate parameter and $q_{mz}$ is the number of event occurrences for mark $m$ at location $z$ over a total observation time $T$ in the data.

The FOMC and MSTHP models have conjugate prior distributions and therefore their posteriors are readily obtained. Details on those prior distributions and the derivation of their posterior distributions are given in Section S1 of the online supplementary material.

## 6 Explanatory modelling

### 6.1 Training

Samples from the posterior distributions for the parameters of the S$\beta$, V$\beta$, M$\beta$, and M$\beta$A models of Section 5.2, and of the FOMC and MSTHP baseline models of Section 5.6 are obtained using all event sequences from the first 20 games of the league season played between 17 August 2013 and 26 August 2013, which constitute $\boldsymbol{X}^{(\text{train})}$. The training data involves $S = 40$ game periods involving of 27,660 events. Each of the 20 teams participating in the league plays in two of the 20 games, one at their home and one at their away venue. As is also described in Section 2.1, there are $M = 30$ marks and $Z = 3$ zones. Table 5 gives the zone-wise event frequencies across marks in the training data used for modelling, reflecting the large variability in the frequencies both across marks and zones.

For the M$\beta$ and M$\beta$A models, the association rule learning screening procedure of Section 5.5 is employed for all combinations of $W \in \{5, 10\}$ and $N \in \{50, 100\}$ to eliminate some of the model parameters and reduce the model complexity before posterior sampling.

The hyper-parameters for the prior distributions specified in Section 5.3 are as follows. The hyper-parameters $a'$ and $b'$ for the exponential priors on the parameters of the Gamma distributions in equaiton (8) are both set to 0.01. The Dirichlet prior on the background mark probabilities $\boldsymbol{\delta}$ has concentration hyper-parameter $\delta' = 1$. The exponential prior on the decay rates $\boldsymbol{\beta}$ has a rate

**Table 5.** Zone-wise frequencies for each event type in the training data

| Home | | | | | Away | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Mark | | Zone | | | Mark | | Zone | | |
| *M* | Label | 1 | 2 | 3 | *M* | Label | 1 | 2 | 3 |
| 1 | Home_Win | 236 | 257 | 41 | 16 | Away_Win | 25 | 204 | 301 |
| 2 | Home_Dribble | 17 | 96 | 96 | 17 | Away_Dribble | 76 | 65 | 19 |
| 3 | Home_Pass_S | 1,699 | 4,633 | 1,658 | 18 | Away_Pass_S | 1427 | 4,390 | 2,030 |
| 4 | Home_Pass_U | 541 | 825 | 725 | 19 | Away_Pass_U | 542 | 811 | 702 |
| 5 | Home_Shot | 0 | 0 | 292 | 20 | Away_Shot | 193 | 2 | 0 |
| 6 | Home_Keeper | 155 | 0 | 0 | 21 | Away_Keeper | 0 | 0 | 192 |
| 7 | Home_Save | 106 | 0 | 0 | 22 | Away_Save | 0 | 0 | 149 |
| 8 | Home_Clear | 557 | 141 | 32 | 23 | Away_Clear | 27 | 124 | 660 |
| 9 | Home_Lose | 122 | 287 | 368 | 24 | Away_Lose | 323 | 349 | 142 |
| 10 | Home_Goal | 0 | 0 | 22 | 25 | Away_Goal | 20 | 0 | 0 |
| 11 | Home_Foul | 62 | 126 | 64 | 26 | Away_Foul | 46 | 112 | 69 |
| 12 | Home_Out_Throw | 97 | 184 | 163 | 27 | Away_Out_Throw | 143 | 173 | 110 |
| 13 | Home_Out_GK | 149 | 0 | 0 | 28 | Away_Out_GK | 0 | 0 | 220 |
| 14 | Home_Out_Corner | 0 | 0 | 112 | 29 | Away_Out_Corner | 76 | 0 | 0 |
| 15 | Home_Pass_O | 7 | 19 | 20 | 30 | Away_Pass_O | 13 | 11 | 5 |

hyper-parameter $\beta' = 0.1$. The Normal priors on the excitation factor $\alpha$ and the baseline-category logit model parameters $\boldsymbol{\phi}$ and $\boldsymbol{\omega}$ have hyper-parameters $\sigma_\alpha, \sigma_\gamma = 10$. The location-specific background mark probability vectors in the M$\beta$ and M$\beta$A models are assigned independent Dirichlet priors with concentration hyper-parameter $\delta'' = 1$.

The ability parameters in the baseline-category logit specification for the conversion rates of the M$\beta$A are not directly identifiable. In order to make them so, we set the abilities $\omega_{cm}$ for West Ham United to 0 ($m = 1, \ldots, 30$). Then, $\omega_{cm} > 0$ indicates that for team $c$, a previous event is more likely to trigger an event of mark $m$ when compared to the reference team.

Samples from the posterior distributions are obtained by running four parallel chains using the Hamiltonian Monte Carlo procedures implemented in Stan. The Stan templates we used are all provided in the online supplementary material. Each chain is initialised with different starting values and runs for a total of 500 iterations post the warm-up phase. Table 6 gives posterior summaries along with convergence diagnostics for some of the parameters of the M$\beta$A model with $W = 5$ and $N = 100$; the corresponding chain-wise trace plots are provided in the online supplementary material.

The convergence of the algorithm is assessed using the potential scale reduction factor $\hat{R}$ proposed by Gelman et al. (1992), which is the ratio of the average variance within each chain to the variance of the aggregated samples across chains. If the chains have converged to the stationary distribution, the expected value of $\hat{R}$ is 1. All parameters have $\hat{R} < 1.1$, which, as recommended in Gelman et al. (1992), is evidence for convergence. Table 6 also gives the effective sample size (see, for example, Gelman et al., 2013, Section 11.5) for the samples for each of the posterior marginals, which indicate that the sampler returned samples with acceptable autocorrelation. For some parameters the effective sample size is larger than the sample size due to negative autocorrelations. This, as pointed out in Vehtari et al. (2021), is a consequence of the Hamiltonian Monte Carlo algorithm used in Stan being an antithetic Markov chain which has negative autocorrelations on odd lags. The impact of the prior distributions in Section 5.3 on the posterior samples is minimal as seen, for a selection of parameters, in Figure 6 indicating that the posterior distributions of the parameters have concentrated after accounting for the likelihood.

**Table 6.** Posterior summaries and convergence diagnostics from 2,000 posterior samples for selected parameters from the M$\beta$A model after screening with $(W, N) = (5, 100)$

| Parameter | Mean | sd | $\hat{R}$ | $N^{(eff)}$ | Parameter | Mean | sd | $\hat{R}$ | $N^{(eff)}$ |
|---|---|---|---|---|---|---|---|---|---|
| $\beta_{3\to3\|1}$ | 0.52 | 0.04 | 1.00 | 1309.81 | $\phi_{3\to4\|2}$ | 1.81 | 0.41 | 1.01 | 540.41 |
| $\beta_{27\to8\|1}$ | 1.97 | 0.86 | 1.00 | 1953.81 | $\phi_{3\to5\|3}$ | 1.38 | 0.35 | 1.01 | 576.16 |
| $\beta_{24\to1\|2}$ | 1.51 | 0.09 | 1.00 | 2042.84 | $\phi_{3\to10\|3}$ | −1.31 | 0.59 | 1.01 | 1098.83 |
| $\beta_{3\to4\|2}$ | 0.65 | 0.03 | 1.01 | 913.52 | $\delta_{3\|1}$ | 0.56 | 0.03 | 1.00 | 1805.80 |
| $\beta_{3\to5\|3}$ | 0.63 | 0.04 | 1.01 | 1933.43 | $\delta_{3\|2}$ | 0.24 | 0.08 | 1.00 | 1356.85 |
| $\beta_{3\to10\|3}$ | 0.81 | 0.24 | 1.01 | 882.31 | $\delta_{3\|3}$ | 0.03 | 0.01 | 1.00 | 2207.39 |
| $\phi_{3\to3\|1}$ | 1.70 | 0.50 | 1.01 | 792.46 | $\alpha$ | 6.30 | 0.09 | 1.01 | 866.96 |
| $\phi_{27\to8\|1}$ | -0.74 | 0.87 | 1.00 | 2159.96 | $\omega_{9,3}$ | 0.59 | 0.26 | 1.03 | 334.15 |
| $\phi_{24\to1\|2}$ | 3.29 | 0.37 | 1.02 | 539.25 | $\omega_{10,3}$ | 0.67 | 0.25 | 1.01 | 341.24 |

## 6.2 Model evaluation

The S$\beta$, V$\beta$, M$\beta$, and M$\beta$A models are compared with each other and with the FOMC and MSTHP baseline models of Section 5.6 in terms of their log point-wise predictive density (14) computed on test data. The test data $X^{(\text{test})}$ includes all events from the five games immediately following the games in the training data, played between 31 August 2013 and 1 September 2013. The test data involves $S = 10$ game periods involving 27,660 events.

Table 7 gives the log predictive densities $\widehat{lpd}$, summed over all the events in the 10 game periods in the test data. Table 7 also provides the number of parameters in each model as a measure of their complexity. The three top-performing models are the M$\beta$ model after screening with $(W, N) = (5, 100)$, followed by M$\beta$ model after screening with $(W, N) = (10, 100)$, and M$\beta$A with $(W, N) = (5, 100)$. Notably, the M$\beta$ model after screening with $(W, N) = (5, 100)$ performs the best among the list of fitted models, significantly outperforming also models of similar complexity, such as the S$\beta$, V$\beta$, and FOMC models. The slightly poorer performance of the M$\beta$A model with $(W, N) = (5, 100)$ is most probably due to the fact that, in the training data, each team plays just one game at their home and one at their away venue. Nevertheless, in order to illustrate the full explanatory potential of the modelling framework in Section 4, we focus on inferences based on the posterior samples from the M$\beta$A model.

## 6.3 Background mark probabilities

Table 8 gives the posterior means of the background probabilities $\delta_{m|z}$ for all marks $m \in \{1, \ldots, 30\}$ and all locations $z \in \{1, 2, 3\}$. The background mark probabilities for the home and away team events in zone 1 are almost equal to those in zone 3 for the away and home team events, respectively. This is as expected because the attacking zone for the home team is the defensive zone for the away team and vice-versa.

The similar probabilities for the home and away background mark probabilities could indicate that the background process of the game is not influenced by home advantage. To confirm this, we fit another M$\beta$A model after constraining all the corresponding home and away background mark probabilities to be equal, for example, $\delta_{\text{Home\_Pass\_S}|1} = \delta_{\text{Away\_Pass\_S}|3}$, $\delta_{\text{Home\_Foul}|3} = \delta_{\text{Away\_Foul}|1}$, $\delta_{\text{Home\_Dribble}|2} = \delta_{\text{Away\_Dribble}|2}$ and so on. The constrained M$\beta$A model has 45 fewer parameters to be estimated as compared to the full M$\beta$A model.

The formal method to test our hypothesis is to calculate the Bayes factor, defined as the ratio of the marginal likelihood of the constrained M$\beta$A model to the marginal likelihood of the full M$\beta$A model. Then a Bayes factor greater than 1 would indicate that there is no evidence in favour of the full M$\beta$A model and therefore, the background mark probabilities do not capture home advantage. However, as a consequence of both M$\beta$A models being high-dimensional (~1,500 parameters), calculating their marginal likelihoods proved computationally infeasible.

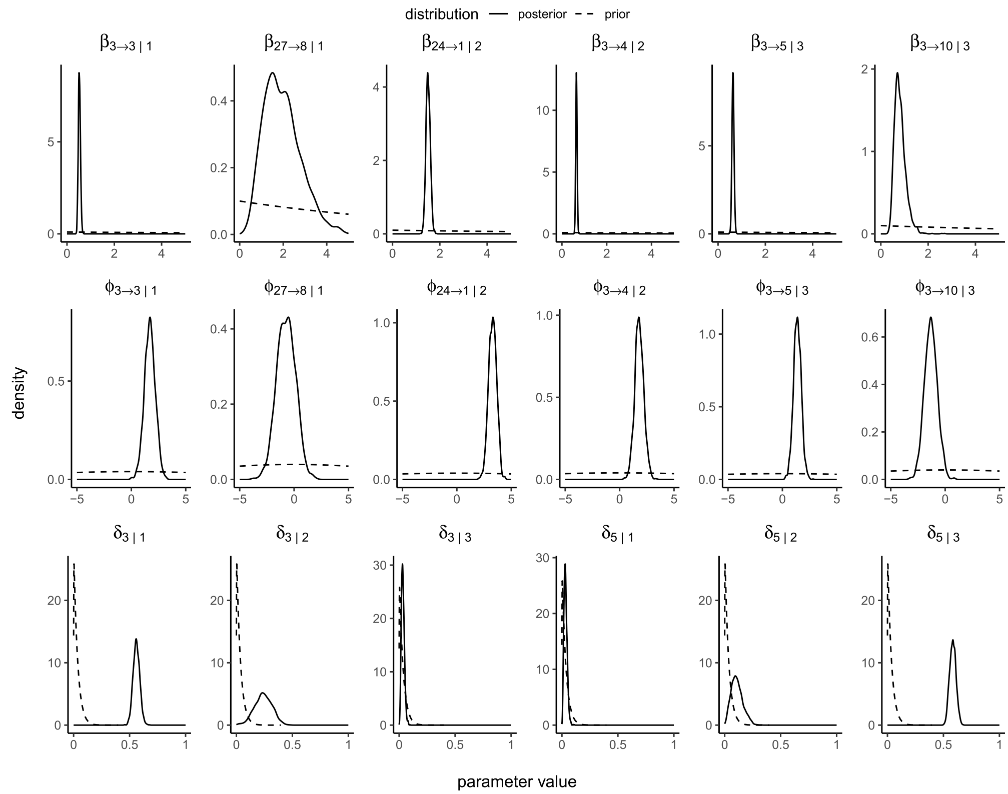



**Figure 6.** Visualising the impact of prior specifications by overlaying the posterior and prior densities for selected model parameters for the M$\beta$A model with $(W, N) = (5, 100)$.

**Table 7.** Cumulative log posterior densities $\widehat{lpd}$ over 10 game periods in the test data for all fitted models along with the number of estimated parameters $d^{(\text{par})}$ in each model

| Model | Abbreviation | $d^{(\text{par})}$ | $\widehat{lpd}$ |
|---|---|---|---|
| Homogeneous Poisson process (Baseline) | MSTHP | 90 | −35,469.50 |
| Matrix $\beta$ $(W, N) = (10, 50)$ | M$\beta$2 | 538 | −22,288.64 |
| Matrix $\beta$ $(W, N) = (5, 50)$ | M$\beta$1 | 538 | −22,152.08 |
| First-order Markov chain (Baseline) | FOMC | 870 | −21,898.31 |
| Scalar $\beta$ | S$\beta$ | 902 | −21,838.04 |
| Vector $\beta$ | V$\beta$ | 915 | −21,829.81 |
| Matrix $\beta$ with abilities $(W, N) = (5, 100)$ | M$\beta$A | 1539 | −21,599.81 |
| Matrix $\beta$ $(W, N) = (10, 100)$ | M$\beta$4 | 988 | −21,496.56 |
| Matrix $\beta$ $(W, N) = (5, 100)$ | M$\beta$3 | 988 | −21,342.57 |

*Note*. For the M$\beta$ models, $W$ is the number of transient events and $N$ is the number of significant event pairs identified in the rule-based framework for reducing model complexity.

As an alternative, for the constrained M$\beta$A model, we calculate its out-of-sample log predictive density on the same test data as carried out for all the other fitted models in Section 6.2. In fact, the constrained M$\beta$A model ($\widehat{lpd} = -21{,}589.28$) turns out with better predictive performance than the full M$\beta$A model ($\widehat{lpd} = -21{,}599.81$), supporting our claim that the background process of the game is not influenced by home advantage.

We also observe that the successful Pass events account for the majority of the background probability mass, while events like Shots and Goals have nearly zero probability. This suggests that the Shot and Goal events are highly unlikely to originate solely from the background component, but are instead triggered by excitations from previous events.

### 6.4 Excitation factor

The excitation factor $\alpha$ in expression (12) is a scaling factor applied to the contributions from the previous occurrences to the event mark probability. In equation (12), the background component has a weight of 1, while previous occurrences are weighted by $\exp(\alpha)$.

The 95% highest posterior density interval for $\exp(\alpha)$ is (451.35, 642.54), providing evidence that the contributions from previous occurrences carry substantially higher weight relative to the background component. In other words, this indicates that event sequences in football have a significant dependence on their history.

### 6.5 Decay rates

As mentioned in Section 4.2 the decay rate $\beta_{m\to m'|z}$ in expression (12) is the exponential decay rate of the excitation caused by an event of mark $m$ on an event of mark $m'$ at location $z$. By allowing the decay rates depend on the pair of marks involved in the excitation, we had hoped to account for scenarios like a Corner event exciting a Pass_S event in the short term and a Shot event in the longer term. Indeed, the 95% highest posterior density interval for $\beta_{\text{Home\_Corner}\to\text{Home\_Pass\_S}|3}$ is (1.34, 2.36) and $\beta_{\text{Home\_Corner}\to\text{Home\_Shot}|3}$ is (0.16, 0.44) illustrating that the Corner → Shot excitation decays at a much slower rate compared to the Corner → Pass_S excitation.

### 6.6 Conversion rates

The parameter $\gamma_{m\to m'|z}$ in expression (12) is the probability the excitation from an event of mark $m$ triggers an event of mark $m'$ at location $z$.

Table 9 gives the posterior means and standard deviations of $\gamma_{m\to m'|z}$ in the midfield region ($z = 2$) for Manchester United. The probabilities for the Home_Win → Home_Pass_S, Home_Dribble → Home_Pass_S and Home_Pass_S → Home_Pass_S conversions are higher compared to their away team counterparts, indicating that Manchester United is better at retaining possession of the ball when playing at home compared to away.

Figure 7 provides a ridge-line plot of the log-odds ratio for the home vs. the away ability of a team to convert a Win → Pass_S (Figure 7a) and Pass_S → Pass_S (Figure 7b). The teams are listed in decreasing order of the means of their respective posterior log-odds ratios which are indicated by vertical lines. The percentage values by each plot, indicate the fraction of the distribution greater than 0. All but two teams in Figure 7a and five teams in Figure 7b have greater than 50% of their distribution greater than 0, confirming that the vast majority of teams possess a higher ability to retain possession while playing at home.

In this way, we not only confirm the well-known home advantage effect, but also quantify team performance for games played at home as well as away.

### 6.7 Team abilities

Figure 8 provides a ridge-line plot of the posterior distribution of the parameters $\omega_{c,\text{Home\_Pass\_S}}$ and $\omega_{c,\text{Away\_Pass\_S}}$. The teams are listed in the decreasing order of the means of their respective posterior distributions which are indicated by vertical lines. We observe that Manchester United, the team with the highest ability to retain possession in home games (Figure 8a), drop significantly down in the rankings for the away games (Figure 8b). This is evidence that when Manchester United plays away they seem to deviate from the possession-based strategy they seem to adopt in the home games.

Figure 9a provides a ridge-line plot of the posterior distribution of the cumulative ability of a team to attempt a shot on goal. A higher $\omega_{c,\text{Home\_Shot}}$, for example, indicates that for the team $c$, an event like Home_Pass_S is more likely to trigger a Home_Shot. We do not expect the cumulative abilities $\omega_{c,\text{Home\_Shot}} + \omega_{c,\text{Away\_Shot}}$ of the dominant teams to be high, as they might prefer to make additional passes to create better goal-scoring opportunities. A weaker team, on the

**Table 8.** Posterior means and standard deviations (in parenthesis) of the zone dependent background mark probabilities $\delta_{m|z}$ for $z \in \{1, 2, 3\}$ from the M$\beta$A model

| Mark label | 1 | 2 | 3 | Mark label | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| Home_Win | . | . | . | Away_Win | . | . | . |
| Home_Dribble | . | 0.01 (0.01) | . | Away_Dribble | . | 0.02 (0.02) | . |
| Home_Pass_S | 0.56 (0.03) | 0.24 (0.08) | 0.03 (0.01) | Away_Pass_S | 0.03 (0.01) | 0.11 (0.05) | 0.58 (0.03) |
| Home_Pass_U | 0.06 (0.01) | 0.04 (0.02) | 0.03 (0.01) | Away_Pass_U | 0.02 (0.01) | 0.07 (0.04) | 0.05 (0.01) |
| Home_Shot | . | . | . | Away_Shot | . | . | . |
| Home_Keeper | 0.04 (0.01) | . | . | Away_Keeper | . | . | 0.05 (0.01) |
| Home_Save | . | . | . | Away_Save | . | . | . |
| Home_Clear | 0.05 (0.02) | . | . | Away_Clear | . | . | . |
| Home_Lose | . | 0.04 (0.02) | . | Away_Lose | 0.02 (0.01) | 0.07 (0.03) | 0.06 (0.01) |
| Home_Goal | . | . | . | Away_Goal | . | . | . |
| Home_Foul | 0.03 (0.01) | 0.09 (0.04) | 0.02 (0.01) | Away_Foul | 0.02 (0.01) | 0.06 (0.03) | . |
| Home_Out_Throw | . | . | . | Away_Out_Throw | . | . | . |
| Home_Out_GK | . | . | . | Away_Out_GK | . | . | . |
| Home_Out_Corner | . | . | . | Away_Out_Corner | . | . | . |
| Home_Pass_O | . | 0.08 (0.02) | . | Away_Pass_O | . | 0.05 (0.02) | . |

*Note*. The dots (·) denote means and standard deviations less than 0.005.

**Table 9.** Posterior means and standard deviations (in parenthesis) of the event conversion probabilities $\gamma_{m_j \to m_i|z_i}$ corresponding to the location $z = 2$ from the MβA model for Manchester United

| | Home_Win | Home_Dribble | Home_Pass_S | Home_Pass_U | Home_Shot | Home_Keeper | Home_Save | Home_Clear | Home_Lose | Home_Goal | Home_Foul | Home_Out_Throw | Home_Out_GK | Home_Out_Corner | Home_Pass_O | Away_Win | Away_Dribble | Away_Pass_S | Away_Pass_U | Away_Shot | Away_Keeper | Away_Save | Away_Clear | Away_Lose | Away_Goal | Away_Foul | Away_Out_Throw | Away_Out_GK | Away_Out_Corner | Away_Pass_O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Home_Win | . | . | **0.41 (0.1)** | 0.03 (0.02) | . | . | . | . | 0.01 (0.01) | . | . | . | . | . | . | 0.01 (0.02) | . | . | 0.03 (0.02) | . | . | . | 0.02 (0.02) | . | . | . | 0.44 (0.11) | . | . | . |
| Home_Dribble | . | . | **0.72 (0.14)** | 0.07 (0.05) | . | . | . | . | 0.05 (0.03) | . | . | . | . | . | . | 0.07 (0.09) | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.08 (0.05) |
| Home_Pass_S | . | . | **0.8 (0.06)** | 0.08 (0.01) | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.03 (0.04) | . | . | . | . | . | . | . | . | 0.02 (0.03) | . | . | . |
| Home_Pass_U | 0.02 (0.02) | . | 0.02 (0.02) | 0.01 (0.01) | . | . | . | 0.02 (0.02) | . | . | . | 0.02 (0.02) | . | . | . | 0.17 (0.04) | . | 0.22 (0.03) | 0.07 (0.01) | . | . | . | 0.15 (0.03) | . | . | . | 0.23 (0.05) | . | . | . |
| Home_Shot | . | . | 0.49 (0.24) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.31 (0.21) | . | . | . | . | . | . | . | . | . | . | . | 0.2 (0.12) |
| Home_Keeper | . | . | 0.4 (0.23) | 0.15 (0.14) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.15 (0.15) | . | . | . | 0.17 (0.16) | . | . | . | . | . | . | 0.13 (0.07) |
| Home_Save | . | . | 0.4 (0.23) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.4 (0.23) | . | . | . | . | . | . | . | . | . | . | . | 0.2 (0.12) |
| Home_Clear | . | . | 0.04 (0.03) | 0.02 (0.01) | . | . | . | . | . | . | . | 0.04 (0.03) | . | . | . | 0.02 (0.01) | . | 0.13 (0.03) | 0.04 (0.02) | . | . | . | . | 0.02 (0.02) | . | . | 0.64 (0.09) | . | . | . |
| Home_Lose | 0.02 (0.02) | . | 0.02 (0.02) | 0.01 (0.01) | . | . | . | 0.02 (0.02) | . | . | . | 0.02 (0.02) | . | . | . | 0.59 (0.08) | . | 0.07 (0.02) | . | . | . | . | . | 0.1 (0.03) | . | . | 0.1 (0.04) | . | . | . |
| Home_Goal | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.55 (0.22) | . | . | . | . | . | . | . | . | . | . | . | 0.45 (0.22) |
| Home_Foul | . | . | | | . | . | . | . | . | . | . | . | . | . | . | . | . | . | | . | . | . | . | . | . | . | . | . | . | |

(*continued*)

**Table 9.** Continued

| | Home_Win | Home_Dribble | Home_Pass_S | Home_Pass_U | Home_Shot | Home_Keeper | Home_Save | Home_Clear | Home_Lose | Home_Goal | Home_Foul | Home_Out_Throw | Home_Out_GK | Home_Out_Corner | Home_Pass_O | Away_Win | Away_Dribble | Away_Pass_S | Away_Pass_U | Away_Shot | Away_Keeper | Away_Save | Away_Clear | Away_Lose | Away_Goal | Away_Foul | Away_Out_Throw | Away_Out_GK | Away_Out_Corner | Away_Pass_O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0.48 (0.15) | 0.24 (0.11) | | | | | | | | | | | | | | | 0.09 (0.09) | | | | | | | | | | | 0.19 (0.1) |
| Home_Out_Throw | . | . | 0.86 (0.05) | 0.09 (0.03) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| Home_Out_GK | . | . | . | 0.06 (0.05) | . | . | . | . | 0.07 (0.09) | . | 0.04 (0.05) | 0.25 (0.2) | . | . | . | . | . | . | 0.14 (0.13) | . | . | . | 0.09 (0.06) | 0.07 (0.08) | . | . | 0.2 (0.18) | . | . | 0.07 (0.03) |
| Home_Out_Corner | . | . | 0.7 (0.18) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.3 (0.18) |
| Home_Pass_O | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.4 (0.19) | 0.25 (0.15) | . | . | . | . | . | . | . | . | . | . | 0.35 (0.16) |
| Away_Win | 0.05 (0.04) | . | . | 0.01 (0.01) | . | . | . | 0.02 (0.02) | 0.03 (0.02) | . | . | 0.59 (0.1) | . | . | . | . | . | **0.22** **(0.07)** | 0.02 (0.01) | . | . | . | . | . | . | . | . | . | . | 0.03 (0.01) |
| Away_Dribble | 0.31 (0.24) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | **0.49** **(0.23)** | 0.06 (0.05) | . | . | . | . | 0.03 (0.03) | . | . | . | . | . | 0.11 (0.06) |
| Away_Pass_S | 0.03 (0.04) | . | 0.06 (0.09) | 0.02 (0.02) | . | . | . | 0.02 (0.03) | 0.02 (0.03) | . | . | 0.06 (0.09) | . | . | . | . | . | **0.68** **(0.11)** | 0.06 (0.01) | . | . | . | . | . | . | . | . | . | . | . |
| Away_Pass_U | 0.25 (0.05) | . | 0.2 (0.03) | 0.07 (0.01) | . | . | . | 0.14 (0.03) | . | . | . | 0.28 (0.05) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| Away_Shot | . | . | 0.44 (0.22) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.36 (0.21) | . | . | . | . | . | . | . | . | . | . | . | 0.2 (0.12) |
| Away_Keeper | . | . | . | 0.16 (0.15) | . | . | . | 0.19 (0.16) | . | . | . | . | . | . | . | . | . | 0.42 (0.22) | 0.11 (0.11) | . | . | . | . | . | . | . | . | . | . | 0.13 (0.07) |
| Away_Save | . | . | 0.47 (0.22) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.31 (0.2) | . | . | . | . | . | . | . | . | . | . | . | 0.22 (0.13) |

(*continued*)

**Table 9.** Continued

| | Home_Win | Home_Dribble | Home_Pass_S | Home_Pass_U | Home_Shot | Home_Keeper | Home_Save | Home_Clear | Home_Lose | Home_Goal | Home_Foul | Home_Out_Throw | Home_Out_GK | Home_Out_Corner | Home_Pass_O | Away_Win | Away_Dribble | Away_Pass_S | Away_Pass_U | Away_Shot | Away_Keeper | Away_Save | Away_Clear | Away_Lose | Away_Goal | Away_Foul | Away_Out_Throw | Away_Out_GK | Away_Out_Corner | Away_Pass_O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Away_Clear | 0.02 (0.01) | . | 0.15 (0.04) | 0.05 (0.02) | . | . | . | 0.02 (0.01) | . | . | . | 0.69 (0.07) | . | . | . | . | . | 0.03 (0.02) | . | . | . | . | . | . | . | . | 0.01 (0.01) | . | . | . |
| Away_Lose | 0.69 (0.06) | . | 0.05 (0.01) | . | . | . | . | . | 0.08 (0.02) | . | . | 0.1 (0.03) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| Away_Goal | . | . | 0.73 (0.17) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.27 (0.17) |
| Away_Foul | . | . | . | 0.13 (0.12) | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.5 (0.15) | 0.18 (0.09) | . | . | . | . | . | . | . | . | . | . | 0.18 (0.1) |
| Away_Out_Throw | 0.02 (0.02) | . | 0.02 (0.02) | . | . | . | . | 0.01 (0.01) | . | . | . | 0.01 (0.02) | . | . | . | . | . | 0.86 (0.06) | 0.05 (0.02) | . | . | . | . | . | . | . | . | . | . | . |
| Away_Out_GK | . | . | . | 0.09 (0.08) | . | . | . | 0.14 (0.08) | 0.13 (0.12) | . | . | 0.19 (0.14) | . | . | . | . | . | . | 0.11 (0.08) | . | . | . | . | 0.1 (0.09) | . | 0.06 (0.06) | 0.1 (0.09) | . | . | 0.08 (0.03) |
| Away_Out_Corner | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.6 (0.2) | . | . | . | . | . | . | . | . | . | . | . | 0.4 (0.2) |
| Away_Pass_O | . | . | 0.74 (0.15) | 0.1 (0.08) | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 0.16 (0.11) |

*Note*. The $\gamma_{m \to m'|z}$s are computed by setting the team identifier $c = 11$ (corresponding to Manchester United), for both the home as well as away events. The highlighted cells (in bold) illustrate the superior ability of Manchester United when playing at home compared to away. The dots ($\cdot$) denote means and standard deviations less than 0.005.

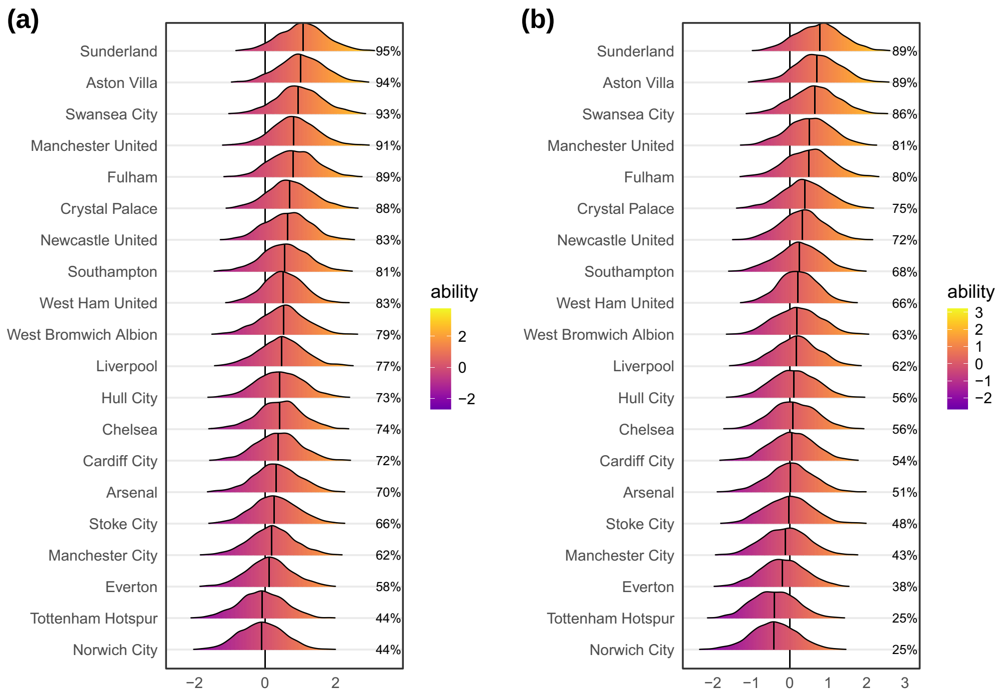


**Figure 7.** Posterior distribution of $\phi_{\text{Home\_Win}\rightarrow\text{Home\_Pass\_S}|2} + \omega_{c,\text{Home\_Pass\_S}} - \phi_{\text{Away\_Win}\rightarrow\text{Away\_Pass\_S}|2} - \omega_{c,\text{Away\_Pass\_S}}$ in (a) and $\phi_{\text{Home\_Pass\_S}\rightarrow\text{Home\_Pass\_S}|2} + \omega_{c,\text{Home\_Pass\_S}} - \phi_{\text{Away\_Pass\_S}\rightarrow\text{Away\_Pass\_S}|2} - \omega_{c,\text{Away\_Pass\_S}}$ in (b), from the baseline logit specification for incorporating team abilities in equation (13). Interpreted as the relative ability of a team to convert a Win to a Successful Pass when playing at home compared to away in (a) and similarly from one Successful Pass to another Successful Pass in (b). Teams are ranked in the decreasing order of the means of their respective posterior distributions shown by the overlaid vertical lines.

other hand typically has fewer opportunities to attack and therefore, is more likely to attempt a shot on goal when possible. Indeed, this is what we observe in Figure 9, where we compare the team rankings based on their cumulative ability $\omega_{c,\text{Home\_Shot}} + \omega_{c,\text{Away\_Shot}}$ with the number of shots per pass completed in the attacking third (S/P column in Figure 9b) in the training data. The comparison between Cardiff City and Norwich City is an interesting example of two teams that appear to be similar with 18 and 19 shots on goal attempted, respectively, in their two games in the training data. However, the two teams are at opposite ends of the ranking based on their cumulative ability $\omega_{c,\text{Home\_Shot}} + \omega_{c,\text{Away\_Shot}}$, capturing the clear difference between their attacking styles.

Column (a) in Table 10 shows the team rankings based on the cumulative ability to trigger five different event types. For example, the Pass column ranks teams in the decreasing order of their posterior means of $\omega_{c,\text{Home\_Pass\_S}} + \omega_{c,\text{Away\_Pass\_S}}$. The teams are ordered in Column (a) in Table 10 by the rankings based on their cumulative passing ability. Despite training on just the first 20 out of 380 games of the 2013/2014 season, the rankings based on the passing ability is a good indicator of the positions the teams finished in the final league table of the 2013/2014 season in Column (b) in Table 10.

## 6.8 Event genealogy

The branching structure $u_{si}$ indicates whether the $i$th event in $s$th sequence is an 'immigrant' ($u_{si} = 0$) or an 'offspring' of a previous event with index $j$ ($u_{si} = j$). Given an observed event sequence $\mathcal{F}_{st_{sn_s}}$, the conditional branching structure probabilities $\mathbb{P}(u_{si} \mid \mathcal{F}_{st_{si}})$ based on the model specification in expression (12) are

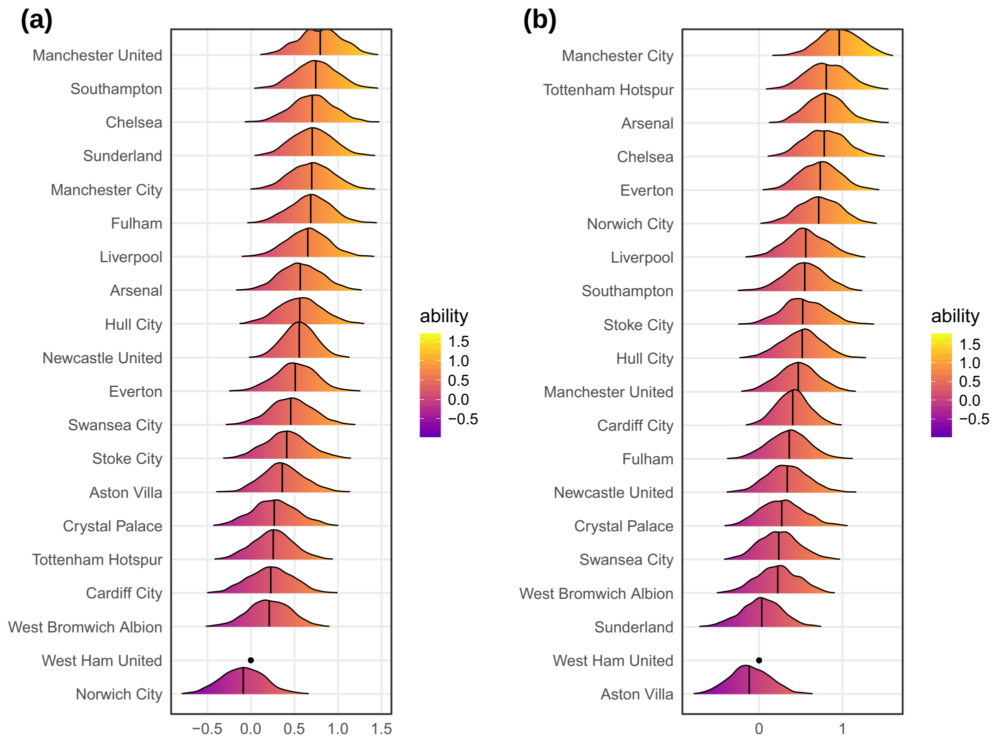
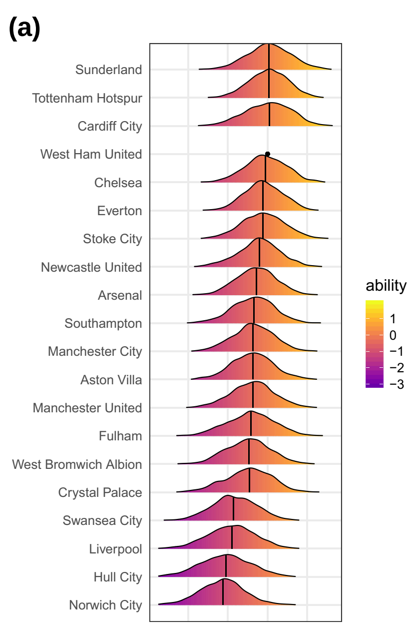



**Figure 8.** Posterior distribution of the parameters $\omega_{c,\text{Home\_Pass\_S}}$ in (a) and $\omega_{c,\text{Away\_Pass\_S}}$ in (b), from the baseline logit specification for incorporating team abilities in expression (13). Teams are ranked in the decreasing order of the means of their respective posterior distributions shown by the overlaid vertical lines.




(b)

| Team | Shots | Passes | S/P |
|---|---|---|---|
| Cardiff City | 18 | 115 | 0.16 |
| Sunderland | 28 | 195 | 0.14 |
| Tottenham Hotspur | 37 | 261 | 0.14 |
| Aston Villa | 20 | 153 | 0.13 |
| Newcastle United | 22 | 180 | 0.12 |
| Crystal Palace | 18 | 148 | 0.12 |
| Stoke City | 24 | 199 | 0.12 |
| Everton | 40 | 342 | 0.12 |
| Chelsea | 29 | 248 | 0.12 |
| Fulham | 19 | 163 | 0.12 |
| West Ham United | 22 | 194 | 0.11 |
| West Bromwich Albion | 18 | 159 | 0.11 |
| Swansea City | 23 | 214 | 0.11 |
| Arsenal | 30 | 280 | 0.11 |
| Southampton | 23 | 223 | 0.10 |
| Liverpool | 28 | 292 | 0.10 |
| Manchester United | 22 | 238 | 0.09 |
| Manchester City | 30 | 330 | 0.09 |
| Norwich City | 19 | 227 | 0.08 |
| Hull City | 13 | 191 | 0.07 |

**Figure 9.** (a) Posterior distribution of $\omega_{c,\text{Home\_Shot}} + \omega_{c,\text{Away\_Shot}}$, the cumulative ability of a team $c$, relative to West Ham (baseline), to attempt a shot on goal. (b) The number of shots and passes completed in the attacking third and shots per pass completed in the attacking third (S/P) for each team in the training data.

**Table 10.** Team rankings based on the cumulative ability to trigger a particular event type

| (a) | | | | | | (b) | |
|---|---|---|---|---|---|---|---|
| Team | Pass | Shot | Goal | Win | Save | League Position | Team |
| Manchester City | 1 | 11 | 1 | 15 | 11 | 1 | Manchester City |
| Chelsea | 2 | 5 | 11 | 20 | 4 | 2 | Liverpool |
| Arsenal | 3 | 9 | 3 | 5 | 7 | 3 | Chelsea |
| Southampton | 4 | 10 | 8 | 7 | 19 | 4 | Arsenal |
| Manchester United | 5 | 13 | 4 | 2 | 18 | 5 | Everton |
| Everton | 6 | 6 | 17 | 11 | 14 | 6 | Tottenham Hotspur |
| Liverpool | 7 | 18 | 12 | 9 | 5 | 7 | Manchester United |
| Hull City | 8 | 19 | 15 | 1 | 10 | 8 | Southampton |
| Tottenham Hotspur | 9 | 2 | 14 | 3 | 6 | 9 | Stoke City |
| Fulham | 10 | 14 | 7 | 14 | 3 | 10 | Newcastle United |
| Stoke City | 11 | 7 | 9 | 12 | 8 | 11 | Crystal Palace |
| Newcastle United | 12 | 8 | 19 | 16 | 17 | 12 | Swansea City |
| Sunderland | 13 | 1 | 13 | 8 | 2 | 13 | West Ham United |
| Swansea City | 14 | 17 | 18 | 17 | 12 | 14 | Sunderland |
| Cardiff City | 15 | 3 | 2 | 19 | 15 | 15 | Aston Villa |
| Norwich City | 16 | 20 | 10 | 4 | 20 | 16 | Hull City |
| Crystal Palace | 17 | 16 | 16 | 13 | 16 | 17 | West Bromwich Albion |
| West Bromwich Albion | 18 | 15 | 20 | 10 | 1 | 18 | Norwich City |
| Aston Villa | 19 | 12 | 5 | 6 | 13 | 19 | Fulham |
| West Ham United | 20 | 4 | 6 | 18 | 9 | 20 | Cardiff City |

*Note*. For example, the column Pass ranks teams in the decreasing order of their posterior means of $\omega_{c,\text{Home\_Pass\_S}} + \omega_{c,\text{Away\_Pass\_S}}$. The teams are ordered in (a) by the rankings based on their passing ability, which is a good indicator of the final position in the league table of the 2013/2014 season in (b).

$$\mathbb{P}(u_{si} = 0 \mid \mathcal{F}_{st_{si}}) = \frac{\delta_{m_{si}|z_{si}}}{\delta_{m_{si}|z_{si}} + \sum_{t_{sk} < t_{si}} e^{\alpha - \beta_{m_{sk} \to m_{si}|z_{si}}(t_{si} - t_{sk})} \gamma_{m_{sk} \to m_{si}|z_{si}}},$$
$$\mathbb{P}(u_{si} = j \mid \mathcal{F}_{st_{si}}) = \begin{cases} \dfrac{e^{\alpha - \beta_{m_{sj} \to m_{si}|z_{si}}(t_{si} - t_{sj})} \gamma_{m_{sj} \to m_{si}|z_{si}}}{\delta_{m_{si}} + \sum_{t_{sk} < t_{si}} e^{\alpha - \beta_{m_{sk} \to m_{si}|z_{si}}(t_{si} - t_{sk})} \gamma_{m_{sk} \to m_{si}|z_{si}}} & \text{for } t_{sj} < t_{si} \\ 0 & \text{for } t_{sj} \geq t_{si}. \end{cases} \tag{17}$$

The branching structure probabilities in equation (17) quantify the relative contributions of the background process and previous occurrences in the mark probability of the $i$th event in the $s$th sequence. Figure 10 shows the posterior means of the branching structure probabilities for all events in the first 4 min of the game between Chelsea and Hull City on 18 August 2013. To illustrate the flexibility of the model to account for dependence between events over arbitrary durations of time, we highlight the event Home_Shot showing a higher probability of being an offspring of the event Home_Out_Corner than being an offspring of the more recent Home_Pass_S event.

## 7 Model-based predictions

Finally, we illustrate how the mechanistic modelling framework presented in this paper can be used to simulate event sequences in football and obtain predictions of event probabilities in real time. We split the game between Arsenal and Tottenham Hotspur (1 September 2013) in the test data into 30-s intervals. For each interval, given the history of events up to but not including

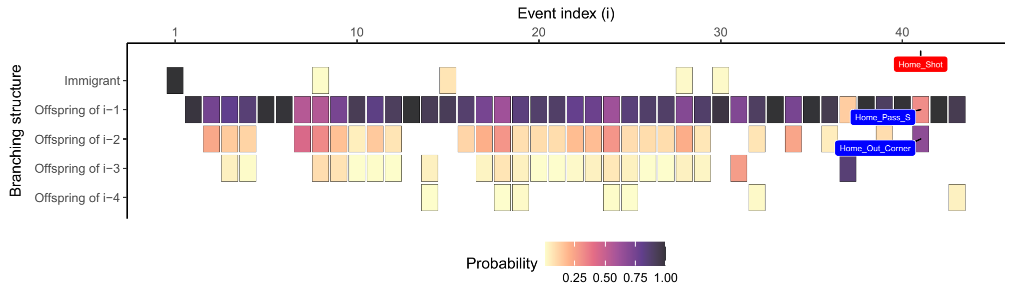



**Figure 10.** Posterior means of branching structure probabilities for events in the first 4 min of the game between Chelsea and Hull City on 18 August 2013. The highlighted event Home_Shot has a higher probability of being an offspring of the event Home_Out_Corner than being an offspring of the more recent Home_Pass_S event.

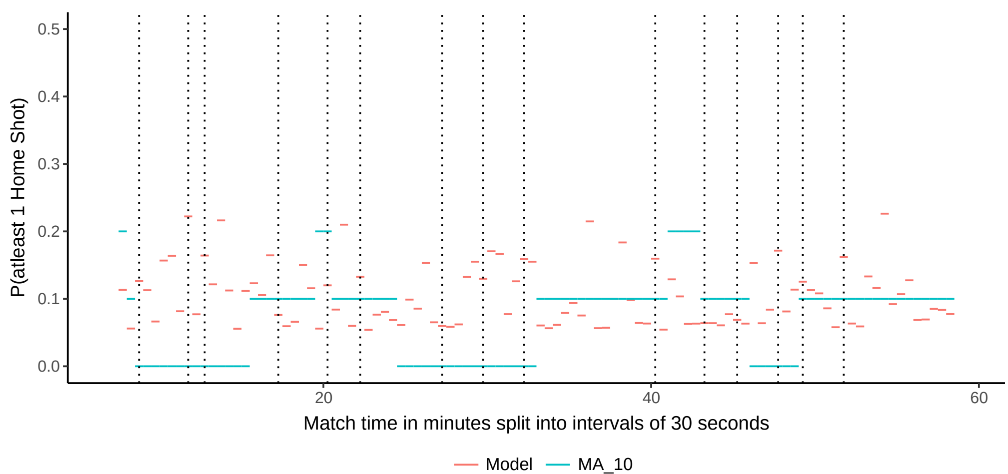



**Figure 11.** Forecasting the probability of observing at least one Home Shot event in 30-s intervals during the game between Arsenal and Tottenham Hotspur (1 September 2013) in the test data. Intervals with observed Home Shot events are highlighted using dotted lines. MA_10 is a 10-step moving average model used as a benchmark for comparison.

the interval, we simulate events over the next 30 s $Q = 100$ times for each of the $R = 500$ posterior samples from the M$\beta$A model with the tuning parameter setting ($W = 5$, $N = 100$).

In Figure 11, we plot the proportion of all simulations within each interval where at least one Home_Shot event was simulated, and use dotted lines to denote the intervals where a Home_Shot event was actually observed. We also include a 10-step moving average model (MA_10) as a benchmark for comparison. We excluded the first 5 min of the game to ensure that we have predictions from both the models being compared. A quick inspection reveals that in 11 of the 15 intervals in which a Home_Shot is observed, the model predicts a shot probability greater than the 10-step moving average model.

In Figure 12, we formally validate the performance of the model against three moving average models for the classification task of whether a shot will be observed in an interval. For this purpose, we use data from the first 20 games in the test set, where we excluded the first 15 intervals of each game to ensure that we have predictions from all four models being compared. To validate the models, we had a total of 1,959 intervals out of which 202 intervals had at least one Home_Shot event. The area under the receiver operating characteristic (ROC) curve is a performance measure that evaluates the performance of a classification model across all classification thresholds. The area under the curve (AUC) values are given in the legend and clearly confirm the superior performance of the model.

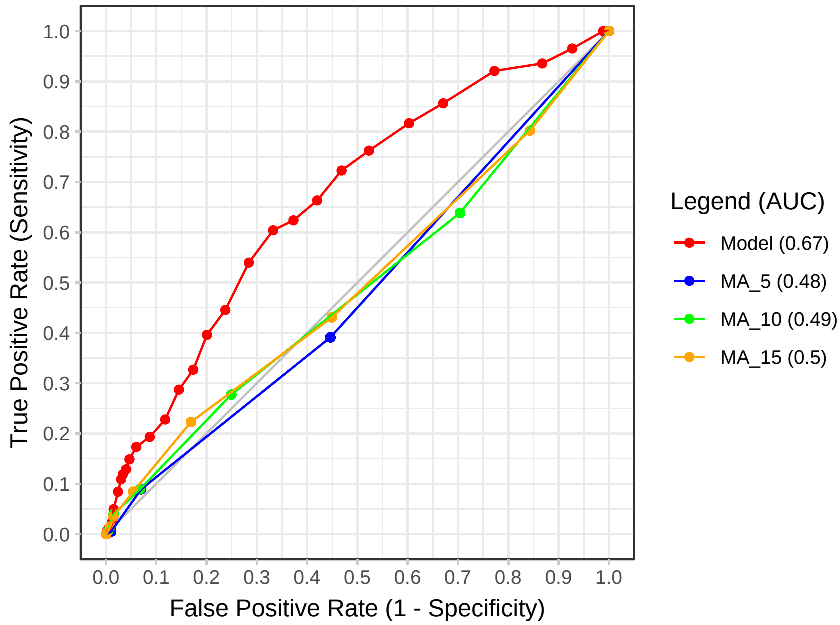



**Figure 12.** Validating the model performance against three moving average models for the task of whether a shot will be observed in each 30-s interval over the first 20 games in the test set. MA_5 is a five-step moving average model and so on. The ROC curve evaluates the performance of a classification model across all classification thresholds. The area under the curve (AUC) values in the legend clearly confirm the superior performance of the model.

## 8 Discussion and concluding remarks

Building on the decomposition of a multivariate distribution function, we showed how the joint modelling in classical point process models like Hawkes processes can be decoupled. The introduced flexible modelling framework can, for example, retain the characteristic property of excitation in Hawkes processes in the model for the marks while avoiding the clustering of event times. A comprehensive Bayesian approach for the modelling of flexible marked spatio-temporal point processes was developed including an approach to evaluate the predictive accuracy of the fitted Bayesian models using the out-of-sample log predictive density.

We presented a case study showing how the modelling framework developed in this paper can be tailored to separately model the components of the events in football, namely, the times, the locations and the event types. We were also able to incorporate team information into the model in a direct way that captured the relative abilities of the teams for each event type. We developed a method based on association rules to reduce the increased model complexity introduced by model extensions. The rule-based approach identified significant event interactions within sequences by placing thresholds on measures of significance. We then evaluated the accuracy of the excitation-based models by comparing them against two baseline models and confirmed the superior performance of the models with excitation effects.

We provided a detailed parameter description showing how the model parameters can be used to gain valuable insight into football. The excitation framework of the best-performing model captured both the magnitudes and the durations of all pairwise event interactions across different locations. From the conversion rate parameters, we were able to quantify the well-known home advantage effect. We also discussed how the team ability parameters can be used to obtain rankings for the teams by event type, which can be used as predictors for team performance. The team ability parameters also captured some interesting differences in the playing styles of the teams that were not immediately apparent just by looking at the data. In this way, the model along with its parameters can be used to develop a deeper understanding of the game-play by the coaching staff and inform strategic decision-making. The proposed model can also be used to simulate the sequence of events in a game to obtain real-time predictions of event probabilities. We believe these predictions would enhance, among other aspects, the viewing experience of televised games.

The data set we used consists of events from a single English Premier League season, which has a total of 380 games. However, as described in Section 6.1, we only used the first 20 games of the season as training data for the modelling exercise, over which every team in the league plays exactly one game each at Home and Away venues. This represents the minimum number of games required to ensure the identifiability of all model parameters, specifically the team abilities. Even though the volume of data was kept to a minimum for computational reasons, our results illustrate that the model can provide valuable insights with limited data. So, the methodology developed in this paper can be readily applied to other team sports like rugby, hockey, basketball, American football, etc., where there may be fewer events per game or fewer games in a season.

Multiple seasons can be modelled together as if they were just one season using the modelling framework we propose, as long as the game rules, and hence, the definition of the events being considered does not change. Another aspect of the tournament to note is the relegation and promotion of teams within the league, which results in some teams not playing the same number of games over multiple seasons. A limitation of the proposed model is that the game periods are exchangeable, because the likelihood is invariant to the order in which the game periods and the games occur. It would be more natural to allow for the team ability parameters in equation (6) to be time-varying, especially over multiple seasons during which team players and managers are likely to change. Due to computational reasons, we were not able to utilise most of the data even within a single season and current work focuses on overcoming this computational barrier using variational inference (Blei et al., 2017).

As none of the methods have been tailored specifically to football or even sports for that matter, they can be applied to a wide range of applications that generate event data streams. Specifically, the conversion rate parameters can be used to capture the triggering structure between different event types, for example, the probability of large earthquakes triggering smaller aftershocks. Also, the team ability parameters can be used for other multi-agent environments, for example, accidents by car type, countries in the analyses of financial events, individuals in identity systems, and so on.

## Acknowledgments

The authors thank Stratagem Technologies Ltd for giving access to all touch-ball events from all English Premier League games in the 2013/2014 season.

*Conflict of interest:* None declared.

## Data availability

The raw data that motivated this work has been provided by Stratagem Technologies Ltd and consists of all touch-ball events for the 2013/2014 season of the English Premier League. The raw data cannot be disclosed as the authors do not have the license to do so. The GitHub repository https://github.com/ForeStats/flexible-msttp-football provides all the computer code used for the data preprocessing and the Stan templates used to carry out the analyses presented in this paper. We also provide guidance on how the code can be used to apply our methods to similar data sets, like the publicly available 2020/2021 FA Women's Super League Data provided by StatsBomb Inc. at https://github.com/statsbomb/open-data. Section S1 of the online supplementary material provides details on the prior distributions and the derivation of their posterior distributions for the FOMC and MSTHP models. Section S2 of the online supplementary material gives details on the association rule-based screening procedure used to identify the most significant event interactions. We also provide the chain-wise trace plots for some of the parameters of the M$\beta$A model with $W = 5$ and $N = 100$.

## Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series C*.

## References


Agrawal R., Imieliński T., & Swami A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, *22*(2), 207–216. https://doi.org/10.1145/170036.170072

Agresti A. (2007). *An introduction to categorical data analysis* (2nd ed.). Wiley.

Blei D. M., Kucukelbir A., & McAuliffe J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Borrie A., Jonsson G. K., & Magnusson M. S. (2002). Temporal pattern analysis and its applicability in sport: An explanation and exemplar data. *Journal of Sports Sciences*, *20*(10), 845–852. https://doi.org/10.1080/026404102320675675

Bowsher C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, *141*(2), 876–912. https://doi.org/10.1016/j.jeconom.2006.11.007

Clemente F. M., Couceiro M. S., Martins F. M. L., & Mendes R. S. (2015). Using network metrics in soccer: A macro-analysis. *Journal of Human Kinetics*, *45*(1), 123–134. https://doi.org/10.1515/hukin-2015-0013

Cox D. R. (1975). Partial likelihood. *Biometrika*, *62*(2), 269–276. https://doi.org/10.1093/biomet/62.2.269

Daley D. J., & Vere-Jones D. (2003). *An introduction to the theory of point processes* (2nd ed., Vol. *I*). Springer-Verlag.

Decroos T., Bransen L., Van Haaren J., & Davis J. (2018). Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD international conference on Knowledge Discovery & Data Mining (KDD'19)* (pp. 1851–1861). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3292500.3330758

Decroos T., Dzyuba V., Haaren J. V., & Davis J. (2017). Predicting soccer highlights from spatio-temporal match event streams. *Proceedings of the AAAI Conference on Artificial Intelligence*, *31*(1), 1302–1308. https://doi.org/10.1609/aaai.v31i1.10754

Diggle P. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *34*(2), 138–147. https://doi.org/10.2307/2347366

Diggle P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns* (3rd ed.). CRC Press.

Duane S., Kennedy A. D., Pendleton B. J., & Roweth D. (1987). Hybrid Monte Carlo. *Physics Letters B*, *195*(2), 216–222. https://doi.org/10.1016/0370-2693(87)91197-X

Duch J., Waitzman J. S., & Amaral L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PLOS One*, *5*(6), e10937. https://doi.org/10.1371/journal.pone.0010937

Gelman A., Carlin J., Stern H., Dunson D., Vehtari A., & Rubin D. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.

Gelman A., & Rubin D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

González J. A., Rodríguez-Cortés F. J., Cronie O., & Mateu J. (2016). Spatio-temporal point process statistics: A review. *Spatial Statistics*, *18*(Part B), 505–544. https://doi.org/10.1016/j.spasta.2016.10.002

Grund T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, *34*(4), 682–690. https://doi.org/10.1016/j.socnet.2012.08.004

Gudmundsson J., & Horton M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys (CSUR)*, *50*(2), 1–34. https://doi.org/10.1145/3054132

Hawkes A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90. https://doi.org/10.1093/biomet/58.1.83

Hawkes A. G., & Oakes D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, *11*(3), 493–503. https://doi.org/10.2307/3212693

Hoffman M. D., & Gelman A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(47), 1593–1623. http://jmlr.org/papers/v15/hoffman14a.html

Lindqvist B. H. (2006). On the statistical modeling and analysis of repairable systems. *Statistical Science*, *21*(4), 532–551. https://doi.org/10.1214/088342306000000448

Mackay N. (2017). *Predicting goal probabilities for possessions in football* [Master's thesis]. Vrije Universiteit Amsterdam.

Mohler G. O., Short M. B., Brantingham P. J., Schoenberg F. P., & Tita G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100–108. https://doi.org/10.1198/jasa.2011.ap09546

Narayanan S. (2021). *Bayesian modelling of flexible marked point processes with applications to event sequences from association football* [Ph.D. thesis]. University of Warwick, Coventry. http://webcat.warwick.ac.uk/record=b3520069~S15

Norris J. R. (1997). *Markov chains*. Cambridge University Press.

Ogata Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402. https://doi.org/10.1023/A:1003403601725

Passos P., Davids K., Araújo D., Paz N., Minguéns J., & Mendes J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport*, *14*(2), 170–176. https://doi.org/10.1016/j.jsams.2010.10.459

Pena J. L., & Touchette H. (2012). 'A network theory analysis of football strategies', arXiv, arXiv:1206.6904, preprint: not peer reviewed.

Rasmussen J. G. (2013). Bayesian inference for Hawkes processes. *Methodology and Computing in Applied Probability*, *15*(3), 623–642. https://doi.org/10.1007/s11009-011-9272-5

Ripley B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(2), 172–192. https://doi.org/10.1111/j.2517-6161.1977.tb01615.x

Robberechts P., Van Haaren J., & Davis J. (2019). 'Who will win it? An in-game win probability model for football', arXiv, arXiv:1906.05029, preprint: not peer reviewed.

Routley K., & Schulte O. (2015). A Markov game model for valuing player actions in ice hockey. In *Proceedings of the thirty-first conference on Uncertainty in Artificial Intelligence (UAI'15)* (pp. 782–791). AUAI Press, Arlington, VA, USA. https://dl.acm.org/doi/10.5555/3020847.3020928

Stan Development Team. (2020). *CmdStan: The command-line interface to Stan*. Version 2.22.1.

Van Haaren J., Hannosset S., & Davis J. (2016). Strategy discovery in professional soccer match data. In *Proceedings of the KDD-16 workshop on large-scale sports analytics* (pp. 1–4). https://lirias.kuleuven.be/retrieve/395200

Veen A., & Schoenberg F. P. (2008). Estimation of space–time branching process models in seismology using an em-type algorithm. *Journal of the American Statistical Association*, *103*(482), 614–624. https://doi.org/10.1198/016214508000000148

Vehtari A., Gelman A., & Gabry J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vehtari A., Gelman A., Simpson D., Carpenter B., & Bürkner P.-C. (2021). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, *16*(2), 667–718. https://doi.org/10.1214/20-BA1221

Wang Q., Zhu H., Hu W., Shen Z., & Yao Y. (2015). Discerning tactical patterns for professional soccer teams: An enhanced topic model with applications. In *Proceedings of the 21th ACM SIGKDD international conference on Knowledge Discovery and Data mining (KDD'15)* (pp. 2197–2206). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2783258.2788577

*Journal of the Royal Statistical Society Series C: Applied Statistics*, 2023, **72**, 1127–1138

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS Journal of the Statistics Society Series C Applied Statistics

**Discussion Paper Contribution**

# Proposer of the vote of thanks to Narayanan, Kosmidis and Dellaportas and contribution to the Discussion of 'Flexible marked spatio-temporal point processes with applications to event sequences from association football'

**Dimitris Karlis**

Department of Statistics, Athens University of Economics and Business, Athens, Greece

*Address for correspondence:* Dimitris Karlis, Department of Statistics, Athens University of Economics and Business, 76 Patission Street, 10434 Athens, Greece. Email: karlis@aueb.gr

I am very happy to discuss a valuable and innovative contribution on modelling soccer data. Recent advances in technology have provided a wealth of data from a soccer match. The paper, for the first time, combines such data in a novel way. In particular, the authors combine event data (data about particular events in soccer, like passes, fouls etc.) and spatial data about the location of players and the ball.

To summarise the paper, it offers an interesting modelling approach for in-game event sequences from soccer. The model is based on a marked spatio-temporal point process and its estimation is carried out via a Bayesian approach. The model is based on a huge amount of rather complex data and can provide quite powerful insights about soccer and the performance of teams in different aspects. Data from the Premier League of the period 2012–2013 are used to illustrate the potential of the methodology. Overall, it provides a novel and fresh look on soccer data that, I believe, will find a lot of applications and extensions in the near future.

Data availability on soccer has long history. Starting from data collected and annotated by hand and personal observation (Reep & Benjamin, 1968), we can now have huge volumes of data collected automatically using player tracking technologies. Early research on soccer data focused on game-level events such as home advantage, the number of goals and the relative strength of individual teams and leagues. Data were available in an aggregated (per match) tabulation. Later, the collection of data for match events led to analyses of different events and tactics (Decroos et al., 2019). Nowadays, it is possible to track player and ball movements over the whole pitch, throughout the course of a game several times per second. This creates very detailed data for all actions during the match, including off-ball actions, and allows to calculate and create different statistics for the performance of the teams and players. Such data can be used in multiple ways, measuring, for example, passing skill by looking at the geometry of players within passing lanes and quantifying how players create space for themselves and teammates (Bornn et al., 2018). Till now, availability of such data is not easy, but it is even possible to derive such data from broadcasting with some limitations. While the analysis of tracking data and event data is now known, the present paper is the first attempt to combine them together and create a

**Received:** July 19, 2023. **Accepted:** July 20, 2023

more holistic view of measuring performances in soccer taking into account the spatial and temporal aspects of the game.

The paper contributes both in the soccer modelling and the statistical methodology. For the former, the main contributions are the use of tracking and event data together and the valuable insights that one can see from such an approach. Also player and team abilities can be derived from the model and create interpretable estimates for them. The model allows a simulation framework that can be the basis to simulate games in a very detailed manner.

For the latter, some new spatio-temporal marked point processes are defined that generalise classical Hawkes process (exciting process). The authors provide a Bayesian approach to estimate the parameters with an implementation in STAN.

There is a number of potential extensions of the recent model. The one direction is about the implementation. The current version is slow and needs to be improved. This would help to use more data, the current version uses few matches only. Also time varying abilities could be introduced if estimation can be done faster. Since the model is based on some tuning parameter choices, as for example, the number of zones, a more detailed examination is needed for such choices. An inherent advantage of the model is that it can forecast in a short horizon the events we expect to see and this makes the model very interesting for in-game modelling. As such, some more covariate information could be used that describes the in-game characteristics, like the current score, the remaining time, cards (yellows and perhaps more importantly red ones) etc. A final, important point, relates to the expansion of the model for simulating games, and hence, running scenarios or allow for in-game forecasts. Also note that while the model is described for soccer, I can see important extensions for other team sports.

I conclude by congratulating the authors on their innovative approach that combines recent advances in data collection in sports with an interesting stochastic model that can produce valuable insights for all stakeholders. I enthusiastically propose the vote of thanks.

*Conflict of interest:* None declared.

## References

Bornn L., Cervone D., & Fernandez J. (2018). Soccer analytics: Unravelling the complexity of the beautiful game. *Significance*, *15*(3), 26–29. https://doi.org/10.1111/j.1740-9713.2018.01146.x

Decroos T., Bransen L., Van Haaren J., & Davis J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1851–1861). Association for Computing Machinery, New York.

Reep C., & Benjamin B. (1968). Skill and chance in association football. *Journal of the Royal Statistical Society. Series A (General)*, *131*(4), 581–585. https://doi.org/10.2307/2343726

https://doi.org/10.1093/jrsssc/qlad071
Advance access publication 28 July 2023

# Seconder of the vote of thanks to Narayanan, Kosmidis, and Dellaportas and contribution to the Discussion of 'Flexible marked spatio-temporal point processes with applications to event sequences from association football'

**Leonardo Egidi**

Department of Economics, Business, Mathematics, and Statistics 'Bruno de Finetti', University of Trieste, Trieste, Italy

*Address for correspondence:* Leonardo Egidi, Department of Economics, Business, Mathematics, and Statistics 'Bruno de Finetti', University of Trieste, Via Valerio 4/1, 34127 Trieste, Italy. Email: legidi@units.it

First of all, I want to start by warmly thanking the three authors for their valuable and challenging paper and their thorough paper presentation: this is a wonderful and deep scientific work. I will try to briefly challenge it from some distinct sides; on the other hand, using Popper's words, a theory/model is scientific only if is falsifiable.

The paper focuses on modelling of in-game event sequences in association football, the so-called *touch-ball events*, such as goals, corners, passes, etc., and generalises the Hawkes marked point processes to better capture the game dynamics. In fact, events in football tend to not cluster in time and are characterised by a non-negligible degree of correlation/self-excitation. The proposed Bayesian model, combined with an efficient MCMC Stan enginery, is applied to events data from the English Premier League 2013/2014: it allows the inclusion of team-information covariates, provides a deep game understanding, and informs strategic decision-making in real time through real-time predictions of event probabilities. Furthermore, team-ability parameters are used to obtain rankings for the teams by event type and capture different teams' styles.

The paper is really well written and the methodology is clearly proposed; I also much appreciated how the authors communicated their results through nice plots and tables. I want to mention some possible weak criticisms and eventual further developments.

First, the authors could try to extend the volume of training data: at this stage, they use only 20 games out of 380, the 5% fraction of the whole dataset. To this aim, what about using variational inference methods in order to provide faster real-time predictions and estimates? I guess these kinds of techniques could dramatically improve the computational times, if compared with the standard MCMC sampling.

Second, what about the amount of overfitting in these predictions?

As a third point, I wonder whether and how much the partition of the field in three rectangular zones influences the final model estimates.

Fourth, I would consider some posterior predictive checks to assess model accuracy.

Fifth, I wonder whether the model is 'scalable' when the number of training matches increases. Moreover, is a model with only two games for each of the 20 teams stable enough?

Sixth, passing ability seems a very discriminant predictor, according to your results. Maybe a model-free passing ability statistics could be even good to create rankings?

I tried to challenge the paper from some computational, predictive, and interpretative viewpoints. However, as you can see I have found the proposal in this paper very stimulating and exciting. For such reasons, it is with great pleasure that I second the vote of thanks for what will be, I am sure, a very influential paper in the field of in-game sports events, and potentially even in other applied settings.

*Conflicts of interest:* none declared.

**The vote of thanks was passed by acclamation.**

https://doi.org/10.1093/jrsssc/qlad072
Advance access publication 27 July 2023

# Paul Smith's contribution to the Discussion of 'Flexible marked spatio-temporal point processes with applications to event sequences from association football' by Narayanan, Kosmidis, and Dellaportas

**Paul A. Smith**

Department of Social Statistics and Demography, University of Southampton, Southampton, UK

*Address for correspondence*: Paul A. Smith, Department of Social Statistics and Demography, University of Southampton, Highfield, SO17 1BJ Southampton, UK. Email: p.a.smith@soton.ac.uk

It was a pleasure (at a time when the World Cup is being televised) to hear some analytically based evidence on football rather than the usual pundit-based discussions. I enjoyed reading the paper, and both the paper and the proposer and seconder prompted some thoughts.

Splitting the field into three parts is clearly a pragmatic approach, but this is done deterministically and taken as a fixed part of the analysis. On the other hand, we can imagine that different areas of the field are related to different events (or chains of events). Is it possible to split the field into areas based on the observed data, and then to make inferences about the importance of different parts of the field?

I was intrigued that events in football are more regular than a Poisson process, and it seems to me that this is most likely because the shortest inter-event times are missing. Could you therefore model the inter-event times more effectively with a truncated Poisson process?

The paper indicates that predicting future events is challenging, but shows some encouraging results on prediction of shots at goal. Can you extend the idea of simulating passages of play to simulate whole games? And if you do, is the scoreline for a simulated game plausible (e.g. 2–0) or implausible (e.g. 14–3)? This would give some evidence for the usefulness of the fitted parameters in defining the important features of the game.

Finally, I read that statistical modelling for football is difficult because a lot of activity important to the outcome takes place off the ball. But the method in the paper focuses exclusively on ball-related events. Could you incorporate other information about the positions or qualities of players or the team in order to improve your model as a game descriptor?

*Conflicts of interest:* None declared.

**The following contributions were received in writing after the meeting.**

https://doi.org/10.1093/jrsssc/qlad070
Advance access publication 26 July 2023

# Jorge Mateu's contribution to the Discussion of 'Flexible marked spatio-temporal point processes with applications to event sequences from association football' by Narayanan, Kosmidis, and Dellaportas

**Jorge Mateu**

Department of Mathematics, University Jaume I, Castellón, Spain

*Address for correspondence:* Jorge Mateu, Department of Mathematics, University Jaume I, E-12071 Castellón, Spain. Email: mateu@uji.es



Modelling sports is by large a difficult task as human behaviour and many other, perhaps uncontrolled, factors may unexpectedly change the course of the dynamics. This article, however, proposes a flexible family of Hawkes processes in combination with a Bayesian framework for inference. The authors are to be congratulated on this valuable and thought-provoking contribution. I would like to focus my attention on several aspects that are arguably of interest for the community.

Motivated by the evidence that football events do not cluster in time but only in marks, the authors build their approach on decoupling the joint modelling of the times and the marks providing a family of marked point processes that generalises the classical Hawkes process, A first point is how to model periodic effects in the background rate under this formulation. Under the standard equation (2), these effects can be easily added, but under (4), further development is needed. It is easy to guess that some sort of periodic or repetitive movements are due in the game. A second point is the role of relaxation parameters in (4). These parameters are basically essential to balance between the background rate and the triggering effects, as noted in Zhuang and Mateu (2019), and also a correct way of defining them permits reducing problems of identifiability, as the authors comment in relation to Hawkes processes. I wonder how in the authors' framework, we can obtain the probability that an event comes from the background or if it is triggered by another event. Although the authors rely on a Bayesian framework, I would like to draw their attention to an inference technique called stochastic declustering and reconstruction, by which we can recover both background terms and triggering effects. This method has been implemented for Hawkes processes so far, but adding marks would not mean a more complicated inferential strategy. It is indeed helpful in a semiparametric setup.

A final comment has to do with a promising extension of the flexible authors' mechanistic model into two directions. One is considering a bivariate self-exciting mechanistic model, where each component is coming from one of the teams. In this context, negative interactions between the teams are better delineated. Another aspect is that of considering Hawkes models for trajectories, in case we aim for considering the trajectories of the ball conditioned on the player and team. This would be trickier but also interesting.

*Conflicts of interest:* None declared.

## Reference

Zhuang, J., & Mateu, J. (2019). A semi-parametric spatiotemporal Hawkes-type point process model with periodic background for crime data. *Journal of the Royal Statistical Society A*, *182*(3), 919–942. https://doi.org/10.1111/rssa.12429

https://doi.org/10.1093/jrsssc/qlad073
Advance access publication 26 July 2023

# Mattia Stival and Lorenzo Schiavon's contribution to the Discussion of 'Flexible marked spatio-temporal point processes with applications to event sequences from association football' by Narayanan, Kosmidis, and Dellaportas

**Mattia Stival and Lorenzo Schiavon**

Department of Economics, Ca' Foscari University of Venice, Venice, Italy

*Address for correspondence:* Mattia Stival, Department of Economics, Ca' Foscari University of Venice, San Giobbe - Cannaregio 873, Fondamenta San Giobbe, 30121 Venice, Italy. Email: mattia.stival@unive.it



In association football, there exist two types of in-game data: event-sequence data provide qualitative information on the succession of ball-related events in time and space outlining a *local* view of the match; tracking data report with fine temporal granularity the positions of the ball and every player, allowing for a *global* view of the match, in which the ball is just one of many interacting objects. Using event-sequence data, the authors place themselves within a local perspective, with the possible undesirable consequence of missing part of relevant information. Figure 1 illustrates this idea, by showing two alternative situations in which red Player 1 has different probabilities of scoring a goal due to the different positions of the defenders.

By construction, event-sequence data do not report the positions of defenders, producing estimates potentially affected by the presence of unobserved information correlated with the observed outcome. To mitigate the impact of this shortcoming, multiple solutions can be accounted for. One option consists in enriching the event-sequence data with additional qualitative knowledge regarding the game situation. If also player tracking data are available, an alternative solution would be merging any observed event in the event-sequence data with the corresponding snapshot of player and ball positions of the tracking data set.

When this extra information is missing, a careful model specification is required. The process defined by the authors represents a brilliant answer to this challenge. Since any sequence of ball-related events is partially determined by the player's locations on the pitch, the observation of a certain sequence carries with it additional implicit information about team positioning. Unlike most of the recent literature based on Markovian assumptions (Schulte et al., 2017; Rudd, 2011; Singh, 2019), the Hawkes-fashioned specification in equations (10)–(12) recognises all past events as relevant factors in determining match evolution. Recognising event sequences as partially linked to the positioning of players may justify why, according to the authors, 'event sequences in football have a significant dependence on their history'.

With this model, predictive probability density functions of the occurrence of any marked event can be derived. Hence, one could reconstruct via simulation the distribution of the number of any event combination observable in a limited amount of time. This is allowed by joint modelling the event sequence and the time between subsequent events, with temporal modelling standing as a crucial feature to formulate in-game forecasts, and representing a key difference with respect to other frameworks based on a discrete-time game-states representation (Decroos et al., 2019; Fernández et al., 2021). To exploit such potentialities, fast updates of the parameters are needed, requiring new computational approaches (Panos et al., 2021).

A final remark concerns how model complexity is addressed. It would be interesting to compare the association rule learning method with alternative strategies in which the modelling assumptions (e.g. Su et al., 2016) or prior distributions (Ishwaran & Rao, 2005) directly account for sparsity.

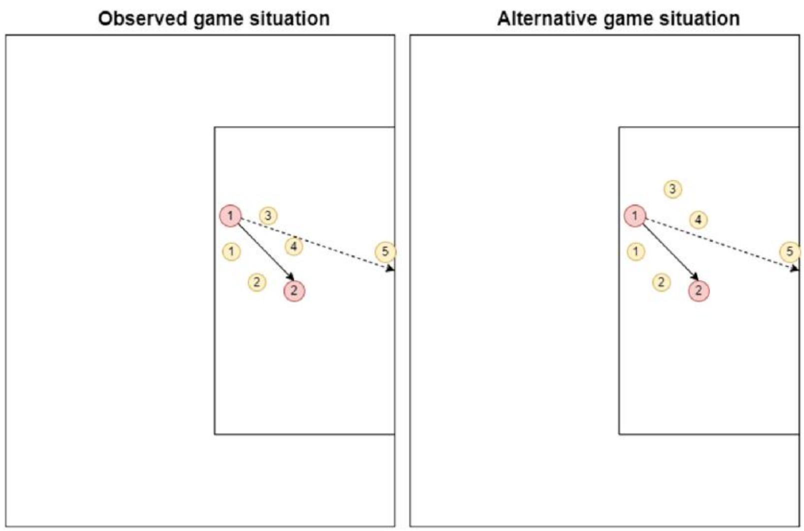



**Figure 1.** On the left, a sketch of the pass (continuous arrow) that led to the goal scored by Jack Wilshere (red circle, player 2) for Arsenal against Norwich, recalled in Figure 1 of the paper. Norwich player positions are reported with smaller yellow circles. On the right, a sketch of an alternative game situation, in which defenders 3 and 4 of Norwich City are placed, arbitrarily, in different positions with respect to the observed ones. Dashed arrows denote unobserved shot events.

To conclude, we thank the authors for their contribution, hoping that our thoughts may enrich the discussion.

*Conflict of interest:* None declared.

## References

Decroos T., Bransen L., Van Haaren J., & Davis J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1851–1861). Association for Computing Machinery. https://doi.org/10.1145/3292500.3330758

Fernández J., Bornn L., & Cervone D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Machine Learning*, *110*(6), 1389–1427. https://doi.org/10.1007/s10994-021-05989-6

Ishwaran H., & Rao J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, *33*(2), 730–773. https://doi.org/10.1214/009053604000001147

Panos A., Kosmidis I., & Dellaportas P. (2021). 'Scalable and interpretable marked point processes', arXiv, arXiv:2105.14574, preprint: not peer reviewed.

Rudd P. (2011, September 24). A framework for tactical analysis and individual offensive production assessment in soccer using Markov Chains. In *New England Symposium on Statistics in Sports*. Cambridge, MA: The Harvard University Science Center.

Schulte O., Khademi M., Gholami S., *et al.* (2017). A Markov Game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery*, *31*, 1735–1757. https://doi.org/10.1007/s10618-017-0496-zt

Singh K. (2019). *Introducing expected threat (xt)*. https://karun.in/blog/expected-threat.html.

Su Z., Zhu G., Chen X., & Yang Y. (2016). Sparse envelope model: Efficient estimation and response variable selection in multivariate linear regression. *Biometrika*, *103*(3), 579–593. https://doi.org/10.1093/biomet/asw036

https://doi.org/10.1093/jrsssc/qlad068
Advance access publication 26 July 2023

# Ron Yurko and Rebecca Nugent's contribution to the Discussion of 'Flexible marked spatio-temporal point processes with applications to event sequences from association football' by Narayanan, Kosmidis, and Dellaportas

**Ron Yurko and Rebecca Nugent**

Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA USA

*Address for correspondence:* Rebecca Nugent, Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA, United States. Email: rnugent@stat.cmu.edu



We thank the authors for their novel contribution to modelling the game of football in a dynamic manner that allows for flexible parametrisation. More generally, we are excited to read another methodological contribution that leverages the rich, complex data that are becoming increasingly available across all sports. As the field of sports analytics has matured in recent years, we are seeing more and more substantive methodological research also pushing forward the statistics and data science disciplines (see e.g. *Journal of Quantitative Analysis in Sports*). We encourage other researchers to pursue similar interdisciplinary work tackling complex problems in sports analytics. While, as the authors note, this approach can be applied to other scientific disciplines, these comments will focus on the impact of their approach in sports analytics.

Sports have a natural hierarchical structure across multiple facets, e.g. football teams consist of players with observed performances across plays within games across seasons. Statisticians need to make modelling choices about which levels to estimate or aggregate. The authors provide a clear demonstration of how this approach can provide team ability assessments. However, this of course results in an aggregate view, collapsing information distinguishing individual players. Does this approach comport to player-level parametrisation? Is it a matter of data? computational burden? A more granular view at the player-level could reveal invaluable insights that are relevant to teams at understanding their roster composition.

Additionally, the authors discuss how simulating in-game events with their proposed model estimates event probabilities. This is an exciting feature of their work, enabling the calculation of an expected goal value at a precise moment within a game. Seminal work by Cervone et al. (2016) introduced this idea in basketball using high-resolution spatio-temporal tracking data. More recently, Fernández et al. (2021) directly apply machine learning methods to estimate the expected goal value in football. While the methodology is inherently different, the end use case for analysts in sports is similar. Have the authors explored using their approach to provide a real-time value for player decision making? What are the advantages of their model in comparison to the simpler machine learning framework by Fernández et al. (2021)? This discussion would be advantageous for both statisticians and data scientists as well as sports analysts looking to implement these approaches.

Again, we thank the authors for their insightful contribution. We look forward to future work, possibly addressing computational burdens, and are excited at the prospect of applying this approach in other sports applications.

*Conflict of interest:* None declared.

## References

Cervone D., D'Amour A., Bornn L., & Goldsberry K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association*, *111*(514), 585–599. https://doi.org/10.1080/01621459.2016.1141685

Fernández J., Bornn L., & Cervone D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Machine Learning*, *110*(6), 1389–1427. https://doi.org/10.1007/s10994-021-05989-6

https://doi.org/10.1093/jrsssc/qlad069
Advance access publication 27 July 2023

# Authors' reply to the Discussion of 'Flexible marked spatio-temporal point processes with applications to event sequences from association football'

**Santhosh Narayanan**[1], **Ioannis Kosmidis**[2] and **Petros Dellaportas**[3,4]

[1]The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, UK
[2]Department of Statistics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
[3]Department of Statistical Science, University College, Gower Street, London WC1E 6BT, UK
[4]Department of Statistics, Athens University of Economics and Business, 76 Patission Str., Athens 10434, Greece

*Address for correspondence:* Santhosh Narayanan, The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK. Email: snarayanan@turing.ac.uk

## 1 Introduction

We are grateful to all discussants for the insightful comments, thoughts, pointers, and proposals for extending our work. We identified that the discussion contributions evolve around the themes: (i) Estimation and scalability, (ii) Evaluation of predictive performance and overfitting, (iii) Choice of field zones, (iv) Inclusion of extra covariate information, (v) Game simulation and in-game forecasts, (vi) Model outputs, (vii) Data considerations, and (viii) Model extensions. In what follows, we structure our reply according to those themes directly referring to the relevant contributions.

## 2 Estimation and scalability

Karlis (2023) and Egidi (2023) note that the current Hamiltonian Monte Carlo implementation can be slow and would benefit from improvement. We acknowledge these observations and note that this has been dealt with in the recent work by Panos et al. (2023) that develops a variational inference framework to provide a highly scalable procedure for training the models we introduced. Panos et al. (2023) also extend the model to allow for time-varying abilities within their proposed variational inference framework and report computational times of a few hours for a whole season's worth of touch-ball data.

Stival and Schiavon (2023) suggested using sparsity-inducing priors as an alternative to our work's association rule learning method to deal with model complexity. That is an excellent suggestion that we plan to pursue as part of future work. The main challenge we faced in our limited attempts with sparsity-inducing priors within the current vanilla posterior sampling framework is again the dimension of the parameter space. Nevertheless, we believe such prior structures can prove helpful alongside the variational inference framework of Panos et al. (2023).

## 3 Evaluation of predictive performance and overfitting

Egidi (2023) enquires about posterior predictive checks to assess model accuracy and the amount of overfitting in the model predictions.

In our work, we evaluate the predictive performance of the models using the log-pointwise predictive density on test data; see Section 6.2 of the main text. Of course, if prediction is the aim of the modelling exercise, we recognise that it is helpful to evaluate the models' predictive performance using additional evaluation criteria that compare predictions against observables more explicitly. Examples include out-of-sample root mean square error for the ability to predict the times of future events, and the wealth of classification performance measures for the ability to predict future marks (see, for example, Sokolova & Lapalme, 2009, for a systematic analysis of such). Within their variational framework, Panos et al. (2023) evaluated the model's performance using such criteria and found that it offers highly competitive computational and predictive performance against other state-of-the-art methods, which typically involve high-dimensional structures through semi- or non-parametric components.

We did not observe any severe overfitting of the models. The order of the models in terms of increasing *out-of-sample* log-pointwise predictive density in Table 7 is similar to the one obtained by naively evaluating the *in-sample* log-pointwise predictive density. The latter is an overestimate of the expected log-pointwise predictive density with respect to future data. A notable difference is with the M$\beta$A model (matrix $\beta$ with team abilities), which has the largest number of parameters, and the largest in-sample log-pointwise predictive density. Despite being close to the best and second-best models, the M$\beta$A model is the third best in the *out-of-sample* evaluations in Table 7. That may indicate a small degree of overfitting for that model, most probably because, in the training data, each team plays just one game at their home and one at an away venue.

## 4 Choice of field zones

Karlis (2023), Egidi (2023), and Smith (2023) comment on the choice of partitioning the field into three zones of known area and enquire how that choice influences the model estimates.

The event triggering parameters of the M$\beta$ parameterisation of the model (see expression (12) in the main text) depend on specifying a partition. That dependence enables us to readily infer the importance of different zones for particular actions, substantially enhancing interpretability (see Sections 6.1–6.4 for such interpretations). For example, we can compute the chance of completing a successful pass or attempting a shot on goal in a particular zone. The choice of zones and their areas represent an idealisation of our understanding of the game, where the playing strategies have a natural dependence on whether the ball is in the defensive, midfield or attacking third of the field.

The recommendation of Smith (2023) for a more data-driven approach to determine the partitions is fruitful and an exciting area for future work. For example, we can consider a team-dependent, continuous spatial process that respects the field boundaries (see, for example, Solin & Kok, 2019) for $h(z_i \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\eta})$ in expression (7) of the main text, and threshold the field adaptively into a fixed number of zones. This way, zones will have an adaptive area that depends on how each team realises its strategy.

## 5 Inclusion of extra covariate information

Karlis (2023), Smith (2023), and Yurko and Nugent (2023) suggested including covariate information, such as in-game characteristics, off-the-ball player positions and player qualities to improve the model performance.

Our current case study uses only the team information as covariates to demonstrate the modelling framework. Nevertheless, the modelling framework readily allows for including other covariates to drive the cross-excitation of the marks; see Section 4.3 of the main text. Including covariates about the game's current state, such as the current score, number of cards, etc., may be particularly beneficial in predictive ability and is the topic of ongoing investigations. The cross-excitation of the marks can also incorporate information about the positions and qualities of players if that information is available.

Other parameters, such as the background process parameters, excitation factors and decay rates, can also be appropriately linked with covariate information through regression structures. The inferential or predictive benefits of including such regression structures should be weighed against the increased model complexity.

## 6 Game simulation and in-game forecasts

Karlis (2023) and Smith (2023) enquire about the model's ability to simulate whole games. The experiments in the paper only dealt with 30-second simulations.

The simulation framework of Section 7 of the manuscript and its implementation through the codebase we provide allows for the simulation of an arbitrary period or even the whole game using the in-game forecasts. We did run a limited number of simulations for entire games, and the forecasted scorelines were in the plausible spectrum. Still, more work is needed to study the usefulness of the proposed model for such applications.

## 7 Model outputs

Egidi (2023) notes that passing ability appears to be a markedly discriminant predictor of team rankings and wonders whether model-free passing ability statistics could be good for generating rankings. The model is not specified with the direct intent to predict team rankings. Our rankings are a bonus, but perhaps unsurprising, output from using a highly interpretable parameterisation. Further work is required to study the utility of the model parameters for generating team rankings using data from multiple seasons and leagues.

Yurko and Nugent (2023) asked if we explored using our approach to provide a real-time value for player decision-making. That is an exciting area of application, which we did not consider and for which our approach is well suited. Including player-level covariate information is necessary in that direction, and we intend to explore it in future work.

## 8 Data considerations

Egidi (2023) enquires whether a model trained on only two games for each of the 20 teams is stable enough. Our decision to use only the first two games for each of the 20 teams resulted from our attempt to demonstrate the wealth of insights that can be generated from our proposal using a limited amount of information, also accounting for the computational limitations we have been facing when fitting the models. A potential stability assessment could come by fitting the models over different sets of games. Note that the variational inference framework in Panos et al. (2023) overcomes the computational limitations and can fit the M$\beta$A model with time-varying abilities in a whole season's worth of touch-ball events in a few hours.

Smith (2023) suggests that the observations that events in football are more regular than Poisson may be because the shortest inter-event times are missing. That is true in the data analysed. Certain kinds of events in football, such as off-the-ball events like player runs, are not recorded. We found that the gamma model was adequate for modelling the inter-event times for the available data. Still, the truncated Poisson, as suggested by Smith (2023), is a valid alternative to compare with.

## 9 Model extensions

Mateu (2023) presents some helpful model extensions we have yet to consider and plan to investigate, including (i) using periodicity on background rates and (ii) using Hawkes models for the ball's trajectories. Naturally, and as identified by the discussant, the former extension is more direct through the conditional intensity function than through the decomposition of a multivariate distribution function in (4). A remedy is to employ similar specifications as in Zhuang and Mateu (2019) and add periodic effects on the definition background mark probabilities in (5). Extensions in direction (ii) are directly possible through the appropriate specification of $h(z_i \mid t_i, \mathcal{F}_{t_{i-1}}; \boldsymbol{\eta})$ in expression (7) of the main text. See Section 4 for relevant discussion.

Another direction for extension mentioned by Mateu (2023) is accommodating negative interactions between events. Indeed, our development here cannot capture inhibition behaviour (i.e. having the occurrence of an event decrease the likelihood of another event to occur). We define a multivariate process on composite event types, where each event type is tracked for both the home and the away team (see Table 3 in the main text) to capture excitations from events both within each team and between teams. Capturing inhibition through our model specification is an exciting direction and valuable for the diverse applications of those models, for which recent developments such as Costa et al. (2020) and Bonnet et al. (2021) can be helpful.

Mateu (2023) also enquires how the probability of an event coming from the background or being triggered by another event can be computed. That is possible by calculating the posterior conditional branching structure probabilities in expression (17), which we use for deriving event genealogies in Section 6.8 of the main text.

*Conflict of interests:* None declared.

## References

Bonnet A., Herrera M. M., & Sangnier M. (2021). Maximum likelihood estimation for Hawkes processes with self-excitation or inhibition. *Statistics & Probability Letters*, *179*, 109214. https://doi.org/10.1016/j.spl.2021.109214

Costa M., Graham C., Marsalle L., & Tran V. C. (2020). Renewal in Hawkes processes with self-excitation and inhibition. *Advances in Applied Probability*, *52*(3), 879–915. https://doi.org/10.1017/apr.2020.19

Egidi, L. (2023). Seconder of the vote of thanks to Narayanan, Kosmidis and Dellaportas and contribution to the Discussion of "Flexible marked spatio-temporal point processes with applications to event sequences from association football". *Journal of the Royal Statistical Society Series C: Applied Statistics*, *72*(5), 1129–1130. https://doi.org/10.1093/jrsssc/qlad072

Karlis, D. (2023). Proposer of the vote of thanks to Narayanan, Kosmidis and Dellaportas and contribution to the Discussion of "Flexible marked spatio-temporal point processes with applications to event sequences from association football". *Journal of the Royal Statistical Society Series C: Applied Statistics*, *72*(5), 1127–1128. https://doi.org/10.1093/jrsssc/qlad071

Mateu, J. (2023). Jorge Mateu's contribution to the Discussion of "Flexible marked spatio-temporal point processes with applications to event sequences from association football" by Narayanan, Kosmidis and Dellaportas. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(5), 1131. https://doi.org/10.1093/jrsssc/qlad073

Panos A., Kosmidis I., & Dellaportas P. (2023). Scalable marked point processes for exchangeable and non-exchangeable event sequences. In F. Ruiz, J. Dy, & J.-W. van de Meent (Eds.), *Proceedings of the 26th international conference on artificial intelligence and statistics: Vol. 206. Proceedings of machine learning research* (pp. 236–252). PMLR.

Smith, P. A. (2023). Paul Smith's contribution to the Discussion of "Flexible marked spatio-temporal point processes with applications to event sequences from association football" by Narayanan, Kosmidis and Dellaportas. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(5), 1130. https://doi.org/10.1093/jrsssc/qlad070

Sokolova M., & Lapalme G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Solin A., & Kok M. (2019). Know your boundaries: Constraining gaussian processes by variational harmonic features. In K. Chaudhuri & M. Sugiyama (Eds.), *Proceedings of the twenty-second international conference on artificial intelligence and statistics: Vol. 89. Proceedings of machine learning research* (pp. 2193–2202). PMLR.

Stival, M., & Schiavon, L. (2023). Mattia Stival and Lorenzo Schiavon's contribution to the Discussion of "Flexible marked spatio-temporal point processes with applications to event sequences from association football" by Narayanan, Kosmidis and Dellaportas. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(5), 1132–1133. https://doi.org/10.1093/jrsssc/qlad068

Yurko, R., & Nugent, R. (2023). Ron Yurko and Rebecca Nugent's contribution to the Discussion of "Flexible marked spatio-temporal point processes with applications to event sequences from association football" by Narayanan, Kosmidis and Dellaportas. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *72*(5), 1134–1135. https://doi.org/10.1093/jrsssc/qlad069

Zhuang J., & Mateu J. (2019). A semiparametric spatiotemporal Hawkes-type point process model with periodic background for crime data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *182*(3), 919–942. https://doi.org/10.1111/rssa.12429

https://doi.org/10.1093/jrsssc/qlad101
Advance access publication 3 November 2023