Journal of the Royal Statistical Society Series C: Applied Statistics, 2026, 00, 1–29
https://doi.org/10.1093/jrsssc/qlag032
Published: 9 June 2026
Original article

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

Bayesian weighted discrete-time dynamic models for association football prediction

Roberto Macrì-Demartino, Leonardo Egidi, Nicola Torelli

Department of Economics, Business, Mathematics, and Statistics 'Bruno de Finetti', University of Trieste, Via A. Valerio 4/1, 34127 Trieste, Italy
Address for correspondence: Roberto Macrì-Demartino, Department of Economics, Business, Mathematics, and Statistics 'Bruno de Finetti', University of Trieste, Via A. Valerio 4/1, 34127 Trieste, Italy.
Email: roberto.macridemartino@deams.units.it

Abstract

In recent years, great emphasis has been placed on the prediction of association football. Due to this, several studies have proposed different types of statistical models to predict the outcome of a football match. However, most existing approaches usually assume that the offensive and defensive abilities of teams remain static over time. We introduce a Bayesian dynamic approach for football goal-based models that uses commensurate priors to flexibly weight the evolution of attacking and defensive abilities. Our approach assigns separate, time-varying precision parameters to each team and ability in every period, controlled via spike-and-slab hyperpriors. This adaptive shrinkage borrows information about teams' strength when past and current performance align and allow rapid adjustments when teams experience substantial changes (e.g. transfer windows or coaching changes). We integrate this framework into five standard goal-based models evaluating predictive performance using data from the last five seasons of the German Bundesliga, English Premier League, and Spanish La Liga. Compared with three leading dynamic approaches, our adaptive approach yields better predictive performance. The proposed methodology has also been implemented in the free and open source R package `footBayes`.

Keywords commensurate prior, hierarchical models, historical borrowing, posterior predictive, sport analytics

1 Introduction

Quantitative analysis of association football (soccer), hereafter referred to as football, and more specifically, matches' prediction is a rapidly evolving field, increasingly valued by participants, coaches, owners, and gamblers looking to gain a competitive advantage. This development is part of a broader revolution in sports analytics, in which statistical and machine learning (ML) methods have transformed decision-making in a wide range of sports. Notable examples include basketball (Deshpande & Jensen, 2016; Lam, 2018; Reich et al., 2006, among others), baseball (Jensen et al., 2009; Phelan & Whelan, 2018, among others), and tennis (McHale & Morton, 2011; Spanias & Knottenbelt, 2013, among others). For more comprehensive reviews, see Cattelan (2012), Santos-Fernandez et al. (2019), and Glickman and Jones (2025). Consequently, there is a growing demand for information that supports better decision making.

Received: August 7, 2025. **Revised:** March 5, 2026. **Accepted:** May 26, 2026.

The outcome of football matches can be predicted using two main statistical modelling frameworks. In the goal-based (or direct) approach, the actual number of goals scored by each team is a count variable—most commonly modelled via Poisson or negative binomial (NB) regression. The expected goal counts are functions of team attributes (e.g. offensive and defensive abilities) and, when relevant, home-field advantage. In contrast, result-based (or indirect) models predict one of the three outcomes—home win, draw, or away win—typically through ordered probit (Koning, 2000) or logit (Carpita et al., 2019, 2015) regressions. Furthermore, the widespread popularity of large datasets promoted the use of ML tools, yielding a fundamentally different modelling approach based on a random forest (Breiman, 2001). The potential of using this as a new results-based model was first explored by Schauberger and Groll (2018) to assess the predictive performance of different types of random forests compared to classical Poisson regression methods on data containing all matches of the FIFA World Cups 2002–14. Along these lines, Groll et al. (2019a), Groll et al. (2019b), Groll et al. (2021), and Groll et al. (2024) further expand this framework. It is worth noting that the result-based framework is formally nested within the goal-based one. Specifically, match results are derived from the underlying goal counts, while knowing only the three-way result provides no information about the actual goals scored, potentially misestimating team strength (Egidi & Torelli, 2021). Furthermore, Ley et al. (2019) provide a comprehensive comparison of 10 strength-based models, including result-based approaches such as Bradley–Terry (Bradley & Terry, 1952) and Thurstone–Mosteller (Mosteller, 1951; Thurstone, 1927) models, as well as goal-based Poisson models estimated via weighted maximum likelihood with adjustments for match importance and time decay. Their findings indicate that Poisson-based models provide superior predictive performance. Therefore, we focus on the richer goal-based structure, which not only yields three-way predictions but also captures the magnitude of outcomes.

In the simplest goal-based formulation, team-specific goal counts are assumed conditionally independent given team abilities or covariates, resulting in a double Poisson (DP) model (Baio & Blangiardo, 2010; Egidi et al., 2018; Groll & Abedieh, 2013; Maher, 1982, among others). To relax the strong independence assumption, several generalizations introduce score dependence. Dixon and Coles (1997) extended the work of Maher (1982) by allowing (a slightly negative) correlation between scores and incorporating a dependence parameter into their model to account for it. Karlis and Ntzoufras (2003) introduced in a frequentist framework the bivariate Poisson model, designed to account for positive goal dependencies. Furthermore, Ntzoufras (2011) extended it from a Bayesian perspective, whereas McHale and Scarf (2011) used copulas to allow dependence between the two Poisson random variables under the assumption that the dependence parameter can be expressed as a linear function of the rank difference between the two teams, thus providing a flexible solution that permits both positive and negative correlation. In this direction, Boshnakov et al. (2017) introduced a model that combines a Weibull inter-arrival-times-based count process with a copula construction to obtain a flexible bivariate distribution for home and away goals. More recently, Michels et al. (2025) demonstrated that the Dixon and Coles model is a special case of the Sarmanov family of distributions, using this connection to extend the dependence structure beyond the four low-scoring outcomes and to accommodate marginal distributions other than Poisson, such as the NB. Furthermore, Petretta et al. (2025) proposed a marginal-conditional (Mar-Co) model that specifies the joint distribution through carefully chosen marginal and conditional distributions, allowing for a more comprehensive dependence structure that spans the entire support of possible match outcomes while preserving interpretability.

A main assumption of previous models is the invariance of team-specific parameters, implying static offensive and defensive abilities over time. However, it is recognized that team performance is inherently dynamic, fluctuating over years and possibly within seasons. One simple approach is the time decay weighting used by Dixon and Coles (1997), in which past match outcomes are downweighted so that recent games have more influence on estimated abilities. However, a more formal approach is to treat team abilities as time-varying parameters. Specifically, a continuous-time dynamic extension of the DP model was introduced by Rue and Salvesen (2000), while Koopman and Lit (2015) and Koopman and Lit (2019) integrated bivariate Poisson models in a state-space framework, allowing the abilities of the team to vary according to a state vector. Alternatively, Owen (2011) proposed a

Bayesian discrete-time approach based on an evolution component that describes the stochastic behaviour of time-dependent parameters. Here, the evolution component is specified as a random walk prior distribution structure for both the attack and defence parameters. Recent developments include Egidi et al. (2018), which incorporates betting odds and other refinements, and Macrì Demartino et al. (2025), which evaluate predictive performance improvements when the ranking of a team is added as a covariate in dynamic models. However, a key limitation of the approach proposed by Owen (2011) is the assumption of a constant and common evolution precision for both attack and defence parameters. This constraint may limit the predictive performance of the goal-based model, ignoring the fact that a team's performance can fluctuate more during certain periods (e.g. early season transfers, mid-season managerial changes). Furthermore, forcing the same evolution precision for both attack and defence neglects that these abilities can have different rates of change—defensive abilities may adapt more slowly than offensive ones or vice versa.

In the present work, we try to fill this gap by proposing a Bayesian weighted discrete-time approach that incorporates team-specific and time-specific commensurate priors (Hobbs et al., 2011, 2012) for both attack and defence parameters. This provides a formal mechanism for letting the prior distributions at a specific time to adaptively borrow information from the previous time, but only to the extent that the data support it. By introducing flexible and dynamic evolution precision, we obtain a more accurate and adaptive modelling of team abilities over time, improving the predictive performances.

The article is organized as follows. Section 2 presents the Poisson and NB goal-based models used in this study. Furthermore, Section 3 describes the commensurate prior framework and introduces our proposed dynamic weighted approach for the offensive and defensive abilities of the teams in the goal-based models. In Section 4, we apply our methodology to some of the top European leagues, namely, the German Bundesliga, English Premier League (EPL), and Spanish La Liga. A total of five seasons from 20/21 to 24/25 are used from each league to perform the study. Finally, Section 5 provides concluding remarks that outline limitations, advantages, and potential future research directions.

2 Goal-based models

This section presents the statistical goal-based models used to predict the chosen competition outcomes. Through an in-depth analysis, our aim is to provide a comprehensive overview of the methodologies applied to predict football matches, highlighting both their statistical foundations and practical implementations in sports analytics.

2.1 Poisson-based models

Let $(x_{i,n}, y_{j,n})$ represent the observed number of goals scored by the home and the away team in the nth match, with $i \neq j = 1, \ldots, N_T$, where N_T denotes the number of teams in the league, and $n = 1, \ldots, N$. A simple DP model assumes that the goal counts follow two conditionally independent Poisson distributions:

$$\begin{aligned} X_{i,n} \mid \lambda_{1,n} &\sim \text{Poisson}(\lambda_{1,n}), \\ Y_{j,n} \mid \lambda_{2,n} &\sim \text{Poisson}(\lambda_{2,n}), \\ X_{i,n} \perp\!\!\!\perp Y_{j,n} &\mid \lambda_{1,n}, \lambda_{2,n}, \end{aligned} \tag{1}$$

where the (nonnegative) parameters $\lambda_{1,n}$ and $\lambda_{2,n}$ are the expected scoring rates of the home and away teams, respectively, in the nth match. The rate at which a team is expected to score is a function of both its own offensive ability and the defensive ability of its opponent,

$$\begin{aligned} \log(\lambda_{1,n}) &= \beta_0 + \text{home} + \beta^{(\text{att})}_{h_n} + \beta^{(\text{def})}_{a_n}, \\ \log(\lambda_{2,n}) &= \beta_0 + \beta^{(\text{att})}_{a_n} + \beta^{(\text{def})}_{h_n}, \end{aligned} \tag{2}$$

where the parameter β_0 is a common intercept, 'home' captures the well-known home-field advantage, and $\beta^{(\text{att})}$ and $\beta^{(\text{def})}$ represent the unknown attacking and defensive abilities of the home team h_n and the away team a_n in the nth match.

Dixon and Coles (1997) proposed an important refinement to the DP model by introducing a dependence correction for low-scoring outcomes. They observed that the independent Poisson assumption tends to underestimate the frequency of draws and certain low-scoring results, particularly the 0–0, 1–0, 0–1, and 1–1 scorelines. To address this, they modified the joint probability of goal counts by applying a multiplicative correction factor $\tau(x_{i,n}, y_{j,n}; \rho)$ to the product of the two independent Poisson probabilities,

$$\mathbb{P}_{X_{i,n},Y_{j,n}}(x_{i,n}, y_{j,n}) = \tau(x_{i,n}, y_{j,n}; \rho)\, \frac{\lambda_{1,n}^{x_{i,n}} \exp(-\lambda_{1,n})}{x_{i,n}!} \frac{\lambda_{2,n}^{y_{j,n}} \exp(-\lambda_{2,n})}{y_{j,n}!}, \tag{3}$$

where the correction factor is defined as

$$\tau(x_{i,n}, y_{j,n}; \rho) = \begin{cases} 1 - \rho\lambda_{1,n}\lambda_{2,n} & \text{if } x_{i,n} = 0, \quad y_{j,n} = 0, \\ 1 + \rho\lambda_{1,n} & \text{if } x_{i,n} = 0, \quad y_{j,n} = 1, \\ 1 + \rho\lambda_{2,n} & \text{if } x_{i,n} = 1, \quad y_{j,n} = 0, \\ 1 - \rho & \text{if } x_{i,n} = 1, \quad y_{j,n} = 1, \\ 1 & \text{otherwise,} \end{cases}$$

and ρ is a parameter governing the degree and direction of dependence between the two score counts for low-scoring matches. The scoring rates $\lambda_{1,n}$ and $\lambda_{2,n}$ are again modelled as in (2). When $\rho = 0$, the correction factor equals one for all scorelines, and the model reduces to the DP specification in (1).

Similarly, Karlis and Ntzoufras (2003) argued that the independence assumption in the DP model may be too restrictive, as the scores of competing teams could exhibit dependence. To allow for potential positive correlation between goal counts, they introduced the bivariate Poisson model, which generalizes the DP by including a covariance parameter. When this parameter is zero, the goal counts become independent, and the bivariate Poisson model reduces to the DP model in (1).

2.2 NB and Skellam alternatives

Poisson models assume equal mean and variance, which may not hold in real-world football data—especially in competitions where overdispersion (sample variance exceeds the sample mean) is observed in the number of goals. To handle this, a common approach is to replace each Poisson marginal with an NB distribution (Reep et al., 1971), which produces a double negative binomial model. That is,

$$X_{i,n} \sim \text{NB}(\lambda_{1,n}, \gamma_1), \qquad Y_{j,n} \sim \text{NB}(\lambda_{2,n}, \gamma_2),$$

where $\lambda_{1,n}$ and $\lambda_{2,n}$ follow the same log-linear structure introduced in Section 2.1 and $\gamma_1, \gamma_2 > 0$ are overdispersion parameters for home and away goals, respectively. Using the mean-overdispersion parametrization, the probability mass function of $X_{i,n}$ is

$$\mathbb{P}_{X_{i,n}}(x_{i,n}) = \binom{x_{i,n} + \gamma_1 - 1}{x_{i,n}} \left(\frac{\lambda_{1,n}}{\lambda_{1,n} + \gamma_1}\right)^{x_{i,n}} \left(\frac{\gamma_1}{\lambda_{1,n} + \gamma_1}\right)^{\gamma_1},$$

with an analogous form for $Y_{j,n}$. Under this parametrization, $\mathbb{E}(X_{i,n}) = \lambda_{1,n}$ and $\text{Var}(X_{i,n}) = \lambda_{1,n} + \lambda_{1,n}^2/\gamma_1$. The double negative binomial model directly captures the overdispersion in the goal count of each team, with lower values of γ_1 and γ_2 indicating greater overdispersion relative to the Poisson model. As $\gamma_1, \gamma_2 \to \infty$, the variance approaches the mean, recovering the DP model described in (1).

Alternatively, Karlis and Ntzoufras (2009) suggest using the Skellam distribution (Skellam, 1946), which directly models the difference in goals. Notably, the Skellam distribution captures not only the overdispersion but also the intrinsic dependence between the teams' scoring outcomes, without requiring explicit correlation modelling. Specifically, let $X_{i,n}$ and $Y_{j,n}$ represent independent Poisson counts for goals scored by teams T_i and T_j in the nth match, respectively, with $i \neq j = 1, \ldots, N_T$ and $n = 1, \ldots, N$. The Skellam model (SM) is then given by the difference of the two goal counts,

$$Z_n = X_{i,n} - Y_{j,n}.$$

The corresponding probability mass function is given by

$$\mathbb{P}_{Z_n}(z_n) = \exp\{-(\lambda_{1,n} + \lambda_{2,n})\} \left(\frac{\lambda_{1,n}}{\lambda_{2,n}}\right)^{h/2} I_h\left(2\sqrt{\lambda_{1,n}\lambda_{2,n}}\right), \quad h \in \mathbb{Z}, \tag{4}$$

where $\mathbb{E}(Z_n) = \lambda_{1,n} - \lambda_{2,n}$ and $\text{Var}(Z_n) = \lambda_{1,n} + \lambda_{2,n}$. Furthermore, $I_h(\cdot)$ is the modified Bessel function of order h (Skellam, 1946). The parameters $\lambda_{1,n}$ and $\lambda_{2,n}$ adopt the same log-linear structure as in the previous cases.

2.3 Inflating the draws probability

Poisson goal-based models often underestimate the incidence of draws, which are the diagonal elements in goal probability matrices. To address excess draws in goal differences, the zero-inflated Skellam model (ZISM) (Karlis & Ntzoufras, 2009) can be adopted,

$$\mathbb{P}_{Z_n}(z_n) = \begin{cases} (1-\omega)\text{SM}\left(\lambda_{1,n}, \lambda_{2,n}\right) & \text{if } z_n \neq 0 \\ (1-\omega)\text{SM}\left(\lambda_{1,n}, \lambda_{2,n}\right) + \omega D(0, \xi) & \text{if } z_n = 0 \end{cases},$$

where $\text{SM}(\cdot)$ is the Skellam probability mass function as in (4), $\omega \in [0, 1]$ controls the inflation weight, and $D(0, \xi)$ is a discrete distribution that places extra mass at zero. Similarly, Karlis and Ntzoufras (2003) introduced a diagonally inflated bivariate Poisson model when we can assume a positive dependence between the two scores.

2.4 Dynamic prior distributions and identifiability constraints

A structural limitation in the previous models is the assumption of static team-specific parameters, namely, teams are assumed to have a constant performance over time, determined by attack and defence abilities $\beta^{(\text{att})}$ and $\beta^{(\text{def})}$, respectively. However, the performance of teams tends to be dynamic—between seasons and even from week to week—due to factors such as summer and winter transfer windows reshaping lineups, injuries benching key players, or midseason coaching changes because of unsatisfactory results.

Several approaches have been proposed to dynamically model team-specific abilities (Koopman & Lit, 2015, 2019; Owen, 2011; Rue & Salvesen, 2000, among others). In particular, Owen (2011) extended the static framework by introducing a discrete-time evolution for team-specific effects. Specifically, the evolution component is specified as a random walk for both the attack and defence parameters by centring the effect of seasonal time τ on the lagged effect in $\tau - 1$. This allows the attack and defence parameters to vary between seasons or weeks. Therefore, for each team T_i, where $i = 1, \ldots, N_T$, and each period τ, where $\tau = 2, \ldots, \mathcal{T}$, the prior distributions for the attack and defence abilities are usually defined as follows:

$$\begin{aligned} \beta_{i,\tau}^{(\text{att})} \mid \beta_{i,\tau-1}^{(\text{att})}, \sigma &\sim \text{N}\left(\beta_{i,\tau-1}^{(\text{att})}, \tfrac{1}{\sigma^2}\right), \\ \beta_{i,\tau}^{(\text{def})} \mid \beta_{i,\tau-1}^{(\text{def})}, \sigma &\sim \text{N}\left(\beta_{i,\tau-1}^{(\text{def})}, \tfrac{1}{\sigma^2}\right). \end{aligned} \tag{5}$$

While for the initial period $\tau = 1$, the prior distributions are initialized as

$$\beta_{i,1}^{(\text{att})} \mid \mu^{(\text{att})}, \sigma \sim \text{N}\left(\mu^{(\text{att})}, \tfrac{1}{\sigma^2}\right),$$
$$\beta_{i,1}^{(\text{def})} \mid \mu^{(\text{def})}, \sigma \sim \text{N}\left(\mu^{(\text{def})}, \tfrac{1}{\sigma^2}\right), \tag{6}$$

where $\mu^{(\text{att})}$ and $\mu^{(\text{def})}$ are the initial prior means representing the expected baseline levels of attack and defensive ability, respectively, across all teams at the start of the observation period. In practice, these hyperparameters are typically set to zero. The parameter σ^2 denotes the common evolution precision, assumed constant over time and identical between all teams and both team-specific abilities. To ensure identifiability, a zero-sum constraint (Baio & Blangiardo, 2010; Owen, 2011) on the random effects within each period is required,

$$\sum_{i=1}^{N_T} \beta_{i,\tau}^{(\text{att})} = 0, \quad \sum_{i=1}^{N_T} \beta_{i,\tau}^{(\text{def})} = 0, \quad \tau = 1, \ldots, \mathcal{T}. \tag{7}$$

As a matter of parameter interpretation, once the models have been estimated, a larger team-attack parameter indicates stronger attacking quality, while a smaller team-defence parameter corresponds to stronger defensive performance.

3 A weighted dynamic proposal

As described in Section 2.4, a key assumption of the discrete-time evolution approach as in (5) is a single constant evolution precision σ^2 shared by all teams and by both their attack and defence parameters. However, this assumption can compromise predictive accuracy by either overborrowing (underborrowing) strength from one period to the next. Specifically, some periods—such as the summer transfer window or a midseason coaching change—can cause rapid shifts in team abilities, justifying the discount of earlier performance information; while other periods, when the teams' abilities are stable, borrowing more past information can improve the predictive performances. Furthermore, offensive and defensive abilities may often evolve at different rates. In this section, we propose a weighted dynamic approach based on commensurate priors, which employ separate, time-varying evolution precisions for attack and defence. By treating the matches played during a specific period as 'current data' with respect to the 'historical data' from the previous period, this approach offers an intuitive framework in which the prior at each time point adaptively borrows information from the previous period, but only to the extent that the data justify it.

3.1 Commensurate priors

In the Bayesian framework, adaptively informative priors are valuable for synthesizing results across studies, particularly in the clinical setting, where the appropriate borrowing of historical knowledge can be critical. By providing a coherent statistical framework that incorporates all relevant sources of information, these methods can substantially reduce the required sample sizes, increase statistical power, and lower both costs and ethical risks.

Here, we focus on hierarchical models that employ commensurate priors (Hobbs et al., 2011, 2012) as the main mechanism to weight prior information according to its consistency (commensurability) with data from previous studies. Hobbs et al. (2011) consider the case in which data from a single historical study inform the analysis of a new study by defining the commensurate prior for the parameter of interest θ as follows:

$$\theta \mid \theta_0, \phi \sim \text{N}\left(\theta_0, \frac{1}{\phi^2}\right), \tag{8}$$

where θ_0 is the estimate from the historical study and ϕ^2 is the precision or commensurability parameter. The formulation in (8) follows from the insight in Pocock (1976) for which historical parameters may be biased representations of their current counterparts. By modelling the unknown bias as $\epsilon = \theta - \theta_0$, the commensurate prior quantifies how much a current study parameter is allowed to vary with respect to the historical estimate in the absence of strong evidence of heterogeneity. Thus, a lack of evidence for substantial bias implies that the historical and current parameters are commensurate.

Hobbs et al. (2012) extended this framework by proposing both empirical and fully Bayesian methods to estimate or assign ϕ^2. Notably, by incorporating prior uncertainty when estimating ϕ^2, the fully Bayesian approach reduces the risk of overstating commensurability. Hobbs et al. (2012) proposed two families of priors for ϕ^2, a family of gamma distributions that leads to a full conditional posterior distribution, as well as a variant of the 'spike-and-slab' distribution introduced by Mitchell and Beauchamp (1988) for Bayesian variable selection based on a mixture prior with two components, which can provide robust borrowing. Specifically, the spike-and-slab prior distribution is a discrete mixture distribution defined as locally uniform between two limits $0 \le \alpha_1 < \alpha_2$ (the slab component), and with a probability mass concentrated at a point $\mathcal{S} > \alpha_2$ (the spike component), such that

$$\begin{aligned} &\mathbb{P}(\phi^2 < \alpha_1) = 0, \\ &\mathbb{P}(\phi^2 < u) = (1 - p_s) \times \frac{u - \alpha_1}{\alpha_2 - \alpha_1}, \quad \alpha_1 \le u \le \alpha_2, \\ &\mathbb{P}(\phi^2 > \alpha_2) = \mathbb{P}(\phi^2 = \mathcal{S}) = p_s, \end{aligned} \tag{9}$$

where p_s is the probability of a spike, which can be interpreted as the prior probability of commensurability. The spike component concentrates the probability mass near θ_0, encouraging strong borrowing from historical data, while the slab component allows for greater deviation when current data conflict with historical evidence. Thus, commensurate priors provide a mechanism for selectively borrowing information from historical data by using the adaptive shrinkage properties of spike-and-slab distributions. Notably, when the current and historical parameters appear commensurate, the prior strongly shrinks the current parameter towards the historical estimate, improving efficiency. Conversely, when there is substantial disagreement, the prior has minimal influence on θ, limiting bias (Murray et al., 2015). Indeed, with appropriate calibration, the spike-and-slab commensurate prior approach achieves desirable frequentist properties—such as controlled Type I error and high power—while adaptively borrowing information when it is commensurate and downweighting it when it is not (Hobbs et al., 2012).

3.2 Weighted dynamic prior distributions

Let $\beta_{i,\tau}^{(k)}$ denote team T_i's ability of type k in period τ, where $k \in \{\text{att}, \text{def}\}$ and $\tau = 1, 2, \ldots, \mathcal{T}$ denotes the time period (e.g. seasons or weeks). In our weighted dynamic approach, each team's ability in period τ has a prior centred on its ability from the previous period $\tau - 1$, rather than assuming a fixed random walk precision across all periods. Therefore, for each team T_i, where $i = 1, \ldots, N_T$, and each period τ, where $\tau = 2, \ldots, \mathcal{T}$, the prior distributions for the attack and defence abilities are

$$\begin{aligned} \beta_{i,\tau}^{(\text{att})} \mid \beta_{i,\tau-1}^{(\text{att})}, \phi_{i,\tau}^{(\text{att})} &\sim \mathrm{N}\left(\beta_{i,\tau-1}^{(\text{att})}, \frac{1}{\left(\phi_{i,\tau}^{(\text{att})}\right)^2}\right), \\ \beta_{i,\tau}^{(\text{def})} \mid \beta_{i,\tau-1}^{(\text{def})}, \phi_{i,\tau}^{(\text{def})} &\sim \mathrm{N}\left(\beta_{i,\tau-1}^{(\text{def})}, \frac{1}{\left(\phi_{i,\tau}^{(\text{def})}\right)^2}\right), \end{aligned} \tag{10}$$

where each team's offensive (defensive) ability in period τ has a normal prior distribution centred on the offensive (defensive) ability of that team in period $\tau - 1$, with a commensurability (precision) parameter $(\phi_{i,\tau}^{(k)})^2$ that governs how closely the agreement is with previous information at time $\tau - 1$. If the commensurability parameter is large, the prior is tightly concentrated around the previous value—effectively assuming the team's ability has not changed much—which leads to a heavy borrowing of strength from the previous period. Conversely, if it is near zero, the prior is diffuse, indicating that we allow the current data to have a dominant influence while minimizing the contribution of the previous data. Notably, we introduce separate attack and defence precisions for each period τ and, similarly to Egidi et al. (2018), for each ability type, rather than a common evolution precision. This means that the model can adjust how much it learns from the attack strength of the previous period independently of how much it learns from the defence strength of the previous period.

To complete the model specification, we assign spike-and-slab hyperpriors to each precision parameter. Rather than using a discrete spike-and-slab with a point mass at $\mathcal{S}$ and a uniform slab on $[\alpha_1, \alpha_2]$ as in (9), we employ a continuous two-component mixture consisting of a highly concentrated spike and a diffuse slab (Hong et al., 2018). For each team T_i, where $i = 1, \ldots, N_T$, period and ability of type k, where $k \in \{\text{att}, \text{def}\}$, we let

$$\phi_{i,\tau}^{(k)} \mid \mu_s, \mu_l, \psi_s, \psi_l, p_{i,s} \sim \mathrm{N}^+(\mu_s, \psi_s^2) \times p_{i,s} + \mathrm{N}^+(\mu_l, \psi_l^2) \times (1 - p_{i,s}), \tag{11}$$

where $\mathrm{N}^+(\mu, \psi^2)$ denotes a normal distribution with mean μ and variance ψ^2 truncated from below at zero. Specifically, μ_s and μ_l represent the means of the spike-and-slab components, respectively, while ψ_s^2 and ψ_l^2 are the corresponding variances, with $0 < \psi_s^2 < \psi_l^2$. Each $\phi_{i,\tau}^{(k)}$, with $\tau \geq 2$, thus has a chance to be larger—implying a high commensurability with the previous period—or to be near zero—suggesting low commensurability and allowing more variability with respect to the previous period. The model will estimate an appropriate value for $\phi_{i,\tau}^{(\text{att})}$ and $\phi_{i,\tau}^{(\text{def})}$ based on the degree to which the new match results align with the trend of the previous period. If the performance of the teams in period τ looks very similar to that of period $\tau - 1$, the posterior for $\phi_{i,\tau}^{(k)}$ will likely favour the spike, implying strong borrowing and shrinkage towards the past. In contrast, if the performance of the teams changes unexpectedly, the posterior of $\phi_{i,\tau}^{(k)}$ will move towards the slab, implying weak borrowing. For the initial period $\tau = 1$, no past information is available, so we use diffuse but proper priors as in (6). Furthermore, to ensure the identifiability of the model, we impose the zero sum constraints of each period as in (7).

Figure 1 provides a schematic illustration of this adaptive mechanism. The upper panel displays the spike-and-slab prior for $\phi_{i,\tau}^{(k)}$, as in (11). The lower panels show the implied priors for the current ability $\beta_{i,\tau}^{(k)}$, centred on the previous period's ability $\beta_{i,\tau-1}^{(k)}$, indicated by the vertical dashed line. When the spike component is favoured, as shown in the right panel in green, the resulting tight prior induces substantial borrowing of information from the past. Conversely, when the slab component is favoured, as shown in the left panel in purple, the more diffuse prior allows the current data to have a greater influence, thus effectively discounting historical information.

Our weighted dynamic approach extends the Bayesian dynamic goal-based models framework by introducing adaptive time-specific shrinkage for team abilities. By allowing the data to decide how much information about attack and defence abilities to borrow from the previous period, the model can reflect real-world changes more responsively. This yields a more flexible evolution of team strengths over time, which should improve predictive performance on match outcomes.

4 Application

Our analysis uses data from the five most recent seasons (2020/2021 through 2024/2025) of three major European football leagues: the German Bundesliga, the English Premier League (EPL), and the Spanish La Liga. We compare the performance of our proposal with several existing approaches. First, we consider the discrete-time model of Owen (2011), which employs a single evolution precision shared between all teams and for both offensive and defensive abilities. We also include the extension proposed by Egidi et al. (2018), which allows for separate but time-invariant evolution precisions for attack and defence, while still assuming that these parameters are shared between teams. Finally, we examine models in the spirit of Koopman and Lit (2015), which account for seasonal heterogeneity by

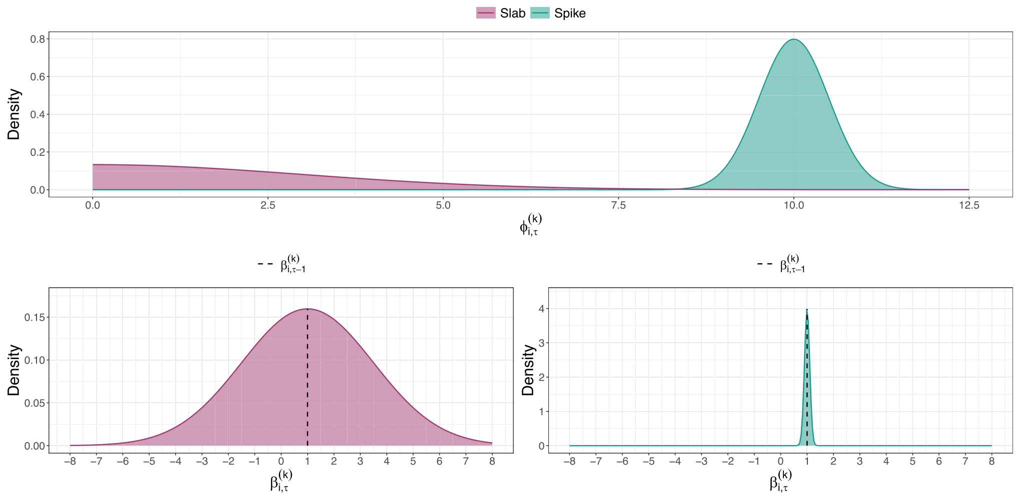

Figure 1. Continuous spike-and-slab commensurate prior mechanism example. Top panel shows the spike and slab priors for $\phi^{(k)}_{i,\tau}$. Bottom left panel displays the resulting prior for current ability $\beta^{(k)}_{i,\tau}$ under the spike scenario. Bottom right panel displays the resulting prior for $\beta^{(k)}_{i,\tau}$ under the slab scenario. The vertical dashed line in the bottom panels represents the previous ability $\beta^{(k)}_{i,\tau-1}$.

inflating the state innovation variance during periods associated with structural breaks, such as summer and winter breaks. Each season in our datasets is divided into two discrete-time periods based on the official football calendar, yielding a total of 10 time periods for analysis. Period boundaries are defined by key structural events within the season. Specifically, for the Bundesliga and La Liga, the first period consists of all matches from the start of the season up to the winter break, while the second period covers matches from the end of the winter break to the end of the season. For the EPL, which does not feature a traditional winter break, we use the opening of the winter transfer window (typically in early January) as the boundary between periods. For the 2022/2023 season, we account for the atypical schedule caused by the FIFA World Cup, which resulted in an extended break from mid-November to late December. This calendar-based partitioning is designed to capture midseason structural events that can significantly alter team composition and performance, including roster changes through transfers, managerial turnovers, and player recovery from injuries.

To comprehensively assess predictive performance, we consider prediction scenarios across the second half of each of the last two seasons in every league. For each half-season, all preceding data are used as the training set, allowing the models to incorporate the results up to the midpoint before generating forecasts. This setup provides a broad view that captures the cumulative impact of postmidseason changes (e.g. transfers, coaching changes, tactical adjustments) and tests model performance across a substantial number of matches under realistic conditions. By evaluating over two half-seasons rather than one, we also gain insight into the consistency of each model's predictive ability across different competitive contexts. An additional analysis focusing on the last round of the most recent season—representing the most uncertain scenario, in which teams have widely varying motivations and surprising outcomes are common—is reported in Appendix B.

The models are implemented using the probabilistic programming language Stan (Carpenter et al., 2017), employing Markov Chain Monte Carlo (MCMC) sampling via the R package `footBayes` (Egidi et al., 2025). For posterior sampling, we run four independent chains, each consisting of 4,000 iterations, with the initial 2,000 iterations discarded as burn-in. In the spike-and-slab formulation as in (11), the spike component has mean $\mu_s = 9$ and standard deviation $\psi_s = 1.5$, while the slab component has mean $\mu_l = 0$ and standard deviation $\psi_l = 3$, that is,

$$\phi^{(\text{att})}_{i,\tau}, \phi^{(\text{def})}_{i,\tau} \mid p_{i,s} \sim p_{i,s} \times \text{N}^+(9, 1.5^2) + (1 - p_{i,s}) \times \text{N}^+(0, 3^2),$$

where $p_{i,s}$ is the team-specific prior probability of drawing from the spike, which is assigned a uniform prior $p_{i,s} \sim \text{Beta}(1, 1)$. The chosen truncated-normal slab prior allows either slight borrowing or complete discounting of prior information when necessary. In contrast, the spike prior is structured to yield complete pooling when the past information aligns closely with current observations (Chen et al., 2018; Hobbs et al., 2012; Ouma et al., 2022; Zheng & Wason, 2022). A sensitivity analysis examining the robustness of these choices is provided in the online supplementary material. For the approaches proposed by Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015), the evolution standard deviations are modelled as

$$\sigma^{-1}, \sigma_{\text{att}}^{-1}, \sigma_{\text{def}}^{-1} \sim t_4^+(0, 1),$$

where $t_\nu^+(0, s)$ denotes the half Student-t distribution with ν degrees of freedom, location 0, and scale s. The half Student-t distribution with moderate degrees of freedom provides a flexible and weakly informative prior for scale parameters in hierarchical models, with advantageous behaviour near zero and minimal influence on posterior estimates (Gelman, 2006). Finally, for all models, weakly informative priors (Gelman et al., 2008) are assigned to the home-effect parameters

$$\text{home} \sim \text{N}(0, 5^2).$$

4.1 Models' assumptions

Before presenting the results, we examine important empirical properties of the data across all three leagues and five seasons that motivate our modelling choices. The diagnostics presented in Figure 2 motivated the inclusion of the five goal-based models considered in Section 2. Notably, Figure 2A displays the evolution of home advantage, measured as the mean goal difference between home and away teams. A positive home advantage is present across all leagues and seasons, though its magnitude varies considerably—most notably in the EPL, where home advantage was nearly absent in the 2020/2021 season, likely due to matches being played behind closed doors during the COVID-19 pandemic, before recovering in subsequent seasons. Figure 2B reports the dispersion index (variance-to-mean ratio) for home and away goals separately. Under the Poisson assumption, this ratio should equal one. The dispersion indices exceed one in some of the league/season combinations. The effect is most pronounced in the Bundesliga and in the first seasons of the EPL, where the ratio is above 1.2, while La Liga displays values closer to the equidispersion threshold. These patterns provide empirical support that it may be appropriate to consider models that account for possible overdispersion in the data, as described in Section 2.2. Figure 2C presents posterior predictive checks for draw frequency using the double Poisson model as a baseline. The observed proportion of draws (horizontal bars) is compared against the posterior predictive distribution of simulated draw frequencies (points). Some seasons exhibit low posterior predictive p-values, indicating that the standard Poisson model tends to underestimate the frequency of draws. Thus, it may be beneficial to consider a draw-inflated model as discussed in Section 2.3. Finally, Figure 2D shows the Pearson correlation between home and away goal counts. The correlations are predominantly negative across leagues and seasons, particularly in the Bundesliga and EPL. Since the bivariate Poisson model of Karlis and Ntzoufras (2003) only accommodates nonnegative dependence through its covariance parameter, this confirms that it is not a suitable specification for our data.

4.2 Predictive performance

One of the key aspects of sports analytics is the ability to generate accurate future predictions. Bayesian models naturally provide posterior probabilities for future matches. Considering the posterior predictive distribution for future observable data $\tilde{\mathcal{D}}$, we incorporate the predictive uncertainty of the model, propagated from the uncertainty of the posterior parameter. Predictions are generated by conditioning future observable values on the posterior parameter estimates:

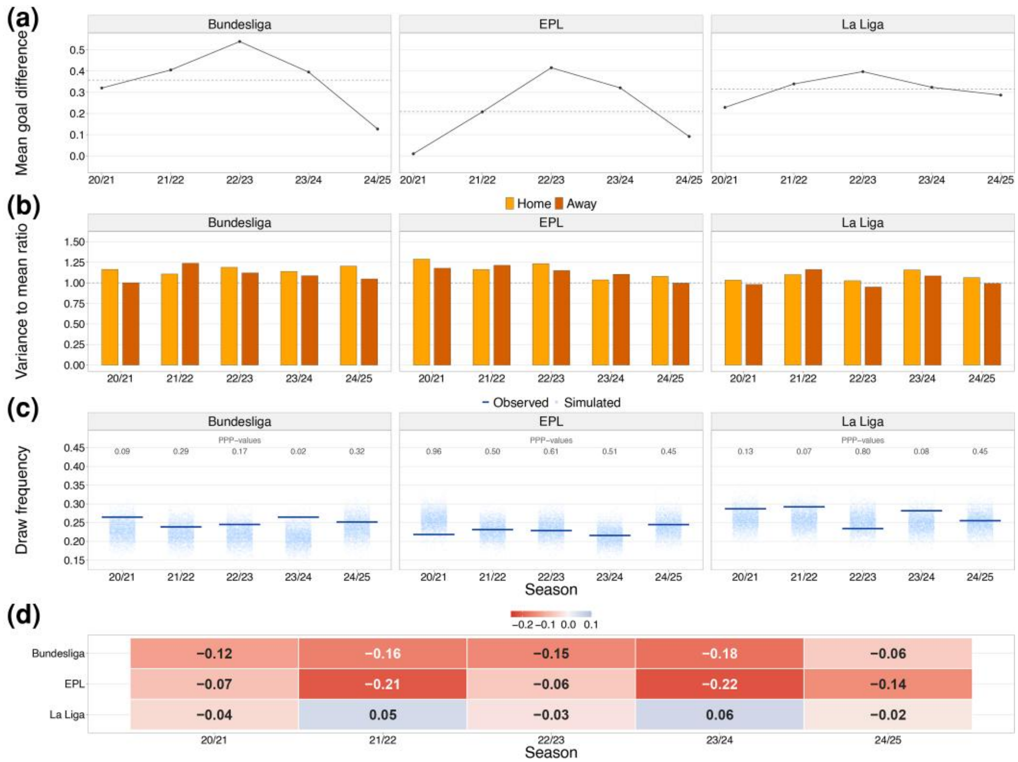

Figure 2. Descriptive statistics across the three evaluated leagues. (a) Mean home-away goal difference. Dashed lines indicate the five-season average. (b) Dispersion index (variance-to-mean ratio) for home and away goals; the dashed line denotes the Poisson equidispersion reference at one. (c) Posterior predictive checks for draw frequency under the double Poisson model: observed proportions (horizontal bars) versus posterior predictive distributions (points), with posterior predictive *p*-values (PPP) annotated. (d) Pearson correlation between home and away goals.

$$p(\tilde{\mathcal{D}}|\mathcal{D}) = \int p(\tilde{\mathcal{D}}|\theta)\pi(\theta|\mathcal{D})d\theta.$$

In practice, the predicted probability for each match outcome is computed as the proportion of MCMC samples that predict that outcome, thus approximating the integral above. After obtaining predictions from the models, evaluating their performance is crucial to assessing their predictive power and reliability. Specifically, we focus on two predictive metrics to rigorously examine the predictive performance of the models described in Section 2. Additional analyses with two other predictive metrics are provided in Appendix A.

4.2.1 Predictive metrics

The evaluation of probabilistic forecasts typically involves scoring rules metrics, which assess forecast performance by comparing predictions with the corresponding outcomes. The Brier score (Brier, 1950), recommended by Spiegelhalter and Ng (2009), is a nonlocal and distance-insensitive scoring rule, essentially acting as a mean squared error for forecasts. It is defined as

$$\text{Brier} = \frac{1}{M}\sum_{m=1}^{M}\sum_{r=1}^{3}(p_{r,m} - \delta_{r,m})^2,$$

where $p_{r,m}$ denotes the predicted probability of outcome r, with $r \in \{\text{home win}, \ \text{draw}, \ \text{away win}\}$, for the mth match played during the forecast period. Here, $\delta_{r,m}$ is the Kronecker delta, that is, 1 if the outcome r occurs in the mth match. The Brier score ranges from 0, indicating perfect prediction accuracy, to a maximum of 2 when predictions consistently assign probability 1 to incorrect outcomes.

For discrete outcomes, it can be beneficial for a scoring rule to account for the proximity or ordering of potential outcomes. In football, a draw is closer to a home win than an away win. The ranked probability score (RPS) (Epstein, 1969) is a distance-sensitive scoring measure that evaluates the degree to which the forecast probability distribution matches the observed outcome, assigning higher scores to probabilistic forecasts that allocate higher probabilities to outcomes near the actual result. The RPS is defined as

$$\text{RPS} = \frac{1}{3-1}\sum_{r=1}^{3-1}\left(\sum_{l=1}^{r} p_{l,m} - \sum_{l=1}^{r}\delta_{l,m}\right)^2.$$

As with the Brier score, lower values indicate better predictive performance.

Table 1 summarizes the Brier scores and RPSs for each of the five goal-based models, comparing our weighted dynamic forecasts with those of Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015) over the second half of 2023/2024 season. Among the three leagues, the weighted dynamic approach consistently achieves the lowest Brier score and RPS values, reflecting more accurate predictions. In the Bundesliga, our weighted dynamic approach achieves the lowest Brier score of 0.596 under the Skellam model and the lowest RPS of 0.195 under the NB model, improving upon all competing methods. For instance, compared to the models in the spirit of Koopman and Lit (2015), the SM reduces the Brier score from 0.602 to 0.596. In the EPL, the weighted dynamic models again outperform the other approaches. The weighted dynamic Dixon and Coles specification yields the lowest Brier score of 0.540 and the lowest RPS of 0.177, compared to 0.544 and 0.180 under Owen (2011), and 0.549 and 0.182 under the method proposed by Egidi et al. (2018). In La Liga, the differences across methods are comparatively small. The weighted dynamic approach achieves the lowest Brier score of 0.580 under the NB model, matching the performance of the Dixon and Coles specification in Koopman and Lit (2015) and the DP model under Egidi et al. (2018). Regarding the RPS, the best value of 0.195 is shared by the weighted dynamic DP model and the Koopman and Lit (2015) specification under the NB model.

Table 2 presents the corresponding comparison for the second half of the 2024/2025 season. Overall, the weighted dynamic framework again demonstrates strong predictive performance across all three leagues. In the Bundesliga, the weighted dynamic SM achieves the best predictive performance, achieving the lowest Brier score of 0.627 and the lowest RPS of 0.213 among all competing approaches. Compared with Egidi et al. (2018), this represents a reduction in the Brier score from 0.630 to 0.627 and in the RPS from 0.215 to 0.213. In the EPL, the weighted dynamic DP model performs best, yielding the lowest Brier score of 0.576 and the lowest RPS of 0.204. These values improve upon the results reported by Owen (2011), which are 0.578 and 0.205, by Egidi et al. (2018), which are 0.579 and 0.205, and by Koopman and Lit (2015), which are 0.579 and 0.205, respectively. Similarly, in La Liga, the weighted dynamic Dixon and Coles model achieves the best results, with a Brier score of 0.576 and an RPS of 0.197. This clearly outperforms the models of Owen (2011), which obtain values of 0.584 and 0.201, and Egidi et al. (2018), which report corresponding scores of 0.583 and 0.200.

An additional comparison with the betting market implied that probabilities are provided in Appendix C. Furthermore, since the goal-based framework produces full predictive distributions for the number of goals scored rather than just final match outcomes, we also evaluate score-level predictive accuracy for the predicted goal difference. The results are reported in the online supplementary material.

4.3 Team abilities

One of the crucial aspects of our proposal is how the weighted dynamic approach in (10) influences the evolution of the attacking and defensive abilities of the teams in the evaluated periods. Figure 3 illustrates the trajectories of these abilities for the best performing model identified in Section 4.2, when forecasting the second half of the 2024/2025 season. Specifically, for the Bundesliga, the best model is the SM, for the EPL the DP, and for La Liga the Dixon and Coles model. Within each league, we selected

Table 1. Brier score and ranked probability score (RPS) for the proposed weighted dynamic method, Owen (2011) method, Egidi et al. (2018), and Koopman and Lit (2015) method, evaluated on the second half of 2023/2024 season for the Bundesliga, English Premier League (EPL), and La Liga

League	Model	Weighted dynamic		Owen (2011)		Egidi et al. (2025)		Koopman and Lit (2015)	
		Brier score	RPS	Brier score	RPS	Brier score	RPS	Brier score	RPS
Bundesliga	Dixon and Coles	0.604	0.199	0.614	0.204	0.614	0.204	0.614	0.204
	Double Poisson	0.604	0.199	0.614	0.204	0.614	0.204	0.613	0.203
	Negative binomial	0.598	0.195	0.617	0.205	0.614	0.204	0.614	0.204
	Skellam model	0.596	0.197	0.603	0.201	0.601	0.199	0.602	0.200
	Zero-infl. Skellam model	0.599	0.198	0.606	0.202	0.603	0.200	0.602	0.200
EPL	Dixon and Coles	0.540	0.177	0.544	0.180	0.549	0.182	0.548	0.182
	Double Poisson	0.541	0.178	0.546	0.180	0.547	0.181	0.548	0.182
	Negative binomial	0.543	0.179	0.547	0.181	0.550	0.183	0.551	0.183
	Skellam model	0.554	0.184	0.553	0.184	0.556	0.186	0.555	0.186
	Zero-infl. Skellam model	0.556	0.185	0.553	0.184	0.555	0.186	0.557	0.186
La Liga	Dixon and Coles	0.581	0.196	0.581	0.196	0.581	0.196	0.580	0.196
	Double Poisson	0.581	0.195	0.581	0.196	0.580	0.196	0.582	0.196
	Negative binomial	0.580	0.196	0.583	0.196	0.582	0.196	0.580	0.195
	Skellam model	0.597	0.202	0.594	0.201	0.590	0.200	0.590	0.200
	Zero-infl. Skellam model	0.597	0.202	0.592	0.200	0.588	0.200	0.591	0.200

two teams: one with relatively stable performance during the study period and another showing inconsistent behaviour. The plot shows that, for stable teams, the weighted dynamic model captures the temporal trends more accurately, with clearer distinctions between attacking and defensive strength. For instance, Bayern Munich, typically dominant in the Bundesliga, showed a slight slump in performance during the 2023/2024 season (23/24-H1 and 23/24-H2) when they finished third. This fluctuation is reflected in the weighted dynamic model, which shows a drop in attacking ability alongside a noticeable increase in defensive vulnerability compared to earlier periods. The competing approaches of Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015) display smoother trajectories that largely mask this situation. Similarly, for Real Madrid and Manchester City, both of which won their respective leagues in 2023/2024 but finished second and third in 2024/2025 (24/25-H1 and 24/25-H2), the weighted dynamic approach captures a noticeable drop in attacking ability and a relative increase in defensive ability. These changes are more distinctly represented in our model compared to the alternatives of Owen (2011) and Egidi et al. (2018).

The benefits of adaptive shrinkage are even more evident for teams with inconsistent performance. For instance, Manchester United, after finishing second in 2020/2021, showed a gradual decline in subsequent seasons, finally placing 15th in the 2024/2025 season. The weighted dynamic model reflects this trend with a notable reduction in attacking ability, while the defence parameter has risen above zero, indicating growing defensive vulnerability. Similarly, Girona FC experienced an exceptional 2023/2024 season, reaching a third place in La Liga, followed by a disappointing 16th place finish in 2024/2025. The weighted dynamic approach effectively captures this volatility, showing the highest attacking ability during the best season (23/24-H1 and 23/24-H2) and a marked decline thereafter. The model even shows a crossover point in the final period (24/25-H2), where Girona's defensive vulnerability is higher than its attacking ability, highlighting a shift in team abilities that other models fail to capture.

Table 2. Brier score and ranked probability score (RPS) for the proposed weighted dynamic method, Owen (2011) method, Egidi et al. (2018), and Koopman and Lit (2015) method, evaluated on the second half of 2024/2025 season for the Bundesliga, English Premier League (EPL), and La Liga

League	Model	Weighted Dynamic		Owen (2011)		Egidi et al. (2025)		Koopman and Lit (2015)	
		Brier score	RPS	Brier score	RPS	Brier score	RPS	Brier score	RPS
Bundesliga	Dixon and Coles	0.629	0.214	0.644	0.221	0.639	0.219	0.640	0.219
	Double Poisson	0.630	0.214	0.642	0.220	0.641	0.219	0.639	0.219
	Negative binomial	0.629	0.214	0.642	0.220	0.640	0.219	0.639	0.219
	Skellam model	0.627	0.213	0.633	0.217	0.630	0.215	0.629	0.214
	Zero-infl. Skellam model	0.631	0.215	0.633	0.217	0.630	0.215	0.629	0.214
EPL	Dixon and Coles	0.577	0.205	0.579	0.205	0.579	0.205	0.578	0.205
	Double Poisson	0.576	0.204	0.578	0.205	0.579	0.205	0.579	0.205
	Negative binomial	0.577	0.205	0.579	0.206	0.578	0.205	0.578	0.205
	Skellam model	0.595	0.212	0.598	0.212	0.601	0.214	0.599	0.212
	Zero-infl. Skellam model	0.596	0.212	0.598	0.212	0.599	0.213	0.601	0.214
La Liga	Dixon and Coles	0.576	0.197	0.584	0.201	0.583	0.200	0.583	0.200
	Double Poisson	0.583	0.200	0.585	0.201	0.582	0.200	0.583	0.200
	Negative binomial	0.584	0.201	0.585	0.202	0.584	0.201	0.585	0.201
	Skellam model	0.586	0.202	0.588	0.203	0.591	0.205	0.590	0.204
	Zero-infl. Skellam model	0.585	0.201	0.588	0.203	0.592	0.205	0.590	0.205

5 Discussion

This work introduced a Bayesian weighted dynamic framework for football predictions that flexibly models the evolution of team-specific abilities over time. By using commensurate priors with spike-and-slab hyperparameters, our approach allows each team's attack and defence strength in a given period to adaptively borrow information from past performance. This yields a more responsive and nuanced dynamic model compared to previous approaches that assume static abilities or a single constant evolution precision. Among five different goal-based distributions and three major European leagues, we found that this adaptive shrinkage mechanism leads to consistent improvements in predictive accuracy. The weighted dynamic models captured team performance trajectories more realistically—for instance, they sharply reflected midseason form fluctuations and major transitions—while maintaining or improving prediction accuracy. Furthermore, all models achieved satisfactory convergence across every league and prediction scenario; detailed diagnostics are reported in Appendix D and in the online supplementary material.

Because each season in our analysis is divided into two calendar-based periods, transitions between seasons involve substantial temporal gaps of three to four months and often coincide with major squad changes during the summer transfer window, while transitions within the season are typically more gradual (Koopman & Lit, 2015). The proposed commensurate prior framework is designed to accommodate this heterogeneity automatically. By allowing the data to determine the extent of information borrowing through the commensurability parameters, the model can downweight historical information when team composition changes substantially. Koopman and Lit (2015) addressed this issue by introducing additional variance parameters that inflate the innovation variance at prespecified break points corresponding to summer breaks. Although effective, this approach requires the estimation of an additional parameter specific to those periods. In contrast, the commensurate prior framework proposed here adaptively accounts for structural discontinuities without introducing such break-specific parameters. The spike-and-slab hyperprior on each period's precision parameter allows the model to learn from the data when historical information should be discounted.

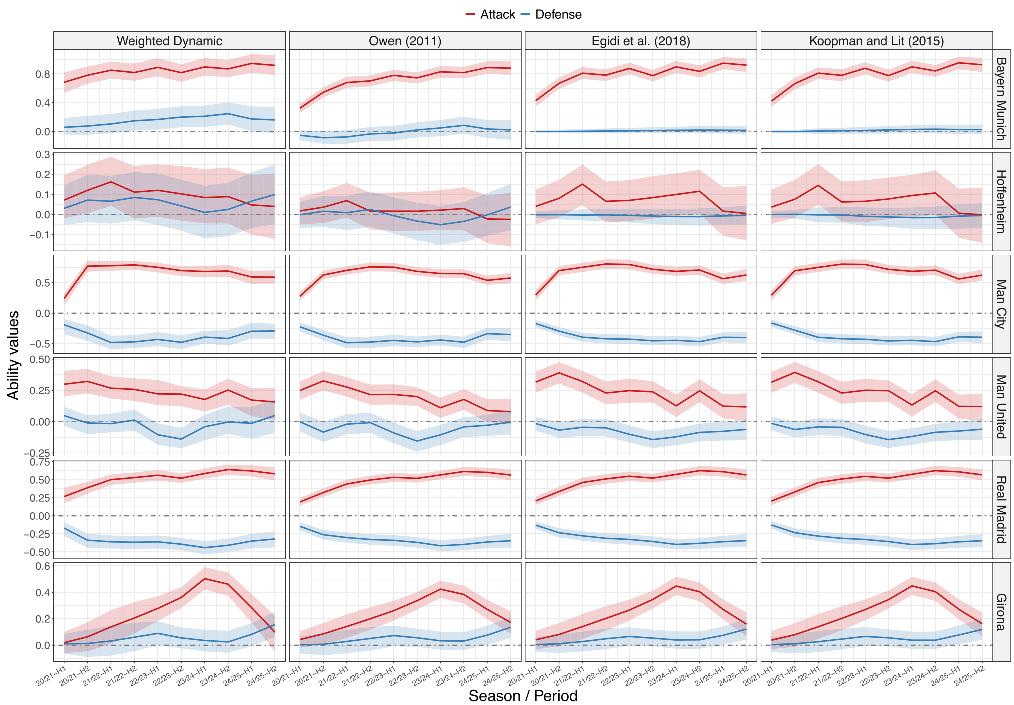

Figure 3. Trajectories of estimated attacking and defensive abilities with their 50% credible intervals over ten periods for two representative teams in each league. Each season is split into two half-season periods: H1 (from the start of the season to the winter break) and H2 (from the winter break to the end of the season). Results from the proposed weighted dynamic method are shown alongside corresponding estimates from Owen (2011) and Egidi et al. (2018), and Koopman and Lit (2015).

Consequently, between-season transitions for teams undergoing substantial changes due to summer transfers naturally receive lower precision estimates, favouring the slab component, whereas more stable teams receive higher precision estimates, favouring the spike component. Importantly, this adaptive behaviour emerges without the need to introduce additional parameters.

It should be noted that the extended parametrization, which introduces team-specific and time-specific precision parameters, implies a moderate increase in computational cost relative to the approaches of Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015). However, this additional computational effort—well within practical limits—is more than compensated by the consistent improvements in predictive performance observed across all leagues and prediction scenarios.

In this article, we used only past match results to infer team abilities; however, information such as team market value, recent injuries, or even in-game statistics (shots, expected goals) could be included as covariates influencing the scoring rate parameters. Future extensions might also consider team-specific or hierarchical evolution parameters so that traditionally inconsistent teams are allowed more variation than stable teams.

Another important direction is the integration with result-based models and other comparative approaches. Although we focused on goal-based distributions (which naturally yield the three-way process as a consequence), our methodology could be extended to models that predict match outcomes directly. For instance, in an ordered probit/logit model or a multiclass logistic model for win–draw–loss, one could let each team's latent strength parameter vary over time using the same weighted random-walk prior. In addition, a weighted dynamic Bradley–Terry–Davidson model for paired comparisons is a natural extension. Our approach could provide a fully Bayesian weighted dynamic Bradley–Terry–Davidson model by allowing each team's strength to evolve with our weighted dynamic approach. Conceptually, this would let the probability of one team beating another adapt rapidly after major changes (e.g. if a traditionally weak team suddenly improves, the model would downweight its past information). We expect that such a model would be computationally even simpler (since it has only one strength per team rather than separate attack/defence), yet still benefit from our weighted dynamic approach.

We emphasize that the Bayesian weighted dynamic approach presented here is quite general and may find use in other sport domains. Many sports and competitive systems (e.g. basketball, volleyball, or handball) involve teams whose skills change over time. By calibrating the commensurate prior to the specific domain, our strategy of time-specific shrinkage could be applied wherever one has sequential performance data and expects occasional shifts in the underlying ability. Furthermore, while our focus here has been on domestic leagues with complete round-robin schedules, future research will turn to high-profile tournaments, including group stages followed by knockout rounds or hybrid formats such as the UEFA Champions League, FIFA World Cup, and UEFA European Championship. Finally, the proposed method is implemented in the free and open source R package `footBayes`.

Acknowledgments

We express our sincere gratitude to the Associate Editor and the two anonymous reviewers for their constructive comments and helpful suggestions, which greatly improved the quality of this manuscript.

Conflicts of interest

None declared.

Funding

This work has been supported by the project SMARTsports: ‘Statistical Models and AlgoRiThms in sports. Applications in professional and amateur contexts, with able-bodied and disabled athletes’, funded by the MIUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2022—Grant No. 2022R74PLE (CUP J53D23003860006).

Data availability

All analyses were performed in the R programming language version 4.4.3 (R Core Team, 2025). Data are freely available online at football-data.co.uk. The code for reproducing this manuscript is openly available at https://github.com/RoMaD-96/BayesWDFM. The proposed methodology has also been implemented in the free and open source R package `footBayes` (version 2.1.0).

Supplementary material

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

Appendix A Average of correct probabilities and *pseudo-R^2*

While proper scoring rules penalize squared errors, mean-based metrics provide direct, interpretable summaries of predictive accuracy. The average of correct probabilities (ACP) is defined as the arithmetic mean of the probabilities assigned to outcomes that actually occurred, that is

$$\text{ACP} = \frac{1}{M}\sum_{m=1}^{M} p_{o,m},$$

where $p_{o,m}$ is the probability assigned to the observed outcome of the mth match. Being an arithmetic mean, the ACP measures the confidence of the average forecast directly on the original probability scale. ACP values near 1 indicate that the model consistently assigns high probabilities to the true outcomes, whereas values near 0 reflect weaker predictive performance.

The pseudo-R^2 (Dobson et al., 2001) is defined as the geometric mean of the probabilities assigned to the actual result of each match:

$$\text{Pseudo-}R^2 = \left(\prod_{m=1}^{M} p_{o,m}\right)^{1/M}.$$

The geometric mean penalizes low-probability predictions more severely than the arithmetic mean. Similarly to ACP, a pseudo-R^2 close to 1 indicates high predictive accuracy, while values approaching 0 suggest weaker performance.

Table A1. Average of correct probabilities (ACP) and pseudo-R^2 for the proposed weighted dynamic method, Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015) methods, evaluated on the second half of 2023/2024 season for the Bundesliga, English Premier League (EPL), and La Liga

League	Model	Weighted dynamic		Owen (2011)		Egidi et al. (2025)		Koopman and Lit (2015)	
		ACP	Pseudo-R^2	ACP	Pseudo-R^2	ACP	Pseudo-R^2	ACP	Pseudo-R^2
Bundesliga	Dixon and Coles	0.425	0.365	0.425	0.360	0.421	0.360	0.422	0.361
	Double Poisson	0.424	0.365	0.424	0.360	0.421	0.360	0.422	0.361
	Negative binomial	0.432	0.369	0.417	0.359	0.415	0.360	0.416	0.361
	Skellam model	0.414	0.370	0.412	0.367	0.414	0.367	0.413	0.367
	Zero-infl. Skellam model	0.414	0.368	0.412	0.365	0.414	0.366	0.414	0.367
EPL	Dixon and Coles	0.448	0.400	0.447	0.399	0.445	0.396	0.446	0.396
	Double Poisson	0.446	0.399	0.447	0.398	0.446	0.397	0.445	0.396
	Negative binomial	0.442	0.398	0.444	0.397	0.441	0.394	0.440	0.394
	Skellam model	0.422	0.390	0.425	0.392	0.422	0.390	0.423	0.391
	Zero-infl. Skellam model	0.421	0.389	0.424	0.392	0.423	0.391	0.422	0.390
La Liga	Dixon and Coles	0.415	0.377	0.414	0.377	0.414	0.377	0.414	0.377
	Double Poisson	0.415	0.377	0.414	0.377	0.414	0.377	0.414	0.377
	Negative binomial	0.412	0.377	0.410	0.376	0.410	0.376	0.411	0.378
	Skellam model	0.386	0.368	0.386	0.370	0.388	0.372	0.388	0.372
	Zero-infl. Skellam model	0.386	0.369	0.387	0.371	0.390	0.373	0.388	0.371

Table A2. Average of correct probabilities (ACP) and pseudo-R^2 for the proposed weighted dynamic method, Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015) methods, evaluated on the second half of 2024/2025 season for the Bundesliga, English Premier League (EPL), and La Liga

League	Model	Weighted dynamic		Owen (2011)		Egidi et al. (2025)		Koopman and Lit (2015)	
		ACP	Pseudo-R^2	ACP	Pseudo-R^2	ACP	Pseudo-R^2	ACP	Pseudo-R^2
Bundesliga	Dixon and Coles	0.398	0.351	0.396	0.344	0.398	0.346	0.398	0.346
	Double Poisson	0.399	0.351	0.396	0.345	0.396	0.346	0.398	0.346
	Negative binomial	0.396	0.351	0.393	0.345	0.393	0.346	0.393	0.346
	Skellam model	0.383	0.352	0.381	0.348	0.383	0.350	0.383	0.350
	Zero-infl. Skellam model	0.383	0.350	0.382	0.349	0.384	0.350	0.384	0.351
EPL	Dixon and Coles	0.420	0.381	0.422	0.380	0.422	0.380	0.423	0.380
	Double Poisson	0.421	0.381	0.423	0.380	0.422	0.380	0.422	0.380
	Negative binomial	0.418	0.380	0.420	0.379	0.419	0.380	0.420	0.380
	Skellam model	0.395	0.370	0.395	0.369	0.394	0.367	0.395	0.368
	Zero-infl. Skellam model	0.395	0.370	0.395	0.369	0.395	0.368	0.394	0.367
La Liga	Dixon and Coles	0.421	0.378	0.416	0.374	0.417	0.375	0.417	0.375
	Double Poisson	0.418	0.375	0.416	0.374	0.417	0.375	0.417	0.375
	Negative binomial	0.414	0.374	0.412	0.374	0.414	0.374	0.413	0.374
	Skellam model	0.399	0.374	0.395	0.372	0.394	0.370	0.396	0.371
	Zero-infl. Skellam model	0.400	0.374	0.396	0.372	0.395	0.370	0.396	0.371

Table A1 reports the ACP and Pseudo-R^2 values for the five goal-based models, comparing the proposed weighted dynamic forecasts with those of Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015) over the second half of the 2023/2024 season. In the Bundesliga, the weighted dynamic specification obtains the highest ACP of 0.432 under the negative binomial model and the highest Pseudo-R^2 of 0.370 under the SM, exceeding the corresponding values from all competing approaches. In the EPL, the weighted dynamic Dixon and Coles model achieves the best overall performance, with an ACP of 0.448 and a Pseudo-R^2 of 0.400, improving on Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015). In La Liga, differences are negligible across methods. The weighted dynamic DP and Dixon and Coles models share the highest ACP at 0.415, while the largest Pseudo-R^2 is obtained by the Koopman and Lit (2015) NB specification at 0.378, closely followed by the weighted dynamic NB model at 0.377.

Table A2 presents the corresponding ACP and Pseudo-R^2 values for the second half of the 2024/2025 season. In the Bundesliga, the highest ACP of 0.399 is obtained by the weighted dynamic DP model, while the weighted dynamic SM achieves the highest Pseudo-R^2 of 0.352. These results improve upon all competing methods; for instance, Owen (2011) achieves at most 0.396 and 0.348 for the ACP and Pseudo-R^2, respectively. In the EPL, the weighted dynamic DP and Dixon and Coles models share the highest Pseudo-R^2 of 0.381, while the highest ACP of 0.432 is achieved by both the DP of Owen (2011) and the Dixon and Coles model of Koopman and Lit (2015). In La Liga, the weighted dynamic Dixon and Coles model clearly outperforms all alternatives, yielding the highest ACP of 0.421 and the highest Pseudo-R^2 of 0.378, compared to 0.416 and 0.374 under Owen (2011) and 0.417 and 0.375 under Egidi et al. (2018).

Appendix B Last round prediction

As a complementary analysis to the half-season forecasts presented in Section 4.2, we examine an additional prediction scenario: forecasting only the final round of the 2024/2025 season. This

Table B1. Brier score and ranked probability score (RPS) for the proposed weighted dynamic method, Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015) methods, evaluated on the last round of the 2024/2025 season for the Bundesliga, English Premier League (EPL), and La Liga

League	Model	Weighted dynamic		Owen (2011)		Egidi et al. (2025)		Koopman and Lit (2015)	
		Brier score	RPS	Brier score	RPS	Brier score	RPS	Brier score	RPS
Bundesliga	Dixon and Coles	0.552	0.218	0.586	0.234	0.572	0.227	0.571	0.226
	Double Poisson	0.549	0.218	0.584	0.233	0.572	0.227	0.568	0.226
	Negative binomial	0.555	0.220	0.584	0.234	0.575	0.228	0.577	0.229
	Skellam model	0.575	0.231	0.600	0.241	0.587	0.232	0.588	0.234
	Zero-infl. Skellam model	0.575	0.232	0.596	0.240	0.591	0.235	0.592	0.235
EPL	Dixon and Coles	0.558	0.198	0.573	0.204	0.580	0.207	0.575	0.205
	Double Poisson	0.555	0.197	0.574	0.204	0.571	0.203	0.583	0.209
	Negative binomial	0.554	0.196	0.577	0.207	0.581	0.209	0.580	0.208
	Skellam model	0.538	0.194	0.564	0.204	0.577	0.209	0.581	0.209
	Zero-infl. Skellam model	0.537	0.192	0.560	0.202	0.577	0.208	0.578	0.209
La Liga	Dixon and Coles	0.482	0.146	0.491	0.151	0.491	0.149	0.488	0.148
	Double Poisson	0.483	0.147	0.494	0.152	0.491	0.150	0.488	0.149
	Negative binomial	0.486	0.148	0.501	0.154	0.500	0.153	0.497	0.152
	Skellam model	0.508	0.166	0.525	0.171	0.530	0.172	0.532	0.174
	Zero-infl. Skellam model	0.511	0.166	0.532	0.173	0.533	0.172	0.532	0.172

Table B2. Average of correct probabilities (ACP) and pseudo-R^2 for the proposed weighted dynamic method, Owen (2011), Egidi et al. (2018), and Koopman and Lit (2015) methods, evaluated on the last round of the 2024/2025 season for the Bundesliga, English Premier League (EPL), and La Liga

League	Model	Weighted dynamic		Owen (2011)		Egidi et al. (2025)		Koopman and Lit (2015)	
		ACP	Pseudo-R^2	ACP	Pseudo-R^2	ACP	Pseudo-R^2	ACP	Pseudo-R^2
Bundesliga	Dixon and Coles	0.426	0.392	0.409	0.372	0.414	0.382	0.415	0.382
	Double Poisson	0.428	0.395	0.410	0.374	0.414	0.382	0.416	0.383
	Negative binomial	0.422	0.391	0.408	0.373	0.409	0.380	0.409	0.378
	Skellam model	0.401	0.378	0.388	0.364	0.396	0.371	0.395	0.370
	Zero-infl. Skellam model	0.402	0.379	0.391	0.366	0.394	0.369	0.394	0.368
EPL	Dixon and Coles	0.434	0.391	0.427	0.383	0.423	0.379	0.424	0.382
	Double Poisson	0.436	0.392	0.426	0.382	0.426	0.384	0.421	0.378
	Negative binomial	0.435	0.391	0.422	0.381	0.419	0.378	0.419	0.379
	Skellam model	0.429	0.404	0.417	0.391	0.408	0.381	0.406	0.378
	Zero-infl. Skellam model	0.431	0.404	0.419	0.393	0.407	0.381	0.408	0.380
La Liga	Dixon and Coles	0.457	0.432	0.451	0.428	0.451	0.427	0.453	0.429
	Double Poisson	0.457	0.431	0.449	0.426	0.452	0.427	0.453	0.429
	Negative binomial	0.453	0.429	0.445	0.421	0.444	0.422	0.446	0.423
	Skellam model	0.427	0.417	0.417	0.407	0.414	0.405	0.412	0.404
	Zero-infl. Skellam model	0.426	0.416	0.413	0.404	0.412	0.403	0.412	0.403

represents the most uncertain setting, as teams have widely varying motivations in the closing round—some are competing for championships, European qualification, or survival from relegation, while others have little at stake—which often leads to surprising outcomes. By isolating this single round, we can assess how well each model adapts to the heightened unpredictability that characterizes the end of the season.

Table B1 reports the Brier score and RPS for the five goal-based models under the four dynamic approaches. In the Bundesliga, the weighted dynamic DP model achieves the lowest Brier score of 0.549 and the lowest RPS of 0.218, improving substantially upon all competing methods. In the EPL, the weighted dynamic ZISM yields the lowest Brier score of 0.537 and the lowest RPS of 0.192, outperforming all alternatives. In La Liga, the weighted dynamic Dixon and Coles model obtains the lowest Brier score of 0.482 and the lowest RPS of 0.146. Overall, the gains from the weighted dynamic framework are particularly pronounced in this single-round scenario, where the ability to adjust quickly to recent changes is most valuable.

Table B2 presents the corresponding ACP and Pseudo-R^2 values. In the Bundesliga, the weighted dynamic DP model achieves the highest ACP of 0.428 and the highest Pseudo-R^2 of 0.395, exceeding the corresponding values from all competing approaches. In the EPL, the weighted dynamic DP model delivers the highest ACP of 0.436, while the highest Pseudo-R^2 of 0.404 is achieved by the weighted dynamic SM, improving on Owen (2011) (0.389), Egidi et al. (2018) (0.381), and Koopman and Lit (2015) (0.379). In La Liga, the weighted dynamic Dixon and Coles model yields the highest ACP of 0.457 and the highest Pseudo-R^2 of 0.432.

Appendix C Comparison with betting market probabilities

To provide an additional external benchmark, we compare the predictive performance of the best weighted dynamic model for each league and scenario against betting market implied probabilities. Specifically, we use prematch odds from bookmakers available on football-data.co.uk, converting decimal odds to implied probabilities via basic normalization (i.e. dividing each implied probability by the sum of the three implied probabilities to remove the bookmaker's overround). We report results for a market average computed as the elementwise mean of the normalized probabilities across all bookmakers with complete coverage on each test set. It is important to note that betting odds incorporate a substantially richer information set than the models considered in this article: bookmakers exploit detailed squad information, injury reports, tactical matchups, market sentiment, and proprietary models. Consequently, outperforming the betting market is not expected and is not the goal of this comparison. Rather, we include this benchmark to contextualize the predictive accuracy of the proposed framework relative to a well-calibrated external source.

Table C1 reports the Brier score and RPS for the best weighted dynamic model in each league/scenario combination alongside the market average benchmark. In the 2023/2024 season, the market average achieves a Brier score of 0.580 and an RPS of 0.189 in the Bundesliga, whereas the best weighted dynamic model in terms of Brier score, namely, the SM, obtains 0.596, and the best model in terms of RPS, namely, the NB model, obtains 0.195. In the EPL, the gap is larger: the market average achieves a Brier score of 0.513 and an RPS of 0.167, compared with 0.540 and 0.177 for the best weighted dynamic model, the Dixon and Coles specification. In La Liga, the difference is narrower, with the weighted dynamic NB model achieving a Brier score of 0.580 compared with 0.572 for the market average and the weighted dynamic DP model obtaining an RPS of 0.195 compared with 0.191 for the market average. In the 2024/2025 season, a similar pattern emerges. In the Bundesliga, the weighted dynamic Skellam model achieves a Brier score of 0.627 and an RPS of 0.213, compared with 0.621 and 0.212 for the market average. In the EPL, the market average obtains a Brier score of 0.565 and an RPS of 0.198, while the best weighted dynamic model, the DP specification, achieves 0.576 and 0.204. In La Liga, the best weighted dynamic model, Dixon and Coles, obtains a Brier score of 0.576 and an RPS of 0.197, compared with 0.555 and 0.188 for the market average. For the last round of the 2024/2025 season, a notably different pattern emerges. In the Bundesliga, the best weighted dynamic model, the DP

Table C1. Brier score and ranked probability score (RPS) for the best weighted dynamic model in each league compared with the bookmaker market average, evaluated on the second half of the 2023/2024 and 2024/2025 seasons and on the last round of the 2024/2025 season

		Weighted dynamic		Market average	
Scenario	**League**	**Brier**	**RPS**	**Brier**	**RPS**
Half season 2023/2024	Bundesliga	0.596	0.195	0.580	0.189
	EPL	0.540	0.177	0.513	0.167
	La Liga	0.580	0.195	0.572	0.191
Half season 2024/2025	Bundesliga	0.627	0.213	0.621	0.212
	EPL	0.576	0.204	0.565	0.198
	La Liga	0.576	0.197	0.555	0.188
Last round 2024/2025	Bundesliga	0.549	0.218	0.634	0.263
	EPL	0.537	0.192	0.605	0.222
	La Liga	0.482	0.146	0.439	0.120

Table C2. Average of correct probabilities (ACP) and pseudo-R^2 for the best weighted dynamic model in each league compared with the bookmaker market average, evaluated on the second half of the 2023/2024 and 2024/2025 seasons and on the last round of the 2024/2025 season

		Weighted dynamic		Market average	
Scenario	**League**	**ACP**	**pseudo-R^2**	**ACP**	**pseudo-R^2**
Half season 2023/2024	Bundesliga	0.432	0.370	0.425	0.377
	EPL	0.448	0.400	0.464	0.414
	La Liga	0.415	0.377	0.415	0.383
Half season 2024/2025	Bundesliga	0.399	0.352	0.399	0.356
	EPL	0.421	0.381	0.431	0.386
	La Liga	0.421	0.378	0.433	0.391
Last round 2024/2025	Bundesliga	0.428	0.395	0.394	0.353
	EPL	0.436	0.404	0.423	0.357
	La Liga	0.457	0.432	0.500	0.462

specification, achieves a Brier score of 0.549 and an RPS of 0.218, outperforming the market average of 0.634 and 0.263 by a substantial margin. A similar reversal is observed in the EPL, where the best weighted dynamic model, the Skellam specification, obtains a Brier score of 0.537 and an RPS of 0.192, compared with 0.605 and 0.222 for the market average. In La Liga, the best weighted dynamic model, Dixon and Coles, achieves a Brier score of 0.482 and an RPS of 0.146, compared with 0.439 and 0.120 for the market average, the only scenario in which the market average holds a slight advantage on the RPS.

Table C2 presents the corresponding comparison in terms of ACP and pseudo-R^2. These metrics largely confirm the pattern observed in Table C1. In the 2023/2024 season, the Bundesliga presents an interesting case. The best weighted dynamic model, the NB specification, achieves a higher ACP of 0.432 than the market average of 0.425, suggesting that the model assigns a higher average probability to the correct outcome, while the market average holds a clear advantage in pseudo-R^2, with 0.377 against 0.360 for the weighted dynamic SM. In the EPL, the market average outperforms more decisively, with an ACP of 0.464 and a pseudo-R^2 of 0.414, compared with 0.448 and 0.400 for the best weighted dynamic model. In La Liga, the two approaches are nearly indistinguishable on the ACP, with both achieving 0.415, while the market average holds a slight advantage in pseudo-R^2, recording 0.383

Table D1. MCMC convergence diagnostics: $\hat{R}$, bulk and tail effective sample sizes (ESS) mean of the $\boldsymbol{\beta}^{(\text{att})}$, $\boldsymbol{\beta}^{(\text{def})}$, $\boldsymbol{\phi}^{(\text{att})}$ and $\boldsymbol{\phi}^{(\text{def})}$ parameters for the proposed weighted dynamic method, evaluated on the second half of 2023/2024 season for the Bundesliga, English Premier League (EPL), and La Liga

		$\beta^{(\text{att})}$			$\beta^{(\text{def})}$			$\phi^{(\text{att})}$			$\phi^{(\text{def})}$		
League	**Model**	$\hat{\bar{R}}$	**Bulk ESS**	**Tail ESS**	$\hat{\bar{R}}$	**Bulk ESS**	**Tail ESS**	$\hat{\bar{R}}$	**Bulk ESS**	**Tail ESS**	$\hat{\bar{R}}$	**Bulk ESS**	**Tail ESS**
Bundesliga	Dixon and Coles	1.01	1,149	2,762	1.01	1,310	2,861	1.01	1,159	1,216	1.01	1,231	1,173
	Double Poisson	1.01	1,355	2,836	1.00	1,454	3,058	1.01	1,419	1,268	1.01	1,453	1,187
	Negative binomial	1.00	3,819	4,469	1.00	3,675	4,887	1.00	3,517	3,042	1.00	3,829	3,241
	Skellam model	1.00	1,572	2,902	1.00	1,703	2,700	1.00	1,325	1,331	1.00	1,332	1,228
	Zero-infl. Skellam model	1.01	975	2,465	1.01	1,061	2,244	1.01	1,229	1,208	1.01	1,290	1,079
EPL	Dixon and Coles	1.00	1,870	2,839	1.00	2,025	2,693	1.01	1,144	1,296	1.01	1,164	1,241
	Double Poisson	1.00	2,321	3,405	1.00	2,698	3,342	1.01	1,325	1,416	1.00	1,386	1,379
	Negative binomial	1.00	2,579	3,580	1.00	3,096	3,800	1.00	1,349	1,532	1.00	1,405	1,500
	Skellam model	1.00	1,175	2,223	1.00	1,259	2,054	1.00	1,147	1,193	1.00	1,224	1,112
	Zero-infl. Skellam model	1.01	1,051	2,146	1.00	1,104	1,901	1.01	1,105	1,077	1.01	1,118	1,019
La Liga	Dixon and Coles	1.00	4,355	4,063	1.00	4,122	3,980	1.00	1,411	1,287	1.00	1,432	1,282
	Double Poisson	1.01	1,562	2,858	1.01	1,459	2,540	1.01	743	1,107	1.01	801	1,093
	Negative binomial	1.01	2,821	3,683	1.01	3,067	3,726	1.01	1,101	1,201	1.01	1,150	1,241
	Skellam model	1.01	3,015	3,111	1.01	3,218	3,112	1.01	1,110	1,450	1.01	1,102	1,318
	Zero-infl. Skellam model	1.01	1,121	1,866	1.01	1,075	2,023	1.01	619	1,024	1.01	634	1,010

Table D2. MCMC convergence diagnostics: $\hat{R}$ and bulk and tail effective sample sizes (ESS) mean of the $\boldsymbol{\beta}^{(\text{att})}$, $\boldsymbol{\beta}^{(\text{def})}$, $\boldsymbol{\phi}^{(\text{att})}$, and $\boldsymbol{\phi}^{(\text{def})}$ parameters for the proposed weighted dynamic method, evaluated on the second half of 2024/2025 season for the Bundesliga, English Premier League (EPL), and La Liga

		$\beta^{(\text{att})}$			$\beta^{(\text{def})}$			$\phi^{(\text{att})}$			$\phi^{(\text{def})}$		
League	**Model**	$\bar{\hat{R}}$	**Bulk ESS**	**Tail ESS**	$\bar{\hat{R}}$	**Bulk ESS**	**Tail ESS**	$\bar{\hat{R}}$	**Bulk ESS**	**Tail ESS**	$\bar{\hat{R}}$	**Bulk ESS**	**Tail ESS**
Bundesliga	Dixon and Coles	1.01	695	1,929	1.01	870	2,098	1.01	694	793	1.01	704	873
	Double Poisson	1.00	1,456	2,718	1.00	1,611	2,617	1.00	1,263	1,012	1.00	1,282	965
	Negative binomial	1.00	2,246	3,666	1.00	2,616	3,763	1.00	1,764	1,364	1.00	1,850	1,343
	Skellam model	1.01	1,587	2,693	1.01	1,657	2,727	1.01	1,284	1,075	1.01	1,304	1,009
	Zero-infl. Skellam model	1.01	1,134	2,392	1.01	1,133	1,907	1.01	1,451	1,076	1.01	1,503	1,067
EPL	Dixon and Coles	1.01	1,263	2,802	1.01	1,289	2,636	1.01	999	1,083	1.01	1,031	1,132
	Double Poisson	1.00	2,003	3,156	1.00	2,137	3,107	1.01	1,277	1,108	1.01	1,340	1,074
	Negative binomial	1.01	1,559	3,060	1.00	1,758	3,044	1.01	1,082	1,128	1.01	1,141	1,144
	Skellam model	1.01	749	1,876	1.01	716	1,587	1.01	1,107	835	1.01	1,145	796
	Zero-infl. Skellam model	1.01	441	984	1.01	442	1,062	1.01	682	686	1.01	710	679
La Liga	Dixon and Coles	1.00	5,470	4,313	1.00	5,549	4,249	1.00	1,725	1,162	1.00	1,705	1,152
	Double Poisson	1.00	6,319	4,832	1.00	6,471	4,780	1.00	1,951	1,206	1.00	1,949	1,169
	Negative binomial	1.00	5,503	4,241	1.00	5,483	4,206	1.00	1,876	1,200	1.00	1,872	1,196
	Skellam model	1.00	4,875	3,723	1.00	5,388	3,868	1.00	1,477	1,094	1.00	1,538	1,040
	Zero-infl. Skellam model	1.01	1,682	2,511	1.01	1,827	2,714	1.01	897	953	1.01	862	964

Table D3. MCMC convergence diagnostics: $\hat{R}$ and bulk and tail effective sample sizes (ESS) mean of the $\boldsymbol{\beta}^{(\text{att})}$, $\boldsymbol{\beta}^{(\text{def})}$, $\boldsymbol{\phi}^{(\text{att})}$, and $\boldsymbol{\phi}^{(\text{def})}$ parameters for the proposed weighted dynamic method, evaluated on the last round of the 2024/2025 season for the Bundesliga, English Premier League (EPL), and La Liga

		$\boldsymbol{\beta}^{(\text{att})}$			$\boldsymbol{\beta}^{(\text{def})}$			$\boldsymbol{\phi}^{(\text{att})}$			$\boldsymbol{\phi}^{(\text{def})}$		
League	**Model**	$\bar{\hat{R}}$	**Bulk ESS**	**Tail ESS**	$\bar{\hat{R}}$	**Bulk ESS**	**Tail ESS**	$\bar{\hat{R}}$	**Bulk ESS**	**Tail ESS**	$\bar{\hat{R}}$	**Bulk ESS**	**Tail ESS**
Bundesliga	Dixon and Coles	1.00	2,441	3,813	1.00	2,739	3,857	1.00	1,922	1,384	1.00	1,986	1,354
	Double Poisson	1.00	1,641	2,791	1.00	1,816	2,848	1.00	1,605	1,130	1.01	1,651	1,079
	Negative binomial	1.01	1,479	3,103	1.01	1,474	2,944	1.01	1,319	1,072	1.01	1,378	1,062
	Skellam model	1.00	1,460	2,424	1.00	1,403	2,135	1.01	1,305	1,002	1.01	1,245	931
	Zero-infl. Skellam model	1.01	1,345	2,574	1.01	1,374	2,527	1.01	1,600	1,082	1.01	1,649	1,028
EPL	Dixon and Coles	1.00	2,400	3,550	1.00	2,641	3,597	1.00	1,609	1,283	1.00	1,647	1,226
	Double Poisson	1.01	1,550	2,724	1.01	1,627	2,521	1.01	1,184	1,082	1.01	1,307	1,116
	Negative binomial	1.01	1,113	2,311	1.01	1,180	2,345	1.01	873	880	1.01	935	889
	Skellam model	1.01	1,056	2,362	1.01	1,079	2,054	1.01	1,347	928	1.01	1,340	888
	Zero-infl. Skellam model	1.01	964	2,344	1.01	992	2,084	1.00	1,507	1,055	1.00	1,566	1,032
La Liga	Dixon and Coles	1.00	4,534	3,843	1.00	4,697	3,770	1.00	1,721	1,042	1.00	1,720	1,087
	Double Poisson	1.00	2,507	2,814	1.00	2,493	2,790	1.01	1,110	775	1.01	1,136	769
	Negative binomial	1.00	5,631	4,291	1.00	5,698	4,190	1.00	1,934	1,108	1.00	1,956	1,135
	Skellam model	1.00	5,705	4,112	1.00	5,971	4,148	1.00	1,736	1,077	1.00	1,754	1,045
	Zero-infl. Skellam model	1.00	3,793	3,054	1.00	3,909	3,263	1.00	1,508	1,036	1.00	1,556	988

against 0.377. In the 2024/2025 season, the differences vary across leagues. In the Bundesliga, the weighted dynamic DP model matches the market average on the ACP, with both recording 0.399, and the difference in pseudo-R^2 is similarly small at 0.352 versus 0.356. In the EPL, the market average achieves an ACP of 0.431 and a pseudo-R^2 of 0.386, compared with 0.421 and 0.381 for the best weighted dynamic model, DP. In La Liga, the best weighted dynamic model, Dixon and Coles, achieves an ACP of 0.421 and a pseudo-R^2 of 0.378, compared with 0.433 and 0.391 for the market average. For the last round, the weighted dynamic framework substantially outperforms the market average in the Bundesliga and EPL on both metrics. In the Bundesliga, the best weighted dynamic model, DP, achieves an ACP of 0.428 and a pseudo-R^2 of 0.395, compared with 0.394 and 0.353 for the market average. In the EPL, the weighted dynamic DP model obtains an ACP of 0.436 and the SM a pseudo-R^2 of 0.404, against 0.423 and 0.357 for the market average. In La Liga, the market average retains an advantage, with an ACP of 0.500 and a pseudo-R^2 of 0.462, compared with 0.457 and 0.432 for the best weighted dynamic model, Dixon and Coles.

Overall, the betting market maintains a consistent yet moderate advantage across the half-season scenarios, which is expected given the substantially richer information set incorporated in the odds. However, the results from the last round reveal that this advantage can reverse in highly unpredictable settings.

Appendix D Convergence diagnostics

All the models achieved satisfactory convergence diagnostics for all the evaluated scenarios. In particular, we verified that the MCMC chains for each model and method yielded Gelman–Rubin statistic $\hat{R}$ (Gelman & Rubin, 1992) below or equal to the threshold of 1.01 and bulk and tail effective sample sizes exceeding 400—the threshold recommended by Vehtari et al. (2021) for reliable inference. Table D1 shows a summary of the convergence metrics of the weighted dynamic method for the second half of the 2023/2024 season scenario. Specifically, the means of the Gelman–Rubin statistic $\hat{R}$ and bulk and tail effective sample sizes is shown for $\boldsymbol{\beta}^{(\text{att})}$, $\boldsymbol{\beta}^{(\text{def})}$, $\boldsymbol{\phi}^{(\text{att})}$, and $\boldsymbol{\phi}^{(\text{def})}$ parameters.

Similar convergence analyses for the other two predictive scenarios (the second half and the last round of the 2024/2025 season) are presented in Tables D2 and D3.

References

Baio G., & Blangiardo M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, *37*(2), 253–264. https://doi.org/10.1080/02664760802684177

Boshnakov G., Kharrat T., & McHale I. G. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, *33*(2), 458–466. https://doi.org/10.1016/j.ijforecast.2016.11.006

Bradley R. A., & Terry M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*(3/4), 324–345. https://doi.org/10.2307/2334029

Breiman L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brier G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthey Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Carpenter B., Gelman A., Hoffman M. D., Lee D., Goodrich B., Betancourt M., Brubaker M., Guo J., Li P., & Riddell A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Carpita M., Ciavolino E., & Pasca P. (2019). Exploring and modelling team performances of the Kaggle European soccer database. *Statistical Modelling*, *19*(1), 74–101. https://doi.org/10.1177/1471082X18810971

Carpita M., Sandri M., Simonetto A., & Zuccolotto P. (2015). Discovering the drivers of football match outcomes with data mining. *Quality Technology & Quantitative Management*, *12*(4), 561–577. https://doi.org/10.1080/16843703.2015.11673436

Cattelan M. (2012). Models for paired comparison data: A review with emphasis on dependent data. *Statistical Science*, *27*(3), 412–433. https://doi.org/10.1214/12-STS396

Chen N., Carlin B. P., & Hobbs B. P. (2018). Web-based statistical tools for the analysis and design of clinical trials that incorporate historical controls. *Computational Statistics & Data Analysis*, *127*, 50–68. https://doi.org/10.1016/j.csda.2018.05.002

Deshpande S. K., & Jensen S. T. (2016). Estimating an NBA player's impact on his team's chances of winning. *Journal of Quantitative Analysis in Sports*, *12*(2), 51–72. https://doi.org/10.1515/jqas-2015-0027

Dixon M. J., & Coles S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *46*(2), 265--280. https://doi.org/10.1111/1467-9876.00065

Dobson S., Goddard J. A., & Dobson S. (2001). *The economics of football* (Vol. *10*). Cambridge University Press.

Egidi L., Macrì-Demartino R., & Palaskas V. (2025). *footBayes: Fitting Bayesian and MLE Football Models*. R package version 2.1.0.

Egidi L., Pauli F., & Torelli N. (2018). Combining historical data and bookmakers' odds in modelling football scores. *Statistical Modelling*, *18*(5-6), 436–459. https://doi.org/10.1177/1471082X18798414

Egidi L., & Torelli N. (2021). Comparing goal-based and result-based approaches in modelling football outcomes. *Social Indicators Research*, *156*(2-3), 801–813. https://doi.org/10.1007/s11205-020-02293-z

Epstein E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology (1962-1982)*, *8*(6), 985–987. https://doi.org/10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2

Gelman A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Gelman A., Jakulin A., Pittau M. G., & Su Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

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

Glickman M. E., & Jones A. C. (2025). Models and rating systems for head-to-head competition. *Annual Review of Statistics and Its Application*, *12*(1), 259–282. https://doi.org/10.1146/annurev-statistics-040722-061813

Groll A., & Abedieh J. (2013). Spain retains its title and sets a new record—generalized linear mixed models on European football championships. *Journal of Quantitative Analysis in Sports*, *9*(1), 51–66. https://doi.org/10.1515/jqas-2012-0046

Groll A., Hvattum L. M., Ley C., Popp F., Schauberger G., Van Eetvelde H., & Zeileis A. (2021). Hybrid machine learning forecasts for the UEFA EURO 2020. arXiv:2106.05799, preprint: not peer reviewed.

Groll A., Hvattum L. M., Ley C., Sternemann J., Schauberger G., & Zeileis A. (2024). Modeling and prediction of the UEFA EURO 2024 via combined statistical learning approaches. arXiv:2410.09068, preprint: not peer reviewed.

Groll A., Ley C., Schauberger G., & Van Eetvelde H. (2019a). A hybrid random forest to predict soccer matches in international tournaments. *Journal of Quantitative Analysis in Sports*, *15*(4), 271–287. https://doi.org/10.1515/jqas-2018-0060

Groll A., Ley C., Schauberger G., Van Eetvelde H., & Zeileis A. (2019b). Hybrid machine learning forecasts for the FIFA Women's World Cup 2019. arXiv:1906.01131, preprint: not peer reviewed.

Hobbs B. P., Carlin B. P., Mandrekar S. J., & Sargent D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, *67*(3), 1047–1056. https://doi.org/10.1111/j.1541-0420.2011.01564.x

Hobbs B. P., Sargent D. J., & Carlin B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis*, *7*(3), 639–674. https://doi.org/10.1214/12-BA722

Hong H., Fu H., & Carlin B. P. (2018). Power and commensurate priors for synthesizing aggregate and individual patient level data in network meta-analysis. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *67*(4), 1047–1069. https://doi.org/10.1111/rssc.12275

Jensen S. T., McShane B. B., & Wyner A. J. (2009). Hierarchical Bayesian modeling of hitting performance in baseball. *Bayesian Analysis*, *4*(4), 631–652. https://doi.org/10.1214/09-BA424

Karlis D., & Ntzoufras I. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *52*(3), 381–393. https://doi.org/10.1111/1467-9884.00366

Karlis D., & Ntzoufras I. (2009). Bayesian modelling of football outcomes: Using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics*, *20*(2), 133–145. https://doi.org/10.1093/imaman/dpn026

Koning R. H. (2000). Balance in competition in Dutch soccer. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*(3), 419–431. https://doi.org/10.1111/1467-9884.00244

Koopman S. J., & Lit R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *178*(1), 167–186. https://doi.org/10.1111/rssa.12042

Koopman S. J., & Lit R. (2019). Forecasting football match results in national league competitions using score-driven time series models. *International Journal of Forecasting*, *35*(2), 797–809. https://doi.org/10.1016/j.ijforecast.2018.10.011

Lam M. W. Y. (2018). One-match-ahead forecasting in two-team sports with stacked bayesian regressions. *Journal of Artificial Intelligence and Soft Computing Research*, *8*(3), 159–171. https://doi.org/10.1515/jaiscr-2018-0011

Ley C., de Wiele T. V., & Eetvelde H. V. (2019). Ranking soccer teams on the basis of their current strength: A comparison of maximum likelihood approaches. *Statistical Modelling*, *19*(1), 55–73. https://doi.org/10.1177/1471082X18817650

Macrì Demartino R., Egidi L., & Torelli N. (2025). Alternative ranking measures to predict international football results. *Computational Statistics*, *40*(4), 1899–1917. https://doi.org/10.1007/s00180-024-01585-z

Maher M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, *36*(3), 109–118. https://doi.org/10.1111/j.1467-9574.1982.tb00782.x

McHale I., & Morton A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, *27*(2), 619–630. https://doi.org/10.1016/j.ijforecast.2010.04.004

McHale I., & Scarf P. (2011). Modelling the dependence of goals scored by opposing teams in international soccer matches. *Statistical Modelling*, *11*(3), 219–236. https://doi.org/10.1177/1471082X1001100303

Michels R., Ötting M., & Karlis D. (2025). Extending the Dixon and Coles model: An application to women's football data. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *74*(1), 167–186. https://doi.org/10.1093/jrsssc/qlae050

Mitchell T. J., & Beauchamp J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, *83*(404), 1023–1032. https://doi.org/10.1080/01621459.1988.10478694

Mosteller F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, *16*(1), 3–9. https://doi.org/10.1007/BF02313422

Murray T. A., Hobbs B. P., & Carlin B. P. (2015). Combining nonexchangeable functional or survival data sources in oncology using generalized mixture commensurate priors. *The Annals of Applied Statistics*, *9*(3), 1549. https://doi.org/10.1214/15-AOAS840

Ntzoufras I. (2011). *Bayesian modeling using WinBUGS* (Vol. *698*). John Wiley & Sons.

Ouma L. O., Grayling M. J., Wason J. M., & Zheng H. (2022). Bayesian modelling strategies for borrowing of information in randomised basket trials. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *71*(5), 2014–2037. https://doi.org/10.1111/rssc.12602

Owen A. (2011). Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics*, *22*(2), 99–113. https://doi.org/10.1093/imaman/dpq018

Petretta M., Schiavon L., & Diquigiovanni J. (2025). Modelling dependence in football match outcomes: Traditional assumptions and an alternative proposal. *Statistical Modelling*, *25*(3), 255–269. https://doi.org/10.1177/1471082X241238802

Phelan G. C., & Whelan J. T. (2018). Hierarchical Bayesian bradley-terry for applications in major league baseball. *Mathematics for Application*, *7*(1), 71–84. https://doi.org/10.13164/ma

Pocock S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, *29*(3), 175–188. https://doi.org/10.1016/0021-9681(76)90044-8

R Core Team (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reep C., Pollard R., & Benjamin B. (1971). Skill and chance in ball games. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *134*(4), 623–629. https://doi.org/10.2307/2343657

Reich B. J., Hodges J. S., Carlin B. P., & Reich A. M. (2006). A spatial analysis of basketball shot chart data. *The American Statistician*, *60*(1), 3–12. https://doi.org/10.1198/000313006X90305

Rue H., & Salvesen Ø. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *49*(3), 399–418. https://doi.org/10.1111/1467-9884.00243

Santos-Fernandez E., Wu P., & Mengersen K. L. (2019). Bayesian statistics meets sports: A comprehensive review. *Journal of Quantitative Analysis in Sports*, *15*(4), 289–312. https://doi.org/10.1515/jqas-2018-0106

Schauberger G., & Groll A. (2018). Predicting matches in international football tournaments with random forests. *Statistical Modelling*, *18*(5-6), 460–482. https://doi.org/10.1177/1471082X18799934

Skellam J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *109*(3), 296–296. https://doi.org/10.2307/2981372

Spanias D., & Knottenbelt W. J. (2013). Predicting the outcomes of tennis matches using a low-level point model. *IMA Journal of Management Mathematics*, *24*(3), 311–320. https://doi.org/10.1093/imaman/dps010

Spiegelhalter D., & Ng Y.-L (2009). One match to go!. *Significance*, *6*(4), 151–153. https://doi.org/10.1111/j.1740-9713.2009.00387.x

Thurstone L. L. (1927). Psychophysical analysis. *The American Journal of Psychology*, *38*(3), 368–389. https://doi.org/10.2307/1415006

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

Zheng H., & Wason J. M. S. (2022). Borrowing of information across patient subgroups in a basket trial based on distributional discrepancy. *Biostatistics*, *23*(1), 120–135. https://doi.org/10.1093/biostatistics/kxaa019