NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

PREDICTION AND RETROSPECTIVE ANALYSIS OF SOCCER MATCHES IN A LEAGUE

by

Håvard Rue and Øyvind Salvesen

PREPRINT
STATISTICS NO. 10/1997

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL
http://www.math.ntnu.no/preprint/statistics/1997/S10-1997.ps
Håvard Rue has homepage: http://www.math.ntnu.no/~hrue
E-mail: hrue@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

PREDICTION AND RETROSPECTIVE ANALYSIS OF SOCCER MATCHES IN A LEAGUE

HÅVARD RUE* AND ØYVIND SALVESEN
DEPARTMENT OF MATHEMATICAL SCIENCES
NTNU, NORWAY

FIRST VERSION: OCTOBER 1997
REVISED: AUGUST 1998, JUNE 1999

ABSTRACT

A common discussion subject for the male part of the population in particular, is the prediction of next weekend's soccer matches, especially for the local team. Knowledge of offensive and defensive skills is valuable in the decision process before making a bet at a bookmaker. In this article we take an applied statistician's approach to the problem, suggesting a Bayesian dynamic generalised linear model to estimate the time dependent skills of all teams in a league, and to predict next weekend's soccer matches. The problem is more intricate than it may appear at first glance, as we need to estimate the skills of all teams simultaneously as they are dependent. It is now possible to deal with such inference problems using the iterative simulation technique known as Markov Chain Monte Carlo. We will show various applications of the proposed model based on the English Premier League and Division 1 1997-98; Prediction with application to betting, retrospective analysis of the final ranking, detection of surprising matches and how each team's properties vary during the season.

KEYWORDS: Dynamic Models; Generalised Linear Models; Graphical Models; Markov Chain Monte Carlo Methods; Prediction of Soccer Matches.

1 INTRODUCTION

Soccer is a popular sport all over the world, and in Europe and South America it is the dominant spectator sport. People find interest in soccer for various reasons and at different levels, with a clear dominance for the male part of the population. Soccer is an excellent game for different forms of betting. The outcome of a soccer match depends on many factors, among these are home–field/away–field, the effect of injured players and various psychological effects. Good knowledge about these factors only determine the final result up to a significant, but not too dominant, random component.

Different models for prediction of soccer matches for betting have existed for a long time. A quick tour around the Web, perhaps starting at `http://dmiwww.cs.tut.fi/riku`, shows an impressive

*ADDRESS FOR CORRESPONDENCE: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7034 Trondheim, Norway.
E-MAIL: `Havard.Rue@math.ntnu.no`

activity with links to small companies selling ready-to-go PC programs for prediction and betting, discussion groups and bookmakers who operate on the Web. The most popular ideas in the prediction programs are to consider win, draw and loss sequences, goals scored over five matches, and the points difference in the current ranking. Most of the programs have a Bayesian flavour and allow the user to include his/her expert knowledge in various ways. The (pure) statistical side of soccer is not so developed and widespread. Ridder, Cramer and Hopstaken (1994) analyse the effect of a red card in a match, Kuonen (1996) model knock-out tournaments, and Lee (1997) provides a simplified generalised linear model with application to final rank analysis and Dixon and Coles (1997) and Dixon and Robinson (1998) provide a more comprehensive model.

In this article, we model the results of soccer matches played in a league, where the teams play against each other twice (home and away) relatively regularly and during a limited time period. We will use the history of the played matches to estimate what we think are the two most important (time dependent) explanatory variables in a Bayesian dynamic generalised linear model; the attack and defence strength. It is more intricate than one might think to estimate the properties like attack and defence strength for each team. Assume team A and B plays against each other with the result $5-0$. One interpretation of this result is that A has a large attack strength, another that B has a low defence strength. The properties for A and B conditional on the result are therefore dependent. As each team play against all the other teams in a league, we soon reach full dependency between the (time dependent) properties for *all* teams. We can analyse such problems using Markov chain Monte Carlo (MCMC) techniques (Gilks, Richardson and Spiegelhalter, 1996) to generate dependent samples from the posterior density. The proposed model and the power of MCMC can be used to make predictions for the next round in the league and answer other interesting questions as well, like; What is the expected ranking at the end of the season? Was Arsenal lucky to win the Premier League 1997-98?

A simplified and stripped off version of our model is similar to the generalised linear model developed independently by Lee (1997). Dixon and Coles (1997) and Dixon and Robinson (1998) presented a more comprehensive model than Lee (1997), trying to mimic time-varying properties by down-weighting the likelihood. They where not able to provide a coherent model for how the properties develop in time. In this paper we provide a Bayesian model which model the time-variation of all properties simultaneously, present a new parametrisation and ideas in goal-modelling, and show how we can do inference and do retrospective analysis of a season using the power of MCMC. We need a joint model for the properties in time to do retrospective analysis of a season, to be able to estimate each team's properties at time t using data from matches *both* before and after time t. Our approach provides a coherent model which is easy to extend to account for further refinement and development in the art of prediction soccer matches. Other similar sports problems could be approached in the same manner.

The rest of the article is organised as follows. We start in Section 2 with the basic explanatory variables and derive the model step by step. Section 3 describes how we estimate global parameters using historical data and do inference from the model using MCMC techniques. In Section 4 we apply our model analysing the English Premier League and Division 1 1997-98, with focus on prediction with application to betting, retrospective analysis of the final ranking, locating surprising matches and study how each team's properties vary during the season.

2 The Model

In this section we derive our model for analysing soccer matches in a league. We start with the basic explanatory variables, continue by linking these variables up with a goal model and model their dependency in time. In Section 2.4 we collect all pieces in the full model.

2.1 Basic Explanatory Variables

There are a variety of explanatory variables that will influence the result of a forthcoming soccer match. Which factors we should include in the model depend on what kind of data that is available. To keep the problem simple, we will only make use of the result in a match, like home team wins $3-1$ over the away team. We therefore ignore all other interesting data like the number of near goals, corners, free kicks and so on. Our next step is to specify which (hidden) explanatory variables attached to each team that will influence the match result.

The two important properties of each team are their defending and attacking skills. Defence and attack strength is represented as the random variables d and a respectively. A high value of d and a means a good defence and attack respectively. Let $\boldsymbol{e}_A = (a, d)_A$ denote the properties for team A, and further let $\mu_{a,A}$ and $\sigma^2_{a,A}$ be the prior mean and variance for a_A, and similar for the defence strength and the other teams.

2.2 The Goal Model

The next step is to specify how the result $(x_{A,B}, y_{A,B})$ depends on the properties of home team A and away team B. A reasonable assumption, is that the number of goals A makes $(x_{A,B})$ depends on A's attack strength and B's defence strength. Similarly, the number of goals B makes $(y_{A,B})$ depends on B's attack strength and A's defence strength. Additionally, we include a psychological effect; Team A will tend to underestimate the strength of team B if A is a stronger team than B. Let $\Delta_{AB} = (a_A + d_A - a_B - d_B)/2$ measure the difference in strength between A and B. We assume further that

$$
\begin{aligned}
x_{A,B} \,|(\boldsymbol{e}_A, \boldsymbol{e}_B) &\stackrel{d}{=} x_{A,B} \mid a_A - d_B - \gamma \Delta_{AB} \\
y_{A,B} \,|(\boldsymbol{e}_A, \boldsymbol{e}_B) &\stackrel{d}{=} y_{A,B} \mid a_B - d_A + \gamma \Delta_{AB},
\end{aligned}
$$

where γ is a small constant giving the magnitude of the psychological effect. We assume the strength of team A and B are not that different since we analyse teams in the same league, so it reasonable to expect $\gamma > 0$. (The opposite effect ($\gamma < 0$) might occur if team A is so superior compared to team B that the latter develops an inferiority complex facing A, which do not expect will happen in the same league.)

To motivate our probability law for $x_{A,B}$ and $y_{A,B}$, we display in Figure 1 the histogram of the result in 924 matches in the Premier league 1993-95. The histogram and nature of the game itself suggest to a first approximation a Poisson law for $x_{A,B}$ and $y_{A,B}$. Thus, as a first approximation we may assume the number of goals conditioned on the teams' properties to be Poisson distributed with mean $\lambda^{(x)}_{A,B}$

and $\lambda^{(y)}_{A,B}$, where

$$\log \lambda^{(x)}_{A,B} = c^{(x)} + a_A - d_B - \gamma\Delta_{AB}, \qquad \text{and} \qquad \log \lambda^{(y)}_{A,B} = c^{(y)} + a_B - d_A + \gamma\Delta_{AB}. \quad (1)$$

Here, $c^{(x)}$ and $c^{(y)}$ are global constants describing (roughly) the log of the empirical mean of the home and away goals.

Although independence of $x_{A,B}$ and $y_{A,B}$ is verified empirically to be quite reasonable (Lee, 1997), it does not imply that $x_{A,B}$ and $y_{A,B}$ are independent conditional on $(\boldsymbol{e}_A, \boldsymbol{e}_B)$. Dixon and Coles (1997) proposed therefore to use the joint conditional law for $(x_{A,B}, y_{A,B})$

$$\begin{aligned} \pi_{g1}\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}\right) &= \kappa\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}\right) \\ &\quad \times \text{Po}\left(x_{A,B} \mid \lambda^{(x)}_{A,B}\right) \text{Po}\left(y_{A,B} \mid \lambda^{(y)}_{A,B}\right) \end{aligned} \quad (2)$$

where $\text{Po}\left(x_{A,B} \mid \lambda^{(x)}_{A,B}\right)$ is the Poisson law for $x_{A,B}$ with mean $\lambda^{(x)}_{A,B}$, and κ is a correction factor given as

$$\kappa\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}\right) = \begin{cases} 1 + 0.1\lambda^{(x)}_{A,B}\lambda^{(y)}_{A,B} & \text{if } x_{A,B} = 0, y_{A,B} = 0, \\ 1 - 0.1\lambda^{(x)}_{A,B} & \text{if } x_{A,B} = 0, y_{A,B} = 1, \\ 1 - 0.1\lambda^{(y)}_{A,B} & \text{if } x_{A,B} = 1, y_{A,B} = 0, \\ 1.1 & \text{if } x_{A,B} = 1, y_{A,B} = 1, \\ 1 & \text{otherwise.} \end{cases}$$

The correction factor κ increases the probability of $0-0$ and $1-1$ at the cost of $1-0$ and $0-1$. All other joint probabilities remain unchanged. Note further that the (conditional) marginal laws of $x_{A,B}$ and $y_{A,B}$ from Eq. (2) equals $\text{Po}\left(x_{A,B} \mid \lambda^{(x)}_{A,B}\right)$ and $\text{Po}\left(y_{A,B} \mid \lambda^{(y)}_{A,B}\right)$, respectively.

We found it necessary to modify Eq. (2) in two ways; The first modification is about the Poisson assumption, the second is a robustness adjustment.

Although the Poisson model seems reasonable, it may not be the case if one of the teams make many goals. This is highly demotivating for the other team, and in most cases imply a contradiction with our underlying model assumption that the goal intensity does not depend on the goals made during the match. We correct for this by truncating the Poisson law $\text{Po}\left(x_{A,B} \mid \lambda^{(x)}_{A,B}\right)$ (and similar with $y_{A,B}$) in Eq. (2) after 5 goals. Denote by π^*_{g1} the resulting truncated law. The result of $7-0$ and $6-5$ will be interpreted as $5-0$ and $5-5$, respectively. The goals each team makes after their first five are noninformative on the two team's properties.

It is our experience that the match result is less informative on the properties of the teams than Eq. (1) and π^*_{g1} suggest. We found it necessary to use a more model-robust goal model by forming a mixture of laws,

$$\begin{aligned} \pi_g\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}\right) &= (1-\epsilon)\pi^*_{g1}\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}\right) \\ &\quad + \epsilon\, \pi^*_{g1}\left(x_{A,B}, y_{A,B} \mid \exp(c^{(x)}), \exp(c^{(y)})\right). \end{aligned} \quad (3)$$

Here, ϵ is a parameter with the interpretation that only $(1-\epsilon) \times 100\%$ of the "information" in the match result is informative on $\boldsymbol{e}_A$ and $\boldsymbol{e}_B$, and the remaining $\epsilon \times 100\%$ is not informative. The non-informative part of π_g uses the average goal intensities, $\exp(c^{(x)})$ and $\exp(c^{(y)})$, and π_g shrinks therefore π^*_{g1} towards the law for an average match. The value of ϵ is in Section 3.2 found to be around 0.2.

2.3 Time Model

It is both natural and necessary to allow the attack and defence variables to vary with time. For the discussion here, assume that t' and $t'' \geq t'$ are two following time points (number of days from a common reference point) where team A plays a match. Let us consider the attack strength where similar arguments are valid also for the defence strength. We need to specify how $a_A^{t''}$ (superscripts t for time) depends on $a_A^{t'}$ and possibly on previous values. Our main purpose is to predict matches in the near future, so only a reasonable local behaviour for a_A in time is needed. As a first choice, we borrow ideas from dynamic models (West and Harrison, 1997), and use Brownian motion to tie together a_A at the two time points t' and t'', i.e. conditional on $\{a_A^t, \ \ t \leq t'\}$

$$a_A^{t''} \stackrel{d}{=} a_A^{t'} + \left(B_{a,A}(t''/\tau) - B_{a,A}(t'/\tau)\right) \frac{\sigma_{a,A}}{\sqrt{1-\gamma(1-\gamma/2)}}. \tag{4}$$

(Recall that $\sigma^2_{a,A}$ is the prior variance for a_A.) Here, $\{B_{\cdot,\cdot}(t), t \geq 0\}$ is standard Brownian motion starting at level zero and where the subscript marks the belonging to the attack strength for team A. The last factor is a scaling factor which we motivate in the next paragraph. The characteristic time parameter τ is common to all teams and gives the inverse loss of memory rate for a_A^t, $\text{Var}(a_A^{t''} - a_A^{t'}) \propto \sigma^2_{a,A}/\tau$. We model the attack and defence strength for all teams as in (4) and assume the Brownian motion processes are independent.

Using Brownian motion is a weak and flexible way of modelling the development of the properties over time. Although the process itself is non-stationary with a variance increasing to infinity with increasing time, the conditioning on the match results will ensure smooth posterior realisations of the properties and reasonable extrapolation into the near future.

The common parameters γ and τ control the psychological effect and the loss of memory rate. These parameters have a nice interpretation if we consider the conditional (on the past $\{t < t''\}$) expected value and variance in the Gaussian conditional density for $\log\lambda_{A,B}^{(x),t''}$ (and $\log\lambda_{A,B}^{(y),t''}$). If we assume for simplicity that $\sigma^2_{a,A} = \sigma^2_{d,A} = \sigma^2_{a,B} = \sigma^2_{d,B}$, we obtain

$$\mathrm{E}\left(\log\lambda_{A,B}^{(x),t''} \mid \text{past}\right) = c^{(x)} + a_A^{t'} - d_B^{t'} - \gamma\Delta_{AB}^{t'}$$

and

$$\text{Var}\left(\log\lambda_{A,B}^{(x),t''} \mid \text{past}\right) = 2\sigma^2_{a,A}\frac{t''-t'}{\tau}. \tag{5}$$

Thus, γ adjust the conditional expected value, and τ controls the conditional variance of $\log\lambda_{A,B}^{(x)}$. The scaling with γ in (4) makes γ and τ orthogonal in this sense. We interpret τ and γ as the main and secondary parameter of interest, respectively.

2.4 FULL MODEL

We can now build the full model based on the previous assumptions. The properties of each team obey (4) for the time development and (1) for the result of each match. We only need to keep record of which teams that play against each other, when and where. Figure 2 shows the situation schematically when four teams play 3×2 matches at time t_0, t_1 and t_2, and the fourth round at time t_3 is not played. (We choose to play the matches at the same days, so the notation is simplified. In practice this is not the case.) The graph is called the *directed acyclic graph* (DAG) (Whittaker, 1990) and the directed edges in the graph show the flow of information or causal relations between parent and child nodes in the graph. If we construct the corresponding moral graph (Whittaker, 1990) of Figure 2, we will find a path between each node in the graph soon after the league starts. We must therefore do inference for all the properties for each team at all time points, simultaneously.

We write $\boldsymbol{\theta}$ for the all variables in the model and keep for the moment parameters ϵ, γ and τ fixed. The variables are; The properties $\boldsymbol{e}_A^{t_0}$, $\boldsymbol{e}_B^{t_0}$, $\boldsymbol{e}_C^{t_0}$, $\boldsymbol{e}_D^{t_0}$ at time t_0, the result of the match between A and B, C and D at time t_0, the properties $\boldsymbol{e}_A^{t_1}$, $\boldsymbol{e}_B^{t_1}$, $\boldsymbol{e}_C^{t_1}$, $\boldsymbol{e}_D^{t_1}$ at time t_1, the result of the match between A and C, B and D at time t_1, and so on. The joint density for all variables in the model $\boldsymbol{\theta}$ is easy to find if we make use of Figure 2. The joint density of $\boldsymbol{\theta}$ is the product of the conditional density of each node given its parents. By using $\pi(\cdot \mid \cdot)$ as a generic notation for the density of its arguments, and indicate the time through superscripts, we obtain starting from the top of the graph where each line of Eq. (6) corresponds to each row of the graph in Figure 2,

$$\begin{aligned}
\pi(\boldsymbol{\theta}) = \;& \pi(\boldsymbol{e}_A^{t_0})\, \pi(\boldsymbol{e}_B^{t_0})\, \pi(\boldsymbol{e}_C^{t_0})\, \pi(\boldsymbol{e}_D^{t_0}) \\
& \times \pi(x_{A,B}^{t_0}, y_{A,B}^{t_0} \mid \boldsymbol{e}_A^{t_0}, \boldsymbol{e}_B^{t_0})\, \pi(x_{C,D}^{t_0}, y_{C,D}^{t_0} \mid \boldsymbol{e}_C^{t_0}, \boldsymbol{e}_D^{t_0}) \\
& \times \pi(\boldsymbol{e}_A^{t_1} \mid \boldsymbol{e}_A^{t_0})\, \pi(\boldsymbol{e}_B^{t_1} \mid \boldsymbol{e}_B^{t_0})\, \pi(\boldsymbol{e}_C^{t_1} \mid \boldsymbol{e}_C^{t_0})\, \pi(\boldsymbol{e}_D^{t_1} \mid \boldsymbol{e}_D^{t_0}) \\
& \times \pi(x_{A,C}^{t_1}, y_{A,C}^{t_1} \mid \boldsymbol{e}_A^{t_1}, \boldsymbol{e}_C^{t_1})\, \pi(x_{B,D}^{t_1}, y_{B,D}^{t_1} \mid \boldsymbol{e}_B^{t_1}, \boldsymbol{e}_D^{t_1}) \\
& \times \cdots
\end{aligned} \tag{6}$$

Here is $\pi(\boldsymbol{e}_A^{t_0})$ the prior density for $\boldsymbol{e}_A$, which will be commented on in Section 3.2, $\pi(x_{A,B}^{t_0}, y_{A,B}^{t_0} \mid \boldsymbol{e}_A^{t_0}, \boldsymbol{e}_B^{t_0})$ is the mixture law given in Eq. (3). Note that the mixture law depends on $\boldsymbol{e}_A^{t_0}$ and $\boldsymbol{e}_B^{t_0}$ only through $\lambda_{A,B}^{(x),t_0}$ and $\lambda_{A,B}^{(y),t_0}$, as given in Eq. (1) and Eq. (3). Further, $\pi(\boldsymbol{e}_A^{t_1} \mid \boldsymbol{e}_A^{t_0})$ equals $\pi(a_A^{t_1} \mid a_A^{t_0})\pi(d_A^{t_1} \mid d_A^{t_0})$ where

$$a_A^{t_1} \mid a_A^{t_0} \quad \sim \quad \mathrm{N}\left(a_A^{t_0},\; \frac{t_1 - t_0}{\tau}\sigma_{a,A}^2\right)$$

and similar with $d_A^{t_1} \mid d_A^{t_0}$. We denote by $\mathrm{N}(\mu, \sigma^2)$ the Gaussian distribution with mean μ and variance σ^2. The rest of the terms in $\pi(\boldsymbol{\theta})$ have similar interpretations, only the teams and times differ.

We have not included the properties in between time t_0, t_1, t_2 and t_3, as their (conditional) distributions are known to be Brownian bridges conditional on the properties at time $t_0, \ldots, t_3$.

Note that we can view the full model as a Bayesian dynamic model in the general framework of West and Harrison (1997), where Eq. (3) represent the observation equation and Eq. (4) the state space equation (taking all teams and properties into account).

To do inference for the properties of each team conditional on the observed match results, we need the conditional (posterior) density derived from (6). It is hard to analyse the posterior with direct methods due to reasons of complexity and to an intractable normalisation constant. We can however make use of Markov chain Monte Carlo methods to analyse our model, and this will be further discussed in the next section leaving details for the Appendix.

3 INFERENCE

In this section we will discuss how we can do inference from the model making use of Markov chain Monte Carlo methods from the posterior density, for fixed values of the τ, γ and ϵ parameters. We will then discuss how we choose the constants $c^{(x)}$ and $c^{(y)}$, and how we estimate τ, γ and ϵ to maximise the predictive ability of the model.

3.1 THE MCMC ALGORITHM

We can do inference from the posterior density proportional to (6) using (dependent) samples from the posterior produced by Markov chain Monte Carlo methods. There is now an extensive literature on MCMC methods and Gilks et al. (1996) provides a comprehensive overview of theory and the wide range of applications. In brief, in order to generate realisations from some density $f(d\boldsymbol{z})$ we construct a Markov chain using an irreducible aperiodic transition kernel which has $f(d\boldsymbol{z})$ as its equilibrium distribution. The algorithm goes as follows; Suppose the current state of the Markov chain is $\boldsymbol{z}$, and we propose a move of type j that moves $\boldsymbol{z}$ to $d\boldsymbol{z}'$ with probability $q_j(\boldsymbol{z}, d\boldsymbol{z}')$. The move to $\boldsymbol{z}'$ is accepted with probability $\min\{1, R_{\boldsymbol{z},\boldsymbol{z}'}\}$, where

$$R_{\boldsymbol{z},\boldsymbol{z}'} = \frac{f(d\boldsymbol{z}')\, q_j(\boldsymbol{z}', d\boldsymbol{z})}{f(d\boldsymbol{z})\, q_j(\boldsymbol{z}, d\boldsymbol{z}')} \tag{7}$$

otherwise we stay in the original state $\boldsymbol{z}$. When q_j is symmetric, (7) reduces to $f(d\boldsymbol{z}')/f(d\boldsymbol{z})$ and the sampler is known as the Metropolis algorithm. The perhaps most well known construction is the Gibbs sampler; Take q_j to be the conditional density of the component(s) to be updated given the remaining components. Because $R_{\boldsymbol{z},\boldsymbol{z}'}$ is one in this case we always accept the new state.

We need in theory two different move types to implement a MCMC algorithm for our model; Update the result for those matches which are not played, update the attack and defence strength for each team at each time a match is played. However, to ease the implementation of the MCMC algorithm we do a reformulation of the mixture model and attach an independent Bernoulli variable to each match. Each Bernoulli variable is updated during the MCMC algorithm and indicates which one of the distributions on the right hand side of Eq. (3) we currently use. We refer to the Appendix for the details of the MCMC algorithm and a discussion of various MCMC issues. Refer also to Chapter 15 in West and Harrison (1997) for a discussion of simulation-based inference in dynamic models.

The average acceptance rate for the proposed MCMC algorithm, is around 55%. Even if our single site updating algorithm is not that specialised, we obtain quite reasonable computational costs. The algorithm does $1\,000$ updates of all variables in a case with 380 matches, in about 12 seconds on a Sun Ultra 4 with a 296 MHz SUNW UltraSPARC-II CPU.

3.2 INFERENCE FOR $c^{(x)}$, $c^{(y)}$, τ, γ AND ϵ

In this section we will discuss how we choose various constants, validate our Gaussian prior distribution for the properties, and how we estimate the important parameters τ, γ and ϵ using historical data from the Premier League and Division 1.

We used 1684 matches from Premier League 1993-97 and 2208 matches from Division 1 1993-97 to estimate two sets of global constants $c^{(x)}$ and $c^{(y)}$. We used for simplicity the logarithm of the empirical mean of the home and away goals. The estimates are $c^{(x)} = 0.395$, $c^{(y)} = 0.098$ for Premier League, and $c^{(x)} = 0.425$, $c^{(y)} = 0.062$ for Division 1. These values are close to those we obtained with a more accurate and comprehensive Bayesian analysis.

It is tempting to use Gaussian priors for the properties of each team. To validate this assumption, we used 924 matches from the Premier League assuming each of the 924 match-results were realisations from matches with a common distribution for $\log \lambda^{(x)}$ and $\log \lambda^{(y)}$. The posterior density of $\log \lambda^{(x)}$ and $\log \lambda^{(y)}$ were close to Gaussians, so our assumption seems reasonable. We therefore took the prior for a and d for all teams to be (independent) Gaussians with an average variance $1/37$ found from the estimates. Note that this implies a common loss of memory rate for all teams (Section 2.3). Although we expect the attack and defence strength to be (positively) dependent, we choose prior independence for simplicity and to be non-informative. Further, the prior variance is confounded with the τ-parameter for all matches apart from the first in each league (see Eq. (5)).

The conditional mean and variance for the goal log-intensity are controlled by the parameters τ and γ, and the goal model depends on the mixture parameter ϵ. The predictive properties of the model depend on these parameters. It is tempting from the Gaussian structure in the time model to make use of the conjugate Gamma density for the inverse variances (precisions); If the prior for τ is gamma distributed so is the posterior. Our experience with this approach tells that there is not much information in the match results about the loss of memory rate τ, although the parameter is important for the predictive abilities for the model. We chose therefore τ, γ and ϵ to optimise the predictive ability of the model on historical data. We ran our model on the second half of the four seasons from 1993-97 both for Premier League and Division 1, and predicted successive each round the second half of each season. (We need to use the first half of each season to learn about the different teams.) To quantify the quality of the predictions, we computed the geometrical average of the probabilities for the observed results (home, draw and away) for each match played so far. This has a flavour of being a normalised pseudo likelihood measure as it is a product of conditional probabilities. We denote this measure by PL. We repeated this procedure for various values of τ, γ and ϵ on a three-dimensional grid. (This was a computational expensive procedure!) To our surprise, there seems to be a common set of values for the parameters that gave overall reasonable results for both leagues and for all seasons. These values were $\tau = 100$ days, $\gamma = 0.1$ and $\epsilon = 0.2$. (The value of $\tau = 100$ corresponds to prior variance of $\sigma_0^2 = 1/37$, so the "optimal" τ is linked to the prior variance through $\sigma_0^2/\tau = 1/3\,700$.) Although we found parameter values giving higher PL for each of the seasons and leagues, we chose to use the common set of parameter values as these simplify both the interpretation of the model and its use.

4 Applications and Results

This section contains applications of the proposed model for the Premier League and Division 1 for 1997-98, in prediction, betting and retrospective analysis of the season.

For simplicity we selected values uniformly spaced in the interval -0.2 to 0.2 as the prior means ($\mu_{a,A}, \mu_{a,B}$ and so on) for the attack and defence strength in the Premier League, based on our prior ranking of the teams. A similar procedure was chosen for Division 1. Another approach is to use the mean properties (possible adjusted) from the end of last season, at least for those teams that stays in the same league. As the prior mean is only present in the first match for each team, all reasonable prior mean values give approximately the same predictions for the second half of the season. The prior mean values are most important for predicting the first rounds of each season. We further used the parameter values suggested in Section 3.2. In the forthcoming experiments, the number of iterations was checked to give reliable results. We used $100\,000$ iterations in the prediction application, and $1\,000\,000$ iterations in the retrospective analysis.

4.1 Prediction and Betting

To compare our predictions and to simulate the betting experiments, we used the odds provided by one of the largest international bookmakers operating on the web, Intertops. Although the firm is officially located in Antigua in West Indies, it is easily accessible from anywhere in the world via Internet at the URL `http://www.Intertops.com`. From their odds for the Premier League and Division 1, we computed the corresponding (predictive) probabilities for home/draw/away.

4.1.1 Predictions

Figure 3 shows the PL as a function of time for the second half of the 1997-98 season in Premier League and Division 1, using the first half of each season to learn about the different teams. Both leagues are nearly equally predictable from the bookmaker point of view, with a final PL-values of .353 and .357 for Premier League and Division 1, respectively. Our model does surprisingly well compared to the bookmaker with a final PL-values of .357 and .372 for Premier League and Division 1, respectively. The prediction are especially good for Division 1. Recall that the model only make use of the match results and not all other information which is available to those who set the bookmaker odds. It seems like bookmakers provide better odds for Premier League than for the lower divisions, which might be natural as the majority of the players bet on the Premier League.

4.2 Single Bets

We can simulate a betting experiment against Intertops using the above predictions. Assume $\mathcal{B}$ is the set of matches we can bet on. Which matches should we bet on and how much? This depends on our utility for betting, but as we decide ourself which matches to bet on, we have a favourable game to play as the posterior expected profit is positive. The statement is conditional on a "correct" model, and a betting experiment will therefore validate our model. For favourable games, Epstein (1995) suggests to bet on outcomes with a positive expected profit but place the bets so we obtain a low variance of the profit. This strategy will additional to a positive expected profit, also make the probability of ruin, which is an absorbing state, small. We chose therefore to place the bets to maximise the expected profit while we keep the variance of the profit lower than some limit. An equivalent formulation is to maximise the expected profit minus the variance of the profit, which determine how we should place our bets up to a multiplicative constant. This constant can be found if we choose a specific value or an upper limit for the variance of the profit. Let μ_i^j and $\sigma_i^{2,j}$ be the expected profit and variance for betting a unit amount on outcome i in match j, where $i \in \{\text{home}, \text{draw}, \text{away}\}$. These values are found from the probability p_i^j and odds o_i^j for outcome i in match j. Let β_i^j be the corresponding bet, where we for simplicity restrict ourselves to place no more than one bet for each match. The optimal bets are found as

$$\arg\max_{\beta_i^j \geq 0} \; U(\{\beta_i^j\}), \quad \text{where} \quad U(\{\beta_i^j\}) = \text{E(profit)} - \text{Var(profit)} = \sum_{j \in \mathcal{B}} \beta_i^j \left(\mu_i^j - \beta_i^j \sigma_i^{2,j} \right).$$

The solution is $\beta_i^j = \max\{0, \mu_i^j/(2\sigma_i^{2,j})\}$, where additionally we choose the outcome i with maximal $\beta_i^j \mu_i^j$ for match j to meet the "not more than one bet for each match" requirement. Figure 4 shows the profit (scaled to have $\sum_j \beta_i^j = 1$ for all bets made so far) using the predictions and odds in Figure 3, together with an approximate 95% interval given as posterior mean± 2 posterior standard deviation. The results are within the upper and lower bound, although the lower bound at the end of the season is negative indicating still a risk for loosing money. For Premier League the final profit was 39.6% after we won on 15 of a total of 48 bets on 17 home-wins, 5 draw and 26 away-wins. For Division 1 the final profit was 54.0% after we won on 27 of a total of 64 bets on 30 home-wins, 6 draw and 28 away-wins. The final bounds were $(-47.2\%, 87.8\%)$ and $(-29.5\%, 58.0\%)$ for Premier League and Division 1, respectively. The β_i^j's varied from 0.001 to 0.211 with an average of 0.047 for Division 1.

It is not enough to predict the match result just slightly better than the bookmaker to earn money when we bet on single matches. The bookmaker takes some percentage of the bets as tax by reducing their odds to less than one over their probability for home-win, draw and away-win, and the odds from Intertops satisfy $1/o^j_{\text{home}} + 1/o^j_{\text{draw}} + 1/o^j_{\text{away}} \approx 1.2$.

The high profit in end January for the Premier League, is due to a single match, Manchester United versus Leicester City at January 31. Intertops gave high odds for an away-win, 13.8, while our model predicted 0.184 chance for an away-win. As Leicester C won $0-1$, this bet gave an significant pay-off.

4.2.1 Combo Bets

The Intertops also provides the opportunity for "combo"-bets; Bet on the correct results for more than one match simultaneously. We have investigated the profit if we chose this option where we (for simplicity) place our bets on the correct results of three matches simultaneously. The probability for getting three matches correct is $p_i^j p_{i'}^{j'} p_{i''}^{j''}$ (approximately only, as the teams properties are dependent), and this event has odds $o_i^j o_{i'}^{j'} o_{i''}^{j''}$. How should we now place our bets $\beta_{i,i',i''}^{j,j',j''}$? The same argument as for single bets applies, place the bets to maximise the expected profit minus the variance. Although the idea is similar, we have now dependency between the various bets as some matches can be in more than one combination of three matches. Let $\boldsymbol{\beta}$ be the vector of bets, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the vector of expected values and covariance matrix for all the available combos with a unit bet. We should place our bets proportional to

$$\arg\max_{\boldsymbol{\beta}\geq \mathbf{0}} \left(\boldsymbol{\beta}^T\boldsymbol{\mu} - \boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta}\right) \tag{8}$$

This is a standard quadratic linear programming problem which is easily solved through well known algorithms, although the covariance matrix $\boldsymbol{\Sigma}$ is somewhat tedious to calculate. We choose our candidates only among those outcomes which we bet on in the single case to obtain a reasonable dimension of the problem in (8). The simulated combo-betting experiment gave less satisfying results. The final profit were -100% (140.2%) after 35 bets and 80.3% (109.7%) after 63 bets for the Premier League and Division 1, with the posterior standard deviation given in the parentheses. If we merge the two divisions together, the profit were 9.7% with a large variance compared to the variance obtained using single bets. It seems to be both easier and more reliable to bet on single matches compared to combo-bets.

4.3 Retrospective analysis of Premier League 1997-98

According to the model assumptions the match results in Premier League 1997-98 updates information about defending and attacking strength for all teams throughout the season. Given this information, it is interesting to know whether Arsenal was lucky to win the Premier League 1997-98. Similar questions arise for other teams; Was Everton lucky to stay in the league? Did Aston Villa deserve their 7th place? It is easy to provide the answer from the model for such questions using the power of MCMC by playing a new match for each of the 380 matches using samples from the *joint* posterior densities for all properties and at all times. By collecting the points and goals made we can compute a conditional sample of the final ranking. We repeat this procedure after each iteration in the MCMC, and compute from all samples in the end the estimates we are interested in. (In this analysis we increased the prior variance for all properties by a factor of 10, and similar with τ to keep the conditional variance in (5) unchanged. These near non-informative priors makes more sense in a retrospective analysis.)

Table 1 shows the estimated (posterior) probabilities for Arsenal, Manchester U, Liverpool and Chelsea to be the first, second and third in the final ranking. The table shows also the observed rank and the number of points achieved. Manchester U had probability .433 of winning the League while Arsenal had .247. Liverpool and Chelsea have similar probabilities for being first, second and third. It seems like Arsenal was lucky to win the title from Manchester U.

Figure 5 gives a more complete picture of the final ranking and displays the expected final rank for each team with approximate 90% (marginal) credibility intervals. The solid line shows the observed ranking. We see from the graph that Everton would have been unlucky to be relegated, and Aston Villa did it better than expected. The uncertainty in the final ranking is surprisingly large, and the observed rank seems to be well within the uncertainty bounds. Aston Villa, for example, could easily have been 15th instead of 7th. The managers surely have to face a lot of uncertainty. It is interesting to note from the graph that the 20 teams divide themselves into four groups: The top four, upper middle seven, lower middle seven, and the bottom two.

To study more how the top four teams differ and how their skills varied though the season, we compute the (posterior) expected value of the offensive and defensive strength as a function of time. Figure 6 shows the result. The difference in defensive skills of the four teams are prominent, while their offensive skills are more similar. Manchester U had a good and stable defence while their attack strength decreased somewhat in the second half of the season. Denis Irwin was badly injured in December, and might be one reason. Later in the season both Ryan Giggs and Nicky Butt suffered from injuries and suspension causing attacking strength to decrease. The defensive skills of Arsenal improved during the season while their offensive skills were best at the beginning and the end of the season. Manchester U defensive skills are superior to Arsenal's during the hole season, while Manchester U's offensive skills are somewhat better in the period of October to March. In total, Manchester U seems to be the strongest team. Liverpool and Chelsea have similar and stable defensive qualities, while there offensive is monotone increasing for Liverpool and monotone decreasing for Chelsea. Arsenal is clearly ranked ahead of Liverpool and Chelsea mainly due to their strong defence. Liverpool is ranked ahead of Chelsea as they had both slightly better defence and attack properties on average. However, this is not a sufficient condition in general; Also which teams they meet at which time is important.

An amusing application of the model appears when we study the posterior probability for the Bernoulli random variable attached to each match. (These variables indicate which one of the distributions on the right hand side of Eq. (3) we currently use, see the Appendix for details.) We ranked each match in the Premier League after the posterior probability for these Bernoulli variables to be 1. This probability has the interpretation as the (posterior) probability for that match to be unexplainable or an outlier, and hence gives a way of locating those matches that were most surprising taking both the observed past and future into account. Table 2 lists the five most surprising results in Premier League 1997-98. The match between Liverpool and Barnsley November 22 where Barnsley won $0 - 1$, is a clear winner of being the surprise match of the season.

5 Discussion and Further Work

The presented model seems to grab most of the information contained in the match result and provide reasonable predictions. The seemingly stability for the γ, τ and ϵ parameter across seasons, is one confirmation. Although the number of variables are more than the number of data, we are not in the situation of over-fitting the data. We take the (posterior) dependency between the attack and defence strength at different time points as various ways to explain the data.

Further, the presented approach seems superior to the earlier attempts to model soccer games as it $i)$ allows for coherent inference of the properties between the teams also in time, $ii)$ easily account for the joint uncertainty in the variables which is important in prediction (Draper, 1995), $iii)$ allows for doing various interesting retrospective analysis of a season, and finally $iv)$ provides a framework where is it easy to change parts or parametrisation in the model. We do not claim that our parametrisation, goal and time model is optimal and cannot be improved on, but that the presented Bayesian approach with MCMC based inference seems promising for these kinds of problems.

There are several points which could and should be improved in the model.

DATA It is of major importance to include more data than just the final match-result into the model, but this depends on what kind of data are (easily) available and useful. No attempts are done along these lines as far as we are aware of. This will imply a change of the model as well, but the basic ideas and framework will remain.

TIME MODEL Brownian motion is a too simple time-model for the team's properties and does not include the first derivative (local trend) in the predictions. A non-stationary time-model is needed to capture the *local* behaviour needed for prediction in the near feature. An integrated autoregressive process might be suitable if we discretize the time which is quite reasonable. Such choice requires (among others) changes in move type 1 in the MCMC algorithm described in the Appendix.

PARAMETER ESTIMATION We assumed that all teams have a common loss-of-memory rate τ and this is a simplification. We have not succeeded estimating a team-specific τ, or found a good way to group each team into a "high/normal/low" loss of memory rate. More observation data than just the final match-result are most likely needed to make progress in this direction.

GOAL MODEL The goal model could be improved on. The birth-process approach of Dixon and Robinson (1998) is natural and interesting, although one should estimate the goal model simultaneously with the time varying properties, coherently. Further, various parametrisations like the inclusion of the psychological effect and the idea of a mixture model, needs to be investiaged further within their birth-process framework.

HOME FIELD ADVANTAGE Each teams (constant) home field advantage is a natural variable to include in the model. We did not find sufficient support from the match results to include this at the current stage, but hopefully more data will change this.

It seems like the statistical community is making progress in understanding the art of predicting soccer matches, which is of vital importance for two reasons: $i)$ demonstrate the usefulness of statistical modelling and thinking on a problem that most people *really* care about, and $ii)$ make us all rich on betting!

ACKNOWLEDGEMENT

We are grateful the Associated Editor and the two referees for valuable comments that helped us improve the paper, and to our colleagues Georg Elvebakk and Bo Lindqvist for valuable discussions on statistical aspects of soccer.

REFERENCES

ANDERSON, E., BAI, Z., BISCHOF, C., DEMMEL, J., DONGARRA, J., CROZ, J. D., GREENBAUM, A., HAMMARLING, S., MCKENNEY, A., OSTROUCHOV, S., AND SORENSEN, D. (1995). *LAPACK Users' Guide*, 2 edn, Philadelphia: Society for Industrial and Applied Mathematics.

DIXON, M. J. AND COLES, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market, *Applied Statistics* 46(2): 265–280.

DIXON, M. J. AND ROBINSON, M. E. (1998). A birth process model for association football matches, *The Statistician* ??(??): ??–?? (to appear).

DRAPER, D. (1995). Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society, Series B* 57(1): 45–97. With discussion and a reply by the author.

EPSTEIN, R. A. (1995). *The theory of gambling and statistical logic*, Academic Press, New York.

GILKS, W. R., RICHARDSON, S. AND SPIEGELHALTER, D. J. (1996). *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall.

KUONEN, D. (1996). Statistical models for knock-out soccer tournaments, *Technical report*, Department of Mathematics, École Polytechnique Fédérale de Lausanne.

LEE, A. J. (1997). Modeling scores in the Premier League: is Manchester United really the best?, *Chance* 10(1): 15–19.

LIU, J. S., WONG, W. H. AND KONG, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes, *Biometrica* 81(1): 27–40.

RIDDER, G., CRAMER, J. S. AND HOPSTAKEN, P. (1994). Down to ten: Estimating the effect of a red card in soccer, *Journal of the American Statistical Association* 89(427): 1124–1127.

RUE, H. (1999). A fast and exact simulation algorithm for general Gaussian Markov random fields, *Statistics no. 8*, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

SHEPHARD, N. AND PITT, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series, *Biometrika* 84(3): 653–667.

WEST, M. AND HARRISON, J. (1997). *Bayesian forecasting and dynamic models*, 2 edn, Springer Verlag.

WHITTAKER, J. (1990). *Graphical models in applied multivariate statistics*, Wiley series in probability and mathematical statistics. Probability and mathematical statistics, Chichester: Wiley.

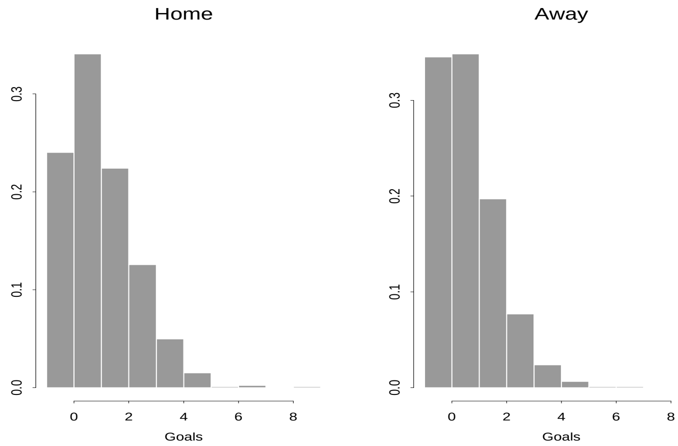


FIGURE 1: Histogram of the number of home and away goals in 924 matches in the Premier League 1993-95.

	P($1st$)	P($2nd$)	P($3rd$)	Rank	Pts
Arsenal	.247	.230	.161	1	78
Manchester U	.433	.230	.131	2	77
Liverpool	.110	.151	.153	3	65
Chelsea	.095	.134	.142	4	63

TABLE 1: The estimated posterior probabilities for each team being the first, second and third in the final ranking in Premier League 1997-98, together with the observed rank and the number of points achieved.

Match			Date	prob(outlier)	Result
Liverpool	–	Barnsley	Nov 22 1997	.76	0 – 1
Newcastle	–	Leicester C	Nov 1 1997	.66	3 – 3
Wimbledon	–	Tottenham	May 2 1998	.61	2 – 6
Sheffield W	–	Manchester U	Mar 7 1998	.60	2 – 0
Sheffield W	–	Arsenal	Nov 22 1997	.59	2 – 0

TABLE 2: The five most surprising results in Premier League 1997-98, ranked according to the posterior probability for being unexplainable or an outlier.

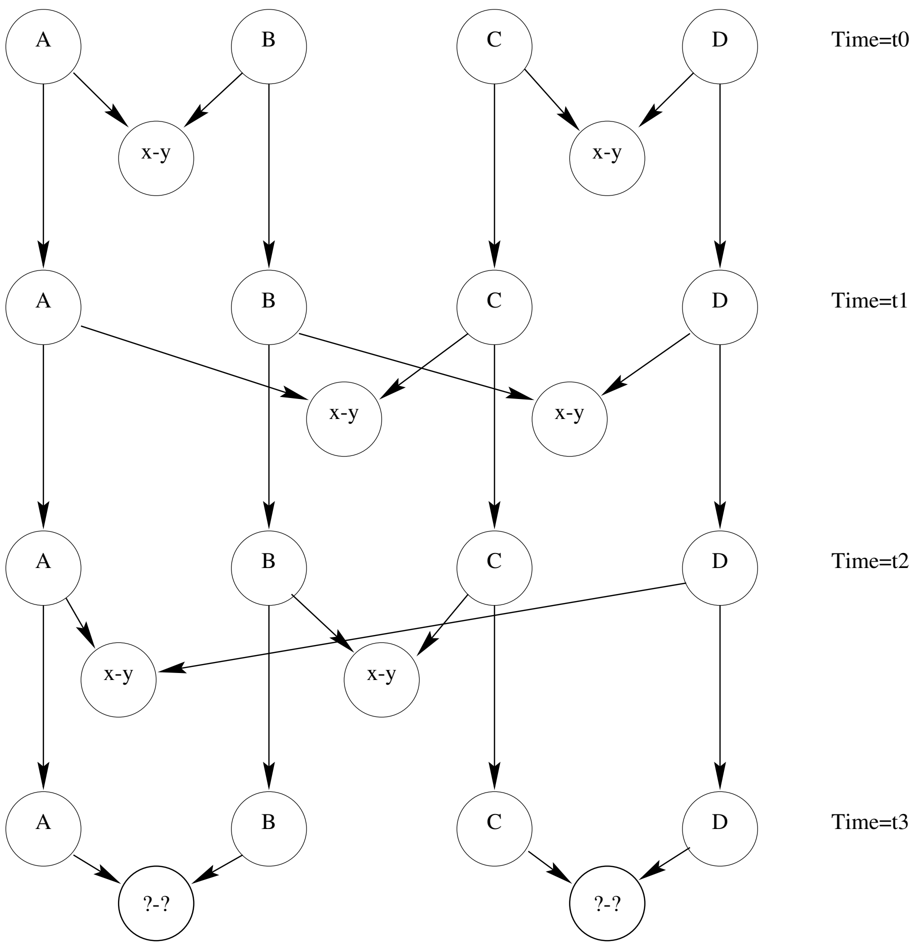


FIGURE 2: The directed acyclic graph describing the causal structure in our model with four teams and eight matches.

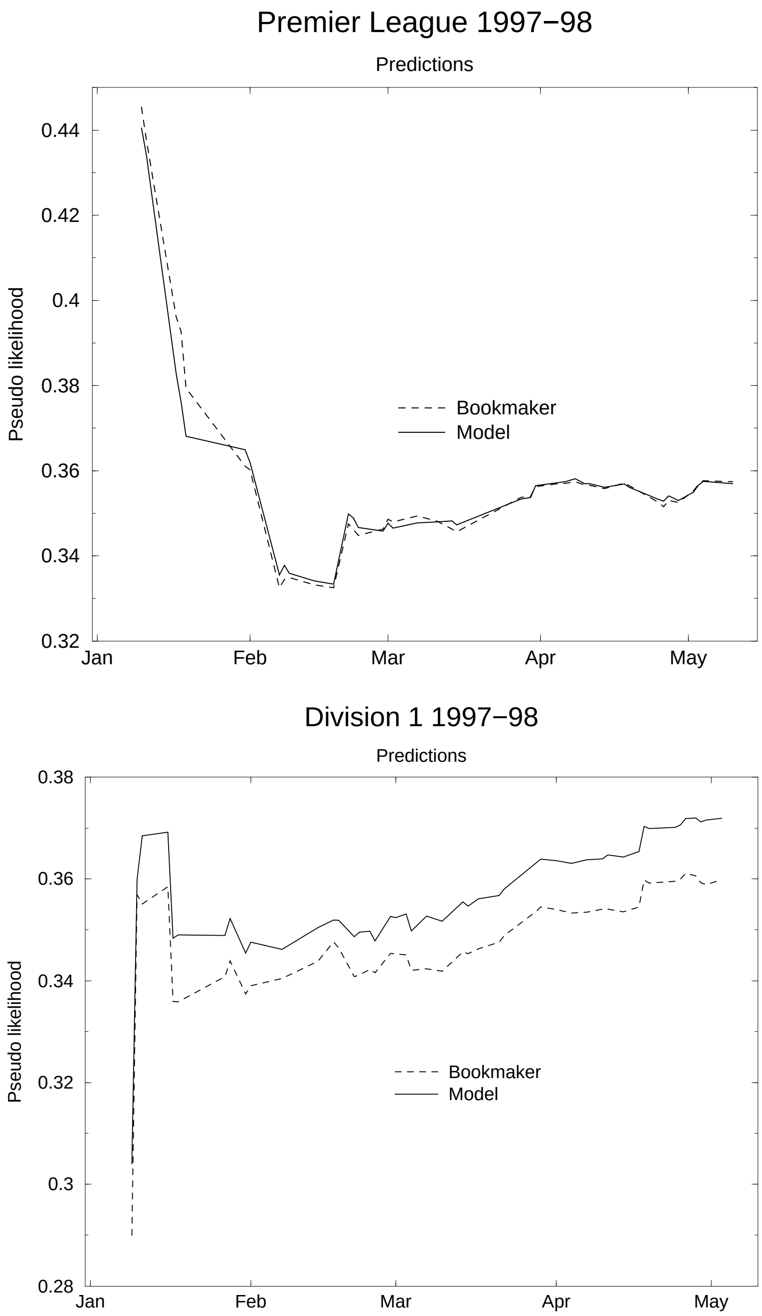


FIGURE 3: The figures display the PL measure for the predictions made by the model and odds from Intertops in Premier League and Division 1.

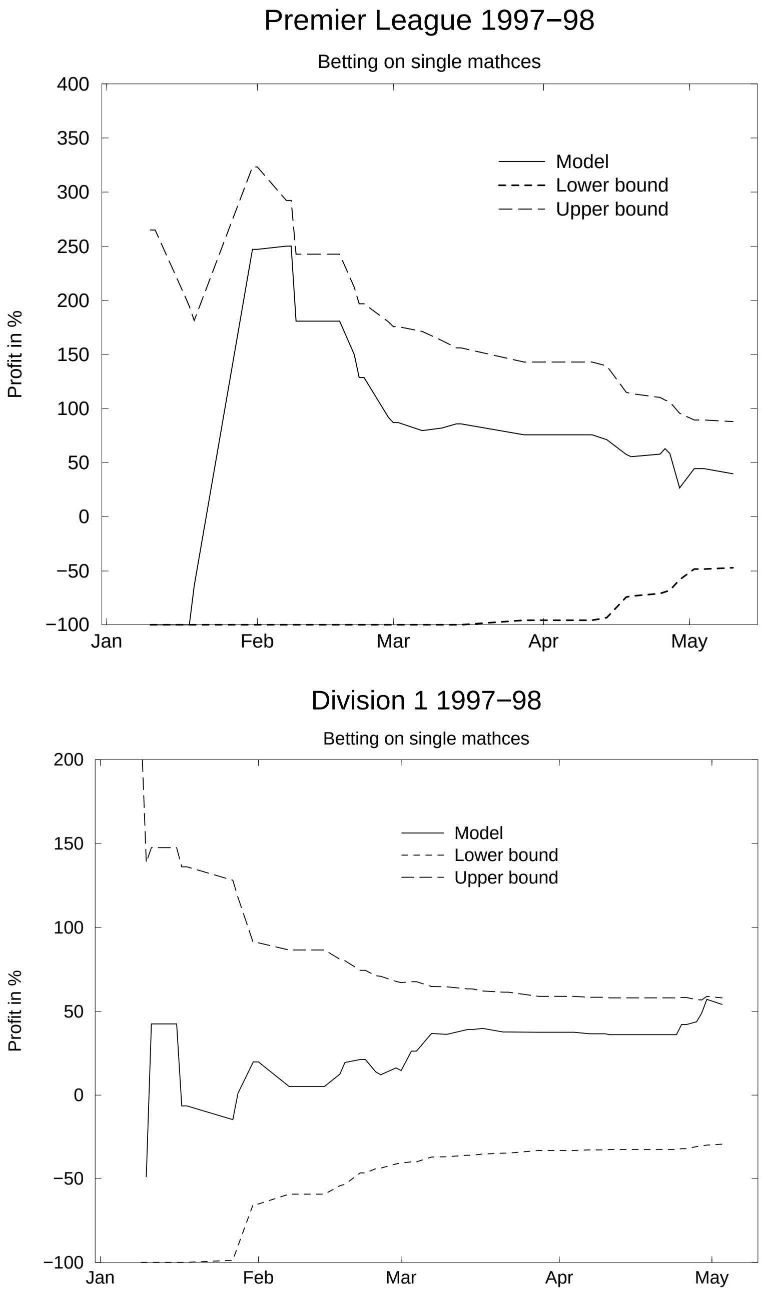


FIGURE 4: The observed profit in the simulated betting experiments for the 1997-98 season in Premier League and Division 1 using the predictions in Figure 3. The bets are on single matches against the odds provided from Intertops.

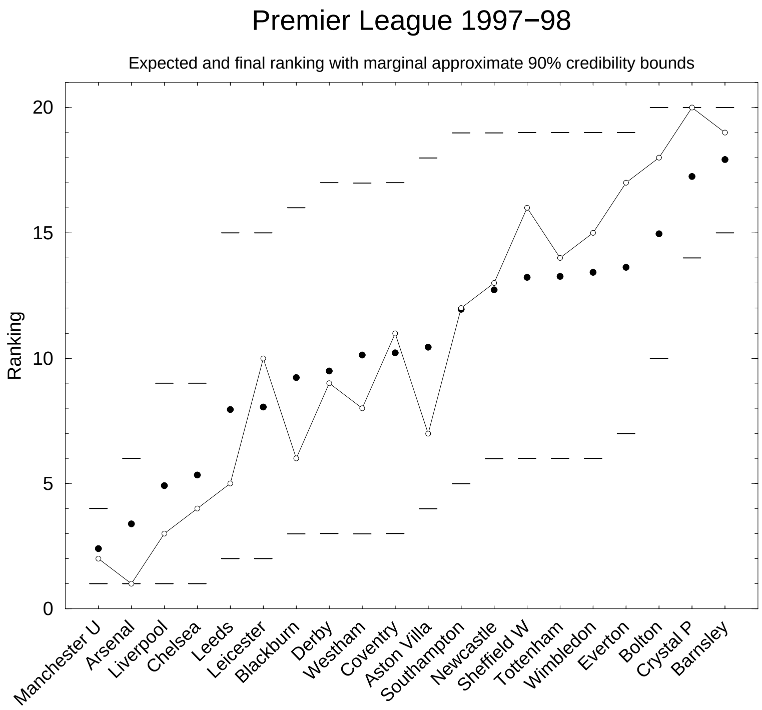
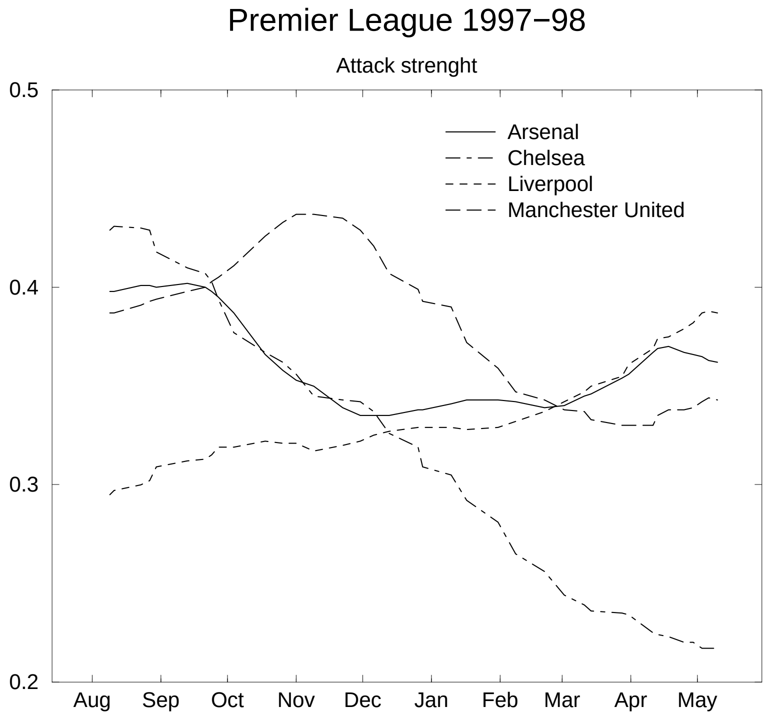


FIGURE 5: The figure show the (posterior) expected final rank (dots) for each team with approximate 90% marginal credibility bounds, together with the observed ranking (solid line). Note the large uncertainty in the final ranking.



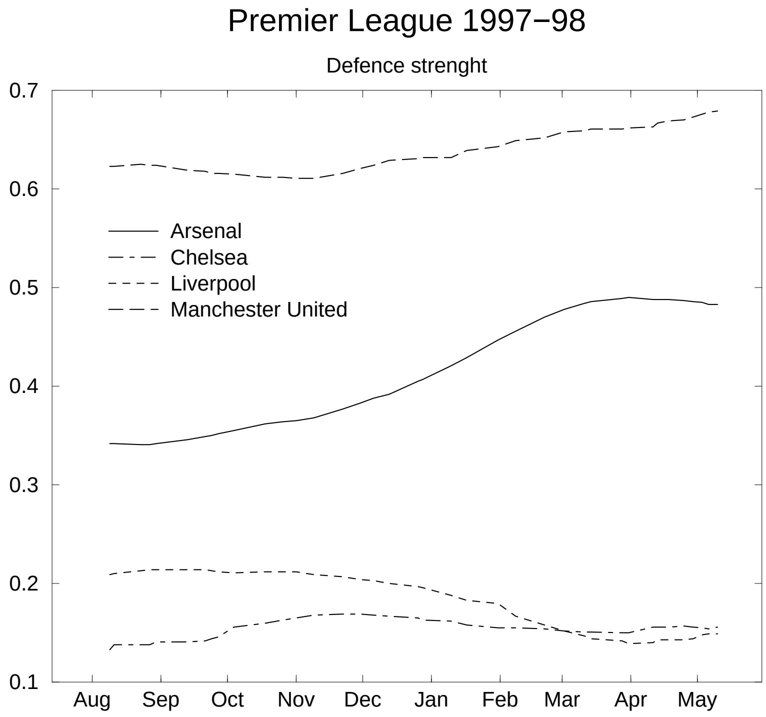


FIGURE 6: The retrospective estimates of the mean offensive and defensive strength for the four best teams during the season.

Appendix Details of the MCMC algorithm

This appendix contains details of our MCMC implementation sketched in Section 3.1. To ease the implementation of the mixture model in Eq. (6), we will use an equivalent reformulation; Define $\delta_{A,B}$ as an independent Bernoulli variable which is 1 with probability ϵ, and define

$$\pi_{g2}\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}, \delta_{A,B}\right) = \begin{cases} \pi^*_{g1}\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}\right) & \text{if } \delta_{A,B} = 0 \\ \pi^*_{g1}\left(x_{A,B}, y_{A,B} \mid \exp(c^{(x)}), \exp(c^{(y)})\right) & \text{if } \delta_{A,B} = 1. \end{cases} \tag{9}$$

Then with obvious notation

$$\pi_g\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}\right) = \mathrm{E}_{\delta_{A,B}}\left[\pi_{g2}\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}, \delta_{A,B}\right)\right].$$

Thus, we can attach one Bernoulli variables to each match and update also these variables in the MCMC algorithm. We ignore their values in the output analysis where we consider only those components of $\boldsymbol{\theta}$ that is of our interest. This yields a correct procedure as the marginal distribution for $\boldsymbol{\theta}$ remains unchanged when we include the Bernoulli variables.

Due to the reformulation of the mixture distribution, we need three different move types to implement an MCMC algorithm for our model; 1) Update the properties for each team every time there is a match, 2) update the match result for each unobserved match, and 3) update the Bernoulli variable for each match. In each full sweep we visit all unobserved (stochastic) variables in a random order and update each one using the appropriate move type.

Move type 1. Updating one of the properties We describe only how we update the attack strength $a^{t''}_A$ for team A at time t'' using a Metropolis step, as the update of the defence strength is similar. Note that all other variables remain constant when we propose an update for $a^{t''}_A$. We assume team A plays a match against team B at time t'' and at A's home-ground, as the acceptance rate when A plays on B's home-ground is similar with obvious changes. Let t' and t''' be the times of the previous and following matches for team A. We will soon return to the case when there is no previous and/or following match. Denote by $(x^{t''}_{A,B}, y^{t''}_{A,B})$ and $\delta^{t''}_{A,B}$ the (current, if not observed) number of goals in the match and the Bernoulli variable attached to that match, respectively.

We sample first a new proposal for $a^{t''}_A$ from a Gaussian (symmetric) distribution, $a^{t'',\text{new}}_A \sim \mathrm{N}(a^{t''}_A, \sigma^2_q)$, where σ^2_q is a fixed constant for all teams, attack and defence. For all our examples in Section 4, we used $\sigma^2_q = 0.05^2$. The new proposal is accepted with probability $\min\{1, R\}$, where

$$R = \frac{\pi(a^{t'',\text{new}}_A \mid a^{t'}_A)}{\pi(a^{t''}_A \mid a^{t'}_A)} \, \frac{\pi(a^{t'''}_A \mid a^{t'',\text{new}}_A)}{\pi(a^{t'''}_A \mid a^{t''}_A)} \, \frac{\pi_{g2}\left(x^{t''}_{A,B}, y^{t''}_{A,B} \mid \lambda^{(x),t'',\text{new}}_{A,B}, \lambda^{(y),t'',\text{new}}_{A,B}, \delta^{t''}_{A,B}\right)}{\pi_{g2}\left(x^{t''}_{A,B}, y^{t''}_{A,B} \mid \lambda^{(x),t''}_{A,B}, \lambda^{(y),t''}_{A,B}, \delta^{t''}_{A,B}\right)} \tag{10}$$

otherwise we remain in the old state. In Eq. (10), $\pi(a^{t'',\text{new}}_A \mid a^{t'}_A)$ denotes the conditional Gaussian density for $a^{t''}_A$ given $a^{t'}_A$ evaluated at $a^{t'',\text{new}}_A$. Further is $\lambda^{(x),t'',\text{new}}_{A,B}$ computed from Eq. (1) using the proposed new value $a^{t'',\text{new}}_{A,B}$ and so on. If there is no previous match, then $\pi(a^{t'',\text{new}}_A \mid a^{t'}_A)$ and $\pi(a^{t''}_A \mid a^{t'}_A)$ is replaced with the prior density for a_A evaluated at $a^{t'',\text{new}}_A$ and $a^{t''}_A$, respectively. If there is no following match, then we remove $\pi(a^{t'''}_A \mid a^{t'',\text{new}}_A)$ and $\pi(a^{t'''}_A \mid a^{t''}_A)$ in the expression for R.

We prefer this simple Metropolis step compared to the more elegant proposal density found from computing the Gaussian approximation to the conditional density, by a second order Taylor expansion of the log conditional density around current values. Although the acceptance rate with a tailored Gaussian proposal increases to well over 90%, it does not seems to be worth the additional computation and implementation costs.

MOVE TYPE 2. UPDATING A MATCH RESULT We update the match result using a Gibbs-step, thus drawing $(x^{t''}_{A,B}, y^{t''}_{A,B})$ from the conditional distribution in Eq. (9). The modifications needed due to truncation and $\kappa(x^{t''}_{A,B}, y^{t''}_{A,B} \mid \lambda^{(x),t''}_{A,B}, \lambda^{(y),t''}_{A,B})$, are easily done by rejection steps. The algorithm is as follows.

1. If $\delta^{t''}_{A,B}$ is 1, then set

$$\lambda^{(x)} = \exp(c^{(x)}), \quad \text{and} \quad \lambda^{(y)} = \exp(c^{(y)})$$

otherwise set

$$\lambda^{(x)} = \lambda^{(x),t''}_{A,B} \quad \text{and} \quad \lambda^{(y)} = \lambda^{(y),t''}_{A,B}.$$

2. Draw x from $\text{Po}(x \mid \lambda^{(x)})$ until $x \leq 5$, and then draw y from $\text{Po}(y \mid \lambda^{(y)})$ until $y \leq 5$.
3. With probability

$$\frac{\kappa\left(x, y \mid \lambda^{(x)}, \lambda^{(y)}\right)}{\max\{1.1, 1 + 0.1\lambda^{(x)}\lambda^{(y)}\}}$$

set $x^{t''}_{A,B} = x$ and $y^{t''}_{A,B} = y$ and return, otherwise go back to 2.

MOVE TYPE 3. UPDATING A BERNOULLI VARIABLE. The Bernoulli variable attached to each match, $\delta^{t''}_{A,B}$, is a random variable with probability ϵ to be 1, and denote the law by $\pi(\delta^{t''}_{A,B})$. We propose always to flip the current value of $\delta^{t''}_{A,B}$ to $\delta^{t'',\text{new}}_{A,B} = 1 - \delta^{t''}_{A,B}$ which is accepted with probability $\min\{1, R\}$,

$$R = \frac{\pi_{g2}\left(x^{t''}_{A,B}, y^{t''}_{A,B} \mid \lambda^{(x),t''}_{A,B}, \lambda^{(y),t''}_{A,B}, \delta^{t'',\text{new}}_{A,B}\right)}{\pi_{g2}\left(x^{t''}_{A,B}, y^{t''}_{A,B} \mid \lambda^{(x),t''}_{A,B}, \lambda^{(y),t''}_{A,B}, \delta^{t''}_{A,B}\right)} \frac{\pi(\delta^{t'',\text{new}}_{A,B})}{\pi(\delta^{t''}_{A,B})}.$$

A COMMENT ON IMPLEMENTATION The model is easiest programmed as a graph-model (see Figure 2) parsing matches with teams, dates, etc. from external input files. The described MCMC algorithm can be modified in several ways to achieve significant speedup. Our approach was to tabulate the truncated Poisson distribution for a large set of λ's, and then use a table lookup with interpolation to obtain values. The normalisation constants for the joint density $\pi^*_{g1}(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B})$ is also needed as a function of $(\lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B})$, which we once more tabulate. It is also possible to tabulate $\pi_g\left(x_{A,B}, y_{A,B} \mid \lambda^{(x)}_{A,B}, \lambda^{(y)}_{A,B}\right)$ directly if ϵ is the same for all matches (as in our examples). In this case we can avoid the above reformulation of the mixture distribution using the Bernoulli variable attached to each match. On the other hand, this prohibit the analysis of the most surprising matches presented in Table 2.

A COMMENT ON RAO-BLACKWELLISATION To predict a future match, A against B say, it is natural to consider the simulated result $(x_{A,B}, y_{A,B})$ of that match and estimate the probability for A wins against B by counting how many times $x_{A,B}$ is grater than $y_{A,B}$ and divide by the total number. However, we can decrease the variance of this estimator by Rao-Blackwellisation; We compute $\Pr(A \text{ wins over } B \mid \boldsymbol{e}_A, \boldsymbol{e}_B)$ and use the empirical mean of this conditional probability as our estimate for the probability that A wins against B. (Again, we tabulate these probabilities and use table lookup with interpolation.) We refer to Liu, Wong and Kong (1994) for a theoretical background of Rao-Blackwellisation in this context.

A COMMENT ON BLOCK-UPDATING The structure of the full model makes it possible to use block-updates of the properties of each team within the MCMC algorithm. A first natural choice is to update the attach strength, say, for team A at all time-points $t_0, \ldots, t_{n-1}$ where team A plays a match, $\boldsymbol{a}_A = (a_A^{t_0}, \ldots, a_A^{t_{n-1}})^T$. A-priori, $\boldsymbol{a}_A$ is a (non-stationary) Gaussian vector (with length n) with a Markov property, and it is then tempting to update $\boldsymbol{a}_A$ using a Gaussian proposal in the MCMC algorithm found from a second-order Taylor expansion of the log-full conditional of $\boldsymbol{a}_A$ around the current state,

$$\begin{aligned} \log \pi(\boldsymbol{a}_A \mid \text{the rest}) &= C + \sum_{i=0}^{n-1} \left(\log \pi(a_A^{t_i} \mid a_A^{t_{i-1}}) + \log \pi(x_{A,B_i}^{t_i}, y_{A,B_i}^{t_i} \mid \lambda_{A,B_i}^{(x)}, \lambda_{A,B_i}^{(y)}) \right) \\ &\approx C' + \boldsymbol{b}^T \boldsymbol{a}_A - \frac{1}{2} \boldsymbol{a}_A^T \boldsymbol{Q} \boldsymbol{a}_A. \end{aligned} \tag{11}$$

Here, $\pi(a_A^{t_0} \mid a_A^{t_{-1}})$ is the prior, C and C' are (arbitrary) constants, $\boldsymbol{b}$ is a $n \times 1$ vector and $\boldsymbol{Q}$ a symmetric $n \times n$ tri-diagonal (precision) matrix. Due to notational convenience we supposed in Eq. (11) that team A played only home-matches against team B_i at time t_i. Note that the computation of Eq. (11) require only $\mathcal{O}(n^2)$ flops due to the Markov property. Assume $\boldsymbol{Q}$ is positive definite, then we can sample (and compute the normalisation constant) from the Gaussian proposal with a Markov property using only $\mathcal{O}(n)$ flops, see Rue (1999) for details. The algorithm goes as follows: Compute the band-Cholesky factorisation of $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^T$, where $\boldsymbol{L}$ has lower-bandwidth equal to 1 due to the Markov property of $\boldsymbol{a}_A$. Solve by forward and backward substitution, $\boldsymbol{L}\boldsymbol{u} = \boldsymbol{b}$ and then $\boldsymbol{L}^T \boldsymbol{\mu} = \boldsymbol{u}$, to obtain the mean $\boldsymbol{\mu}$ of the proposal density. We can now sample $\boldsymbol{a}_A^{\text{new}}$ from the Gaussian propsal density by first solving by backward substitution $\boldsymbol{L}^T \boldsymbol{v} = \boldsymbol{z}$, where $\boldsymbol{z}$ is a vector with independent standard Gaussian variates, and then add the mean, $\boldsymbol{a}_A^{\text{new}} = \boldsymbol{\mu} + \boldsymbol{v}$. Note that high-performance implementations of all required numerical algorithms exists in the freely available Lapack-library (Anderson, Bai, Bischof, Demmel, Dongarra, Croz, Greenbaum, Hammarling, McKenney, Ostrouchov, and Sorensen, 1995). Computation of the acceptance rate for this block-update is straight forward after computing the approximation Eq. (11) around $\boldsymbol{a}_A^{\text{new}}$ to obtain the proposal density of the reverse move from $\boldsymbol{a}_A^{\text{new}}$ to $\boldsymbol{a}_A$.

Shephard and Pitt (1997) used similar ideas for non-Gaussian measurement time series with good results, and report simulation experiments studing the efficiency of using block-updates in the MCMC algorithm. The efficiency of blocking compared to sing-site updating can be large if $\boldsymbol{a}_A$ is highly dependent. For our application, the dependency is not as high that single-site updating does not perform reasonable well. However, an experimental study using block-updates of $\boldsymbol{a}_A$ or even better, $(\boldsymbol{a}_A, \boldsymbol{d}_A)^T$ simultaneously, would be interesting. Similar linearisation ideas like Eq. (11) could be used more frequently as (very) fast samplings algorithms (including evaluation of the normalisation constant) exists for Gaussian proposals (also on lattices) with a Markov property, see Rue (1999) for algorithmic details.