

Dynamic Bayesian forecasting of AFL match results using the Skellam distribution

A. A. Manderson^{1*}, K. Murray², and B. A. Turlach³

University of Western Australia

Summary

The scoring and defensive abilities of Australian Rules Football teams change over time as a result of evolving player rosters, tactics and other management factors. We develop a dynamic model based on the Poisson difference (Skellam) distribution which simultaneously models the two different point scoring mechanisms in Australian Rules Football, the motivation for which comes from work on predicting outcomes in soccer matches. Our model is developed in a Bayesian framework and is fitted using the Stan modelling language. Model validation is performed on the 2015 Australian Football league (AFL) home and away season.

Key words: Australian rules football; dynamic models; Bayesian hierarchical models; Stan; Skellam

1. Introduction

Australian Rules Football is one of the most popular sports in Australia and as such it has a considerable fan base and extensive media coverage. Speculating on the outcome of such sporting events is widespread, with many punters betting on the outcomes of football matches each week, thus providing motivation to develop statistical models to help make informed bets. Previously, work on modelling and predicting the outcomes of Australian Football League (AFL) matches has been attempted by Clarke (1993) and Ryall & Bedford (2010). They assigned a singular rating to each team, and updated this rating as the results from each round became available, similar to that of the Elo (1978) system used in competitive chess. While such models have also been used for sports other than AFL, these sports also saw the development of various *attack and defence strength* models. In particular several of these models have been proposed to predict results in association football leagues, including the

*Author to whom correspondence should be addressed.

¹School of Mathematics and Statistics (M019), University of Western Australia, 35 Stirling Highway, Crawley WA 6009

Email: andrew.manderson@research.uwa.edu.au

²School of Population and Global Health (M431), University of Western Australia, 35 Stirling Highway, Crawley WA 6009

³Centre for Applied Statistics (M019), University of Western Australia, 35 Stirling Highway, Crawley WA 6009

double Poisson model (Maher 1982), the bivariate Poisson (Karlis & Ntzoufras 2003), its stochastic parameter counterpart (Koopman & Lit 2015) and the Bayesian version (Baio & Blangiardo 2010); as well as ordinal models (Knorr-Held 2000; Brillinger 2007, 2008).

We propose a modelling framework similar to that of Baio & Blangiardo (2010) and Koopman & Lit (2015), which is based on the assumption that each team's score follows a Poisson distribution. Koopman & Lit (2015) allow for the two Poisson distributions to be correlated in a specific way such that the difference in goals is given by a Skellam distribution. However, given the high scoring nature of AFL games, modelling the scores (or differences in scores) directly as Poisson distributions could be problematic, due to overdispersion. For example, the English premier league football data used by Karlis & Ntzoufras (2003) showed that in approximately 85% of matches, each team scores two or fewer goals. In contrast to this, between the start of the 2013 season and the end of the 2015 AFL season approximately 36% of matches have ended with a score difference of more than 40 points. These high scores are due to the two types of scoring mechanisms in Australian Rules Football, where a goal is worth six points, a behind is worth one point, and the winning team is the one with the higher aggregate score at the end of the match. Our experience in attempting to model the scores and/or differences directly with Poisson/Skellam random variables led to very poor model fit and predictions. Consequently, we propose a model that differs from the soccer modelling approach of Koopman & Lit (2015) in order to account for these two types of scoring mechanisms in AFL and to accommodate the potentially larger score differentials between the teams. This model is distinct from many previous sports modelling approaches as it simultaneously models different scoring mechanisms, here two, thus encapsulating the highly variable nature of AFL scoring.

In this paper we use the Skellam distribution to model both the difference in goals and the difference in behinds, scored between two teams. However, in order to ensure we can fit our model without the need to develop application specific algorithms, such as in Karlis & Ntzoufras (2003), we express our model as a Bayesian hierarchical model. The hierarchical and Bayesian nature of the model allows us to use a readily available Bayesian inference language such as Stan (Stan Development Team 2016), which also allows us to fully specify the model in the Stan language, run MCMC chains, obtain parameter estimates, and make model predictions for the outcome of future matches.

The remainder of the paper is structured as follows: Section 2 contains the model details including a description of the Skellam distribution, construction of the dynamic Bayesian hierarchical model and the constraints required for identifiability. In Section 3 we describe the application of our methodology to data from the 2015 AFL home and away season, and provide some results. Finally, we provide some conclusions in Section 4. All code required to implement our methodology is freely available at <https://github.com/hhau/DynBayes>.

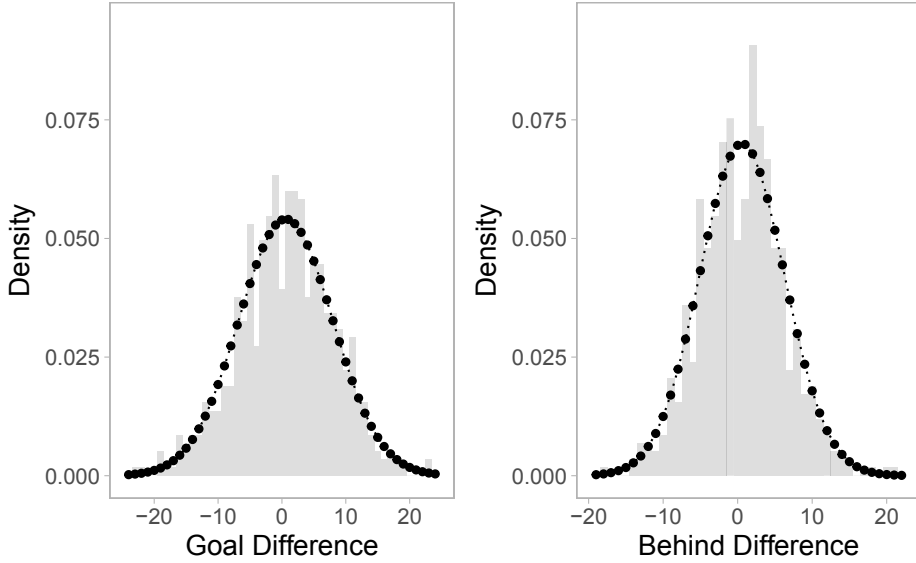


Figure 1. The goal and behind difference for the 2013-2015 AFL home and away seasons, with Skellam distributions fitted. The parameter values of the Skellam distributions are $\lambda_1 = 27.63, \lambda_2 = 27.02$ for the goal difference, and $\lambda_1 = 16.66, \lambda_2 = 16.07$ for the behind difference.

56

2. Model framework

The Skellam distribution (Skellam 1946) serves as the basis for our modelling, and is defined as the distribution of the difference between two independent Poisson distributed random variables. Specifically, if X and Y follow Poisson distributions with parameters λ_1 and λ_2 , respectively, then $Z = X - Y$ has the following probability mass function:

$$f(z | \lambda_1, \lambda_2) = e^{-(\lambda_1 + \lambda_2)} \left(\frac{\lambda_1}{\lambda_2} \right)^{z/2} I_z \left(2 \sqrt{\lambda_1 \lambda_2} \right),$$

where $I_z(x)$ is the modified Bessel function of the first kind (Abramowitz & Stegun 1964, p. 1046), and the score difference z , has support over all the integers; see Figure 1 for examples of Skellam distributions. Note that if correlation between individual scores of teams are modelled by expressing the scores as $X = X_1 + X_0$ and $Y = X_2 + X_0$, where the X_i s are independent Poisson distributed random variables (see, for example, Koopman & Lit 2015), then the difference between X and Y will still follow a Skellam distribution. This property allows us to assume that factors affecting the score of each team equally, such as the weather and stadium conditions, could be appropriately modelled via X_0 and hence ignored when modelling score differences.

2.1. Dynamic elements

While the motivation for our model comes from the framework of Koopman & Lit (2015), one of the novel aspects of our model is that we propose to model a paired response, namely the goal difference and the behind difference from each match. While the distribution for the goal difference is governed by parameters λ_1 and λ_2 , another Skellam distribution with parameters λ_3 and λ_4 is used to model the behind difference.

However, it is desirable to have some dynamic structure on these parameters to allow them to vary over time, as well as change depending on the two opposing teams. Consider two teams i and j , who play each other at time point t , at team i 's home ground. We would like the goal scoring ability of each team to depend both on the 'goal-attack' strength of that team, as well as the 'defence' strength of the opposition. The same set up is imposed on each teams behind scoring ability. We would also like to include some form of home ground advantage. The 'goal-attack' strength of the home team i at time point t is denoted by $\alpha_{i,t}$, its 'behind-attack' strength by $\gamma_{i,t}$ and its 'defence' strength by $\beta_{i,t}$. The home ground advantage is denoted by δ , which is assumed to be constant for all teams, and constant over time. We relate this structure to $\lambda_1, \dots, \lambda_4$ with the log link function:

$$\begin{aligned}\lambda_{1,i,j,t} &= \exp(\delta + \alpha_{i,t} - \beta_{j,t}), & \lambda_{2,i,j,t} &= \exp(\alpha_{j,t} - \beta_{i,t}), \\ \lambda_{3,i,j,t} &= \exp(\delta + \gamma_{i,t} - \beta_{j,t}), & \lambda_{4,i,j,t} &= \exp(\gamma_{j,t} - \beta_{i,t}).\end{aligned}$$

Each team is assumed to have its own 'attack' strength and 'defence' strength, which represent team performance at each point in time. Note that we use the same δ in both distributions, which is a data motivated choice based on our data exploration that indicated on average the home team scored 0.603 and 0.621 more goals and behinds respectively, with standard errors of 0.288 and 0.184. We ruled out the use of team specific δ parameters, since such parameters could arguably vary from season to season, necessitating the introduction of a large number of additional parameters that would have to be excessively regularised to avoid overfitting. We use a common β for both the goal difference and behind difference, as we expect teams that concede fewer goals to concede fewer behinds. This also implicitly establishes a correlation between the number of goals and behinds, and also keeps the number of latent variables at a reasonable level. Furthermore, we are choosing to model goal difference and behind difference ostensibly as a method to reduce variability in our data compared to directly modelling the score difference. This also results in the effective doubling of the number of data inputs at each time point compared to just the score difference. In order to allow for variation in attack and defence strengths over time, autoregressive processes are

placed on both the ‘goal-attack’, ‘behind-attack’ and ‘defence’ strengths. Explicitly:

$$\begin{aligned}\alpha_{i,t} &= \mu_{\alpha,i} + \phi_{\alpha}\alpha_{i,t-1} + \varepsilon_{\alpha,i,t} \\ \gamma_{i,t} &= \mu_{\gamma,i} + \phi_{\gamma}\gamma_{i,t-1} + \varepsilon_{\gamma,i,t} \\ \beta_{i,t} &= \mu_{\beta,i} + \phi_{\beta}\beta_{i,t-1} + \varepsilon_{\beta,i,t},\end{aligned}$$

with $\mu_{\alpha,i}$, $\mu_{\gamma,i}$ and $\mu_{\beta,i}$ autoregressive means, ϕ_{α} , ϕ_{γ} and ϕ_{β} the autoregressive coefficients and $\varepsilon_{\alpha,i,t}$, $\varepsilon_{\gamma,i,t}$ and $\varepsilon_{\beta,i,t}$ the normally distributed error terms. The autoregressive coefficients are assumed to be constant for each team. This maintains comparability to Koopman & Lit (2015), and also reduces the number of parameters to ensure identifiability. We expect the week to week change amongst teams to be similar across teams. The independent error terms are assumed to come from a normal distribution:

$$\varepsilon_{j,i,t} \sim N(0, \sigma_k^2), \quad k = \alpha, \beta, \quad j = \alpha, \gamma, \beta,$$

and are assumed to be the same for all teams’ ‘attack’ and ‘defence’ processes. Note that this setup means that the errors of α and γ have the same $N(0, \sigma_a^2)$ distribution, and β has a separate $N(0, \sigma_{\beta}^2)$ distribution. This is again to reduce parameter count and is suitable for the assumption that all teams are equally variable from round to round. This relationship of $\alpha, \gamma \rightarrow a$ and $\beta \rightarrow \beta$ will continue throughout the rest of this paper.

2.2. Complete Bayesian specification

In order to complete the description of our model, we need to specify our priors. It should be noted that these are in no way objective, in fact they are considerably influenced by other implementations of similar models. We place uniform priors on the standard deviation parameter:

$$\sigma_k \sim U[0.005, 5] \quad k = \alpha, \beta \quad (1)$$

This approach seems reasonable compared to Baio & Blangiardo (2010), who use a Gamma(0.1,0.1) prior on the variance and Owen (2011), who uses a Gamma(2,250), or a Gamma(2,500) prior on the variance, as we expect significantly more variance in a higher scoring sport such as AFL. We place the following Beta priors on the ϕ parameters of the autoregressive process:

$$\phi_k \sim \text{Beta}(100, 20) \quad k = \alpha, \gamma, \beta$$

as we expect a reasonably large estimate for ϕ , but not as large as those found in soccer matches. This also serves to avoid any potential numerical issues that might be encountered from ϕ estimates that are close to zero. Koopman & Lit (2015) estimated values for ϕ_{α} and ϕ_{β}

of 0.9985 and 0.9992 in their English Premier League soccer data. In view of the considerably larger variability in AFL scores, we would expect lower estimates. Given the large number of data points, spread out across several distinct time series, this specification for ϕ is also not as informative as it may appear, as it can easily result in an estimate for ϕ in a region of low probability.

2.3. Identifiability constraints and consequences

The time series structure that is imposed on the strengths of each team introduced a large number of latent variables that can make fitting the model difficult and inconsistent. To ensure the fitting method converges to an appropriate model, and converges to this same model each time the fitting procedure is run, we implement some identifiability constraints.

For every time point in our data set, we have to consider three parameters unique to each team ($\alpha_{i,t}$, $\gamma_{i,t}$, $\beta_{i,t}$). Combining these with the three autoregressive means per team ($\mu_\alpha, \mu_\gamma, \mu_\beta$), the autoregressive variables ($\phi_\alpha, \phi_\gamma, \phi_\beta$), the home team team advantage (δ) and other ancillary latent variables results in nearly 3800 parameters and latent variables for our largest data set. Given we only add two observations per time point per match (goal difference and behind difference between teams i and j), we need some additional constraints in order to ensure the model is identifiable. It should be noted here that all values of $\alpha_{i,t}$, $\gamma_{i,t}$, $\beta_{i,t}$ for t other than the current time point are of no statistical interest, in terms of making inference about the outcome of matches at the next time point, and as such are latent variables.

As our interest is about predicting the difference between two teams, the only important quantity is the relative difference between autoregressive estimates, for example how much stronger is the Sydney Swans' 'goal-attack' strength than Carlton's. As such we can apply the sum-to-zero constraint on all the autoregressive parameters:

$$\sum_{i=1}^n \alpha_{i,t} = 0, \quad \sum_{i=1}^n \gamma_{i,t} = 0, \quad \sum_{i=1}^n \beta_{i,t} = 0, \quad (2)$$

for all t in $\{1, \dots, T\}$ where T is the total number of rounds in the data set and n is the number of teams in the AFL. See Section 2.5 for the specifics of these values. The Stan manual (Stan Development Team 2016, Chapter 6.6) discusses several possibilities on how to implement such sum-to-zero constraints, the most appealing of which in our context would be to define the n th parameter to be:

$$\alpha_{n,t} = -\sum_{i=1}^{n-1} \alpha_{i,t}, \quad \gamma_{n,t} = -\sum_{i=1}^{n-1} \gamma_{i,t}, \quad \beta_{n,t} = -\sum_{i=1}^{n-1} \beta_{i,t},$$

for a given time point t . However, this approach results in estimates for the n th parameters which were considerably higher or lower than any other parameter estimates at that given time point. A possible explanation is that with this approach the n th parameter estimate is forced to be equal to the negative sum of the other estimates. As discussed in Kruschke (2011, Chapter 18), this is fine if the prior on the parameters (which is only applied here to the first $n - 1$ parameters) does not depend on hyperpriors, as the hyperpriors should be informed by all n parameters, and not just $n - 1$ of them. However, in our time series specification the prior on the parameters does depend on hyperpriors. Further discussion of the issue of combining sum-to-zero constraints with a time series specification for parameters can be found in Knorr-Held (2000).

In order to correct this numerical anomaly, as well as increase computational efficiency, the following ideas from Knorr-Held (2000) and Owen (2011) are implemented. Consider just the goal attack strength parameters α_i for $i \in \{1, \dots, n\}$. The evolution of these parameters can be expressed in terms of a multivariate normal distribution:

$$\boldsymbol{\alpha}_t \sim N(\phi_{\alpha}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\mu}_{\alpha}, \mathbf{D}),$$

where $\mathbf{D}$ is a diagonal matrix with entries σ_{α}^2 . Stepping backwards through time until the beginning of the time series leads us to the initial values, $\boldsymbol{\alpha}_1 \sim N(\phi_{\alpha}\boldsymbol{\mu}_{\alpha}, \mathbf{D})$, where $\boldsymbol{\mu}_{\alpha}$ is the vector of autoregressive means for the $\boldsymbol{\alpha}$ process. As we require the sum-to-zero constraints on $\boldsymbol{\alpha}$ to hold at all points in time, including $t = 1$, we are implicitly requiring the elements $\boldsymbol{\mu}_{\alpha}$ to sum to zero. In addition to this constraint, Owen (2011) shows that in order for the original identifiability constraint (2) to hold, the covariance matrix $\mathbf{D}$ must be modified such that it has the following form:

$$\mathbf{R} = \frac{n\sigma^2}{(n-1)} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right),$$

where n is the number of teams, $\mathbf{I}_n$ is the $(n \times n)$ identity matrix and $\mathbf{1}_n$ the $(n \times 1)$ matrix of ones. This construction of $\mathbf{R}$ has its own associated issues, notably $\mathbf{R}$ is not of full rank and hence has no inverse. As such, Stan cannot directly sample from the multivariate normal distribution with $\mathbf{R}$ as its covariance structure. Sampling is instead performed from an $(n - 1)$ dimensional non-degenerate multivariate normal distribution. This distribution has an appropriate mean and covariance structure:

$$\mathbf{S} = \frac{n\sigma^2}{(n-1)} \left(\mathbf{I}_{n-1} + \mathbf{1}_{n-1} \mathbf{1}_{n-1}^T \right),$$

where we refer to variates of this distribution as the *raw* parameters. Transformation of these *raw* parameters into the parameters of interest is handled by the $n \times (n - 1)$ matrix:

$$\mathbf{J} = \begin{pmatrix} \mathbf{I}_{n-1} - \frac{1}{n} \mathbf{1}_{n-1} \mathbf{1}_{n-1}^\top \\ -\frac{1}{n} \mathbf{1}_{n-1}^\top \end{pmatrix},$$

whose structure implements the desired centring constraints. To illustrate this more explicitly, take the $(n - 1)$ dimensional *raw* means of all of the auto regressive processes:

$$\boldsymbol{\mu}_{\alpha,raw} \sim N(\mathbf{0}, \sigma_\mu^2 \mathbf{S}) \quad \boldsymbol{\mu}_{\gamma,raw} \sim N(\mathbf{0}, \sigma_\mu^2 \mathbf{S}) \quad \boldsymbol{\mu}_{\beta,raw} \sim N(\mathbf{0}, \sigma_\mu^2 \mathbf{S}).$$

These means are then transformed into the parameters used for inference by $\mathbf{J}$:

$$\boldsymbol{\mu}_\alpha = \mathbf{J} \boldsymbol{\mu}_{\alpha,raw} \quad \boldsymbol{\mu}_\gamma = \mathbf{J} \boldsymbol{\mu}_{\gamma,raw} \quad \boldsymbol{\mu}_\beta = \mathbf{J} \boldsymbol{\mu}_{\beta,raw}.$$

In a similar manner, the initial values for the team strength parameters come from appropriate zero mean distributions:

$$\boldsymbol{\alpha}_{t,raw} \sim N(0, \sigma_a^2 \mathbf{S}) \quad \boldsymbol{\gamma}_{t,raw} \sim N(0, \sigma_a^2 \mathbf{S}) \quad \boldsymbol{\beta}_{t,raw} \sim N(0, \sigma_\beta^2 \mathbf{S}),$$

which are then used to find the initial values:

$$\boldsymbol{\alpha}_1 = \boldsymbol{\mu}_\alpha + \frac{\sigma_a^2}{\sqrt{1 - \phi_\alpha^2}} \boldsymbol{\alpha}_{1,raw} \quad \boldsymbol{\gamma}_1 = \boldsymbol{\mu}_\gamma + \frac{\sigma_a^2}{\sqrt{1 - \phi_\gamma^2}} \boldsymbol{\gamma}_{1,raw} \quad \boldsymbol{\beta}_1 = \boldsymbol{\mu}_\beta + \frac{\sigma_\beta^2}{\sqrt{1 - \phi_\beta^2}} \boldsymbol{\beta}_{1,raw}.$$

Finally, these are mapped forward in time to create the ‘attack’ and ‘defence’ strengths:

$$\begin{aligned} \boldsymbol{\alpha}_{t+1} &= \phi_\alpha \boldsymbol{\alpha}_t + \boldsymbol{\mu}_\alpha + \sigma_a^2 \mathbf{J} \boldsymbol{\alpha}_{t+1,raw} \\ \boldsymbol{\gamma}_{t+1} &= \phi_\gamma \boldsymbol{\gamma}_t + \boldsymbol{\mu}_\gamma + \sigma_a^2 \mathbf{J} \boldsymbol{\gamma}_{t+1,raw} \\ \boldsymbol{\beta}_{t+1} &= \phi_\beta \boldsymbol{\beta}_t + \boldsymbol{\mu}_\beta + \sigma_\beta^2 \mathbf{J} \boldsymbol{\beta}_{t+1,raw} \end{aligned}$$

To complete this hierarchical specification, σ_μ requires some form of prior distribution, which is again chosen to be uniform:

$$\sigma_\mu \sim U(0.005, 1).$$

110 The slightly smaller bounds on this variance parameter compared to σ_a and σ_β in (1) are due
 111 to decreased variability in the value of $\boldsymbol{\mu}$. This reflects the fact that a single unexpected result
 112 only slightly changes the mean of the underlying autoregressive process. Other Bayesian

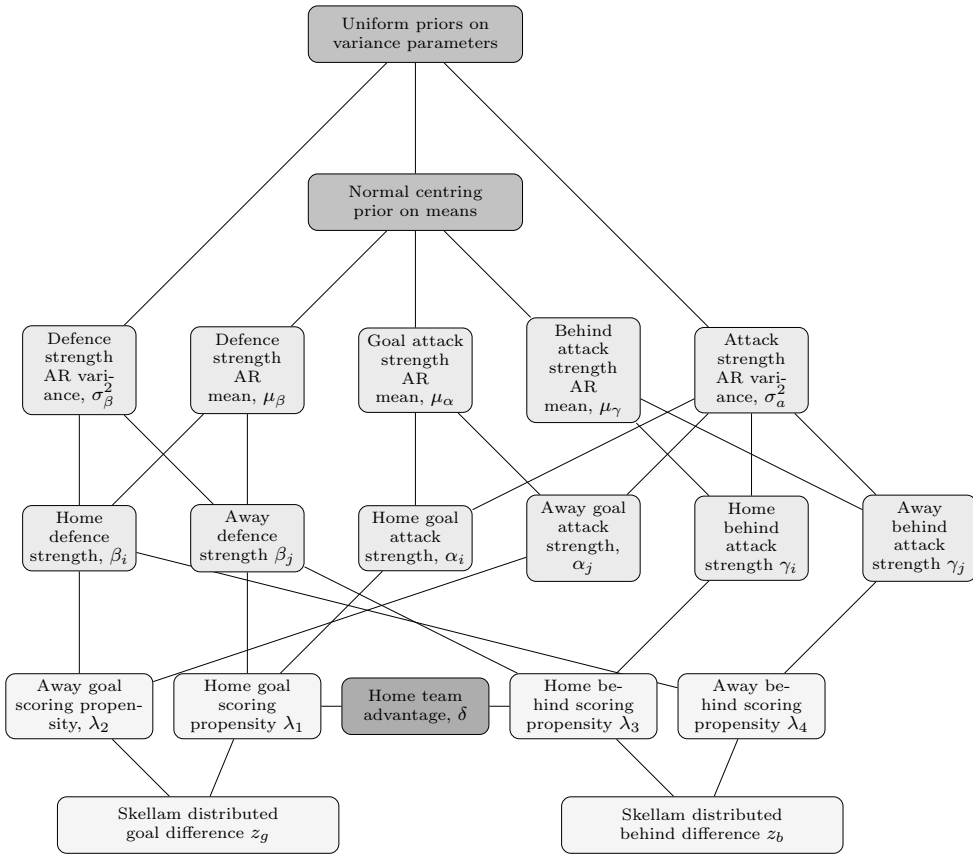


Figure 2. A graphical depiction of the structure of the model described in Sections 2.1 to 2.3

attempts do not allow this parameter to vary, and simply fix it at some value (see, for example, Baio & Blangiardo 2010).

A graphical representation of this model is presented in Figure 2, which is arranged in the following manner: the nodes in the first (bottom) layer represent the data generating distributions, with the nodes in the second layer representing the parameters that govern said distributions, except for the darker grey node in the second layer, which represents the home ground advantage. The nodes in the third and fourth layers pertain to the time series structure, and lastly the nodes in the fifth and sixth layers represent the prior distributions required to complete the specification of the model.

2.4. Start-of-season variance

As teams draft, trade, and train players during the off season, their performance can change substantially from the end of one season to the start of the next. As our time variable t

only indexes the occurrence of matches, it does not include detail about the start and end of seasons. For example, $t = 22$, $t = 23$ and $t = 24$ are all one time lag apart, yet the transition in real time from $t = 22$ to $t = 23$ represents one week, but $t = 23$ to $t = 24$ represents approximately 5 months due to the off season break (the regular AFL season contains 23 rounds; the same is true for the set of time points at the end of each season in our data set, i.e. $t = 45, 46, 47, \dots$).

The change in a team's performance during the off season can be difficult to model, as season breaks are fairly lengthy periods of time and a team's off season strategy may change considerably each year. Some years they may focus on training rookies for future season endeavours, other years they may focus on conditioning so that senior players are in peak condition near the end of the season. As such, there are a small number of data points associated with inter-seasonal performance change, which have large variability. In the model implementation, start-of-season variance is handled by specifying the time points corresponding to matches occurring after season breaks, and inflating the variance of the time series estimates at these points. Subsequent rounds have a similar, albeit reduced inflation factor based on the assumption that the first round contains the most information about a team's change in performance during the off season. As such we use an exponentially decaying inflation factor of $(v - 1) \exp\{-(t - t_{ss})\} + 1$ such that:

$$\alpha_{i,t} \sim N\left(\mu_i + \phi_a \alpha_{i,t-1}, [(v - 1) \exp\{-(t - t_{ss})\} + 1] \sigma_a^2\right),$$

where t_{ss} is the time point that corresponds to the start of the season and v is a chosen variance inflation factor.

2.5. Miscellaneous details

The AFL has 18 teams (i.e. $n = 18$), who each play 22 matches across 23 rounds in a regular home and away season, for a total of 198 matches. This means teams do not play exactly two games against every other team, rather they play once against all teams and twice against a selection of the other teams. For each of these games each team receives four competition points for a win, two for a draw and zero for a loss. The season is then concluded with a four week finals series which reduces the top eight finishing teams to two teams to play in the grand final. However, we do not model this part of the season for two reasons: First, match tactics in finals are markedly different to those employed during the standard home and away season. Second, as only a limited number of teams make the finals series, we would have insufficient data to estimate the attack and defence strengths of each team.

3. Results

3.1. Model fitting, estimation of parameters and prediction generation

The model described in Section 2 was coded in the Stan modelling language and fitted using the `rstan` interface. Specifically, we concern ourselves with the estimation of binary win or loss results, as well as win, draw, and loss (WDL) probabilities for the 2015 AFL home and away season. Historical and current results for the AFL were acquired from www.footywire.com. Home and away data from 2013 to 2015 were used to fit the model and predict results. Each round took approximately 6 to 8 hours on a standard university workstation to fit with acceptable convergence, which enables the use of this model in a round to round forecasting scenario.

Traceplots, density estimates and summary statistics for convergence of MCMC chains were assessed with the most important parameters (or latent variables) for stable model behaviour being the attack ($\alpha_{i,t}, \gamma_{i,t}$) and defence ($\beta_{i,t}$) strengths over time for each team. Figure 3 uses $\alpha_{1,36}$ and $\alpha_{1,37}$ to show typical examples of traceplots of any of these parameters and demonstrates no undesirable behaviour. Note that this traceplot was generated with one chain of 5000 iterations, whilst our reported predictions were generated using 4 chains of 10000 iterations, all of which converged and mixed appropriately.

3.2. Predictive results of the 2015 AFL season and model validation

Predictions for match outcomes were obtained in a one-step-ahead manner, whereby all results from previous rounds were used to predict the current round matches. Their associated probabilities and estimated differences were generated by sampling from the posterior predictive distribution of the parameters. For practical purposes these predictions included a predicted point difference, as well as associated WDL probabilities. We also construct prediction intervals for the score differences in each match and check whether each actual difference was contained within the corresponding interval.

Figure 4 describes the performance with respect to win/loss predictions across the 2015 season, with the dashed line indicating the proportion of predictions that were correct in each round, whilst the solid line is the mean of the performance up to each time point. The total number of correctly predicted matches was 132, which corresponds to 2/3 of the total number of games in the season. This is a reasonably good performance, as top football tipsters with many years of experience rarely exceed 145 correct predictions for the home and away season. There appears to be a slight drop in predictive performance towards the end of the season, which is more difficult to explain but could be due to the complacency of top teams,

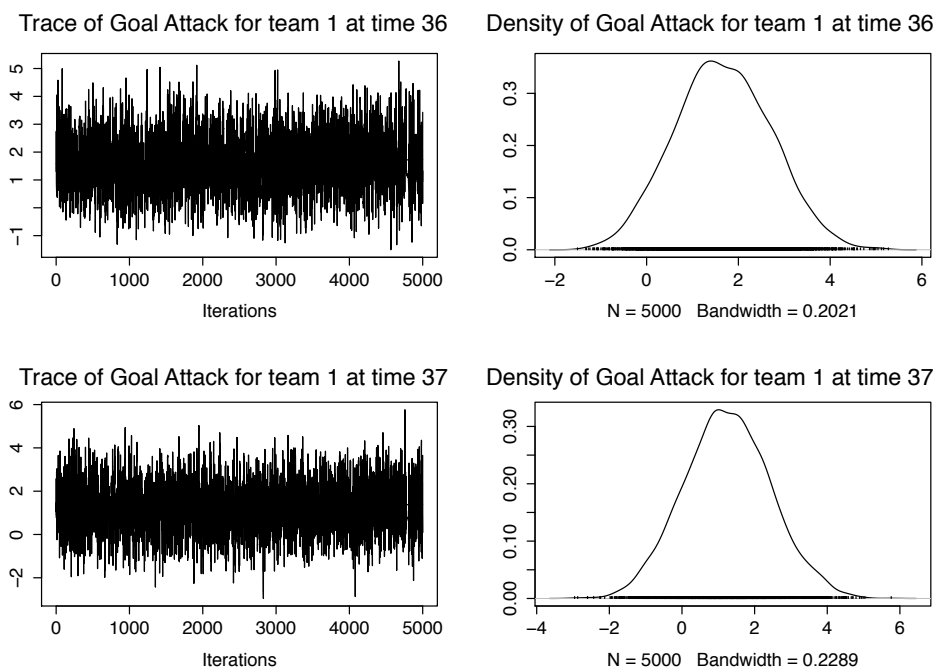


Figure 3. A traceplot and density estimate of the Goal Attack parameter for team 1 and time point 36 and 37. This particular plot was generated with 1 chain of 5000 iterations, which is less than what was used to generate predictions.

or disinterest of the lower ranked teams. However, it is equally possible that fatigue or other unknown factors manifested themselves as random variation.

Whilst these binary win or loss outcomes are definitely of practical interest, they are not necessarily the best means for validating the model. Take for example, a model prediction of a points difference of +1, with an estimated standard deviation of 5, and an actual result of -2. Such a prediction would be considered reasonable, however the dichotomisation into a win and loss would lead to the prediction being classified as incorrect. In order to explore the model's outputs and predictions, we construct a prediction interval using the estimated difference and its standard deviation. As we have a large number of samples, and the estimated distributions from the traceplots of the goal and behind differences appear relatively symmetric, we construct an approximate normal interval.

Figure 5 shows the estimated and actual coverage probability of the $100\left(1 - \frac{\alpha}{2}\right)\%$ prediction intervals as a function of α . For all values of α it depicts a reasonably well behaved model, with actual coverage very closely following nominal coverage. As such, there is some implicit degree of confidence in the estimated differences, standard deviations, and WDL probabilities.

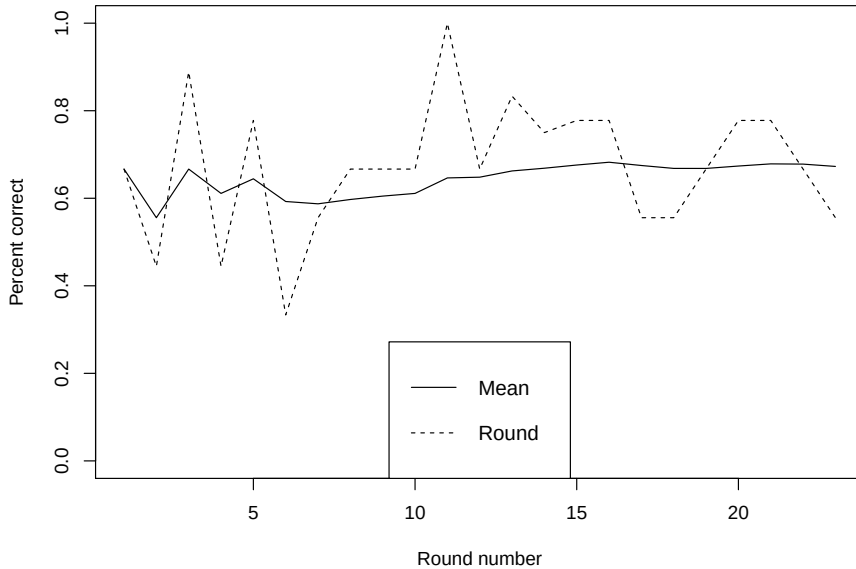


Figure 4. Predictive performance of the model over the 2015 home and away season. The dashed line represents the proportion of games the model would have predicted correctly in each round. The solid line is the average proportion of games predicted correctly over the season up to and including that point (i.e. the point at round two is the mean proportion of games predicted correctly in rounds one and two.).

We also validate the predictive capabilities of the model, particularly the estimated WDL probabilities, by estimating the cumulative number of competition points accrued at each point in time. Multiplying the possible number of competition points that can be earned in a game (namely four for a win, two for a draw and zero for a loss) by our estimated WDL probabilities yields the expected number of competition points in each match. Hence, we obtain a predicted point trajectory for each team over the whole season which is compared to the actual number of competition points received in Figure 6. We note that the observed and predicted trajectories shown are relatively close for most teams. The intra-seasonal changes in performance appear to be well captured by our model; for example, the increase in North Melbourne's performance towards the end of the season. However, as one would expect from hierarchical models, it is also apparent that there is some evidence of shrinkage, with underperforming teams such as Brisbane, Carlton and the Gold Coast Suns having win rates over-predicted by the model and vice versa for the higher performing teams. This is somewhat less true for Hawthorn: as the data used to fit the model encompasses years in which Hawthorn has won the grand final, the output seems very reasonable.

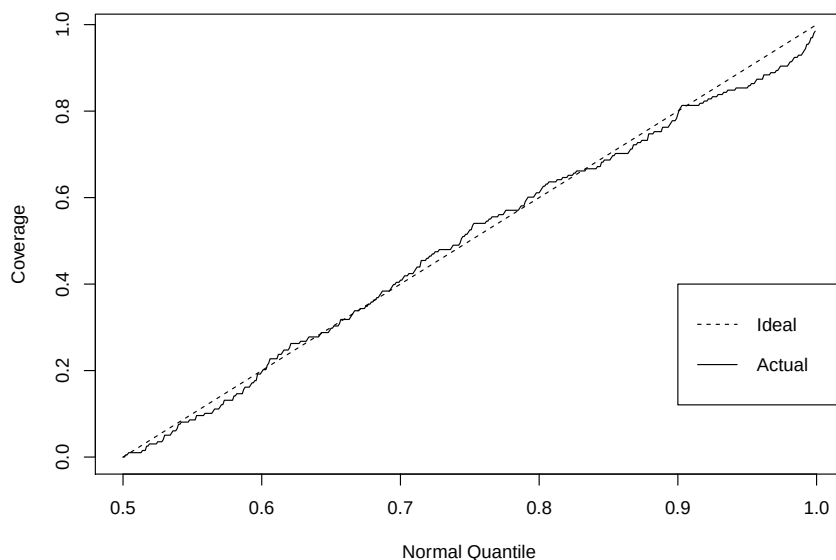


Figure 5. The coverage proportion as a function of the quantile used to construct the prediction interval. The dotted line represents the ideal coverage, whilst the solid line represents the actual coverage.

208 The difficulty in predicting outcomes at the start of the season is apparent, with teams
 209 like Hawthorn and the Western Bulldogs having their predictions and actual curves exhibiting
 210 near parallel behaviour after the initial few weeks of the competition. The major difference
 211 between the two curves is due to teams unexpectedly winning or losing games at the start
 212 of the season where prediction is difficult due to the inter-seasonal changes in each teams'
 213 ability.

4. Conclusions

Currently there is a lack of published methodology for outcome prediction in the AFL, with Clarke (1993) and Ryall & Bedford (2010) appearing to be alone in attempting to forecast the results of AFL matches. We have provided an approach to modelling AFL results using dynamic hierarchical Bayesian models which, to the best of our knowledge of the sports modelling literature, have only been implemented for soccer data (see for example Owen 2011; Baio & Blangiardo 2010). Our models were based on the Skellam distribution and in contrast to previous work, using similar methodology for sports modelling, we have modelled the difference in the two scoring mechanisms simultaneously, in an attempt to encapsulate the highly variable nature of AFL match results. We considered issues in constraining parameters and model identifiability and applied appropriate methods to rectify such problems. Our final model is implemented in the form of a Stan model, which enables the generation of predictive results and we have subsequently applied this to data from the 2015 AFL season, with comparable predictive performance to similar models used for soccer outcome modelling. All code and data used to fit these models is freely available on <https://github.com/hhau/DynBayes>.

Future extensions of modelling AFL match outcomes would benefit from focussing on the problems encountered due to the highly variable nature of AFL scores. Solutions may include: using the difference of negative binomial random variables as a potential alternative to the Skellam distribution, or binning the scores appropriately instead of using exact differences and thus extending our methods to ordinal regression. Furthermore, identifying the best and worst performing teams to provide an extra adjustment to their time-series means could be used to correct for shrinkage issues encountered using current techniques.

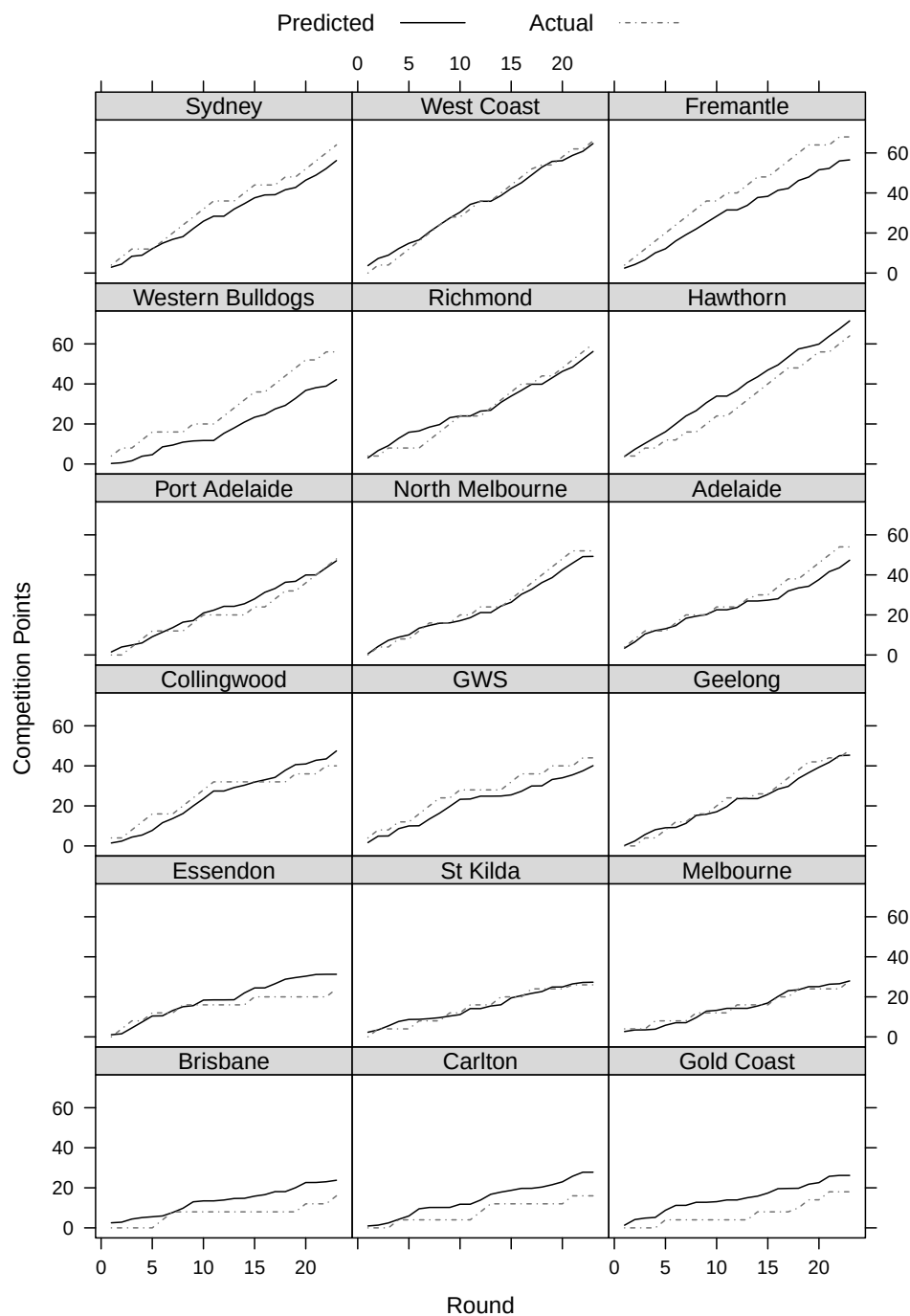


Figure 6. Predicted competition point total (solid) and actual competition point total (dotted) over the 2015 AFL home and away season. The individual plots are sorted in order of total actual competition points at the end of the season, with the plot being read from right to left, top to bottom.

References

- ABRAMOWITZ, M. & STEGUN, I.A. (eds.) (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. NBS Applied Mathematics Series 55, National Bureau of Standards, Washington, D.C.
- BAIO, G. & BLANGIARDO, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics* **37**, 253–264.
- BRILLINGER, D.R. (2007). Modelling some Norwegian soccer data. In *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell A Doksum*, ed. V. Nair, *Series in Biostatistics*, vol. 3. Singapore: World Scientific, pp. 3–20.
- BRILLINGER, D.R. (2008). Modelling game outcomes of Brazilian 2006 Series A championship as ordinal-valued. *Brazilian Journal of Probability and Statistics* **22**, 89–104.
- CLARKE, S.R. (1993). Computer forecasting of Australian Rules Football for a daily newspaper. *The Journal of the Operational Research Society* **44**, 753–759.
- ELO, A.E. (1978). *The rating of chessplayers, past and present*. Arco Pub, New York.
- 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**, 381–393.
- KNORR-HELD, L. (2000). Dynamic rating of sports teams. *Journal of the Royal Statistical Society: Series D (The Statistician)* **49**, 261–276.
- 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**, 167–186.
- KRUSCHKE, J.K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Burlington, MA: Academic Press.
- MAHER, M.J. (1982). Modelling association football scores. *Statistica Neerlandica* **36**, 109–118.
- OWEN, A. (2011). Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics* **22**, 99–113.
- RYALL, R. & BEDFORD, A. (2010). An optimized ratings-based model for forecasting Australian Rules football. *International Journal of Forecasting* **26**, 511 – 517.
- SKELLAM, J.G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society* **109**, 296.
- STAN DEVELOPMENT TEAM (2016). *Stan Modeling Language User's Guide and Reference Manual, Version 2.15.0*. URL <http://mc-stan.org>.