Matthew Heiner*, Gilbert W. Fellingham and Camille Thomas

Skill importance in women's soccer

Abstract: Soccer analytics often follow one of two approaches: 1) regression models on number of shots taken or goals scored to predict match winners, or 2) spatial and/or temporal analysis of plays for evaluation of strategy. We propose a new model to evaluate skill importance in soccer. Play by play data were collected on 22 NCAA Division I Women's Soccer matches with a new skill notation system. Using a Bayesian approach, we model play sequences as discrete absorbing Markov chains. Using posterior distributions, we estimate the probability of 35 distinct offensive skills leading to a shot during a single possession.

Keywords: absorbing state; Bayesian methods; Markov chain.

***Corresponding author: Matthew Heiner,** Statistics, Brigham Young University, 223 TMCB Provo, UT 84602, USA, Tel.: +8012056687, e-mail: mheiner@gmail.com
Gilbert W. Fellingham: Statistics, Brigham Young University, Provo, UT, USA
Camille Thomas: Physical Education and Human Performance, Southern Utah University, Cedar City, UT, USA

1 Introduction

According to the Fédération Internationale de Football Association (FIFA 2006), soccer boasts more than 240 million regular players. Professional matches routinely attract thousands of spectators and in 2006, the final match of the FIFA World Cup drew a cumulative television audience of 715.1 million viewers (FIFA 2007). FIFA (FIFA 2006) further estimates that there are 20 million women players worldwide.

Quantitative analysis of soccer has become increasingly popular in the past few decades. While academic research makes a modest contribution, most soccer analytics are from private enthusiasts and enterprises. Businesses and blogs such as *statDNA.com*, *soccermetrics.net*, *socceranalysts.com*, and *onfooty.com* are well established and produce a high volume of analyses.

Motivations for soccer analytics include developing strategy, predicting match and tournament outcomes, building rosters, improving players, or gaining insight into the dynamics of the game. Many analyses focus on strategy and outcome prediction. Although interesting, prediction is less helpful to coaches who are responsible for the details which lead to the final score.

In this paper we build a statistical model with data gathered via notational analysis of soccer in order to identify which offensive skills contribute most to creating scoring opportunities. These methods can be applied to compare skill importance across strata such as gender or league, or to help individual teams identify important skills for their style of play. This information can assist coaches and managers with decisions on use of practice time, game strategy, and player recruitment.

In our analysis, we model skill sequences from Division I NCAA Women's Soccer matches using absorbing Markov chains. Estimation and inference are performed using the Bayesian paradigm, which yields rich inference on updated parameter densities, conditional on the data. With posterior inference on the probability of 35 different skills leading to a shot, we gain insight into the importance of various skills in collegiate women's soccer.

In Section 2 we give a brief review of some literature in soccer analytics and highlight a few papers which employ methods similar to ours. In Section 3 we specify the methodology used to fit and assess our model. In Section 4 we present the results of the analysis and in Section 5 we provide further discussion and conclusions.

2 Literature review

2.1 Research in soccer analytics

A classic and early quantitative treatment of soccer was given by Reep and Benjamin (1968). They considered a model for sequences of successful passes and demonstrated several regularities in soccer such as ratios of goals from certain field and tactical positions.

A common theme in soccer analytics has been modeling match scores. Maher (1992) investigated the Poisson model for match scores and showed that a correlated bivariate model was reasonable. Dixon and Coles (1997) extended Maher's model and demonstrated that their proposed model could inform a profitable betting strategy.

Hamilton (2011) extended the Pythagorean expectation formula developed originally for baseball by James (1980) to model points per match, allowing for ties.

Various regression methods have been proposed for modeling shot, goal, and game outcomes. Karlis and Ntzoufras (2003) demonstrated bivariate Poisson regression for match scores and explored issues such as correlation and overdispersion. Pollard and Reep (1997) developed a system for discriminating the quality of shot opportunities and modeled the probability of scoring a goal less the probability of conceding a goal. Rue and Salvesen (2000) proposed a Bayesian dynamic generalized linear model to predict game outcomes in the English Premier League and to develop betting strategies. Goddard (2005) used bivariate Poisson and ordered probit regression models to forecast goals and match outcomes respectively.

Several analyses have investigated which skills lead to positive outcomes. McHale and Scarf (2007) fit bivariate discrete distributions to model shots-for and shots-against in English Premier League matches. Using counts of plays involving various skills as covariates, their model for shots indicated that more passes and crosses result in more shot opportunities. Allan (2009) performed a skill importance analysis with a logistic regression model in which estimated coefficients were standardized and reported as importance scores (Fellingham and Reese 2004). Different skills were judged to be most important to different teams. Thomas, Fellingham, and Vehrs (2009) developed a notational system for women's soccer to record and rate discrete actions. Using this system, they rated games for a women's collegiate team and performed a regression analysis which identified dribbling as the most important skill for creating scoring opportunities.

Other analyses focus on details of play. For example, Hughes and Franks (2005) investigated lengths of passing sequences for successful and unsuccessful teams. Brillinger (2007) focused on a single play. He developed and fit a "potential function" from a graphical representation of ball trajectory in a series of passes.

2.2 Markov chain methods applied in sports analytics

Markov process modeling has been applied in sports analytics, including soccer. We present a sample of recent contributions, applied to soccer or using methods similar to ours.

Hirotsu and Wright (2003) incorporated a continuous-time Markov process relating to possession and scoring in soccer and used this model to evaluate offensive and defensive strength of teams in the English Premier League.

Miskin, Fellingham, and Florence (2010) performed a skill analysis of collegiate women's volleyball using absorbing Markov chains to model play sequences. Absorption probabilities for the outcomes were estimated and used to rank skills by importance, allowing them to make strategic recommendations. This paper represents an extension of the work done by Miskin et al. in women's volleyball to women's soccer.

Rudd (2011) analyzed player performance in the English Premier League using absorbing Markov chains. She estimated transition probabilities between spatial and set piece states and rated players by the change in probability of scoring a goal affected by the player's move.

Goldner (2012) proposed an absorbing Markov chain model for drives in the National Football League by relating down states to outcome states such as touchdown. He used the model to calculate expected points from each state, which is useful for efficiency evaluation and strategy development.

3 Methods

3.1 Data

Data were collected on 22 Division I NCAA Women's Soccer matches played among 24 teams in 2006 and 2009. Each match was filmed and notated according to a modification of the system presented in Thomas et al. (2009).[1] We have a discrete chronological sequence of moves taken by the team in possession of the ball. At each move, the following were recorded:

1. *The offensive skill performed on the move.*
2. *A rating based on the outcome of the skill performed.*
3. *Indicators further discriminating pass types.*
4. *Whether the home or visiting team performed the move.*
5. *Whether the move occurred at the end of a half or end of the game.*
6. *Additional notes such as rule violations or indication if the ball went into the mixer (the area encompassed by the penalty box, 18 yards from the end line and 44 yards wide).*

The five notated skills are shot, pass, first-touch pass, dribbling touch, and controlling touch. Details on skill-rating combinations are found in Table 1. Additional information, like team indicator and game, are excluded from the model for simplicity. The data appear as in Table 2 which contains five rows representing two sample sequences from the data:

1 The original system is given in Table A.1 in the appendix.

Table 1 Notation system used for identifying and rating skills performed during soccer matches.

Shot	
Rating	
3	A goal is scored on the shot
2	The shot is on target but saved by the defending team
1	The shot is off target but within six yards of the goal
0	The shot is outside six yards of the goal frame
Pass	The ball is kicked to a teammate
First-touch Pass	A kick, head, thigh, etc. is used to pass the ball on the first touch
Rating	
8	The pass results with the ball in the "mixer"
7	The pass results in a forward (penetrating) play
6	The pass is lateral (perpendicular to the sidelines)
5	The pass is backward (toward own goal)
4	The player passing the ball is fouled
3	The pass is deflected out of bounds by the opposing team
2	The pass results in a 50/50 ball (either team could win possession) and the passing team retains possession
1	The pass results in a 50/50 ball and the opposing team wins possession
0	The ball is passed out of bounds or the opposing team intercepts
Dribbling Touch	Player uses feet, thigh, and/or head to advance the ball after the first controlling touch
Controlling Touch	
Rating	
5	The touch results in a forward play
4	The touch results in a lateral play
3	The touch results in a backward play
2	The player is fouled
1	The touch results in a deflection out of bounds by the opposing team
0	The touch results in loss of possession

Table 2 Five rows from the original data. The first two lines and the last three lines represent distinct play sequences.

	Skill	Rating	Team	Notes	Game
1	PASS	6	HOME	Kickoff	22
2	1-T PASS	0	HOME		22
3	TRAP	5	AWAY		22
4	DRIB	5	AWAY		22
5	DRIB	0	AWAY		22

It is important to acknowledge that the skill ratings are judged on the outcomes of the moves and not by the expertise with which they are performed. Although likely correlated, execution and outcome are distinct. Note also that the majority of the moves are not spatially referenced. Consequently in our model, it is often the case that a move performed in the back half of the pitch is undistinguished from the same move performed in the front half.

3.2 Model specification

Modeling these data as a Markov process is intuitively appealing for two reasons. First, the data are moves that are sequential in time. Second, it is reasonable to assume dependence between consecutive moves.

Our proposed model is a discrete, time-homogeneous, first order Markov chain. It is discrete because there are a finite number of skills to be used as states in the chain. Time-homogeneity is a simplifying assumption allowing estimation of only one transition probability between each pair of states. The chain is assumed to be of first order for two reasons. The first is because, as noted earlier, a move in soccer directly affects the move immediately following. It is less clear whether there is direct impact two or more moves later. The second reason is that estimating a second order chain would dramatically increase the number of parameters in the model.

Each move is identified by a skill and rating combination from Table 1. We use this as the basis for defining states in the Markov chain. We also introduce the following new states which consolidate information:

1. *Turnover* (*TOV*): inserted at each change in team possession (indicated by an offensive move performed by the opposing team) during the regular flow of play.
2. *Bad Turnover* (*BTOV*): defined as a turnover after which the opposing team's first move is a shot, any pass rated 8, or touch rated 5.

3. *Deflected OB*: inserted after any pass rated 3 or touch rated 1, which occur when the opposing team deflects the ball out of bounds and play is restarted with a throw in, corner kick, or goal kick.
4. *Fouled*: inserted after any pass rated 4 or touch rated 2, which occur with fouls and are always followed by a set piece (free kick).

Six states are modeled as absorbing states: *Shot 0–3*, *Turnover*, and *Bad Turnover*. A state is called an *absorbing state* if it transitions to itself with probability one (Lawler 1995: 10, 23–24). All others are *transient* states. We do not include end of half or game as states in the model because they are not states that may occur on any given play. Therefore, each sequence in the data ends in an absorbing state except the ends of playing time. Note that our strict definition of turnover precludes analysis of strategies in which a team advances the ball as far as possible with the hope of regaining possession downfield.

Pass moves which are identified as free kicks (used to restart play), corner kicks (taken by the attacking team near their opponent's goal), goal kicks (taken by the defending team near their own goal), goalie punts, goalie throws, throw-ins, and kick-offs are also considered separate states since they are distinct forms of passing. Their rating scale is the same as for regular passing.

Some skill-rating combinations are extremely rare and/or never occurred in the 22 recorded games. In order to maintain simplicity of this analysis, we consolidated several states. Table 3 gives a list of these states including the states to which they were consolidated and justification for consolidating. In addition, some pass-type skills rated 3 and 4 (deflected or fouled) were not included as states because they are impossible or never occurred (e.g., *Goalie Throw 4*). *Kick-off 0* was also removed because it is never preceded in a sequence and only leads to absorbing states.

As noted in Table 2, all pass type moves 1 and 2 (50/50 passes which could be recovered by either team) were consolidated in order to maintain that these are the same move with different outcomes. By combining them, we maintain the stochastic nature of the move and capture the risk inherent in 50/50 passes.

Most states represent successful completion of a skill. However, because of the potential to score from close proximity to the goal, we are interested not only in successful passes into the mixer, but also marginal ones. Coaches and players may consider advancing the ball into the mixer to create a scoring opportunity even though a player may not be in good position to receive such a pass. Accordingly, skills *Pass 1.5, First-touch Pass 1.5, Free Kick 1.5*, and *Throw-in 1.5* were added. They represent pass skills rated 1 or 2 into the mixer.

We are now prepared to fully specify the model with its states. A complete list of the 57 transient states and their possible transitions in the model is given in Table A.2 in the appendix. If we add our six absorbing states, the resulting transition probability matrix **P** has 63 rows and 63 columns. We can partition **P** as

$$\mathbf{P}=\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{B} & \mathbf{Q} \end{pmatrix} \tag{1}$$

where **I** is a 6×6 identity matrix containing transition probabilities for the absorbing states, **B** (57×6) contains transition probabilities from the transient states to the absorbing states, and **Q** (57×57) is a substochastic matrix containing transition probabilities among the transient states (Lawler 1995: 26).

The parameters in our model are the transition probabilities in **B** and **Q** which we denote $p(s_j, s_k)$ where s_j is the "from" state and s_k is the "to" state. For example, *p*(*Pass*

Table 3 States which were consolidated for this analysis. Some states were rare or very similar to other states. Other states were consolidated to maintain the stochastic properties of a single move which had been separated by the notation system.

States	Consolidated to	Reason
Free Kick 5–6	Free Kick 6	Backward free kicks were rare
Corner Kick 5–7	Corner Kick	All were relatively rare, but distinct from Corner Kick 8
Goal Kicks 1–2, 5–8; Goalie Punts 1–2, 5–8	Goal Kick	Many of these states had little data. Goal kicks are almost always performed by the goal keeper and are similar to punts
Goalie Punt 3	Goal Kick 3	Done for consistency with other goal kicks and goalie punts
Goal Kick 0, Goalie Punt 0	Goal Kick 0	Done for consistency with other goal kicks and goalie punts
Goalie Throw 1–2, 6–7	Goalie Throw	The transition distributions in the data were very similar and Goalie Throws 1 and 2 were very rare
Throw-in 7–8	Throw-in 7	Throw-in 8 was very rare
Kick-off 5–6	Kick-off 6	Kick-off 5 occurred only once
All other pass type moves 1–2	Pass move 1	They are the same move

7, *Shot* 3), the probability of transitioning from a forward pass to a goal-scoring shot, is a parameter we wish to estimate. Probabilities for transitions not listed in Table A.2 (e.g., p(*Pass* 7, *Turnover*)) are fixed at zero and probabilities for transitions which must occur (e.g., p(*Pass* 4, *Fouled*)) are fixed at one.

Defining our model with absorbing and transient states is useful because we can compute marginal probabilities of absorption. That is, we are interested in

$$\lim_{n\to\infty}\mathbf{P}^n=\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{A} & \mathbf{0}\end{pmatrix} \tag{2}$$

where **A** contains the marginal absorption probabilities from the transient to absorbing states. Lawler (1995: 26) gives an elegant solution,

$$\mathbf{A}=(\mathbf{I}-\mathbf{Q})^{-1}\mathbf{B} \tag{3}$$

which always exists for chains of the form given in (1). Another useful result comes directly from (3). Each row of $(\mathbf{I}-\mathbf{Q})^{-1}$ corresponds with a transient state and contains the expected number of visits to each of the other transient states (starting in the state represented by the row) before entering one of the absorbing states. Thus, summing the rows of $(\mathbf{I}-\mathbf{Q})^{-1}$ yields the expected number of transitions to absorption from each transient state (Lawler 1995: 25).

3.3 Estimation

If we let **P** denote the transition probability matrix and $p(s_j, s_k)$ denote the transition probability from state s_j to state s_k, the full sampling distribution for data under the discrete, time-homogenous, finite Markov chain model with states $S=\{s_1, s_2, \ldots, s_m\}$ can be written as

$$\begin{aligned} f(\mathscr{Y}|\mathbf{P})&=\prod_{j,k}p(s_j, s_k)^{n(s_j,s_k)}\\ &\propto\prod_j\left[\frac{n(s_j)!}{\prod_k n(s_j, s_k)!}\prod_k p(s_j, s_k)^{n(s_j,s_k)}\right]\end{aligned} \tag{4}$$

such that $0<p(s_j, s_k)<1$ for all j, k pairs we wish to estimate and $\Sigma_k p(s_j, s_k)=1$ (Anderson and Goodman 1957). Note that $n(s_j, s_k)$ is the number of transitions from s_j to s_k and $n(s_j)=\Sigma_k n(s_j, s_k)$. The $n(s_j, s_k)$ are computed from the data $\mathscr{Y}$, which may consist of several observed independent chains. Anderson and Goodman (1957) observe that the $n(s_j, s_k)$ constitute a set of sufficient statistics for **P**. They further observe that (4) is equivalent to the product of m independent multinomial probability mass functions, each represented by a row of **P**. The second line of (4) demonstrates this relationship.

To specify a prior on the transition probabilities, we follow Strelioff, Crutchfield, and Hubler (2007), who approach inference for Markov chain models including parameter estimation and model comparison in a Bayesian framework. They describe the product of independent Dirichlet distributions as the conjugate prior to (4). The Dirichlet density as a prior for row j of **P** corresponding to state s_j is given by

$$\begin{aligned}\pi(p(s_j,\cdot)|\alpha(s_j,\cdot))&=\frac{\Gamma(\Sigma_k\alpha(s_j,s_k))}{\prod_k\Gamma(\alpha(s_j,s_k))}\prod_k p(s_j,s_k)^{\alpha(s_j,s_k)-1}\\ &\propto\prod_k p(s_j,s_k)^{\alpha(s_j,s_k)-1}\end{aligned} \tag{5}$$

where $p(s_j, \cdot)$ and $\alpha(s_j, \cdot)$ represent vectors across values of k, each $\alpha(s_j, s_k)>0$, and $\Sigma_k p(s_j, s_k)=1$. As a demonstration, consider inference for row j of a transition probability matrix, which we can consider to be a multinomial probability vector. Placing a Dirichlet prior with hyperparameters $\alpha(s_j, s_k)$ on this vector, we obtain the following posterior density

$$\begin{aligned}\pi(p(s_j,\cdot)|\mathscr{Y},\alpha(s_j,\cdot))&\propto f(\mathscr{Y}|p(s_j,\cdot))\times\pi(p(s_j,\cdot)|\alpha(s_j,\cdot))\\ &\propto\prod_k p(s_j,s_k)^{n(s_j,s_k)}\times\prod_k p(s_j,s_k)^{\alpha(s_j,s_k)-1}\\ &=\prod_k p(s_j,s_k)^{n(s_j,s_k)+\alpha(s_j,s_k)-1}\end{aligned} \tag{6}$$

which is the kernel of a Dirichlet density with updated $\alpha^*(s_j, s_k)=n(s_j, s_k)+\alpha(s_j, s_k)$ (Gelman et al. 2013: 69).

This is easily extended to a prior on the full matrix **P** where we consider the product of Dirichlet priors given by

$$\begin{aligned}\pi(p(\cdot,\cdot)|\alpha(\cdot,\cdot))&=\prod_j\left[\frac{\Gamma(\Sigma_k\alpha(s_j,s_k))}{\prod_k\Gamma(\alpha(s_j,s_k))}\prod_k p(s_j,s_k)^{\alpha(s_j,s_k)-1}\right]\\ &\propto\prod_{j,k}p(s_j,s_k)^{\alpha(s_j,s_k)-1}\end{aligned} \tag{7}$$

again subject to $0<p(s_j, s_k)<1$ for all j, k pairs we wish to estimate and $\Sigma_k p(s_j, s_k)=1$. Here $p(\cdot, \cdot)$ and $\alpha(\cdot, \cdot)$ represent all the transition probabilities we are estimating and their corresponding α hyperparameters.

Using the data, we can construct a count matrix of observed transitions in the same form as the 57×63 matrix [**B Q**] from (1). The full sampling distribution is proportional to the product of 46 independent multinomial probability mass functions with each representing a transient state (11 states have only one possible transition for which the probability is fixed at one). Although each row contains 63 transition probabilities, many are fixed at zero and need not be estimated. We therefore consider each row to be multinomial with the number of categories equal to the number of possible transitions from that particular state. Application of Bayes' theorem on (4) and (7) yields our full posterior over all (estimated) transition probabilities in [**B Q**] given by

$$\pi(p(\cdot,\cdot)|\mathscr{Y}, \alpha(\cdot,\cdot)) \propto \prod_j \prod_{s\in S_j} p(s_j, s)^{n(s_j,s)+\alpha(s_j,s)-1} \quad (8)$$

where S_j is the set of possible transitions from state s_j. This posterior is proportional to a product of Dirichlet densities with updated $\alpha^*(s_j, s_k)=n(s_j, s_k)+\alpha(s_j, s_k)$. This Dirichlet-multinomial conjugacy enables straightforward sampling from the joint posterior distribution of [**B Q**]. Specifically, for row j of [**B Q**], we simulate a draw from a Dirichlet distribution with parameter vector $\alpha^*(s_j, \cdot)$. This is done for each row, followed by arrangement of the draws into the appropriate matrix **P**. Sampling is repeated to obtain a simulated approximation to (8).

Without strong a priori beliefs regarding the transition probabilities, we specify relatively non-informative priors. A common Dirichlet prior has hyperparameters $\alpha(s_j, s)=1$ for each $s\in S_j$. However, as the number of categories increases even moderately, this prior becomes increasingly informative and shrinks all transition probability estimates toward $1/K_j$ where $K_j=|S_j|$, the number of possible transitions from state s_j.

The problem of choosing a non-informative Dirichlet prior with a multinomial sampling distribution as the number of categories increases is addressed by de Campos and Benavoli (2011). They define *prior strength* (employing our notation) as $\Sigma_{s\in S_j}\alpha(s_j, s)$ and develop a maximum recommended prior strength for large sample sizes. This maximum prior strength is a decreasing function of the number of categories K_j:

$$\xi(K_j) \equiv \frac{-2(1-2^{1/(K_j-1)})(K_j-(1-2^{1/(K_j-1)}))}{K_j-1} \quad (9)$$

(de Campos and Benavoli 2011). This was implemented in our analysis by applying (9) to the number of possible transitions for each row of [**B Q**] to find the recommended prior strength for each row as $\xi(K_j)$. We then set $\alpha(s_j, s)=\xi(K_j)/K_j$ for all $s\in S_j$. The selected prior strength for transient state rows with only two possible transitions was set to one instead of $\xi(2)=6$.

This procedure results in a relatively non-informative Dirichlet prior which assigns equal prior weight for each transition in a row. With little a priori information as to which transitions are relatively more common, we consider this to be a reasonable prior. Furthermore, this prior yields posterior point estimates similar to the maximum likelihood estimates of the transition probabilities. Table 4 gives the chosen values for all prior hyperparameters.

The probabilities in [**B Q**] are not the ultimate interest of estimation. We seek the marginal probabilities of each transient state being absorbed into each of the absorbing states. For example, if the current state is *Pass 5*, we are interested in the probability that the current chain will be absorbed into *Shot 3* (goal), which we denote p_A (*Pass* 5, *Shot* 3). These probabilities are found in matrix **A** from (2), the closed form solution for which is given in (3). Rather than transform to find the posterior distribution for **A**, we draw independent samples from it using

Table 4 Transient states with corresponding number of possible transitions K_j, chosen total prior strength $\xi(K_j)$, and Dirichlet prior hyperparameter value $\alpha(s_j, s)=\xi(K_j)/K_j$ for each parameter. The selected prior strength for the last group of transient states with two transitions was set to one instead of $\xi(2)=6$.

States	Transitions	Prior strength	$\alpha(s_j, s)$
Pass 5–8; First-touch Pass 7–8; Free Kick 6–8; Corner Kick 8; Corner Kick; Throw-in 5–7; Kick-off 6–7	19	0.08305	0.00437
First-touch Pass 5–6	24	0.06393	0.00266
Dribbling Touch 5	34	0.04377	0.00129
Dribbling Touch 3–4; Controlling Touch 3–5	39	0.03780	0.00100
Fouled	10	0.17933	0.01793
Deflected OB	15	0.10913	0.00728
Pass 1, 1.5; First-touch Pass 1.5; Free Kick 1, 1.5; Corner Kick 1; Goal Kick; Goalie Throw; Throw-in 1, 1.5	21	0.07418	0.00353
First-touch Pass 1	26	0.05854	0.00225
Pass 0; First-touch Pass 0; Dribbling Touch 0; Controlling Touch 0; Free Kick 0; Corner Kick 0; Goal Kick 0; Goalie Throw 0; Throw-in 0	2	1.0	0.5

Monte Carlo simulation. With each draw from the posterior distribution for [**B Q**], we perform the transformation (3) to obtain draws for **A**. This sample then yields a posterior distribution on each absorption probability $p_A(s_j, s_a)$, from which we can make inference and draw conclusions.

3.4 Model assessment

To assess the predictive fitness of our specified model and to check the model against simpler alternatives, we performed the following cross-validation checks (see Hastie et al. 2009: 241). Data from 11 randomly selected games were set aside as a validation set. The model was then fit to a training set (the remaining 11 games) and 95% credible intervals were calculated for each transition probability. Maximum likelihood estimates of the same transition probabilities were calculated from the validation set and compared to the credible intervals from the training set. Our assessment statistic was the interval capture rate (where interval boundaries were rounded to the nearest ten-thousandth to allow estimates of zero to agree with intervals nearly reaching zero). This process was repeated for a total of 30 randomly selected partitions of games. This resulted in an average capture rate of 79.6% with a standard deviation of 1.4%. The average capture rate among non-zero estimates was 70.2%. Although this is much lower than the nominal 95%, we emphasize that the training and validation estimates did not strongly disagree in general. The mean absolute difference between the estimated transition probabilities (for non-zero estimates) was 0.023.

This process was repeated for several competing models. The competing models were obtained by further combining states with similar transitions as in Section 3.2, resulting in substantially smaller transition matrices. Credible interval capture rates were comparable to our full model, generally falling between 70% and 80%. Thus using our criterion, we conclude that our full model is not over-fit in terms of inferior prediction. We proceed with this model because it is informative regarding most skill and rating combinations, allowing a richer comparison of skills.

Additionally, the full model was fit to a subset of the data corresponding only to a team which played in 15 of the 22 games. The resulting inference was very similar to that given in Section 4. This provides some justification for pooling data from different teams as well as evidence for the generalizability of our model to women's collegiate soccer.

4 Results

In this section, we present our results from applying the specified model to the data. Analysis is performed by examining the marginal posterior densities of parameters of interest, in our case transition probabilities. Samples from the full joint posterior density yield posterior probabilities for any statistical hypothesis regarding the parameters. Due to the low cross-validated credible interval coverage reported in Section 3.4, we believe that the following inferences may be somewhat anti-conservative, meaning the posterior distributions are likely too narrow.

Posterior sampling was performed in the R statistical programming language (R Core Team 2012) with additional packages to assist in posterior simulation and reporting (Plummer et al. 2006; Sarkar 2008; Martin, Quinn, and Park 2011; Bates and Maechler 2013; Dahl 2013). The sample consists of 100,000 independent draws. Because the full posterior distribution for [**B Q**] is the product of 46 independent Dirichlet distributions, posterior draws were sampled directly from the Dirichlet distributions. With draws from the full joint posterior, we are able to obtain posterior distributions on functions of the parameters, such as absorption probabilities or expected sequence length.

Recall that we defined sequences to end with either a turnover or shot. With this definition, there are 11,573 total sequences in the data, with mean length 3.46 moves and standard deviation 3.30 moves. This averages to about 526 sequences per game and about 5.8 sequences per minute of play. Most sequences are five moves or less, and the longest sequence is 40 moves. Posterior mean point estimates and credible intervals of highest posterior density (HPD; Carlin and Louis 2009: 48) for expected sequence lengths by skill are consistent with observed sequence lengths in the data and are given in Table A.3.

4.1 Probability of scoring opportunity

Because the probability of scoring a goal on a given offensive sequence is small and we are interested in the skills which create scoring opportunities, we focus attention on shots rather than goals. We define the *Scoring Opportunity* absorption probability for skill s_j as

$$p_{scop}(s_j) \equiv p_A(s_j, Shot\ 3) + p_A(s_j, Shot\ 2) + p_A(s_j, Shot\ 1), \tag{10}$$

the sum of absorption probabilities into viable shots. This quantity is easily obtained from the posterior draws of **A**. The parameter p_{scop} for skill j is properly interpreted as the marginal probability (under the Markov model) that the *current* offensive sequence, in which skill j was the most recent move, will end with a shot within six yards of the goal.

Table A.4 gives posterior mean point estimates and 95% HPD credible intervals for absorption probabilities into *Scoring Opportunity, Turnover*, and *Bad Turnover*. Note that the absorption probabilities from turnover moves (e.g., *Pass 0*) to shot outcomes must be zero because turnovers are absorbing states.

Because summary information is often easier to absorb when presented visually, posterior box plots for each of the skills that can lead to scoring opportunities are given in Figure 1. These box plots summarize the marginal posterior densities of p_{scop} for each skill. The center line of each box plot is the posterior median, the box edges are the 25th and 75th percentiles, and whiskers reach to the boundaries of the 95% HPD credible intervals. Figure 2 shows the same box plots in the low p_{scop} range for easier comparison of skills with lower scoring opportunity probabilities.

Figure 1 reveals that successful passes into the mixer are associated with dramatically higher probability of a scoring opportunity than any other moves. Note the high degree of uncertainty in these estimates; our 95% HPD interval for p_{scop} (*Free Kick* 8) is (0.223, 0.643). Consequently, we lack strong evidence to suggest that any of these passes into the mixer has higher p_{scop} than the others. Successful corner kicks (not into the mixer) also result in a high probability situation, although not as high as successful passes into the mixer.

All the passes into the mixer discussed so far are *successful* passes. Each of *Pass, First-touch Pass*, and *Free Kick* have an associated 1.5 rating that indicates a 50/50 pass into the mixer. It is important to acknowledge that all risky passes into the mixer are associated with much lower p_{scop} than successful passes into the mixer. Not only this, but the point estimates of p_{scop} for passes rated 1.5 are lower than for ordinary pass moves. For example, *Pass 5–7* are each associated with higher probability of a scoring opportunity than *Pass 1.5*.

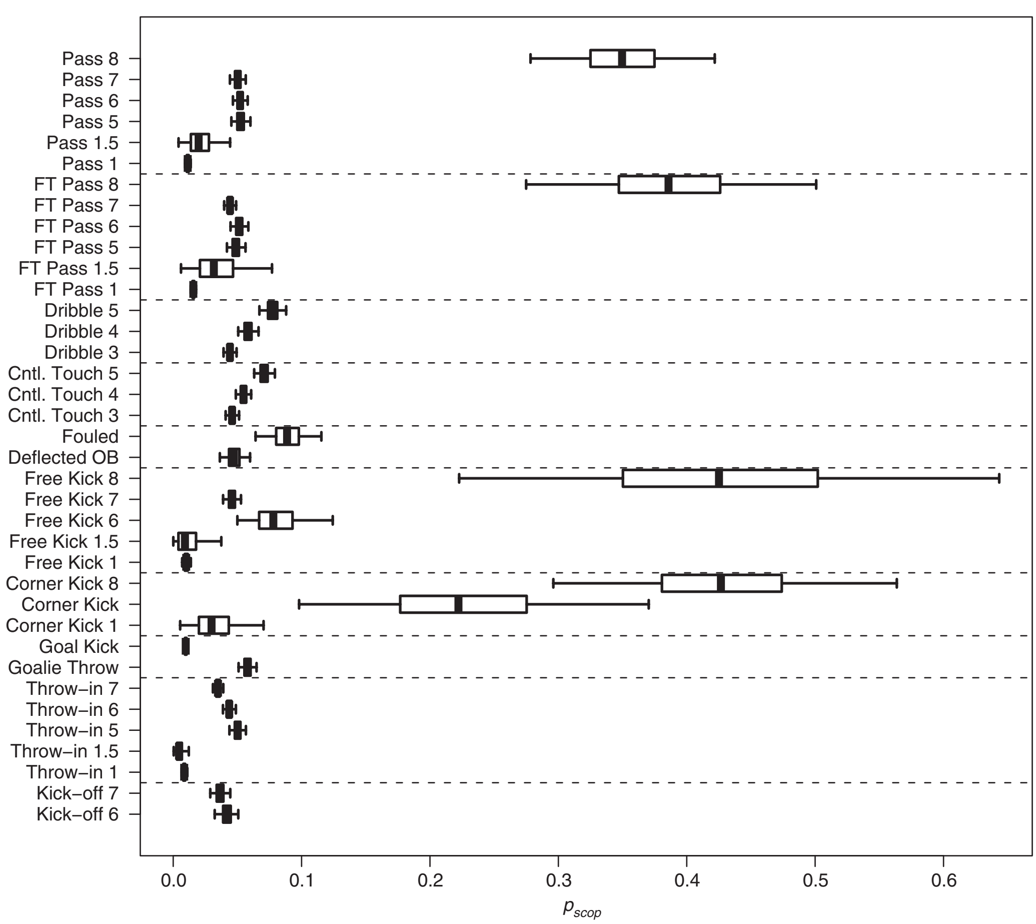

Figure 1 Marginal posterior box plots of p_{scop} (probability of absorption into *Scoring Opportunity*) for skills that can lead to scoring opportunities. The box plot centers are posterior medians, box edges reach to the first and third quartiles, and whiskers reach to the boundaries of 95% HPD intervals. Dashed horizontal lines separate offensive skill classes.

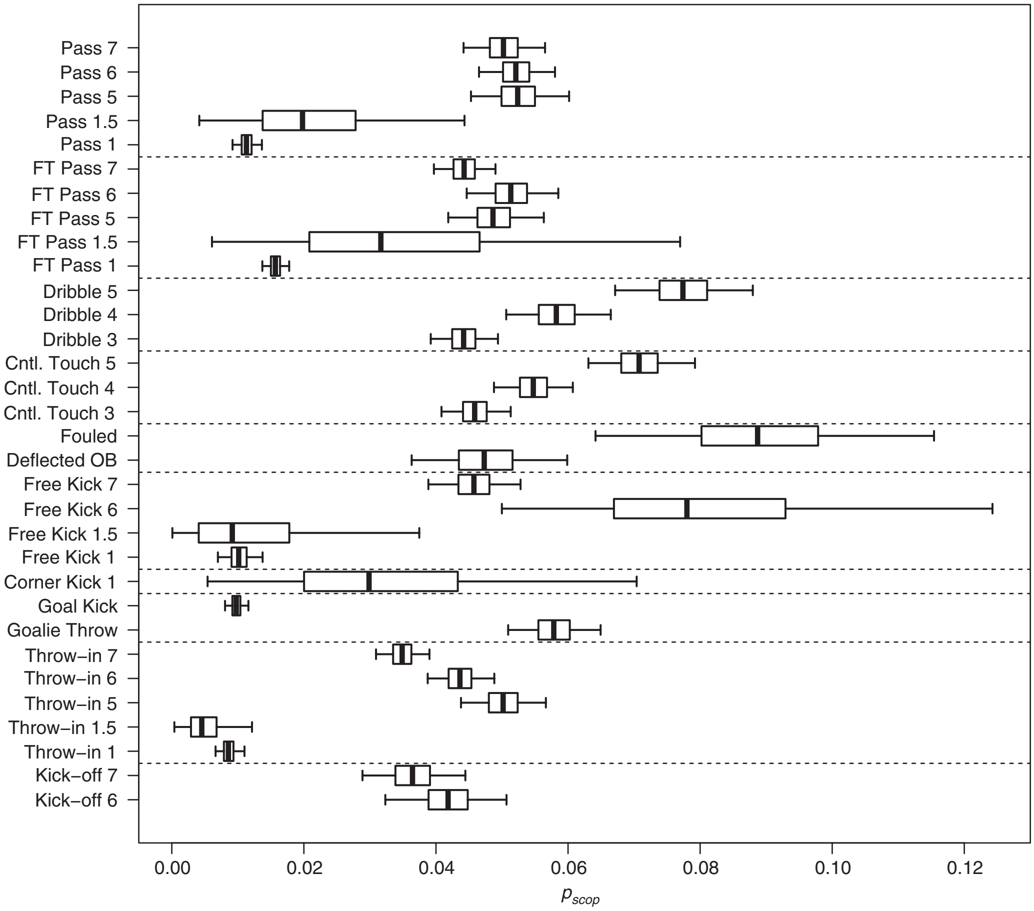

Figure 2 Marginal posterior box plots of p_{scop} (probability of absorption into *Scoring Opportunity*) for skills with p_{scop} estimated to be <0.12. The box plot centers are posterior medians, box edges reach to the first and third quartiles, and whiskers reach to the boundaries of 95% HPD intervals. Dashed horizontal lines separate offensive skill classes. Note that this is identical to Figure 1 except the horizontal axis is restricted and skills with large p_{scop} are removed.

Pass and *First-touch Pass 5–7* have no clear pattern. However, the point estimate for p_{scop} (*First-touch Pass 7*) (forward) is the lowest of the six moves. In contrast with pass moves, touch moves have a clearly discernible progression. Forward dribbling touches are associated with higher p_{scop} than lateral (perpendicular to the sideline) dribbling touches, which are associated with higher p_{scop} than backward dribbling touches. In each case, the posterior probability of separation[2] is >0.999. The same can be said for controlling touch moves. Furthermore, it is clear from Figure 2 that forward dribbling touches are preferable to all pass moves that are not into the mixer or set piece kicks. This result is consistent with those reported by Thomas et al. (2009).

2 When comparing skills, we refer to the posterior probability that p_{scop} is greater for one skill than another as the posterior probability of separation between the two skills.

Although they are not direct offensive actions, *Fouled* and *Deflected OB* can offer strategic insight. From a defensive perspective, our results demonstrate evidence that deflecting the ball out of bounds is preferable to committing a foul, as the posterior probability that p_{scop} (*Fouled*)>p_{scop} (*DeflectedOB*) is >0.999. That is, committing a foul puts the team with possession of the ball in better offensive position. This result can also be seen by comparing p_{scop} among free kicks (associated with fouls) and throw-ins (associated with deflections).

Free kicks offer an interesting and less intuitive result. *Free Kick 6* (lateral) is associated with a higher p_{scop} than *Free Kick 7* (forward). Sparse data for direct transitions from these states to scoring opportunity outcomes may account for this result. Additionally, *Free Kick 6* and *Free Kick 7* transition to different kinds of controlling touches, which also affects the p_{scop} estimates. Goal kicks and goalie throws offer another surprise. When considering the probability of achieving a scoring opportunity on the *current* *uninterrupted possession*, *Goalie Throw* is much preferred over *Goal Kick*. There are two important caveats to this statement. The first is that with short expected possession length (between five and six), the high p_{scop} associated *Goalie Throw* likely has more to do with being followed by stable moves (which in other sequences lead to scoring opportunities) than with *Goalie Throw* leading to scoring opportunities. The second is that the consequence of a poor goalie throw is likely worse than that of a poor goal kick.

Throw-in moves also follow an interesting pattern. *Throw-in 5–7* display a clear progression of *decreasing* p_{scop}. The posterior probability of separation for these moves is >0.99 in each case. One possible explanation for this result is that backward throw-in passes could lead to 1) more stable possession and 2) forward touches associated with higher p_{scop}. Although the posterior probability of separation is smaller, we again see this pattern with kick-offs.

4.2 Incorporating Turnover Risk

To gain more understanding of the relative value of these skills, we can also consider risks, such as the probability that the offensive sequence will end in *Bad Turnover*. Recall that $p_A(s_j, BTOV)$ denotes the marginal probability of a chain currently in state s_j being absorbed into the state *Bad Turnover*. Table A.4 gives posterior mean estimates and credible intervals for this probability for the skills under consideration. Figure 3 is a reproduction of Figure 1 with posterior box plots corresponding to $p_A(s_j, BTOV)$ added for each skill. The box plots corresponding to *Bad Turnover* are rose-colored and thin.

We see from Figure 3 that the marginal probability of *Bad Turnover* is in the 0.08–0.10 range for many skills and does not change with rating. There are two sets of states with particularly interesting results. First, our claim that deflecting the ball out of bounds is better for a defending team than committing a foul is further justified. Fouls are

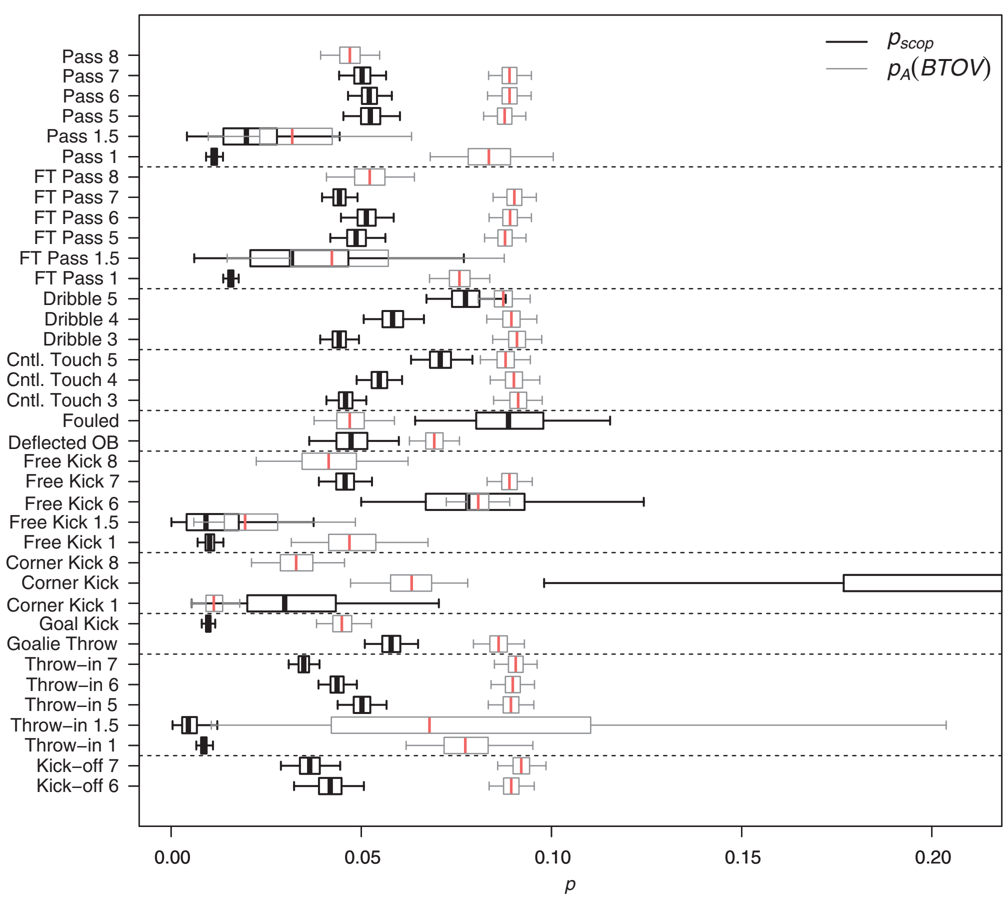

Figure 3 Marginal posterior box plots of p_{scop}, the probability of absorption into *Scoring Opportunity*, and $p_A(\cdot, BTOV)$, the probability of absorption into *Bad Turnover* (rose-colored and thin) for skills that can lead to scoring opportunities. The box plot centers are posterior medians, box edges reach to the first and third quartiles, and whiskers reach to the boundaries of 95% HPD intervals. Dashed horizontal lines separate offensive skill classes.

associated with higher probability of leading to a scoring opportunity for the team in possession and lower probability of *Bad Turnover*. Deflections out of bounds exhibit the reverse, with higher probability of *Bad Turnover* and lower probability of leading to a scoring opportunity.

The second interesting set of states is *Goal Kick* and *Goalie Throw*. It was noted earlier that although p_{scop} is significantly higher for *Goalie Throw*, the consequence of turning the ball over after a goalie throw is likely more severe. We see that p_A(*Goalie Throw, BTOV*) is significantly higher than p_A(*Goal Kick, BTOV*), confirming that successful goal kicks result in fewer bad turnovers than successful goalie throws. To examine the consequence of turning the ball over near the team's own goal, we can compare our estimates associated with unsuccessful goal kicks and goalie throws given in Table A.4. Our estimate for p_A(*Goal Kick* 0, *BTOV*) is 0.063 with 95% HPD interval (0.018, 0.115) and our estimate for p_A(*Goalie Throw* 0, *BTOV*) is 0.417 with 95% HPD interval (0.080, 0.770). The wide intervals are due to sparse data, for which we make no inferential claim. We note, however, that the high estimate for p_A(*Goalie Throw* 0, *BTOV*) is intuitive because it is an immediate turnover near the goal.

Our utility for *Bad Turnover* probabilities is limited due to the absence of spatial data and our *ad hoc* definition of *Bad Turnover*. Our incorporation of turnover risk is primarily illustrative of the potential for an analysis on data free of these difficulties.

5 Discussion

Statistically, our strongest result is that successful passes into the mixer and corner kicks are associated with much higher probability of leading to a scoring opportunity than any other move. However, this comes with a strong qualification. Uncertain passes into the mixer not only have much lower probability of a scoring opportunity than successful passes into the mixer, but often have lower probability than other routine and non-aggressive moves. This discourages the strategy of risky passes into the mixer.

It appears that passing direction, with exception of passing into the mixer, does not change the probability of leading to a scoring opportunity. However, direction is clearly important for touch moves, where forward plays are associated with higher probabilities.

When a team playing defense is forced to break up a play, it appears more advantageous to deflect the ball out of bounds than to commit a foul (of course, committing a foul is preferable to allowing the offensive team to successfully advance the ball into the mixer). The preference of deflecting the ball out of bounds is supported not only by its association with lower probability of leading to a scoring opportunity, but also an increased probability of leading to a bad turnover, of which the recovering team can take advantage.

An interesting result from this analysis that is consistent across multiple types of moves is that conservative inbounding passes are associated with a higher probability of leading to a scoring opportunity than the more aggressive forward in-bounding passes. This is true of free kicks, goal kicks, throw-ins, and kick-offs. A potential explanation for these consistent results (in addition to those given in Section 4.1) could be that in-bounding pass moves allow the defense to become organized. This may reduce the effectiveness of penetrating passes, as the defense perceives a penetrating play as more threatening and defends more aggressively. One interesting special case is for goalie throw-ins, where we see a trade-off between scoring opportunities and bad turnovers. A goalie's choice to instead punt the ball downfield results in lower probability of leading to a scoring opportunity on the same play, but also results in lower probability of leading to a bad turnover.

These results identify association between offensive skills and the probability of a scoring opportunity, and can be used to distinguish skills by importance and aid in strategic planning and decision making. We emphasize that the results apply primarily to the 24 teams represented in the data. There is some evidence that these patterns may generalize to collegiate women's soccer, but not necessarily to soccer in general. It would be interesting to apply these methods of game notation and analysis to other leagues. This would allow us to make broader generalizations and determine whether the patterns observed in this analysis are regularities in soccer or are specific to teams. It would also be insightful to use this model to compare how relative importance of skills changes across leagues or teams that play at different levels of expertise.

Acknowledgments: We are grateful to the Editor-in-Chief, Associate Editor, referees, and all other editing contributors whose suggestions have enhanced this paper.

Appendix

Additional Tables

Table A.1 Original notation system introduced in Thomas et al. (2009) for identifying and rating the effects of skills performed during soccer matches.

Pass		
Performance score		
	7	The pass results in a scoring opportunity
	6	The pass results in a ball in the "mixer"
	5	The pass results in forward (penetrating) play
	4	The pass results in a square play
	3	The pass results in back play
	2	The pass is deflected out of bounds/player is fouled
	1	The pass results in a 50/50 ball to the opponents
	0	The pass results in an immediate loss of possession
Dribble		
Performance score		
	4	The dribble results in a scoring opportunity
	3	The dribble is toward the opponent's goal (penetrating)
	2	The dribble is toward own goal or square
	1	The dribble results in a deflection out of bounds/player is fouled
	0	The dribble results in an immediate loss of possession
First touch		
Performance score		
	5	The first touch results in a scoring opportunity
	4	The first touch results in penetrating play
	3	The first touch results in a square play
	2	The first touch results in a back play
	1	The first touch is deflected out of bounds/player is fouled
	0	The first touch results in an immediate loss of possession
Defensive tactics		
Performance score		
	8	Challenge results in possession won; direct play in the attacking third
	7	Challenge results in possession won; direct play in the middle third
	6	Challenge results in possession won; direct play in the defensive third
	5	Challenge results in possession; indirect play/forced errors out of bounds
	4	Challenge with delay results in a ball being played indirectly
	3	Challenge with delay, but the opponents still penetrate
	2	Challenge but no delay, and the opponents penetrate
	1	Challenged a 50/50 ball, but possession is not regained
	0	Player did not provide immediate chase or chase results in a foul

Note that this system includes defensive skills, which were omitted from our offensive-oriented analysis. Also, the system includes no rating system for shots taken.

Table A.2 All transient states with their possible transitions.

State	Possible transitions from the state
Pass 5–8	Shot 0–3; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Pass 4	Fouled
Pass 3	Deflected OB
Pass 1.5	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Pass 1	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Pass 0	Turnover; Bad Turnover
First-touch Pass 7–8	Shot 0–3; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
First-touch Pass 5–6	Shot 0–3; First-touch Pass 0–1, 3–8; Controlling Touch 0–5; Goal Kick; Goal Kick 0, 3; Goalie Throw; Goalie Throw 0
First-touch Pass 4	Fouled
First-touch Pass 3	Deflected OB
First-touch Pass 1.5	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
First-touch Pass 1	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5; Goal Kick; Goal Kick 0, 3; Goalie Throw; Goalie Throw 0
First-touch Pass 0	Turnover; Bad Turnover
Dribbling Touch 5	Shot 0–3; Pass 0–1, 3–8; First-touch Pass 0–1, 3–8; Dribbling Touch 0–5; Controlling Touch 0–5
Dribbling Touch 3–4	Shot 0–3; Pass 0–1, 3–8; First-touch Pass 0–1, 3–8; Dribbling Touch 0–5; Controlling Touch 0–5; Goal Kick; Goal Kick 0, 3; Goalie Throw; Goalie Throw 0
Dribbling Touch 2	Fouled
Dribbling Touch 1	Deflected OB
Dribbling Touch 0	Turnover; Bad Turnover
Controlling Touch 3–5	Shot 0–3; Pass 0–1, 3–8; First-touch Pass 0–1, 3–8; Dribbling Touch 0–5; Controlling Touch 0–5; Goal Kick; Goal Kick 0, 3; Goalie Throw; Goalie Throw 0
Controlling Touch 2	Fouled
Controlling Touch 1	Deflected OB
Controlling Touch 0	Turnover; Bad Turnover
Fouled	Shot 0–3; Free Kick 0–1, 6–8
Deflected OB	Corner Kick; Corner Kick 0–1, 3, 8; Goal Kick; Goal Kick 0, 3; Throw-in 0–1, 3, 5–7
Free Kick 6–8	Shot 0–3; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Free Kick 1.5	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Free Kick 1	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Free Kick 0	Turnover; Bad Turnover
Corner Kick 8	Shot 0–3; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Corner Kick	Shot 0–3; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Corner Kick 3	Deflected OB
Corner Kick 1	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Corner Kick 0	Turnover; Bad Turnover
Goal Kick	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Goal Kick 3	Deflected OB
Goal Kick 0	Turnover; Bad Turnover
Goalie Throw	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Goalie Throw 0	Turnover; Bad Turnover
Throw-in 5–7	Shot 0–3; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Throw-in 3	Deflected OB
Throw-in 1.5	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Throw-in 1	Shot 0–3; Turnover; Bad Turnover; First-touch Pass 0–1, 3–8; Controlling Touch 0–5
Throw-in 0	Turnover; Bad Turnover
Kick-off 6–7	Shot 0–3; First-touch Pass 0–1, 3–8; Controlling Touch 0–5

Table A.3 Posterior mean estimates and 95% HPD credible intervals for expected sequence length in moves for sequences starting at the skill indicated. The 11,573 observed sequences in the data had a mean length of 3.46 and standard deviation of 3.30 moves.

		Expected sequence length
Pass 8	2.835	(2.498, 3.181)
Pass 7	5.008	(4.897, 5.116)
Pass 6	5.228	(5.129, 5.330)
Pass 5	4.903	(4.785, 5.026)
Pass 1.5	1.171	(1.049, 1.309)
Pass 1	2.092	(1.961, 2.231)
First-touch Pass 8	2.939	(2.410, 3.474)
First-touch Pass 7	5.150	(5.038, 5.257)
First-touch Pass 6	4.962	(4.843, 5.079)
First-touch Pass 5	4.888	(4.759, 5.014)
First-touch Pass 1.5	1.667	(1.330, 2.029)
First-touch Pass 1	2.616	(2.526, 2.706)
Dribbling Touch 5	3.953	(3.847, 4.061)
Dribbling Touch 4	4.489	(4.382, 4.597)
Dribbling Touch 3	5.000	(4.876, 5.124)
Controlling Touch 5	4.613	(4.522, 4.707)
Controlling Touch 4	4.825	(4.738, 4.913)
Controlling Touch 3	5.111	(5.012, 5.208)
Fouled	3.336	(3.135, 3.537)
Deflected OB	4.238	(4.069, 4.411)
Free Kick 8	2.020	(1.406, 2.706)
Free Kick 7	5.224	(4.905, 5.542)
Free Kick 6	4.990	(4.584, 5.360)
Free Kick 1.5	1.153	(1.059, 1.263)
Free Kick 1	2.127	(1.887, 2.376)
Corner Kick 8	2.067	(1.615, 2.558)
Corner Kick	4.476	(3.619, 5.264)
Corner Kick 1	1.405	(1.176, 1.661)
Goal Kick	2.179	(2.063, 2.294)
Goalie Throw	5.506	(5.296, 5.701)
Throw-in 7	4.796	(4.646, 4.950)
Throw-in 6	5.313	(5.131, 5.484)
Throw-in 5	5.401	(5.162, 5.628)
Throw-in 1.5	1.720	(1.136, 2.412)
Throw-in 1	2.019	(1.888, 2.154)
Kick-off 7	4.620	(4.163, 5.060)
Kick-off 6	5.273	(4.762, 5.761)

Note that these estimates are slightly inflated with respect to the data lengths, as the transition matrix **P** includes states *Fouled* and *Deflected OB* in addition to the skills that led to those states (e.g., *Pass 4*).

Table A.4 Posterior mean estimates of absorption probabilities to *Shot Opportunity* (p_{scop}), *Turnover (TOV)*, and *Bad Turnover (BTOV)* with 95% HPD credible intervals.

		p_{scop}		*TOV*		*BTOV*
Pass 8	0.350	(0.278, 0.422)	0.448	(0.381, 0.516)	0.047	(0.039, 0.055)
Pass 7	0.050	(0.044, 0.057)	0.838	(0.829, 0.846)	0.089	(0.083, 0.095)
Pass 6	0.052	(0.047, 0.058)	0.831	(0.822, 0.839)	0.089	(0.083, 0.095)
Pass 5	0.053	(0.045, 0.060)	0.831	(0.821, 0.841)	0.088	(0.082, 0.093)
Pass 1.5	0.022	(0.004, 0.044)	0.942	(0.905, 0.975)	0.034	(0.010, 0.063)
Pass 1	0.011	(0.009, 0.014)	0.898	(0.881, 0.915)	0.084	(0.068, 0.100)
Pass 0			0.889	(0.875, 0.903)	0.111	(0.097, 0.125)
First-touch Pass 8	0.387	(0.275, 0.501)	0.490	(0.388, 0.595)	0.052	(0.041, 0.064)
First-touch Pass 7	0.044	(0.040, 0.049)	0.844	(0.837, 0.852)	0.090	(0.085, 0.096)
First-touch Pass 6	0.052	(0.045, 0.059)	0.837	(0.827, 0.846)	0.089	(0.084, 0.095)
First-touch Pass 5	0.049	(0.042, 0.056)	0.841	(0.832, 0.851)	0.088	(0.082, 0.093)
First-touch Pass 1.5	0.036	(0.006, 0.077)	0.900	(0.837, 0.956)	0.046	(0.015, 0.088)
First-touch Pass 1	0.016	(0.014, 0.018)	0.900	(0.892, 0.909)	0.076	(0.068, 0.084)
First-touch Pass 0			0.901	(0.889, 0.912)	0.099	(0.088, 0.111)
Dribble 5	0.078	(0.067, 0.088)	0.797	(0.784, 0.810)	0.087	(0.081, 0.094)
Dribble 4	0.058	(0.051, 0.066)	0.824	(0.814, 0.835)	0.090	(0.083, 0.096)
Dribble 3	0.044	(0.039, 0.049)	0.841	(0.832, 0.850)	0.091	(0.085, 0.097)
Dribble 0			0.882	(0.861, 0.902)	0.118	(0.098, 0.139)
Controlling Touch 5	0.071	(0.063, 0.079)	0.808	(0.797, 0.819)	0.088	(0.081, 0.094)
Controlling Touch 4	0.055	(0.049, 0.061)	0.829	(0.820, 0.838)	0.090	(0.084, 0.097)
Controlling Touch 3	0.046	(0.041, 0.051)	0.839	(0.831, 0.848)	0.091	(0.085, 0.098)
Controlling Touch 0			0.906	(0.894, 0.917)	0.094	(0.083, 0.106)
Fouled	0.089	(0.064, 0.115)	0.820	(0.788, 0.851)	0.047	(0.038, 0.059)
Deflected OB	0.048	(0.036, 0.060)	0.859	(0.843, 0.875)	0.069	(0.063, 0.076)
Free Kick 8	0.428	(0.223, 0.643)	0.421	(0.228, 0.612)	0.042	(0.022, 0.062)
Free Kick 7	0.046	(0.039, 0.053)	0.843	(0.833, 0.854)	0.089	(0.083, 0.095)
Free Kick 6	0.082	(0.050, 0.124)	0.766	(0.710, 0.817)	0.081	(0.072, 0.089)
Free Kick 1.5	0.013	(0.000, 0.037)	0.964	(0.930, 0.991)	0.023	(0.006, 0.048)
Free Kick 1	0.010	(0.007, 0.014)	0.936	(0.916, 0.955)	0.048	(0.032, 0.068)
Free Kick 0			0.960	(0.918, 0.994)	0.040	(0.006, 0.082)
Corner Kick 8	0.428	(0.296, 0.564)	0.320	(0.204, 0.436)	0.033	(0.021, 0.046)
Corner Kick	0.230	(0.098, 0.370)	0.613	(0.469, 0.745)	0.063	(0.047, 0.078)
Corner Kick 1	0.034	(0.005, 0.070)	0.940	(0.895, 0.978)	0.012	(0.005, 0.018)
Corner Kick 0			0.984	(0.939, 1.000)	0.016	(0.000, 0.061)
Goal Kick	0.010	(0.008, 0.012)	0.940	(0.932, 0.948)	0.045	(0.038, 0.053)
Goal Kick 0			0.937	(0.885, 0.982)	0.063	(0.018, 0.115)
Goalie Throw	0.058	(0.051, 0.065)	0.829	(0.817, 0.840)	0.086	(0.079, 0.093)
Goalie Throw 0			0.583	(0.230, 0.920)	0.417	(0.080, 0.770)
Throw-in 7	0.035	(0.031, 0.039)	0.856	(0.848, 0.864)	0.091	(0.085, 0.096)
Throw-in 6	0.044	(0.039, 0.049)	0.846	(0.837, 0.854)	0.090	(0.084, 0.096)
Throw-in 5	0.050	(0.044, 0.057)	0.837	(0.827, 0.846)	0.089	(0.083, 0.095)
Throw-in 1.5	0.006	(0.000, 0.012)	0.907	(0.789, 0.988)	0.084	(0.011, 0.204)
Throw-in 1	0.009	(0.007, 0.011)	0.909	(0.892, 0.926)	0.078	(0.062, 0.095)
Throw-in 0			0.953	(0.918, 0.985)	0.047	(0.015, 0.082)
Kick-off 7	0.037	(0.029, 0.044)	0.855	(0.843, 0.866)	0.092	(0.086, 0.099)
Kick-off 6	0.042	(0.032, 0.051)	0.849	(0.837, 0.863)	0.089	(0.084, 0.095)

References

Allan, M. L. 2009. *Measuring Skill Importance in Women's Soccer and Volleyball*, Master's thesis, Brigham Young University. (http://hdl.lib.byu.edu/1877/etd2809).

Anderson, T. W. and L. A. Goodman. 1957. "Statistical Inference about Markov Chains." *The Annals of Mathematical Statistics* 28:89–110.

Bates, D. and M. Maechler. 2013. *Matrix: Sparse and Dense Matrix Classes and Methods*. (http://CRAN.R-project.org/package=Matrix, r package version 1.0-12).

Brillinger, D. R. 2007. "A Potential Function Approach to the Flow of Play in Soccer." *Journal of Quantitative Analysis in Sports* 3.

Carlin, B. P. and T. A. Louis. 2009. *Bayesian Methods for Data Analysis*. 3rd ed. Boca Raton: CRC Press.

Dahl, D. B. 2013. *xtable: Export tables to LaTeX or HTML*. (http://CRAN.R-project.org/package=xtable, r package version 1.7-1).

de Campos, C. and A. Benavoli. 2011. "Inference with Multinomial Data: Why to Weaken the Prior Strength." in *International Joint Conference on Artificial Intelligence*, 2107–2112. Retrieved August 6, 2013 (http://www.aaai.org/ocs/index.php/IJCAI/IJCAI11/paper/view/3292).

Dixon, M. J. and S. G. Coles. 1997. "Modelling Association Football Scores and Inefficiencies in the Football Betting Market." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 46:265–280.

Fellingham, G. W. and C. S. Reese. 2004. "Rating Skills in International Men's Volleyball." Brigham Young University. Unpublished report to the USA National Men's Volleyball Team.

FIFA. 2006. "FIFA Survey: Approximately 250 Million Footballers Wordwide." Retrieved August 5, 2013 (http://web.archive.org/web/20060915133001/http://access.fifa.com/infoplus/IP-199_01E_big-count.pdf).

FIFA. 2007. "2006 FIFA World Cup Broadcast Wider, Longer and Farther than ever before." Retrieved August 5, 2013 (http://www.fifa.com/aboutfifa/organisation/marketing/news/newsid=111247/).

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis*. 3rd ed. Boca Raton: Chapman and Hall/CRC Press.

Goddard, J. 2005. "Regression Models for Forecasting Goals and Match Results in Association Football." *International Journal of Forecasting* 21:331.

Goldner, K. 2012. "A Markov Model of Football: Using Stochastic Processes to Model a Football Drive." *Journal of Quantitative Analysis in Sports* 8.

Hamilton, H. H. 2011. "An Extension of the Pythagorean Expectation for Association Football." *Journal of Quantitative Analysis in Sports* 7.

Hastie, T., R. Tibshirani and J. Friedman. 2009. *The Elements of Statistical Learning*. 2nd ed. New York: Springer.

Hirotsu, N. and M. Wright. 2003. "An Evaluation of Characteristics of Teams in Association Football by using a Markov Process Model." *Journal of the Royal Statistical Society: Series D (The Statistician)* 52:591–602.

Hughes, M. and I. Franks. 2005 "Analysis of Passing Sequences, Shots and Goals in Soccer." *Journal of Sports Sciences* 23:509–514.

James, B. 1980. *The Bill James Abstract*. self-published.

Karlis, D. and I. Ntzoufras. 2003. "Analysis of Sports Data by using Bivariate Poisson Models." *Journal of the Royal Statistical Society: Series D (The Statistician)* 52:381–393.

Lawler, G. F. 1995. *Introduction to Stochastic Processes*. New York: Chapman & Hall.

Maher, M. J. 1992. "Modelling Association Football Scores." *Statistica Neerlandica* 36:109–118.

Martin, A. D., K. M. Quinn, and J. H. Park. 2011. "MCMCpack: Markov chain Monte Carlo in R." *Journal of Statistical Software* 42:22. (http://www.jstatsoft.org/v42/i09/).

McHale, I. and P. Scarf. 2007. "Modelling Soccer Matches using Bivariate Discrete Distributions with General Dependence Structure." *Statistica Neerlandica* 61:432–445.

Miskin, M. A., G. W. Fellingham, and L. W. Florence. 2010. "Skill Importance in Women's Volleyball." *Journal of Quantitative Analysis in Sports* 6.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News* 6:7–11. (http://CRAN.R-project.org/doc/Rnews/).

Pollard, R. and C. Reep. 1997. "Measuring the Effectiveness of Playing Strategies at Soccer." *Journal of the Royal Statistical Society: Series D (The Statistician)* 46:541–550.

Reep, C. and B. Benjamin. 1968. "Skill and Chance in Association Football." *Journal of the Royal Statistical Society. Series A (General)* 131:581–585.

Rudd, S. 2011. "A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer using Markov Chains." Retrieved August 6, 2013 (http://onfooty.com/2011/09/nessis-wrap-up-and-slides.html).

Rue, H. and Y. Salvesen. 2000. "Prediction and Retrospective Analysis of Soccer Matches in a League." *Journal of the Royal Statistical Society: Series D (The Statistician)* 49:399–418.

R Core Team. 2012. *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. (http://www.R-project.org), ISBN 3-900051-07-0.

Sarkar, D. 2008. *Lattice: Multivariate Data Visualization with R*. New York: Springer. (http://lmdvr.r-forge.r-project.org), ISBN 978-0-387-75968-5.

Strelioff, C., J. Crutchfield, and A. Hubler. 2007. "Inferring Markov Chains: Bayesian Estimation, model Comparison, Entropy Rate, and Out-of-class Modeling." *Physical Review E* 76:011106.

Thomas, C., G. Fellingham, and P. Vehrs. 2009. "Development of a Notational Analysis System for Selected Soccer Skills of a Women's College Team." *Measurement in Physical Education and Exercise Science* 13:108–121.