## 1. Introduction

A soccer (or football) game is won by the team that has scored the most goals. Thus, the simple aim of soccer is to score more goals than the opponent. Nevertheless, many games end scoreless. This is a universal phenomenon, with scoreless draws occurring in recent finals of Europe's Champions League (Milan AC vs Juventus, 2003), South America's Copa Libertadores (Boca Juniors CF vs Once Caldas, 2004), North America's Champions Cup (Guadalajara vs Pachuca, 2007), Africa's Champions League (Etoile du Sahel vs Al-Ahly, 2005), Asia's Champions League (Suwon Bluewings vs Anyang Cheetahs, 2002), FIFA's World Cup (Brazil vs Italy, 1994), the African Nations Cup (Egypt vs Ivory Coast, 2006), the Asian Nations Cup (Saudi Arabia vs UAE, 1996), and North America's Gold Cup (USA vs Panama, 2005). We did not find scoreless draws in the finals of the European Championships and the Copa America. According to Gaviria (2000), a major drop in goal scoring occurred in the 1960s, possibly due to tactic shifts in the approach of the game. But, apart from the fact that not many goals are scored on average, the scoreless draw is a result with a special status. It is typically considered as an unwanted result – often jeopardizing the spectacle value of the game – even though it happens regularly (but note that players or fans often prefer a scoreless game over a loss). In this paper, we aim at investigating the relationship between a few performance indices of a soccer team and the proportion of scoreless draws obtained by that team. We do not aim to predict scoreless draws or soccer results in general. We merely aim to check if and how some measures are related to one specific game result.

We randomly selected teams from more than fifty domestic leagues for each of three seasons between 2003 and 2006. We recorded the number of goals scored and allowed per game (GPG and GAPG, respectively) as indices of offensive and defensive strength, the mean number of points earned per game (PPG) as an index of overall relative performance, and the percentage of games that ended in a scoreless draw (%SD). The total number of goals per game (TGPG) was simply obtained as GPG + GAPG, and could be interpreted as a team's spectacle or entertainment level. Using Bayesian networks (Heckerman et al., 1995), least squares support vector machines (LS-SVMs, Suykens & Vandewalle, 1999, Suykens et al., 2002), and hybrid Monte Carlo multi-layer perceptrons (HMC-MLP, Duane et al., 1987, Neal, 1996) we related the percentage of scoreless draws to a team's defensive and offensive qualities, overall performance, and spectacle level.

An obvious hypothesis is that %SD depends on GPG and GAPG, with teams scoring less goals and/or allowing less goals being more prone to scoreless draws. However, it is not evident that indices like GPG and GAPG by themselves are strongly related to one specific game result since the particular end score of a game is also influenced by, for our study at least, random variation (weather, injuries, bad luck, midweek matches, etc). Thus, the strength of the relationships is of particular interest to us. Obviously, expecting relationships with GPG and GAPG also indicates that a relationship with TGPG is expected. The more goals that are scored in a team's games, the less likely a particular game will end in 0-0.

With respect to PPG, it is much harder to have a clear hypothesis. One possible rule to enhance goal scoring could be to award no points for a scoreless draw. This would be most effective when PPG is not related to %SD: suppose that bad teams (i.e., those with low PPG ranked at the bottom of the league table irrespective of their goal scoring results) have many more scoreless draws than good or average teams, then not awarding points for a scoreless draw will result in large punishments for bad teams such that the impact of such a rule change is not entirely what one would hope for. One hypothesis concerning the relationship between PPG and %SD, however, could be that average teams would have more scoreless draws than bad or good teams. Average teams play more teams that are similar in strength and play fewer teams that differ greatly in strength as opposed to teams winning many or few points. This would suggest a curvilinear relationship between PPG and %SD. On the contrary, PPG and GPG/GAPG will be highly related such that taking into account information on goal scoring may leave PPG unrelated with %SD.

## 2. Previous research

Previous research has focused on events that may have considerable influence on the likelihood of a scoreless draw in a specific game: tactical decisions (Hirotsu and Wright, 2002, 2003, 2006), professional fouls and red cards (Ridder et al., 1994, Wright and Hirotsu, 2003), and penalty kicks (Jordet et al., 2007). Jordet et al. (2007) investigated to what extent the outcome of a penalty kick shootout depends on stress, player skill, and fatigue, using data from three tournaments (World Cup, European Championships, and Copa America). All indices were related to penalty kick success in a logical way: more stress and fatigue were negatively related to success, whereas skill was positively related to success. The researchers concluded that stress was most strongly linked to penalty kick success, such that psychological factors appear to matter more than skill or physical factors.

Ridder et al. (1994) investigated the effect of a red card on the game result. Using a Poisson model, they estimated that a red card increases the 'scoring intensity' or 'scoring rate' with 88% for the team that has numerical advantage while the scoring rate slightly decreases (5%) for the team that has a player expelled. Based on these estimates, Ridder et al. (1994) show that, the

earlier the expulsion, the higher the probability that the 11-player team wins the game. Also, they find that the number of scored goals increases when the expulsion occurs early in the game, indicating that the red card system achieved the aim of increasing the number of goals scored. Based on their analyses, teams of equal strength should have a 25% chance of ending the match with a draw. Using a more elaborate and realistic model by incorporating additional parameters such as home advantage, Wright and Hirotsu (2003) investigate under what circumstances a professional foul (e.g., a defender who uses his hands to deflect a goal-bound shot the goalkeeper cannot save himself) would be of benefit for the team. The foul benefits the team when it increases the expected number of earned points.

Hirotsu and Wright (2002, 2003) modeled a soccer match as a Markov process incorporating a team's tendency to score goals, conceding goals, gaining possession, and losing possession. These tendencies are obtained using a log-linear (Poisson) model. This allows the computation of the expected number of points earned at any position in any match. The authors show that this mathematical conception of a soccer match is useful for determining, with the use of dynamical programming, the optimal moment for tactical decisions such as substitutions. Wright and Hirotsu (2006) took this research another step forward by incorporating the interaction between both teams' tactical decisions. Using a game-theoretic approach in which a soccer match is modeled as a zero-sum game (i.e., one team's increase of the probability of winning equals the other team's decrease of this probability), the authors modeled the effect of decisions to change the team formation. It is demonstrated that such tactical decisions can influence the probability of winning the game.

Pollard and Reep (1997) investigated the effectiveness of ball possession originating in different situations (area of pitch and open versus set play) and of different strategies when having a specific type of ball possession (e.g., goal kick, throw-in in particular pitch area) by computing the probability of scoring a goal minus the probability of conceding a goal. Ball possession in open play has higher effectiveness as compared to set play, and gaining ball possession close to the target goal is obviously more effective. Long goal kicks are counter effective, and throw-ins close to the target goal are more effective if they are long throws in the direction of the goalmouth. Fernandez-Cantelli and Meeden (2002) investigate an alternative point system for soccer that tries to make teams play conservatively less often. A main characteristic of the system they propose is that big wins/losses are rewarded/punished more strongly than small wins/losses. They showed data suggesting that the switch from the 2-1-0 award system to the 3-1-0 award system, contrary to the red card system, did not achieve its aim: it neither increased scoring nor decreased the number of draws. Finally, Rue and Salvesen (2000) set up a Bayesian model using Markov chain Monte Carlo simulations to predict the outcome of soccer matches. The model makes good predictions based on the results of previous matches to estimate the time-varying strengths of teams' attack and defense.

## 3. Materials and methods

### 3.1. Data

The data were collected from the website of the Rec.Sport.Soccer Statistics Foundation (RSSSF, n.d.). The aim of RSSSF is "to build an archive ... of association football ... statistics, as extensive as possible, and to make this available on the so-called 'World Wide Web'" (RSSSF, n.d.). We selected 54 countries (Table 1) and three seasons (the seasons starting in 2003, 2004, and 2005). From each of these 54*3=162 competitions, we randomly selected one team and recorded that team's number of scoreless draws, goals scored, goals allowed, and points earned. We divided all of these indices by the number of games played to obtain %SD, GPG, GAPG, TGPG, and PPG. The same team could only be selected once such that the three selected teams from one country were always different teams. This way of data collection was chosen in the hope of achieving independent data points. We believe this data strategy is better than recording data from all teams from one or a few leagues. Three such datasets were collected (each with $n$ = 162), one for training, one for validation, and one for testing the regression models. This resulted in 486 teams for which data were collected (the same team could be selected for several data sets but could only be selected once per season). When match results were awarded due to irregularities, the awarded match result was used. These matches were sometimes played in whole or in part with the obtained result annulled, or were not played at all such that using the awarded match result is the best solution. Awarded results make up for a very small portion of all match results. When teams had points deducted due to irregularities, we did not take these into account. Instead, the original number of earned points was used. Again, point deduction occurred in only a very small portion of the data.

Most competitions were of the double, triple, or quadruple round robin type. In South America, often the season is split into two similar competitions (typically called 'apertura' and 'clausura'), each usually applying a round robin system. For our analyses, we always aggregated the final tables from the apertura and clausura tournaments. In some leagues, teams do not play each other team equally often (e.g., USA). Championship or relegation play-off games were not taken into account. A few exceptional situations occurred: the 2004 season in Australia was cancelled; the Colombian apertura and clausura stages consist of a

single round robin plus one extra game such that each team plays twice against only one of its opponents; the 2005 season in Ecuador consisted of an apertura and clausura stage with one team relegated after each stage such that both stages could not be aggregated and were included separately.

Table 1. Countries included in the study

| Association | *N* | List of countries |
|---|---|---|
| UEFA | 29 | Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, England, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Netherlands, Norway, Poland, Portugal, Romania, Russia, Scotland, Serbia & Montenegro, Slovakia, Spain, Sweden, Switzerland, Turkey, Ukraine |
| CONMEBOL | 10 | Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela |
| AFC | 7 | Australia, China, Iran, Japan, Saudi Arabia, South Korea, UAE |
| CAF | 5 | Egypt, Ivory Coast, Morocco, South Africa, Tunisia |
| CONCACAF | 3 | Costa Rica, Mexico, USA |

UEFA: *U*nion of *E*uropean *F*ootball *A*ssociations
CONMEBOL: *Con*federación Sud*a*mericana de Fút*bol*
AFC: *A*sian *F*ootball *C*onfederation
CAF: *C*onfédération *A*fricaine de *F*ootball
CONCACAF: *Co*nfederation of *N*orth, *C*entral *A*merican and *C*aribbean *A*ssociation *F*ootball

### 3.2. Statistical analyses

The data were analyzed univariately using descriptive statistics and scatter plots. Next, rather than relying on bivariate correlation coefficients, interrelations between %SD, GPG, GAPG, and PPG were assessed using Bayesian networks. Further insight into the prediction of %SD using GPG, GAPG, and PPG was obtained by applying LS-SVM regression and Bayesian multilayer perceptrons using a Markov chain Monte Carlo (MCMC) method.

#### 3.2.1. Bayesian networks

Bayesian networks (Pearl, 1988, Heckerman et al., 1995) are graphical probabilistic models that have two elements: the graph showing associations or dependencies between several variables, and the local probability distributions for the variables included. The first part visualizes which variables are related and which are (conditionally) unrelated, the second part explains exactly how the variables are related. In the present work, Bayesian networks are used to obtain the first part. The second part is replaced by regression analyses based on the resulting network. A main advantage of Bayesian networks is that it analyzes associations among several variables simultaneously. Correlation coefficients and scatter plots, for example, usually deal with no more than two variables at once. The graph connects the related variables with an arrow indicating the direction in which to 'run through the network'. These arrows do not have to be interpreted as representing causal relationships, however. Since this is a Bayesian analysis tool, we can provide prior expert information to the network. This prior information deals with both the association graph and the local probability distributions. Since we were interested in the story the data has to tell us about associations between variables, we used so-called non-informative or weak prior information to indicate that we have no clear prior knowledge or beliefs and that we want to see which associations are revealed by the data collected. For example, the prior network we used was the empty network in which no arrow between any two variables was drawn. In this way, we refrained from suggesting any association a priori and we obtained a network based on the data.

The Bayesian network is fitted using the `deal` package for R (Bøttcher and Dethlefsen, 2003). When fitting the Bayesian network, first the association graph is constructed using a heuristic strategy. Often, researchers opt for a *greedy search with random restarts* (Heckerman et al., 1995). For datasets with few variables to be modeled, like the dataset in this study, it is equally feasible to test each possible starting configuration because the likelihood modularity of Bayesian networks allows for local model selection (Sebastiani et al., 2005). A network score is computed that represents the fit between an initial association graph and the data (possibly combined with prior knowledge). Next, minimal changes are made to the initial graph and the change that most increases the network score is retained; this process is repeated until the network score is not improved anymore. Since this strategy may yield good yet suboptimal association graphs, it is run several times, each time with a different initial graph. The graph with the highest network score is selected as the final association graph.

#### 3.2.2. Least squares support vector machines (LS-SVM) for regression

LS-SVMs (Suykens and Vandewalle, 1999, Suykens et al., 2002) are flexible methods that are well able to discover nonlinear regression functions. Let $k$ and $l$ refer to the $k^{th}$ and $l^{th}$ of $N$ training cases and let their input data be denoted as $\mathbf{x}_k$ and $\mathbf{x}_l$. Their outcome value is denoted as $y_k$ and $y_l$. In LS-SVMs for regression, nonlinear regression functions can be obtained by first transforming the input space of dimension $p$ into a feature space of dimension $q$ using the mapping $\varphi(\cdot) = \mathbb{R}^p \rightarrow \mathbb{R}^q$. The feature space can be infinite dimensional. The regression

problem $y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b$, where $\mathbf{w}$ is the vector of length $q$ containing the weight for each feature, is then solved using the following optimization problem

$$\min_{\mathbf{w},b,\mathbf{e}} J_P(\mathbf{w},\mathbf{e}) = 0.5\left(\mathbf{w}^T\mathbf{w} + \gamma\sum_{k=1}^{N} e_k^2\right), \qquad (1)$$

where $y_k = \mathbf{w}^T\varphi(\mathbf{x}_k) + b + e_k$, and $\mathbf{e} = [e_1; \ldots; e_N]$. Minimization of the first term, $\mathbf{w}^T\mathbf{w}$, implements regularization to prevent the model from overfitting the training data by keeping the weights small. Minimization of the second term aims at minimizing the errors. Thus, the LS-SVM model tries to find a good equilibrium between simplicity and accuracy. The hyperparameter $\gamma$ is called the regularization parameter as it controls the level of regularization in the model.

Due to the constraints, the optimization problem can be reformulated by constructing the Lagrangian

$$\mathcal{L}(\mathbf{w},b,\mathbf{e},\boldsymbol{\alpha}) = J_P(\mathbf{w},\mathbf{e}) - \sum_{k=1}^{N}\alpha_k\left\{\mathbf{w}^T\varphi(\mathbf{x}_k) + b + e_k - y_k\right\}, \qquad (2)$$

where the $\alpha$s are the Lagrange multipliers, and $\boldsymbol{\alpha} = [\alpha_1; \ldots; \alpha_N]$. This results in the following linear optimization problem to be solved in $\boldsymbol{\alpha}$ and $b$:

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + I/\gamma \end{bmatrix}\begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}, \qquad (3)$$

where $\mathbf{y} = [y_1; \ldots; y_N]$, $1_v = [1; \ldots; 1]$, and $\Omega_{kl} = \varphi(\mathbf{x}_k)^T\varphi(\mathbf{x}_l) = K(\mathbf{x}_k,\mathbf{x}_l)$, and $k, l = 1, \ldots, N$. $K(\cdot,\cdot)$ is the kernel function which allows us to work in the high dimensional feature space without explicitly constructing it. This alternative formulation results in the following formulation of the regression problem:

$$y(\mathbf{x}) = \sum_{k=1}^{N}\alpha_k K(\mathbf{x},\mathbf{x}_k) + b. \qquad (4)$$

This problem is solved by a linear system as opposed to standard ($\varepsilon$-tube) SVM regression (Vapnik, 1995) that needs to be solved by a more complicated quadratic programming problem. When using a linear kernel, $K(\mathbf{x},\mathbf{x}_k) = \mathbf{x}_k^T\mathbf{x}$, the regression solution will be a hyperplane. A popular kernel function to obtain a nonlinear solution is the radial basis function (RBF) kernel, $K(\mathbf{x},\mathbf{x}_k) = \exp\left(-\|\mathbf{x}-\mathbf{x}_k\|_2^2/\sigma^2\right)$. This kernel is used in the present work. The LS-SVM regression was performed using the LS-SVM$_{lab}$ toolbox for MATLAB (Suykens et al., 2002).

The least squares approach to support vector machines for classification and regression has been applied successfully in many domains such as ovarian cancer prediction (Van Calster et al., 2007), microarray data analysis on ovarian cancer (De Smet et al., 2005), endometrial carcinoma analysis (De Smet et al., 2006), brain tumour recognition (Devos et al., 2005), mental development of preterm newborns (Ameye et al., 2002), marketing (Viaene et al., 2001), electric load forecasting (Espinoza et al., 2007), and financial time series (Van Gestel et al., 2001).

### 3.2.3. Bayesian multi-layer perceptrons using Markov chain Monte Carlo

In a feed-forward multi-layer perceptron (MLP) neural network model (Bishop, 1995) with one hidden layer, the input vector $\mathbf{x}$ is connected to the units of the hidden layer which, in turn, are connected with output unit $y$. The hidden layer allows the detection of nonlinear regression functions. The activation of a hidden layer unit consists of a weighted sum of the inputs augmented with a bias term. The unit's output is the transformation of this linear combination. We used the identity transformation. The activation of the output unit then is a weighted sum of the hidden unit outputs, augmented with a bias term. This result is again transformed using the identity transformation, as is common for regression problems. The MLP's parameter vector $\mathbf{w}$ contains all weights and bias terms of the network. In a Bayesian approach, a posterior distribution $p(\mathbf{w}|D)$ of $\mathbf{w}$ is obtained by normalizing the product of the prior distribution $p(\mathbf{w})$ and likelihood function $p(D|\mathbf{w})$ (Bishop, 1995). Because we expect the underlying regression function to be smooth, we generally expect the network weights to be small. Therefore, we used a Gaussian prior with mean 0 and variance $1/\alpha$. Further, we expect the error to be Gaussian with mean 0 and variance $1/\beta$, such that the likelihood function uses the sum of squared differences between the true and predicted outcomes. The error of a model, $E(\mathbf{w})$, is equal to $-\log p(\mathbf{w}|D)$ and thus to $-\log p(D|\mathbf{w}) - \log p(\mathbf{w})$. Using the prior and error distributions described above, this gives

$$E(\mathbf{w}) = \frac{\beta}{2}\sum_{n=1}^{N}\left(y(\mathbf{x}_n,\mathbf{w}) - y_n\right)^2 + \frac{\alpha}{2}\sum_{i=1}^{W} w_i^2, \qquad (5)$$

where $\alpha$ and $\beta$ are the hyperparameters, $w_i$ is the $i^{th}$ element and $W$ the last element of $\mathbf{w}$. The hyperparameters control the level of regularization in the

model. By integrating over $p(\mathbf{w}|D)$, an outcome prediction distribution $p(y|\mathbf{x},D)$ can be obtained. A single prediction is then given by the mean of this distribution.

The problem of the Bayesian approach is the derivation of the posterior distribution. We used a Markov chain Monte Carlo (MCMC) method. MCMC is used when one cannot directly sample from a desired distribution, in our case the posterior distribution of the network weights (Neal, 1996, Robert and Casella, 1999). In MCMC sampling using the basic Metropolis-Hastings algorithm (Hastings, 1970), a Markov chain of samples $\mathbf{w}_t$ ($t$ = 1, ..., $S$) is generated in which each sample is determined only by the previous one and by the proposal distribution $T(\mathbf{w}_t|\mathbf{w}_{t-1})$. Candidate samples $\mathbf{w}_t^*$ are accepted as $\mathbf{w}_t$ with a probability that is related to $E(\mathbf{w})$. If rejected, $\mathbf{w}_t$ equals $\mathbf{w}_{t-1}$. In this way, the distribution underlying the samples in the chain converges into the desired distribution $p(\mathbf{w}|D)$. Therefore, it is common practice to throw away the first set of samples (burn-in period) such that the remaining samples are drawn from the target distribution. We used 300 samples after a burn-in period of 300 samples. The samples are not independent but can still be used to compute complex integrals.

A disadvantage of Metropolis-Hastings is its random walk behavior with often clear autocorrelation. The result is that the distribution is not explored systematically leading to longer convergence times. One adjustment for overcoming these drawbacks is the hybrid Monte Carlo algorithm (HMC; Duane et al., 1987, Neal, 1996, Husmeier et al., 1999). HMC combines Metropolis-Hastings with a dynamical simulation to make use of the gradient of the posterior distribution. In this way, the Markov chain will tend to systematically explore regions with high density under the posterior. To this end, all network weights ($\mathbf{w}$) are given their own momentum ($\mathbf{p}$), and the joint distribution $p(\mathbf{w},\mathbf{p}|D)$ is defined as

$$p(\mathbf{w},\mathbf{p}\,|\,D) \propto \exp\left(-H(\mathbf{w},\mathbf{p})\right) = \exp\left(-\left(E(\mathbf{w}) + K(\mathbf{p})\right)\right), \qquad (6)$$

where $H(\mathbf{q},\mathbf{p})$ is the Hamiltonian function giving the total energy of the system. It consists of the potential energy $E(\mathbf{w})$, i.e., the negative log-likelihood of the posterior, and the kinetic energy $K(\mathbf{p})$, given by $\mathbf{p}^T\mathbf{p}/2$. Since $\mathbf{p}$ and $\mathbf{w}$ are independent, the marginal distribution of equation (6) is given by

$$p(\mathbf{w}\,|\,D) \propto \exp\left(-E(\mathbf{w})\right). \qquad (7)$$

Thus, we set up a Markov chain that converges into $p(\mathbf{w},\mathbf{p}|D)$, and ignore the samples of $\mathbf{p}$ to obtain samples from the posterior. For each sample of network weights from the posterior, the network output is computed (probability of scoreless draw), and the final probability is the average probability over all samples. This approximates the Bayesian probability, which is equal to the network output integrated over the posterior distribution of network weights. The samples are generated in two steps: first, given a fixed total energy $H$, a candidate sample $\left(\mathbf{w}_t^*,\mathbf{p}_t^*\right)$ is generated using a dynamical transition with the help of the so-called leapfrog method. As in Metropolis-Hastings, the candidate is accepted with probability

$$\min\left\{1, \exp\left(H\left(\mathbf{w}_{t-1},\mathbf{p}_{t-1}\right) - H\left(\mathbf{w}_t^*,\mathbf{p}_t^*\right)\right)\right\}, \qquad (8)$$

because the leapfrog method used to approximate the dynamical transition may result in small changes in $H$. Second, the value of $H$ is changed by updating the momentum $\mathbf{p}$. The update consists of a weighted sum of the old value and a random component (rather than generating a random new value for $\mathbf{p}$). As such, the trajectory of the Markov chain will persist in the same direction. This algorithm is called HMC with persistent momentum (Horowitz, 1991, Neal, 1996). HMC for MLPs is implemented using the NETLAB toolbox for MATLAB (Nabney, 2002). We used five hidden units.

### 3.2.4. Input selection and hyperparameter tuning for the LS-SVM and HMC-MLP models

For LS-SVMs using an RBF kernel, two hyperparameters need to be estimated: the regularization parameter $\gamma$ and the kernel parameter $\sigma$. This was done using 10-fold cross-validation (10CV) on the training set. The 10CV estimation method was repeated ten times and the ten results were used to train the LS-SVM on the training set. The resulting models were applied to the validation set and the mean squared error (MSE) was computed to evaluate each model's performance. We repeated this model fitting strategy for nine different sets of inputs: GPG, GAPG, TGPG, PPG, GPG-GAPG, GPG-PPG, GAPG-PPG, TGPG-PPG, GPG-GAPG-PPG. This resulted in 90 models (ten times 10CV for nine different input selections). The final LS-SVM model was the one with lowest validation MSE. Note that the results for the first four sets of inputs served also as the univariate analysis.

For the HMC-MLP, no input selection was done. We simply used the inputs selected by the LS-SVM analyses. The result of the MLP does differ depending on different initializations of the hyperparameters and the weights such that the MLP was fitted ten times with different initializations. Again, the network with lowest validation MSE was chosen. Hyperparameters do not have to be tuned since, in a MCMC approach to MLPs, one integrates over their posterior distribution.

## 4. Results

Team-based descriptive statistics averaged over all three datasets are presented in Table 2. In that table, country-based descriptive statistics are also presented to get a complete view of the percentage scoreless draws and number of goals per game for the countries included. The team-based statistics only include data on the selected teams. The table shows no trend in %SD, from 8.6% in 2003 to 8.7% in 2005. In general, 8.7% of the games ended without goal (one per 11.5 games). The total number of goals per game decreased from 2.68 to 2.59. Thus, on average there are 0.09 goals less per game, which coincides with one goal every 11.1 games. With respect to team results, some teams manage to have more than 25% of scoreless draws in one season, with a maximum of 30%. Also, 47 teams (9.7% of all 486 teams included) had no scoreless draws. Figure 1 gives the histogram of the proportion of scoreless draws, showing the large variability in this variable.

Table 2. Descriptive statistics

| Team-based descriptives[a] | 2003 | 2004 | 2005 | All | Min | Max |
|---|---|---|---|---|---|---|
| % Scoreless draws (%SD) | 8.3% | 9.0% | 9.3% | 8.9% | 0.0% | 30.0% |
| Goals scored per game (GPG) | 1.32 | 1.28 | 1.27 | 1.29 | 0.45 | 3.00 |
| Goals allowed per game (GAPG) | 1.36 | 1.31 | 1.29 | 1.32 | 0.23 | 3.20 |
| Total goals per game (TGPG) | 2.69 | 2.59 | 2.56 | 2.61 | 1.40 | 4.23 |
| Points per game (PPG) | 1.33 | 1.34 | 1.37 | 1.34 | 0.30 | 2.77 |
| **Country-based descriptives** | **2003** | **2004** | **2005** | **All** | **Min** | **Max** |
| % Scoreless draws | 8.6% | 8.9% | 8.7% | 8.7% | 2.3% | 23.3% |
| Total goals per game | 2.68 | 2.62 | 2.59 | 2.63 | 1.57 | 3.54 |

[a] descriptive statistics for the teams that are included in our data sets.

With respect to country results, the lowest percentage of scoreless draws was observed for Saudi Arabia (2.3% in the 2004/05 season, one per 34.5 games) while the highest percentage was observed for Morocco (23.3% in the 2004/05 season, one per 4.3 games). Table 3 shows additional country-based data on %SD per association. These tables show that there is ample variation in %SD between seasons and countries. Combining the data from all three seasons, the five countries with the highest %SD were Morocco (21.8%), Tunisia (14.5), Iran (13.7), France (12.4), and South Korea (12.3). Sixteen countries had at least one in ten games ending in 0-0. The five countries with the lowest %SD were Saudi Arabia (4.3), United Arab Emirates (4.8), Chile (5.0), Scotland (5.1), and Switzerland (5.6). The high percentage of scoreless draws in Africa is mainly due to Morocco. No consistent shift in %SD was observed over time.

Figure 1. Histogram of the proportion of scoreless draws (486 teams included)

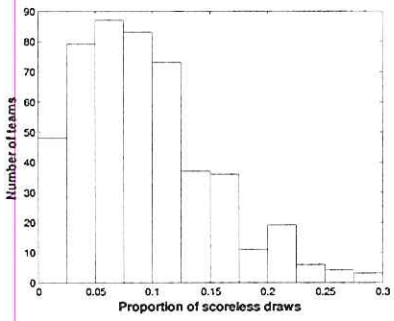



Table 3. Additional country-based descriptive statistics per association

| Association | %SD | Range | %SD | Range | %SD | Range | %SD | Gms /SD |
|---|---|---|---|---|---|---|---|---|
| UEFA | 8.3 | 3.8-12.9 | 8.4 | 4.5–14.2 | 8.8 | 4.9–14.2 | 8.5 | 11.8 |
| CONMEBOL | 7.6 | 2.9–13.9 | 8.1 | 4.2–11.9 | 7.7 | 5.3–9.5 | 7.8 | 12.8 |
| AFC | 9.3 | 5.0–13.7 | 8.7 | 2.3–16.0 | 8.4 | 5.3–13.3 | 8.8 | 11.3 |
| CAF | 12.7 | 6.6–22.5 | 15.0 | 9.3–23.3 | 13.4 | 8.8–19.6 | 13.7 | 7.3 |
| CONCACAF | 7.6 | 7.1–8.3 | 6.3 | 3.9–8.7 | 6.7 | 4.7–7.5 | 6.9 | 14.5 |

Gms/SD: Average number of games per scoreless draw

Figure 2 shows the scatter plots for the training set. The plots show comparably clear relationships between both GPG and GAPG and %SD. The relation between TGPG and %SD appeared particularly strong. Interestingly, the relationship between PPG and %SD is indeed curvilinear with average teams having more scoreless draws. Multivariate analysis, incorporating goal scoring information, will show whether this relation is present irrespective of goals scored or allowed.

Figure 2. Scatter plots of %SD with GPG (panel A), GAPG (B), TGPG (C), and PPG (D), with a scatter plot smoother superimposed

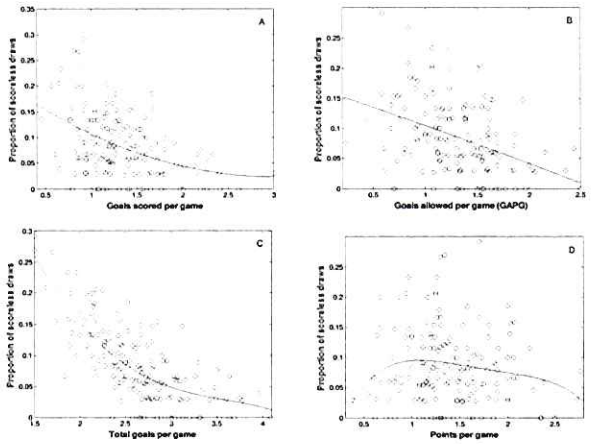
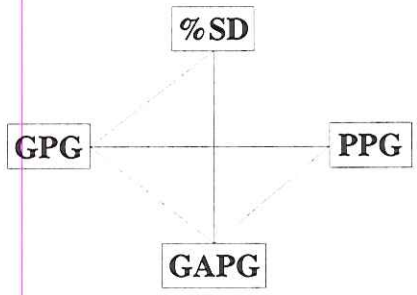



Table 4 gives the bivariate Pearson correlation coefficients for all variables as obtained on the training set. TGPG is most strongly associated with %SD ($r$ = -0.70), GPG and GAPG have an equally strong association with %SD ($r$ = -0.44 and -0.40), and PPG has no linear association with %SD. A Bayesian network was fitted to investigate the associations when looking at all variables simultaneously (Figure 3). PPG was fed to the network once as a continuous variable and once as a categorical variable, with similar results. All variables seem to be interrelated, except for PPG and %SD, for which no unconditional association was suggested. This means that if GPG and GAPG are known, PPG is not suggested to further contribute strongly in predicting the percentage of scoreless draws.

Table 4. Pearson correlation coefficients for the training set ($n$ = 162)

| | %SD | GPG | GAPG | TGPG | PPG |
|---|---|---|---|---|---|
| **%SD** | | | | | |
| **GPG** | -.44 | | | | |
| **GAPG** | -.40 | -.29 | | | |
| **TGPG** | -.70 | .65 | .54 | | |
| **PPG** | -.09 | .80 | -.73 | .13 | |

Figure 3. Bayesian network for the training set




Next, we performed model selection using LS-SVMs for function estimation (Table 5). One of the models using TGPG and PPG had the lowest validation MSE and was thus selected as the final LS-SVM model. This analysis suggests that a small relationship (undetected by the Bayesian network) between PPG and %SD exists that is able to improve the prediction of the percentage of scoreless draws of a team over a season. When applying the final model to the test set, the MSE was 0.001619. The mean absolute deviation of the predictions from the observed percentage of scoreless draws in the test set was 3.23%, and the median absolute deviation was 2.74%. The predictions are shown in Figure 4. Then, an HMC-MLP model was selected using the same two variables. The best model, as judged by the MSE on the validation set, had a test set MSE of 0.0017. For this model, the mean absolute deviation of the predictions from the observed values was 3.72% (median absolute deviation = 3.23%). The predictions of this model are shown in Figure 5. Both models clearly (and obviously) show that less

TGPG is strongly associated with an increase in scoreless draws. Interestingly, it is suggested that average teams and to some extent also less-than-average teams more often have scoreless draws, a fortiori when TGPG is small.

Table 5. Model selection using LS-SVMs by means of a validation set (the set of input variables with best performance is italicized)

| Input variables | Best MSE Validation set | Average MSE Validation set |
|---|---|---|
| GPG | .002927 | .003013 |
| GAPG | .003462 | .003472 |
| TGPG | .002373 | .002393 |
| PPG | .003513 | .003569 |
| GPG – GAPG | .002341 | .002353 |
| GPG – PPG | .002395 | .002417 |
| GAPG – PPG | .002592 | .002612 |
| *TGPG – PPG* | *.002331* | *.002348* |
| GPG – GAPG – PPG | .002353 | .002367 |

Figure 4. Predicted proportion of scoreless draws based on TGPG and PPG as estimated by the LS-SVM model (training data are shown as white crosses)

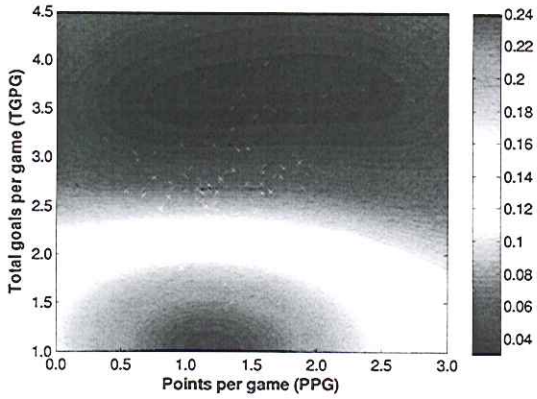



Figure 5. Predicted proportion of scoreless draws based on TGPG and PPG as estimated by the HMC-MLP (training data are shown as white crosses)

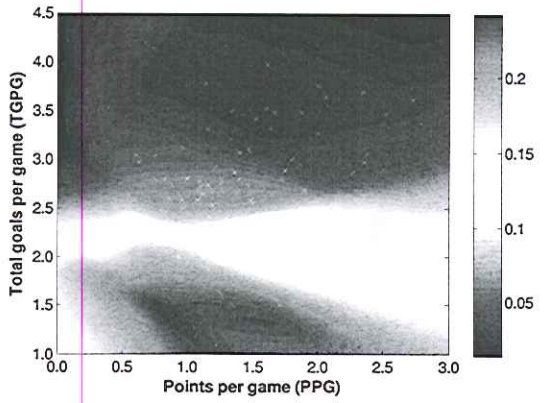
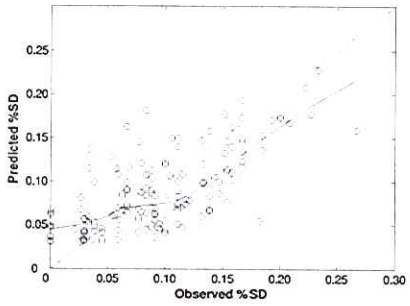



Figure 6 presents the scatter plot of predicted and observed values for the proportion of scoreless draws for the LS-SVM model. The Pearson correlation is 0.71 (for the HMC-MLP it is 0.69), indicating good correspondence between prediction and observation. The robust loess scatter plot smoother that is superimposed is close to the ideal line.

As an additional analysis, we collected season-based average attendance levels for all included UEFA countries except Israel using the 'European Football Statistics' website (EFS, 2007). For each country, we averaged both their attendance level and percentage of scoreless draws over all three seasons. The Spearman correlation coefficient between attendance and the percentage of scoreless draws was -0.26. The number of spectators may thus also have an impact: the more spectators, the fewer scoreless draws. Of course, the relation might as well be reversed: the fewer scoreless draws are produced, the more people are attracted to become spectators. This correlation was not significant ($p$ = .19), possibly due to lack of power (there were only 28 countries included). A simple ordinary least squares regression (after log-transforming the attendance variable) suggested that a country with on average 2000 spectators per game may

end up with 9.1% scoreless draws (one per 11.0 games) while a country with on average 30000 spectators per game may end up with 7.7% scoreless draws (one per 13.0 games).

Figure 6. Predictions of the LS-SVM regression model; a robust loess scatter plot smoother is superimposed (full line) as well as the ideal reference line (dashed)




## 5. Discussion and conclusions

According to our data, almost 9% of soccer matches end in 0-0. There are large differences between teams (0% up to 30%) and countries (4.3% up to 21.8%).

The LS-SVM and HMC-MLP models suggest that teams with low spectacle level (few goals on either side) are prone to many scoreless draws. This is a rather trivial result: few goals result in scoreless games. The models, however, give an indication as to how strong that relationship is, and how many scoreless draws one can expect from a team having $z$ goals per game. More interestingly, the models suggest that, on top of the strong link between goals per game and scoreless draws, average and bad-to-average teams are more likely to collect goalless games than above average teams, in particular when few goals are scored in their matches. This kind of teams may end up having no less than one fifth of their games ending without goal. The LS-SVM model performed better than the HMC neural network in this application.

The fact that an increase in the proportion of scoreless draws is also noted for the bad-to-average teams may indicate that the difference between bottom-ranked and average-ranked teams is smaller than the difference between average-ranked and top-ranked teams. In other words: there may often be more top teams than bottom teams that are clearly separated from the rest.

On a social psychological note, a suggestive additional analysis on country level rather than individual team level showed that the attendance level may also have an impact. Countries with high attendance levels may have less scoreless draws than countries with low attendance levels. However, extensive team-level and country-level analyses are needed to obtain strong conclusions.

A few remarks about our analyses should be made. The proportion of scoreless draws is bounded between 0 and 1. Even though flexible models such as support vector machines and multi-layer perceptrons should be able to cope with this, we also developed models to predict the logit-transformed proportion of scoreless draws (as in Lesaffre et al., 2007). These models were similar (slightly worse) than those mentioned in this manuscript. Also, we used PPG as an index of overall relative performance of a team in its league. Of course, teams with the same PPG may rank differently in their respective leagues. However, in our data PPG has a Spearman correlation of -0.96 with a team's rank in the league (with the rank divided by the number of teams participating in the league): PPG is very strongly related to a team's relative ranking. This supports our use of PPG as an index of overall relative performance. Yet, more complex indices could be constructed.

Finally, it should be noted that we investigated scoreless draws within the framework of soccer leagues. It would also be interesting to investigate scoreless draws within soccer tournaments. For example, our result with respect to PPG could suggest that scoreless draws are more common in the knock-out stage of a tournament (1/8 finals, quarterfinals, semifinals, and finals) versus the preceding group stage: teams advancing to the knock-out stage are on average more similar in strength than teams in the group stage. Therefore, we collected data from the FIFA World Cup (five tournaments between 1990 and 2006), the European Championships (five tournaments between 1998 and 2004), the Copa America (seven tournaments between 1993 and 2007), the African Cup of Nations (nine tournaments between 1990 and 2006), and the AFC Asian Cup (five tournaments between 1992 and 2007). These 31 tournaments involved 952 games, 98 of which ended scoreless after regular playing time (10.3%, 1 per 9.7 games). There were 711 games played in the group stage of these tournaments, 66 of which ended scoreless (9.3%, 1 per 10.8 games). The remaining 241 games were played in the subsequent knock-out stage, 32 of which ended scoreless (13.3%, 1 per 7.5 games). Thus, there are 4% more knock-out games ending scoreless (95% confidence interval = -0.4% to 9.2%). Even the confidence interval's pessimistic

estimation of the difference in percentages (-0.4%) suggests that games from the knock-out stage end scoreless at least as often as games from the group stage. Note that five of the 31 finals ended without a goal (16.1%). Of course, other parameters might influence scoreless draws in tournaments because of the specific circumstances (e.g., national versus club teams, specific tournament tactics, and increased player fatigue during the tournament).

## References


Ameye L, Lu C, Lukas L, De Brabanter J, Suykens JAK, Van Huffel S, Daniels H, Naulaers G, and Devlieger H, Prediction of mental development of preterm newborns at birth time using LS-SVM, In *Proceedings of the European Symposium on Artificial Neural Networks (ESANN)*, Bruges, Belgium, Apr 2002, pp. 167-172.

Bishop CM, *Neural networks for pattern recognition*, 1995, Oxford University Press, Oxford.

Bøttcher SG and Dethlefsen C, Learning Bayesian networks with R, In Hornik K, Leisch F, and Zeileis A (eds), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria, Mar 2003. http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings.

De Smet F, Pochet NLMM, Engelen K, Van Gorp T, Van Hummelen P, Marchal K, Amant F, Timmerman D, De Moor BLR, and Vergote IB, Predicting the clinical behavior of ovarian cancer from gene expression profiles, *International Journal of Gynecological Cancer*, 2005, 15, 1-5.

–, De Brabanter J, Konstantinovic ML, Pochet N, Van den Bosch T, Moerman P, De Moor B, Vergote I, and Timmerman D, New models to predict depth of infiltration in endometrial carcinoma based on transvaginal ultrasonography, *Ultrasound in Obstetrics and Gynecology*, 2006, 27, 664-671.

Devos A, Simonetti AW, van der Graaf M, Lukas L, Suykens JAK, Vanhamme L, Buydens LMC, Heerschap A, and Van Huffel S, The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification, *Journal of Magnetic Resonance*, 2005, 173, 218-228.

Duane S, Kennedy AD, Pendleton BJ, and Roweth D, Hybrid Monte Carlo, *Physics Letters B*, 1987, 195, 216-222.

EFS, Attendances, n.d., available on http://www.european-football-statistics.co.uk/attn.htm (accessed on June 1st, 2007).

Espinoza M, Suykens JAK, Belmans R, and De Moor B, Electric load forecasting: using kernel-based modeling for nonlinear system identification, *IEEE Control Systems Magazine*, 2007, 27, 43-56.

Fernandez-Cantelli E and Meeden G, An improved award system for soccer, *Chance*, 2002, 16, 23-29.

Gaviria A, Is soccer dying? A time series approach, *Applied Economics Letters*, 2000, 7, 275-278.

Hastings WK, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 1970, 57, 97-109.

Heckerman D, Geiger D, and Chickering D, Learning Bayesian networks: the combination of knowledge and statistical data, *Machine Learning*, 1995, 20, 197-243.

Hirotsu N and Wright M, Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions, *Journal of the Operational Research Society*, 2002, 53, 88-96.

– and –, Determining the best strategy for changing the configuration of a football team, *Journal of the Operational Research Society*, 2003, 54, 878-887.

– and –, Modeling tactical changes of formation in association football as a zero-sum game, *Journal of the Quantitative Analysis in Sports*, 2006, 2 (2), article 4.

Horowitz AM, A generalized guided Monte Carlo algorithm, *Physics Letters B*, 1991, 268, 247-252.

Husmeier D, Penny WD, and Roberts SJ, An empirical evaluation of Bayesian sampling with hybrid Monte Carlo for training neural network classifiers, *Neural Networks*, 1999, 12, 677-705.

Jordet G, Hartman E, Visscher C, and Lemmink KAPM, Kicks from the penalty mark in soccer: the roles of stress, skill, and fatigue for kick outcomes, *Journal of Sport Sciences*, 2007, 25, 121-129.

Lesaffre E, Rizopoulos D, and Tsonaka R, The logistic transform for bounded outcome scores, *Biostatistics*, 2007, 8, 72-85.

Nabney IT, *NETLAB: algorithms for pattern recognition*, 2002, Springer, London.

Neal RM,. *Bayesian learning for neural networks* (Lecture notes in statistics, Volume 118), 1996, Springer, New York.

Pearl J, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, 1988, Morgan Kaufmann, San Mateo.

Pollard R and Reep C, Measuring the effectiveness of playing strategies at soccer, *Journal of the Royal Statistical Society – Series D: the Statistician*, 1997, 46, 541-550.

Ridder G, Cramer JS, and Hopstaken P, Down to ten: estimating the effect of a red card in soccer, *Journal of the American Statistical Association*, 1994, 89, 1124-1127.

Robert CP and Casella G, *Monte Carlo statistical methods*, 1999, Springer, New York.

RSSSF, The introduction page of the RSSSF – The Rec.Sport.Soccer Statistics Foundation, n.d., available on www.rsssf.com (accessed on February 16th, 2007).

Rue H and Salvesen Ø, Prediction and retrospective analysis of soccer matches in a league, *Journal of the Royal Statistical Society – Series D: the Statistician*, 2000, 49, 399-418.

Sebastiani P, Abad MM, and Ramoni MF, Bayesian networks, In Maimon O and Rokach L (eds), *Data Mining and Knowledge Discovery Handbook: a complete guide for research scientists and practicioners*, 2005, Springer, New York, pp. 193-230.

Suykens JAK and Vandewalle J, Least squares support vector machine classifiers, *Neural Processing Letters*, 1999, 9, 293-300.

–, Van Gestel T, De Brabanter J, De Moor B, and Vandewalle J, *Least squares support vector machines*, 2002, World Scientific, Singapore.

Van Calster B, Timmerman D, Lu C, Suykens JAK, Valentin L, Van Holsbeke C, Amant F, Vergote I, and Van Huffel S, Preoperative diagnosis of ovarian tumors using Bayesian kernel-based methods, *Ultrasound in Obstetrics and Gynecology*, 2007, 29, 496-504.

Van Gestel T, Suykens J, Baestaens D, Lambrechts A, Lanckriet G, Vandaele B, De Moor B, and Vandewalle J, Financial time series prediction using least squares support vector machines within the evidence framework, *IEEE Transactions on Neural Networks*, 2001, 12, 809-821.

Vapnik V, *The nature of statistical learning*, 1995, Springer, New York.

Viaene S, Baesens B, Van Gestel T, Suykens J, Van den Poel D, Vanthienen J, De Moor B, and Dedene G, Knowledge discovery in a direct marketing case using least squares support vector machines, *International Journal of Intelligent Systems*, 2001, 16, 1023-1036.

Wright M and Hirotsu N, The professional foul in football: tactics and deterrents, *Journal of the Operational Research Society*, 2003, 54, 213-221.