

# pi-football: A Bayesian network model for forecasting Association Football match outcomes

ANTHONY C. CONSTANTINOU\*, NORMAN E. FENTON AND MARTIN NEIL

Risk Information Management (RIM) Research Group,  
Department of Electronic Engineering and Computer Science,  
Queen Mary University of London, UK, E1 4NS

## **THIS IS A PRE-PUBLICATION DRAFT OF THE FOLLOWING CITATION:**

Constantinou, A. C., Fenton, Norman E. & Neil, M. (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, 36: 322, 339.

DOI: [10.1016/j.knosys.2012.07.008](https://doi.org/10.1016/j.knosys.2012.07.008)

CONTACT: Dr. Anthony Constantinou, [anthony@constantinou.info](mailto:anthony@constantinou.info), [a.constantinou@qmul.ac.uk](mailto:a.constantinou@qmul.ac.uk)

## ABSTRACT

A Bayesian network is a graphical probabilistic model that represents the conditional dependencies among uncertain variables, which can be both objective and subjective. We present a Bayesian network model for forecasting Association Football matches in which the subjective variables represent the factors that are important for prediction but which historical data fails to capture. The model (pi-football) was used to generate forecasts about the outcomes of the English Premier League (EPL) matches during season 2010/11 (but is easily extended to any football league). Forecasts were published online prior to the start of each match. We show that:

- a) using an appropriate measure of forecast accuracy, the subjective information improved the model such that posterior forecasts were on par with bookmakers' performance;
- b) using a standard profitability measure with discrepancy levels at  $\geq 5\%$ , the model generates profit under maximum, mean, and common bookmakers' odds, even allowing for the bookmakers' built-in profit margin.

Hence, compared with other published football forecast models, pi-football not only appears to be exceptionally accurate, but it can also be used to 'beat the bookies'.

*Keywords:* Bayesian reasoning, football betting, football predictions, soccer predictions, subjective information

## 1 INTRODUCTION

Association Football (hereafter referred to simply as 'football') is the world's most popular sport (Dunning & Joseph A. M., 1993; Mueller et al., 1996; Dunning E., 1999), and constitutes the fastest growing gambling market (Constantinou & Fenton, 2012b). As a result, researchers continue to introduce a variety of football models which are formulated by diverse forecast methodologies. While some of these focus on predicting tournament

---

\* Corresponding author. E-mail addresses: [anthony@constantinou.info](mailto:anthony@constantinou.info) (Constantinou, A. C.), [norman@eecs.qmul.ac.uk](mailto:norman@eecs.qmul.ac.uk) (Fenton, N. E.), [martin.neil@eecs.qmul.ac.uk](mailto:martin.neil@eecs.qmul.ac.uk) (Neil, M.)

outcomes (Kuonen, 1996; Buchner, et al., 1997; Koning et al., 2003; Halicioglu, 2005a; Halicioglu, 2005b) or league positions (Koning, 2000), our interest is in predicting outcomes of individual matches.

A common approach is the Poisson distribution goal-based data analysis whereby match results are generated by the attack and defence parameters of the two competing teams (Maher, 1982; Dixon & Coles, 1997, Lee 1997; Karlis & Ntzoufras, 2003). A similar version is also reported in (Dixon & Pope, 2004) where the authors demonstrate profitability against the market only at very high levels of discrepancy, but which relies on small quantities of bets against an unspecified bookmaker. A time-varying Poisson distribution version was proposed by (Rue & Salvesen, 2000) in which the authors demonstrate profitability against Intertops (a bookmaker located in Antigua, West Indies), and refinements of this technique were later proposed in (Crowder et al., 2002) which allow for a computationally less demanding model.

In contrast to the Poisson models that predict the number of goals scored and conceded, all other models restrict their predictions to match result, i.e. win, draw, or lose. Typically these are ordered probit regression models that consist of different explanatory variables. For example, (Kuypers, 2000) considered team performance data as well as published bookmakers' odds, whereas (Goddard & Asimakopoulos, 2004; Forrest et al., 2005) considered team quality, recent performance, match significance and geographical distance. (Goddard, 2005) compared goal-driven models with models that only consider match results and concluded that both versions generate similar predictions.

Techniques from the field of machine learning have also been proposed for prediction. In (Tsakonas et. al., 2002) the authors claimed that a genetic programming based technique was superior in predicting football outcomes to other two methods based on fuzzy models and neural networks. More recently, (Rotshtein et al., 2005) claimed that acceptable match simulation results can be obtained by tuning fuzzy rules using parameters of fuzzy-term membership functions and rule weights by a combination of genetic and neural optimisation techniques.

Models based on team quality ratings have also been considered, but they do not appear to have been extensively evaluated. Knorr-Held (2000) used a dynamic cumulative link model to generate ratings for top division football teams in Germany. The ELO rating that was initially developed for assessing the strength of chess players (Elo, 1978) has been adopted to football (Buchdahl, 2003). In (Hvattum & Arntzen, 2010) the authors used the ELO rating for match predictions and concluded that the ratings appeared to be useful in encoding the information of past results for measuring the strength of a team, but the forecasts generated were not on par with market odds. (Leitner et al., 2010) have also assessed an ELO rating based model along with the FIFA/Cocal Cola World rating model and concluded that both were inferior against bookmakers' forecasts for EURO 2008.

Numerous studies have considered the impact of specific factors on match outcome. These factors include: home advantage (Hirotsu & Wright, 2003), ball possession (Hirotsu & Wright, 2003), and red cards (Ridder et al., 1994; Vecer et al., 2009)<sup>†</sup>

---

<sup>†</sup> While this work falls within the scope of our interest, other empirical forecasting studies such as attendance demand (Peel & Thomas, 1989; Peel & Thomas, 1992; Peel & Thomas, 1997; Falter & Perignon, 2000; Forrest & Simmons, 2002), and the effectiveness of football tipsters (Forrest & Simmons, 2000) do not.

Recently researchers have considered Bayesian networks and subjective information for football match predictions. In particular, (Joseph et. al., 2006) demonstrated the importance of supplementing data with expert judgement by showing that an expert constructed Bayesian network model was more accurate in generating football match forecasts for matches involving Tottenham Hotspurt than machine learners of MC4, naive Bayes, Bayesian learning and K-nearest neighbour. A model that combined a Bayesian network along with a rule-based reasoner appeared to provide reasonable World Cup forecasts in (Min et al., 2008) through simulating various predefined strategies along with subjective information, whereas in (Baio & Blangiardo, 2010) a hierarchical Bayesian network model that did not incorporate subjective judgments appeared to be inferior in predicting football results when compared to standard Poisson distribution models.

In this paper we present a new Bayesian network model for forecasting the outcomes of football matches in the distribution form of  $\{p(H), p(D), p(A)\}$ ; corresponding to home win, draw and away win. We believe this study is important for the following reasons:

- a) the model is profitable under maximum, mean and common bookmakers' odds, even by allowing for the bookmakers' introduced profit margin;
- b) the model priors are dependent on statistics derived from predetermined scales of team-strength, rather than statistics derived from a particular team (hence enabling us to maximise historical data);
- c) the model enables us to revise forecasts from objective data, by incorporating subjective information for important factors that are not captured in the historical data;
- d) the significance of recent information (objective or subjective) is weighted using degrees of uncertainty resulting in a non-symmetric Bayesian parameter learning procedure;
- e) forecasts were published online before the start of each match (pi-football, 2010);
- f) although the model has so far been applied for one league (the English Premier League) it is easily applicable to any other football league.

The paper is organised as follows: section 2 describes the historical data and method used to inform the model priors, section 3 describes the Bayesian network model, section 4 describes the assessment methods and section 5 provides our concluding remarks and future work.

## 2 DATA

The basic data used to inform the priors for the model were the results (home, draw or away) of all English Premier League (EPL) matches from season 1993/94 to 2009/10 inclusive (a total of 6244 occurrences). This information is available online at (Football-Data). The forecasts generated by the model were for season 2010/11, a total of 380 EPL matches.

In contrast to previous approaches we use the historical data to generate prior forecasts that are 'anonymous' by using predetermined levels of team-strength, rather than distinct team-names. We achieve this by replacing each team-name in each match in the database with a ranked number that represents the strength of that particular team for a particular season. The team-strength number is derived from the total number of points<sup>‡</sup> that the particular team achieved during that particular season as shown in Table 1.

Table 1. Predetermined levels of team strength

| Total points | >84 | 80-84 | 75-79 | 70-74 | 65-69 | ...(intervals of 5 points) | 30-34 | 25-29 | <25 |
|--------------|-----|-------|-------|-------|-------|----------------------------|-------|-------|-----|
| Strength     | 1   | 2     | 3     | 4     | 5     | ...                        | 12    | 13    | 14  |

This implies that the same team may receive different ranks for different seasons and that different teams may receive identical ranks within the same season.

For example, the Manchester City at home to Aston Villa match in season 2006-07 is classified as a ranked 10 versus a ranked 8 team (because in that season Manchester City totalled 42 points and Aston Villa 50 points), whereas in season 2009-10 the Manchester City at home to Aston Villa match is classified as a ranked 5 versus a ranked 6 team (because in that season Manchester City totalled 67 points and Aston Villa 64 points).

The granularity (of 14 levels of team strength) has been chosen to ensure that for any match combination (i.e. a team of strength  $x$  at home to a team of strength  $y$ ) there are sufficient data points for a reasonably well informed prior for  $\{p(H), p(D), p(A)\}$ . This approach has a number of important advantages:

- a) it enables us to make maximum use of limited data and be able to deal with the fact that every season the set of 20 teams changes (three are relegated and three new teams are promoted). For example, forecasts for teams for which there is little or no historical data (such as those recently promoted) are based on data for different teams but of similar strength;
- b) historical observations do not have to be ignored or weighted since the challenge here is to estimate a team's current strength and *learn* how such a team performed in the past given the specified ground (home/away) and opponent's strength. For example, consider the prior for the Manchester City at home to Aston Villa match in season 2010-11. Because the historical performances of Manchester City and Aston Villa prior to season 2010-11 were in no way representative of their strength in season 2010-11, what matters is not the results of previous matches between Manchester City and Aston Villa (which would be sparse as well as irrelevant), but the results of all previous matches where a rank 4 team played at home to a rank 9 team.
- c) historical observations do not necessarily require weekly updating. The database already consists of thousands of historical match observations, and

---

<sup>‡</sup> In EPL a total of 20 football teams participate and thus, a team can accumulate a minimum of 0 and a maximum of 114 points

adding a few more matches every week will not make a major difference (this can be done once a year).

- d) historical observations from one league can be used to predict match results for teams in another league (as long as the introduced ranking is redefined to accommodate potential discrepancies in the number of teams participating within that league);

### 3 THE MODEL

The model, which we call 'pi-football' (v1.32), generates predictions for a particular match by considering four generic factors for both the home and away team, namely: 1) strength, 2) form, 3) psychology and 4) fatigue. There are model components corresponding to each of the four generic factors. In this sections we describe each of the model components (with further details regarding the assumptions and the different scenarios available for each of the Bayesian network nodes provided in Appendix A), but first we provide a brief overview.

Component 1 provides an estimate of each team's current strength (based on recent data) expressed as a distribution. Using historical outcomes between such ranked teams we get a distribution for the predicted outcome as shown in Figure 1. Here we have a home team with mean strength 65-69 points (or rank 5) and an away team with mean strength 80-84 points (or rank 2). Component 1 is predominantly dependent on objective information for prediction and thus, we will refer to the resulting forecasts as 'objective forecasts'.

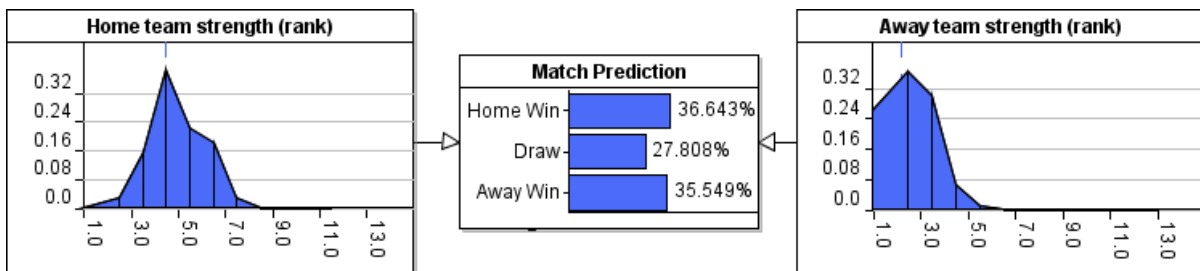


Figure 1. An example of an objective forecast generated at component 1.

Components 2, 3 and 4 are predominantly dependent on subjective information. They are used to revise the forecast from component 1. The outcome of each of the components is mutually summarised in a single value (considering both teams) which we describe as 'subjective proximity'. The subjective proximity is measured on a scale from 0 to 1. A value equal to 0.5 indicates no advantage to either of the teams; a value less than 0.5 indicates an advantage for the home team, while a value greater than 0.5 indicates an advantage for the away team. Since the forecast nodes are ranked in the sense of (Fenton et. al., 2007), the Bayesian Network software we have used (Agena, 2012) automatically updates the forecast taking account of the subjective proximity as shows for different examples in Figure 2. Figure 3 illustrates how the four components are linked. We will refer to the revised (and final) forecasts as 'subjective forecasts'.

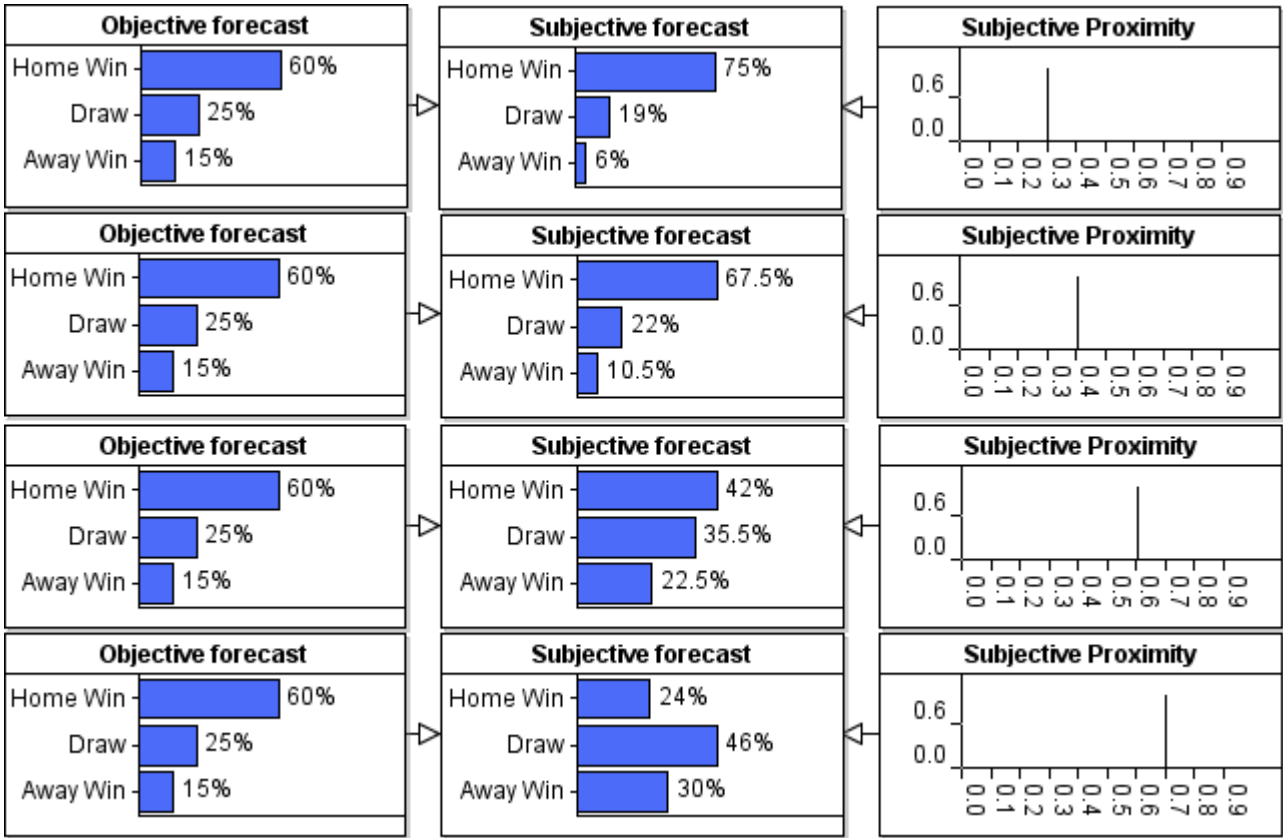


Figure 2. Forecast revision given different indications of subjective proximity.

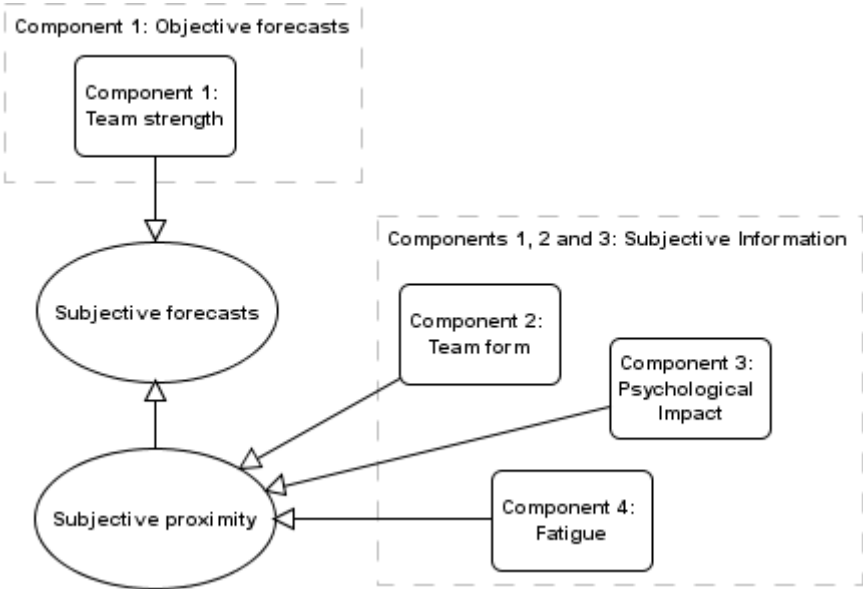


Figure 3. How components 1, 2, 3 and 4 are linked.

### 3.1. Component 1: team strength

The Bayesian network corresponding to the team strength component is shown in Figure 4 and it can be explained in terms of the following information:

- a) *Previous information*: represented by five parameters (nodes 2, 3, 4, 5, and 6), each of which holds the number of total points accumulated in each of the five previous seasons with degrees of uncertainty (higher uncertainty for older seasons);
- b) *Current information*: represented by a single parameter (node 9) that holds an estimate about the strength of the team in total points, and which is measured according to the total points accumulated during the current season and the points expected from residual matches<sup>§</sup> with degrees of uncertainty (lower uncertainty for higher number of matches played).
- c) *Subjective information (optional)*: represented by a single parameter (node 7) that holds the expert's subjective belief about the strength of the team in total points with degrees of uncertainty (reflects the expert's confidence). This information is used in cases where important changes happen before the start of the current season that cannot be captured by the historical data. A good example is Manchester City at the start of seasons 2009/10, 2010/11 and 2011/12, who dramatically improved their strength by spending £160m, £77m and £75m respectively signing some of the world's top players (Soccer Base, 2012).

The degree of uncertainty is modelled by exponential predetermined levels of variance in an attempt to achieve a limited memory process. This process produces a non-symmetric Bayesian parameter learning procedure. Accordingly,

- a) *Previous information*: this indication receives increased rates of variance (and hence become less important) for each previous season, following the exponential growth illustrated in Figure 5a;
- b) *Current information*: this indication receives decreased rates of variance (and hence become more important) after each subsequent gameweek<sup>\*\*</sup>, following the exponential decay illustrated in Figure 5b;
- c) *Subjective Information*: this indication receives decreased or increased rates of variance according to the expert's confidence regarding his indication. The decreased/increased rates of variance follow those of the *previous information*<sup>††</sup> (Figure 5a).

Further information regarding the variables and available scenarios of this process is provided in table A1. An example with observations from the actual match between Man City and Man United dated 10th of November 2010 is illustrated in Appendix B.

---

<sup>§</sup> It is important to appreciate that the resulting parameter summarises a belief about the team's strength in points and not the points the team is expected to have by the end of the proceeding season.

<sup>\*\*</sup> A complete EPL season consists of 38 gameweeks.

<sup>††</sup> For example, the degree of uncertainty when the expert's confidence is "Very Low" (fifth lowest out of five) is equal to the degree of uncertainty introduced for the points accumulated during the 5<sup>th</sup> preceding season.

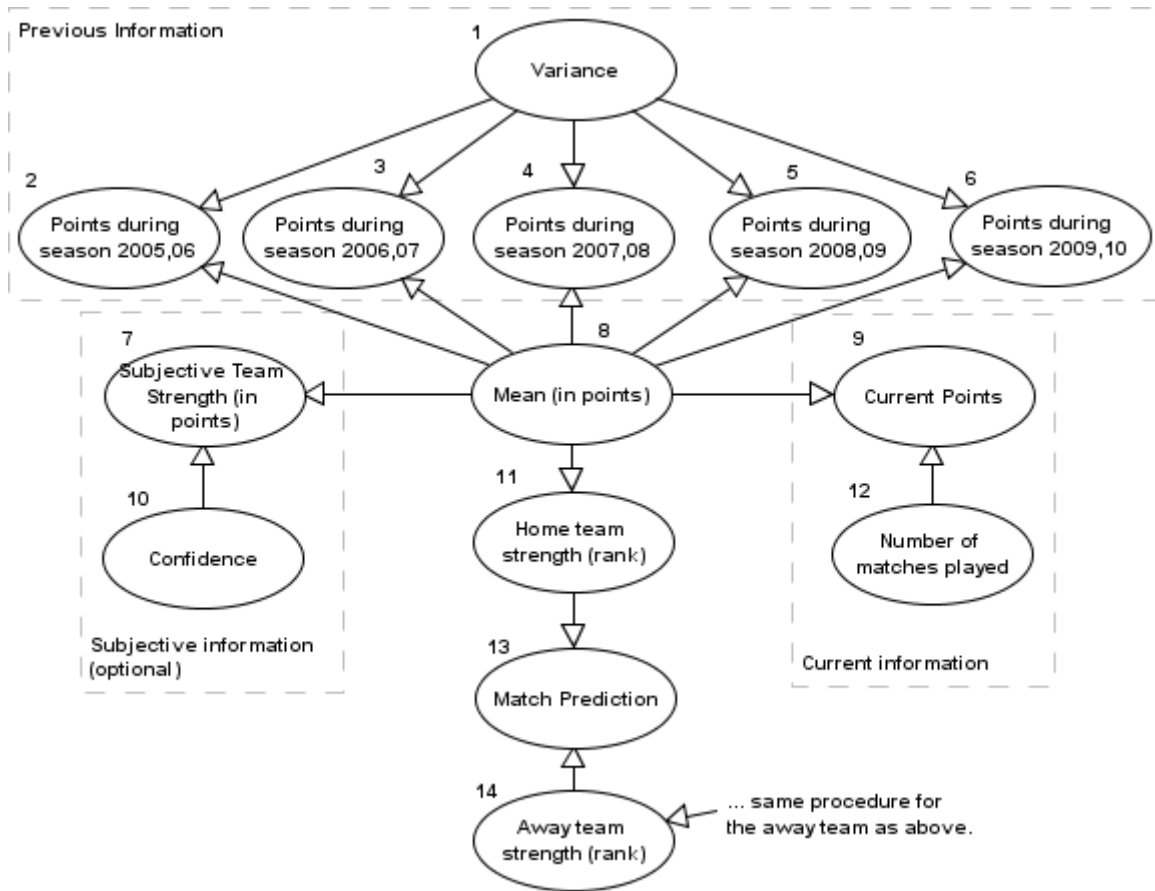


Figure 4. Component 1: Non-symmetric Bayesian parameter learning network for measuring the strength of the two teams and generating objective match predictions.

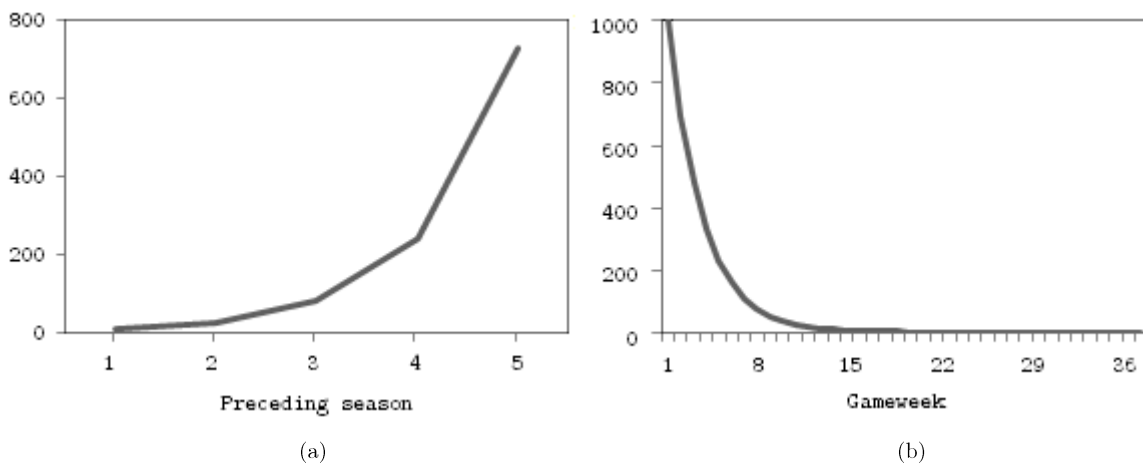


Figure 5. Limited memory process achieved by exponential growth/decay rates of uncertainty for (a) the previous seasons and (b) the gameweeks played under the current season.



### 3.2. Component 2: Team form

This Bayesian network component is shown in Figure 6. The 'form' of a team (node 10 for the home team and 12 for the away team) indicates the particular team's recent performance against expectations, and it is measured by comparing the team's expected performance<sup>‡‡</sup> against its observed performance during the five most recent gameweeks.

The form of a team is represented on a scale that goes from 0 to 1. When the value is close to 0.5 it suggests that the team is performing as expected; a higher value indicates that the team is performing better than expected. Further, if the particular team is playing at home, then the model will consider home form and away form with weights  $[2/3, 1/3]$  respectively (nodes 5, 6, 7; the reverse applies for the away team). The form is revised according to subjective indications about the availability of certain players (nodes 1, 2, 3, 4)<sup>§§</sup>. The expert constructed Bayesian network determines whether one team has an advantage over the other when comparing each other's form. Further information regarding the variables and available scenarios of this process is provided in table A2.



Figure 6. Component 2: Expert constructed Bayesian network for estimating potential advantages in form between the two teams.

<sup>‡‡</sup> Represented by what the model had initially forecasted.

<sup>§§</sup> Form decreases if the team has new first-team injuries and increases when important players return back to action.

### 3.3. Component 3: Psychological impact

This Bayesian network component is shown in Figure 7. The psychology of a team is determined by subjective indications regarding motivation, team spirit, managerial issues and potential head-to-head biases. The Bayesian network estimates the difference in psychological impact between the two teams. This process is divided into two levels; where the information assessed during level 1 (node 6) is updated at level 2 (node 7). This implies that the total information of level 1 (nodes 1, 2) shares identical impact with that of level 2 (node 4). Further information regarding the variables and available scenarios of this process is provided in table A3.

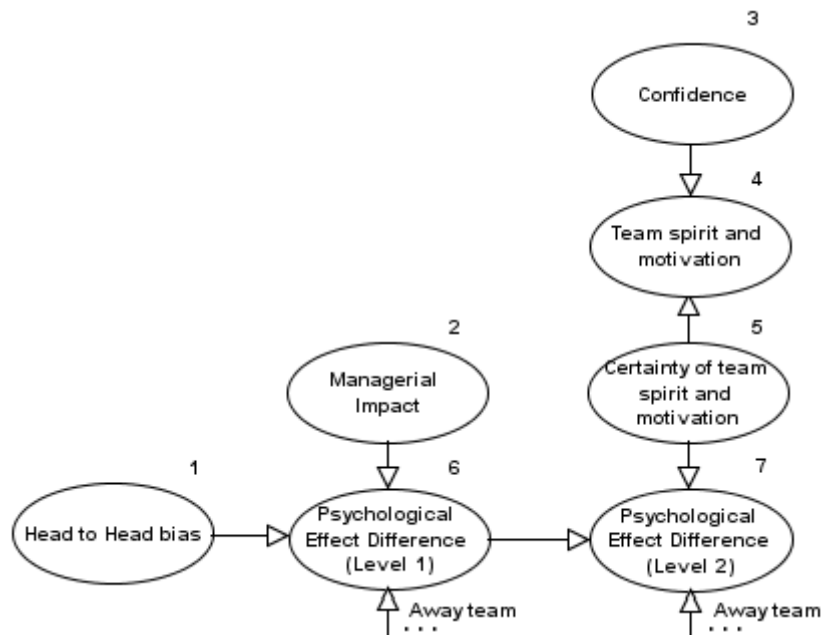


Figure 7. Component 3: Expert constructed Bayesian network for estimating potential advantages in psychological impact between the two teams.

### 3.4. Component 4: Fatigue

This Bayesian network component is shown in Figure 8. The fatigue of a team is determined by the toughness of the previous match, the number of days gap since that match, the number of first team players rested (if any), and the participation of first team players in national team matches (if any). The Bayesian network estimates the difference in the level of fatigue between the two teams. In particular, the resulting tiredness, which is determined according to the toughness of the previous match (node 5), is diminished according to a) the number of days gap since the last match (node 1), and b) the number of first-team players rested during that match<sup>\*\*\*</sup> (node 2). Further, the indication of fatigue may increase up to 50% towards its maximum value depending on the level of participation of first team players in additional matches with their national team<sup>†††</sup> (nodes 6, 7). If there is no national team

<sup>\*\*\*</sup> Where (a) is defined to be twice as important to (b) when calculating 'Restness' (node 3).

<sup>†††</sup> When football teams are given a break due to national matches, top level teams (e.g. Man United) might suffer greater levels of fatigue due to having many players who are first-team regulars with their national team.

participation the fatigue will receive no increase. Further information regarding the variables and available scenarios of this process is provided in table A4.

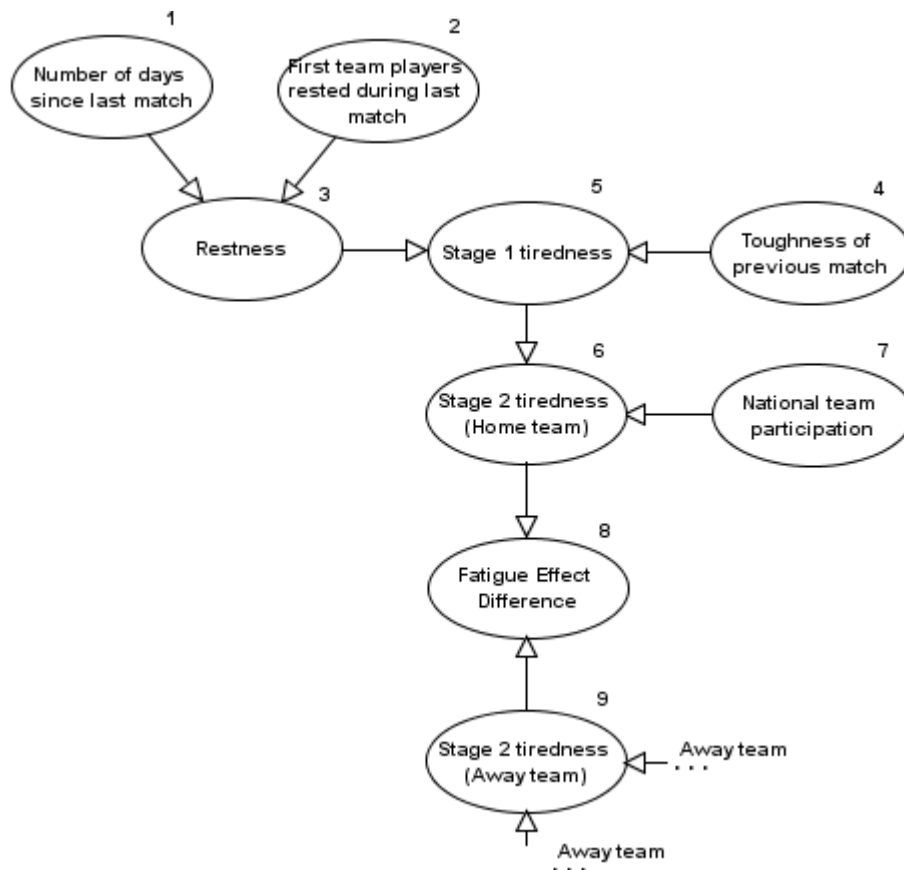


Figure 8. Component 4: Expert constructed Bayesian network for estimating potential advantages in fatigue between the two teams.

## 4 RESULTS & DISCUSSION

There are various ways in which the quality of a forecast model can be assessed. In particular, we can consider *accuracy* (how close the forecasts are to actual results) and *profitability* (how useful the forecasts are when used as the basis of a betting strategy). Researchers have already concluded that there is only a weak relationship between commonly used measures of accuracy and profitability (Leitch & Tanner, 1991) and that a combination of the two might be best (Wing et. al., 2007). Hence we use assessments of both accuracy (Section 4.1) and profitability (Section 4.2) in order to get a more informative picture about the performance of pi-football, whereas Section 4.3 provides an analysis of impact of the subjective components of the model based on the two measures.

### 4.1. Accuracy Measurement

For assessing the accuracy of the forecasts we use of the Rank Probability Score (RPS), a scoring rule introduced in 1969 (Epstein), and which has been described to be particularly appropriate in assessing both interval and ordinal scale probabilistic variables

(Murphy, 1970). We explained why it was the most rational scoring rule of those that have been proposed and used for football outcomes in (Constantinou & Fenton, 2012a). In general, this scoring rule represents the difference between the observed and forecasted cumulative distributions in which a higher difference leads to a higher penalty (Wilks, 1995), which is subject to a negative bias that is strongest for small ensemble size (Jolliffe & Stephenson, 2003). RPS is both strictly proper and sensitive to distance (Murphy, 1969; Murphy, 1970). For a single forecast the RPS is defined as

$$RPS = \frac{1}{r-1} \sum_{i=1}^{r-1} \left( \sum_{j=1}^i (p_j - e_j) \right)^2$$

where  $r$  is the number of potential outcomes, and  $p_j$  and  $e_j$  are the forecasts and observed outcomes at position  $j$ . A lower score indicates a more accurate forecast (lower error).

To determine the accuracy of our model we compute the RPS for the following three forecasts:

- a) the objective forecasts generated at component 1; we will refer to these forecasts as  $f_o$ ;
- b) the subjective (revised) forecasts after considering components 2, 3 and 4; we will refer to these forecasts as  $f_s$ ;
- c) the respective normalised<sup>†††</sup> bookmakers' forecasts; we will refer to these forecasts as  $f_B$ .

Other studies have concluded that the normalised odds of one bookmaker are representative of any other bookmaker (Dixon & Pope, 2004; Forrest et al., 2005; Constantinou & Fenton, 2012b). However, instead of selecting a single bookmaker we make use of the mean<sup>§§§</sup> bookmakers' odds as provided by (Football-Data). Figure C1 demonstrates the RPS generated per forecast under the three datasets.

Figure 9 presents the cumulative RPS difference for a)  $f_B - f_o$ , b)  $f_B - f_s$ , and c)  $f_o - f_s$ . Since a higher RPS value indicates a higher error a cumulative difference for  $A - B$  below 0 indicates that  $A$  is more accurate than  $B$ . Accordingly, the graphs suggest that the accuracy of pi-football improves after considering subjective information. However, the bookmakers appear to have a higher overall accuracy even after the forecasts are revised. We performed 2-tailed paired  $t$ -tests to determine the importance of the above discrepancies. The null hypothesis is that the two datasets are represented by similar forecasts. The results are:

- a) the dataset  $f_o$  is statistically significant to that of  $f_B$  at 99% confidence interval with a  $p$ -value of 0.0023; therefore, the null hypothesis is rejected;

---

<sup>†††</sup> The bookmakers' odds are normalised such so that the sum of probabilities over the possible events is equal to 1 (the introduced profit margin is eliminated). For more information see (Constantinou & Fenton, 2012b).

<sup>§§§</sup> The mean odds are measured by considering a minimum of 28 and a maximum of 40 different bookmakers per match instance (Football-Data).

- b) the dataset  $f_s$  is not statistically significant to that of  $f_B$  at 99% (not even at 90%) confidence interval with a  $p$ -value of 0.1319; therefore, the null hypothesis is accepted.

We conclude that the accuracy of objective forecasts was significantly inferior to bookmakers' forecasts, and that subjective information improved the forecasts such that they were on par with bookmakers' performance. This also suggests that the bookmakers, as in the pi-football model, make use of information that is not captured by the standard statistical football data available to the public. Further, appendix D provides evidence of significant improvements in  $f_o$  by incorporating subjective information. Table D.1. presents match instances in which  $f_o$  and  $f_s$  generate the highest RPS discrepancies, along with indications whether  $f_s$  lead to a more accurate forecast.

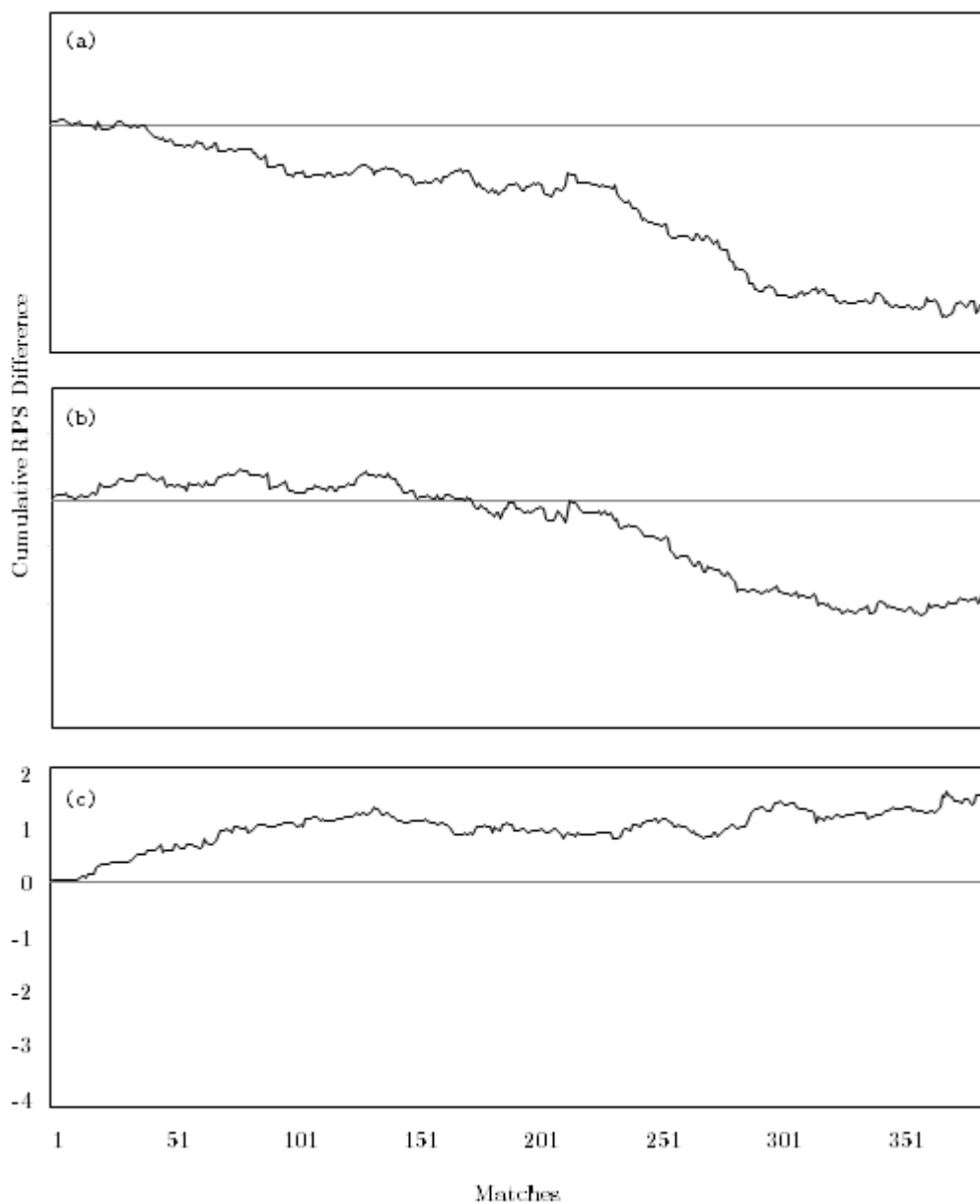


Figure 9. Cumulative RPS difference when (a)  $f_B-f_o$ , (b)  $f_B-f_s$ , (c)  $f_o-f_s$ .

## 4.2. Profitability Measurement

For assessing the profitability of the forecasts we perform a simple betting simulation which satisfies the following standard betting rule: *for each match instance, place a 1-pound bet on the outcome with the highest discrepancy, of which the pi-football model predicts with higher probability, if and only if the discrepancy is greater or equal to 5%.*

This assessment, of course, depends on the availability of an appropriate bookmaker's odds\*\*\*\*. In contrast to previous papers (Forrest & Simmons, 2002; Forrest et al., 2005), the work in (Constantinou & Fenton, 2012b) shows that the published odds of a single bookmaker are not representative of the overall market. Unlike the case of accuracy (Section 4.1) where published odds are normalised and hence the profit margin is eliminated, for profitability we have to consider the published odds (such odds are not normalised and are considered with their profit margins), hence the odds of one bookmaker can be significantly different to another. Accordingly, in determining pi-football's profitability we consider the following three different sets of bookmakers' odds†††:

- a) *the maximum* (best available for the bettor) bookmakers' odds which we are going to refer to as  $f_{maxB}$ . This dataset is used to estimate how an informed bettor, who knows how to pick the best odds by comparing the different bookmakers' odds, could have performed;
- b) *the mean* (average) bookmakers' odds which we are going to refer to as  $f_{meanB}$ . This dataset is used to estimate how an ignorant bettor could have performed, assuming he selects a bookmaker at random;
- c) *the most common* bookmakers' odds which we are going to refer to as  $f_{WH}$ . This dataset is used to estimate how the common UK bettor could have performed. For this, we consider the odds provided by the leading UK bookmaker William Hill, who represents the 25% of the total market throughout the UK and Ireland (William Hill PLC, 2012).

Figure 10 demonstrates the cumulative profit/loss generated against a)  $f_{maxB}$ , b)  $f_{meanB}$  and c)  $f_{WH}$  after each subsequent match, assuming a 1-pound stake when the betting condition is met. The model generates a profit under all of the three scenarios and the simulation almost never leads into a negative cumulative loss even allowing for the in-built bookmakers' profit margin†††. Figure 11 illustrates the *Risk of Ruin* for up to a bankroll 100 times the value of a single bet. A bankroll of ~£55 (or 55 times the value of a single bet) and ~£45 is required to ensure that the probability to lose the specified bankroll under infinite betting is  $\leq 5\%$  for  $f_{maxB}$  and  $f_{WH}$  respectively. In the case of  $f_{meanB}$  the profit rate is not high

---

\*\*\*\* See also the following studies on the football gambling market: (Pope & Peel, 1989; Dixon & Coles, 1997; Kuypers, 2000; Rue & Salvesen, 2000; Forrest & Simmons, 2001; Dixon & Pope, 2004; Goddard & Asimakopoulos, 2004; Forrest & Simmons, 2008; Graham & Stott, 2008).

††† The bookmakers' odds are also provided by (Football-Data).

†††† We have also performed the identical betting simulation given  $f_o$ . Figure E1 demonstrates how the betting simulation results in losses of -13.98% against  $f_{maxB}$ , -19.92% against  $f_{meanB}$  and -12.84% against  $f_{WH}$ . This confirms the accuracy measurement results; that is, the significant improvements in  $f_o$  (which form  $f_s$ ) by incorporating subjective information.

enough to ensure a risk of ruin  $\leq 5\%$  with a bankroll up to 100 times the value of a single bet. Table 2 summarises the statistics of the betting simulation for all of the three scenarios.

Overall, pi-football won approximately 35% of the bets simulated under all of the three scenarios, with the mean odds of winning bets at approximately 3.00. This suggests that the model was able to generate profit via longshot bets; what makes this especially interesting is that longshots are proven to be biased against the bettors (Cain et al., 2000, Forrest & Simmons, 2001; 2002; Forrest et al., 2005; Graham & Stott, 2008; Constantinou & Fenton, 2012b). This implies that the model would have generated even higher profits if the betting market was to provide unbiased odds. Additionally, profits are most likely to have been even higher under scenarios (b) and (c) if we were to eliminate the respective built-in profit margins of 6.09% and 6.50%.

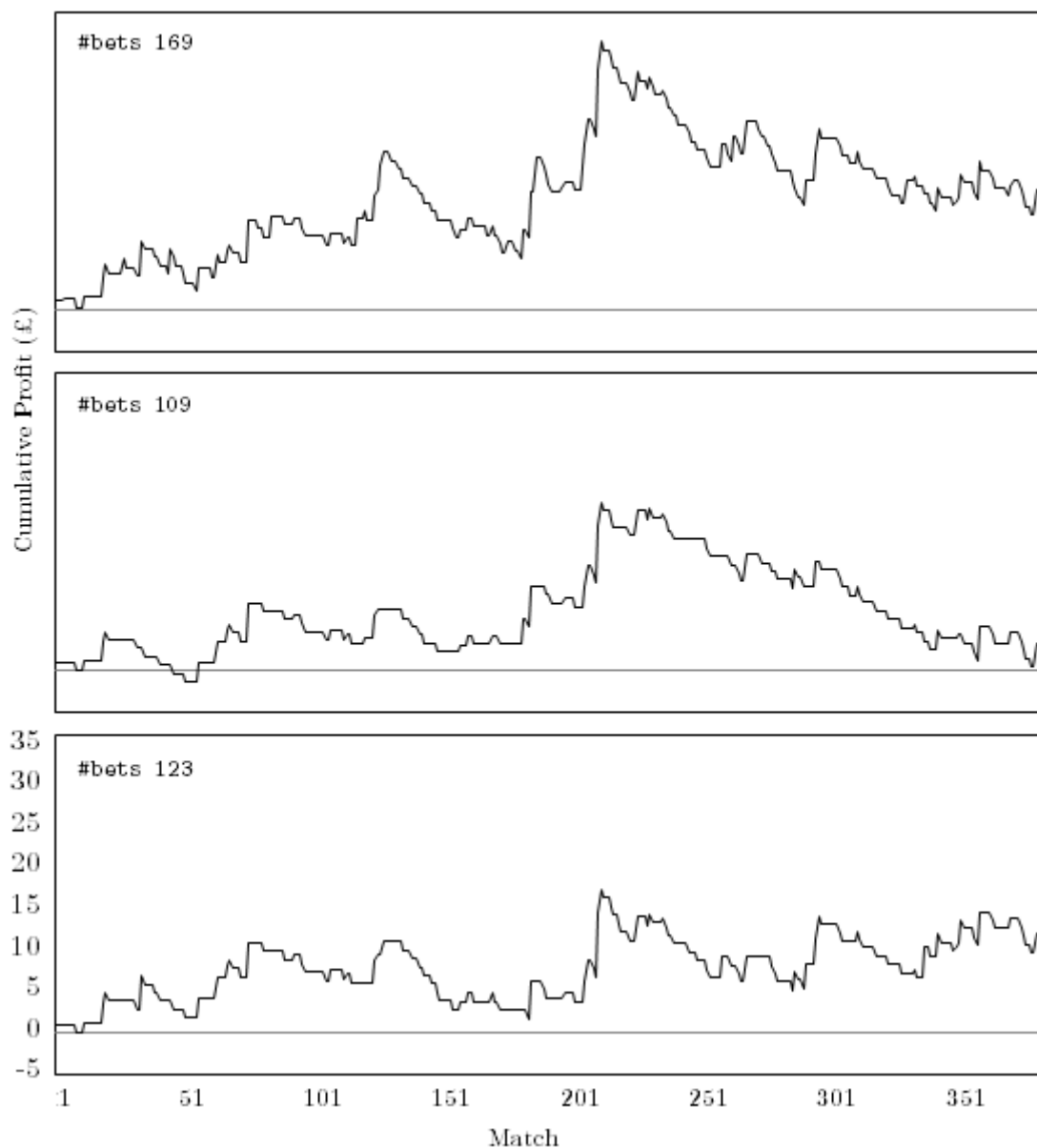


Figure 10. Cumulative profit/loss observed given  $f_s$  when simulating the standard betting strategy at discrepancy levels of  $\geq 5\%$  against a)  $f_{maxB}$ , b)  $f_{meanB}$  and c)  $f_{WH}$ .

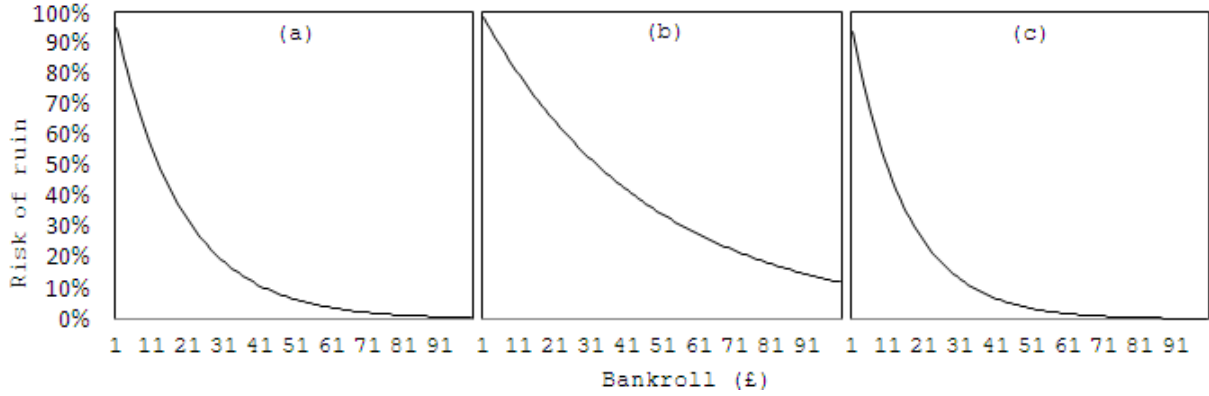


Figure 11. Risk of Ruin given the specified betting simulation against a)  $f_{maxB}$ , b)  $f_{meanB}$  and c)  $f_{WH}$ .

Table 2. Betting simulation stats given  $f_s$  against )  $f_{maxB}$ , b)  $f_{meanB}$  and c)  $f_{WH}$  at discrepancy levels of  $\geq 5\%$

|   | $f_{maxB}$  | $f_{meanB}$ | $f_{WH}$    |
|---|-------------|-------------|-------------|
| Total bets                                  | 169         | 109         | 123         |
| Bets won                                    | 57 (33.72%) | 38 (34.86%) | 44 (35.77%) |
| Total returns                               | £183.19     | £112.13     | £134.66     |
| Min. P/L balance observed                   | £0.28       | -£0.04      | -£0.09      |
| Max. P/L balance observed                   | £30.67      | £19.86      | £16.86      |
| Final P/L balance                           | £14.19      | £3.13       | £11.66      |
| Profit/Loss (%)                             | 8.40%       | 2.87%       | 9.48%       |
| Max. bookmakers considered per instance     | 40          | 40          | 1           |
| Min. bookmakers considered per instance     | 28          | 28          | 1           |
| Mean bookmakers considered per instance     | 35.73       | 35.73       | 1           |
| Max. odds won                               | 9           | 7.73        | 8.5         |
| Min. odds won                               | 1.19        | 1.40        | 1.40        |
| Mean odds won                               | 3.21        | 2.95        | 3.06        |
| Mean profit margin (for all 380 instances)  | 0.63%       | 6.09%       | 6.50%       |
| Arbitrage instances (for all 380 instances) | 62          | 0           | 0           |

Table F1 provides further statistics when performing this betting simulation given  $f_s$  against  $f_{maxB}$ ,  $f_{meanB}$ , and  $f_{WH}$  using discrepancy levels that are different from the standard 5%. In general, pi-football appears to perform much worse at the lowest discrepancy levels (1%-3%) and much better at higher discrepancy levels (4%-11%). Considering a minimum of 30 simulated bets, the maximum profits are observed at discrepancy levels of 11% (35.63%), 9% (8.86%) and 8% (10.07%) against  $f_{maxB}$ ,  $f_{meanB}$ , and  $f_{WH}$  respectively. At discrepancy levels above ~11% there were too few betting instances to be able to derive meaningful conclusions.

#### 4.3. Analysis of Impact of the Subjective Components

Table 3 describes profitability and accuracy performances based on the specified combinations of active components (*fatigue*, *psychology*, and *form*) relative to prior performances given  $f_o$ . Here, a component state is assumed to be *true* for a given match instance when there is  $\geq 5\%$  absolute discrepancy between competing teams in subjective



proximity for that component. For example (scenario 1) there were 40 matches in which the subjective proximity was  $\geq 5\%$  for all three components.

Table 3. Analysis of impact of different subjective components of the model on profitability and accuracy

| Scenario                           | Component/s True    | Relevant Occurrences (given comp. state) | Bets Simulated (given the betting strategy) | Revised profitability relative to $f_0$ (cumulative profit increase/decrease) | Revised accuracy relative to $f_0$ (cumulative profit increase/decrease) | Measures agree in the direction of the revision |
|------------------------------------|---------------------|--|---|---|--|---|
| 1<br>(All components True)         | All                 | 40                                       | 22  | +£16.56   | -0.4823  | TRUE  |
| 2<br>(Exactly two components True) | Fatigue, Psychology | 31                                       | 13*   | +£7.70*   | -0.3394  | TRUE*   |
|                                    | Fatigue, Form       | 9*                                       | 2*  | -£2.60*   | -0.2004*   | FALSE*  |
|                                    | Psychology, Form    | 57                                       | 27  | +£6.28  | -0.4008  | TRUE  |
| 3<br>(Exactly one component True)  | Fatigue             | 32                                       | 20  | +£15.25   | +0.0277  | FALSE   |
|                                    | Psychology          | 79                                       | 31  | +£5.67  | -0.1694  | TRUE  |
|                                    | Form                | 53                                       | 29  | -£11.60   | +0.0569  | TRUE  |

\* sample size too small to contribute to conclusions.

Considering both profitability and accuracy measures, it appears that all three components have contributed significantly in increasing the forecasting capability of this model, but it is dangerous to form strong conclusions about individual component-based performances due to the low numbers of relevant occurrences under the various scenarios. There is weak evidence that, in terms of profitability based on the betting simulation specified in section 4.2, the *Fatigue* component appears to provide the highest overall improvement<sup>§§§§</sup> when active, followed by the *Psychology* component that demonstrates improvements under all scenarios for which is active. The *Form* component appears to provide declines in profitability under scenario 3c. In contrast, the accuracy measure suggests that the *Psychology* component provided the highest reduction in error under all the scenarios for which is involved, whereas components *Fatigue* and *Form* appear to provide very similar error fluctuations for all respective sub-scenarios.

## 5 CONCLUDING REMARKS & FUTURE WORK

We have presented a novel Bayesian network model called pi-football (v1.32) that was used to generate the EPL match forecasts during season 2010/11. The model considers both objective and subjective information for prediction, in which time-dependent data is weighted using degrees of uncertainty. In particular, objective forecasts are generated first and revised afterwards according to subjective indicators. Because of the 'anonymous'

---

§§§§ Evidence of slight decline under scenario 3b are based only on two simulated bets.

underlying approach which generates predictions by only considering the strength of the two competing teams given results data and total points, the entire model is easily applicable to any other football league.

For assessing the performance of our model we have considered both accuracy and profitability measurements since earlier studies have shown conflicting conclusions between the two and suggested that both measurements should be considered. In (Dixon & Coles, 1997) the authors claimed that for a football forecast model to generate profit against bookmakers' odds without eliminating the in-built profit margin it requires a determination of probabilities that is sufficiently more accurate from those obtained by published odds, and (Graham & Stott, 2008) suggested that if such a work was particularly successful, it would not have been published. Ours is the first study to demonstrate profitability against all of the (available) published odds. Previous studies have only considered a single bookmaker, since only recently it was proven that the published odds of a single bookmaker cannot be representative of the overall market (Constantinou & Fenton, 2012b). In fact, pi-football was able to generate profit against maximum, mean, and common bookmakers' odds, even allowing for the bookmakers' in-built profit margin.

We showed that subjective information improved the forecast capability of our model significantly. Our study also emphasises the importance of Bayesian networks, in which subjective information can both be represented and displayed without any particular effort. Because of the nature of subjective information, we have been publishing our forecasts online (pi-football, 2010) prior to the start of each match (earlier studies which incorporated subjective information have not done so). Appendix G provides examples of both objective ( $f_o$ ) and subjective ( $f_s$ ) forecasts for match instances at the beginning of the EPL season 2010/11. At standard discrepancy levels of 5% the profitability of this model ranges from 2.87% to 9.48%, whereas at higher discrepancy levels (8% to 11%) the maximum profit observed ranges from 8.86% to 35.63%, depending on the various bookmakers' odds considered. No other published work appears to be particularly successful at beating all of the various bookmakers' odds over a large period of time, which highlights the success of pi-football.

Clearly the real potential benefits of a model such as this are critically dependent on both the structure of the model and the knowledge of the expert. A perfect BN model would still fail to beat the bookmakers at their own game if the subjective expert inputs are inaccurate. Because of the weekly pressure to get all of the model predictions calculated and published online, there was inevitable inconsistency in the care and accuracy taken to consider all the subjective inputs for each match; in most cases the subjective inputs were provided by a member of the research team who is certainly not an expert on the English premier League. If the model were to be used by more informed experts we feel it would provide posterior beliefs of both higher precision and confidence.

An individual component-based analysis failed to provide us with strong conclusions about their distinct efficiency due to the relative low number of relevant occurrences. Planned extensions of this research will determine the distinct component-based effectiveness by adding further evidence of relevant occurrences, and a reverse engineering approach will help us understand how specific model components help in matching bookmakers' odds (hinting at why bookmakers are indeed experts and also what information they might

consider for prediction). We have already summarised several aspects concerning bookmakers' inefficiency in (Constantinou & Fenton, 2012b). Other extensions of the research will determine whether revising the strength of the team (given subjective information) rather than the probability distribution itself would improve the performance of the model; this is important because the former represents a natural causality whereas the latter does not. Further, since we have not yet assessed the impact of time-dependent uncertainty for weighting the more recent information, we plan to determine the degree of irrelevance to prediction per preceding information, as well as the degree of efficiency of the various time-series methodologies introduced throughout the sports academic literature (none of the previous football studies have attempted to measure their efficiency).

## ACKNOWLEDGEMENTS

We acknowledge the financial support by the Engineering and Physical Sciences Research Council (EPSRC) for funding this research and Agena Ltd for software support. We also acknowledge the anonymous referees whose comments have led to improvements in the paper.

## APPENDIX A: Subjective scenarios and assumptions per specified variable (node)

Table A.1. Team Strength (as presented in Figure 2)

| ID           | Variable (node)                      | Description  | Subjective Scenarios                     |
|--------------|--------------------------------------|--|--|
| <i>I.</i>    | Subjective team strength (in points) | Expert indication regarding the current strength of the team in seasonal points.                               | [0,114]                                  |
| <i>II.</i>   | Confidence                           | Expert indication regarding its confidence about his input (I).  | [Very High, High, Medium, Low, Very Low] |
| <i>III.</i>  | Current Points                       | Assumption: Variance as demonstrated in figure 1, given variable "Number of matches played".                   | -  |
| <i>IV.</i>   | Points during season 2005/06         | Assumption: variance= $(\text{Variance}+3^6)$  | -  |
| <i>V.</i>    | Points during season 2006/07         | Assumption: variance= $(\text{Variance}+3^5)$  | -  |
| <i>VI.</i>   | Points during season 2007/08         | Assumption: variance= $(\text{Variance}+3^4)$  | -  |
| <i>VII.</i>  | Points during season 2008/09         | Assumption: variance= $(\text{Variance}+3^3)$  | -  |
| <i>VIII.</i> | Points during season 2009/10         | Assumption: variance= $(\text{Variance}+3^2)$  | -  |
| <i>IX.</i>   | Predicted mean (in points)           | The predicted team strength after considering all of the seven parameters<br>Assumption: mean=57, variance=300 | -  |

Table A.2. Team Form (as presented in Figure 3)

| ID          | Variable (node)                           | Description   | Subjective Scenarios                     |
|-------------|---|---|--|
| <i>I.</i>   | Primary key-player availability           | Expert indication regarding his confidence about the availability of the primary key-player.                  | [Very High, High, Medium, Low, Very Low] |
| <i>II.</i>  | Secondary key-player availability         | Expert indication regarding his confidence about the availability of the secondary key-player.                | [Very High, High, Medium, Low, Very Low] |
| <i>III.</i> | Tertiary key-player availability          | Expert indication regarding his confidence about the availability of the tertiary key-player.                 | [Very High, High, Medium, Low, Very Low] |
| <i>IV.</i>  | Remaining first team players availability | Expert indication regarding his confidence about the availability of the remaining first-team players.        | [Very High, High, Medium, Low, Very Low] |
| <i>V.</i>   | First team players returning              | Expert indication regarding the potential return of other first team players who missed the last few matches. | [Very High, High, Medium, Low, Very Low] |

Table A.3. Team Psychology (as presented in Figure 4)

| ID          | Variable (node)            | Description   | Subjective Scenarios  |
|-------------|----------------------------|---|---|
| <i>I.</i>   | Team spirit and motivation | Expert indication regarding the team's level of motivation and team spirit                      | [Very High, High, Normal, Low, Very Low]  |
| <i>II.</i>  | Confidence                 | Expert indication regarding its confidence about his input in (I).                              | [Very High, High, Medium, Low, Very Low]  |
| <i>III.</i> | Managerial impact          | Expert indication regarding the impact of the current managerial situation.                     | [Very High, High, Normal, Low, Very Low]  |
| <i>IV.</i>  | Head-to-Head bias          | Expert indication regarding potential biases in a head-to-head encounter between the two teams. | [High advantage for home team, Advantage for home team, No bias, Advantage for away team, High advantage for away team] |

Table A.4. Team Fatigue (as presented in Figure 5)

| ID          | Variable (node)                             | Description   | Subjective Scenarios                                      |
|-------------|---|---|---|
| <i>I.</i>   | Toughness of previous match                 | Expert indication regarding the toughness of previous match.                                    | [Lowest, Very Low, Low, Medium, High, Very High, Highest] |
| <i>II.</i>  | First team players rested during last match | Expert indication regarding the first team players rested during last match.                    | [1-2, 3, 4, 5, 6+]  |
| <i>III.</i> | National team participation                 | Expert indication regarding the level of international participation by the first team players. | [None, Few, Half team, Many, All]                         |

## APPENDIX B: An actual example of component's 1 process (as presented in Fig. 2)

Figure B.1 presents a real component 1 example between Manchester City (home team) and Manchester United, as prepared for the 11<sup>th</sup> of October 2010. The steps for calculating component's 1 forecast are enumerated below:

- 1) **Previous information:** the points accumulated per previous season are passed as five distinct ordered inputs. Starting from the oldest season, the inputs are [43, 42, 55, 50, 67] for Man City, and [83, 89, 87, 90, 85] for Man United. Note that Man City generates a significantly higher variance than that of Man United, with the more recent seasons having greater impact as described and illustrated in section 3.1.
- 2) **Current information:** the points accumulated for the current season, as well as the total number of matches played are passed as a single parameter with the appropriate variance as described and illustrated in section 3.1. For Man City the inputs are [20, 11] and for Man United the inputs are [23, 11], for points accumulated and number of matches played respectively.
- 3) **Subjective information** (optional): the optional subjective indication about the current team's strength in total points, as well as the confidence with reference to that indication are passed as a single parameter. For Man City, we suggested that the team was playing as a 72-point team (a 5-point increase from last season) with "High" confidence (out of "Very High")<sup>\*\*\*\*\*</sup>. On the other hand, we have introduced a 5-point decrease for Man United with "High" confidence<sup>++++</sup>. Accordingly, the inputs were [72, 'High'] and [80, 'High'] for Man City and Man United respectively.
- 4) The model summarises the seven parameters in node "Mean". The impact each parameter has is dependent on its certainty (variance). For Man City the summarised belief in total points (node "Mean") is 68.95 whereas for Man United is 80.78. Note that the variance introduced for Man City is a higher than that of Man United; 26.83 and 21.92 respectively.
- 5) Each team's "Mean" is converted in the predetermined 14-scale ranking. The model suggests that Man City will most likely perform similar to teams ranked 3 to 4 (out of 14), whereas for Man United it mostly suggests ranks 1 and 2.
- 6) The model generates the objective forecast in node "Match Prediction", by considering each teams estimated ranking, before proceeding to potential forecast revisions suggested by the expert constructed component models 2, 3 and 4.

---

<sup>\*\*\*\*\*</sup>A 5-point increase was suggested due to high profile players joining the team during the summer transfer window.

<sup>++++</sup>A 5-point decrease was suggested due to the significant decrease in stamina observed by the older core-team players (e.g. Scholes, Giggs, Ferdinand, Vidic) without taking care of appropriate replacements.

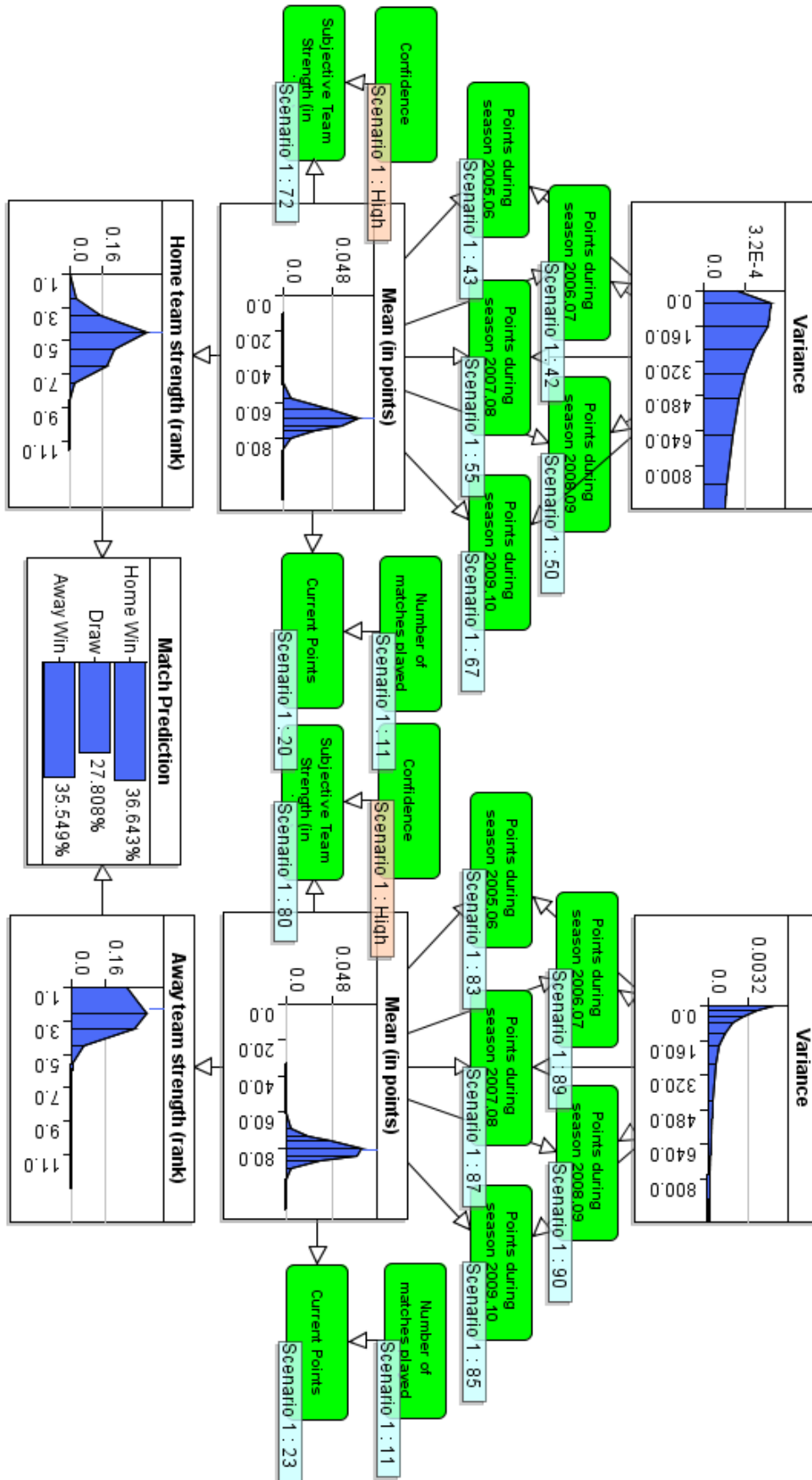


Figure B1. An actual example of the Bayesian network (from Figure 2) at component 1. The parameters represent the actual observations provided from the Man City vs. Man United match, 10th of November, 2010.

## APPENDIX C: Match RPS per dataset

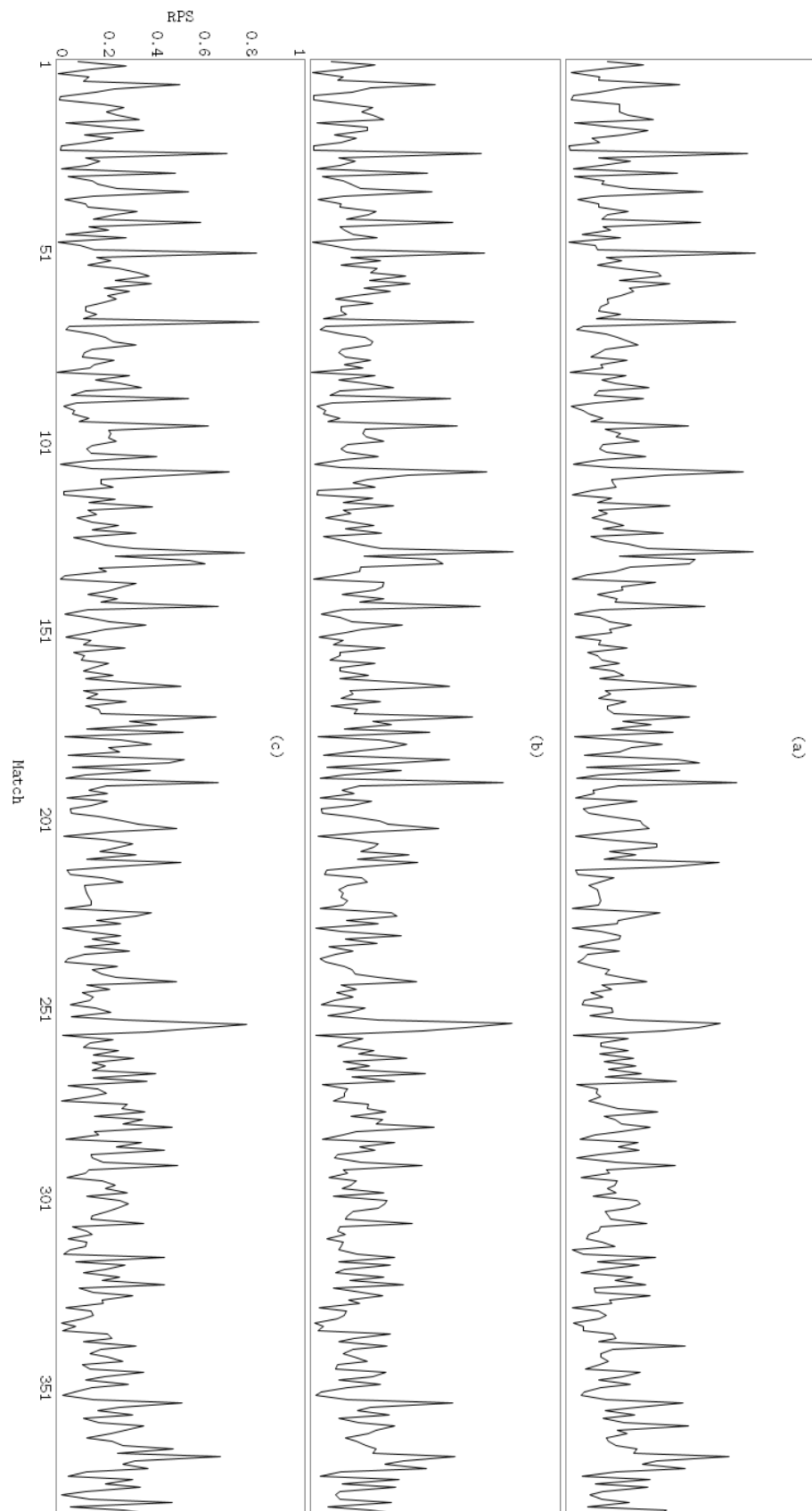


Figure C.1. RPS per match for datasets  $f_o$  (a),  $f_s$  (b), and  $f_B$  (c) respectively.

**APPENDIX D:** Evidence of significant improvements in  $f_o$  by subjective information

In this section we provide evidence of football matches in which subjective information revised  $f_o$  the most. Table D1 presents 17 with the highest absolute RPS discrepancies between  $f_o$  and  $f_s$  forecasts, assuming a minimum discrepancy level of 0.1. The instances are ranked by highest discrepancy and the 'Decision' column indicates whether the subjective information improved  $f_o$ .

Overall, the results appear to be particularly encouraging. Only in 6 out of the 17 cases our subjective information leads to a higher forecast error. The results are even more encouraging when we only concentrate on the first 10 highest discrepancy instances, in which subjective revisions improve 8 out of the 10 instances. Further, in those 17 instances we have observed 15 distinct teams, and no evidence exist that strong subjective indications follow a particular type of a team. A rather surprising and interesting observation is that the observed outcome is a draw in only in 1 out of the 17 instances presented here.

Table D1. RPS discrepancies  $\geq 0.1$  between objective ( $f_o$ ) and revised ( $f_s$ ); ranked by highest discrepancy

| RPS<br>Discrep. | Date       | Home<br>Team | Away<br>Team | R | Objective ( $f_o$ ) |       |       | Revised ( $f_s$ ) |       |       | Decision |
|-----------------|------------|--------------|--------------|---|---------------------|-------|-------|-------------------|-------|-------|----------|
|                 |            |              |              |   | p(H)                | p(D)  | p(A)  | p(H)              | p(D)  | p(A)  |          |
| .2078           | 14/05/2011 | Sunderland   | Wolves       | A | .4942               | .3403 | .1656 | .2627             | .4124 | .3250 | ✓        |
| .1765           | 06/03/2011 | Liverpool    | Man Utd      | H | .2392               | .2219 | .5389 | .3423             | .3691 | .2887 | ✓        |
| .1614           | 03/10/2010 | Liverpool    | Blackpool    | A | .8303               | .1412 | .0285 | .6516             | .2895 | .0589 | ✓        |
| .1582           | 09/04/2011 | Man Utd      | Fulham       | H | .7570               | .1881 | .0549 | .4016             | .4552 | .1432 | ✗        |
| .1421           | 22/05/2011 | Stoke        | Wigan        | A | .5140               | .3023 | .1837 | .3535             | .3684 | .2781 | ✓        |
| .1406           | 02/10/2010 | Sunderland   | Man Utd      | D | .1223               | .1940 | .6837 | .2029             | .3973 | .3998 | ✓        |
| .1322           | 18/09/2010 | Tottenham    | Wolves       | H | .7422               | .1751 | .0827 | .4396             | .4063 | .1541 | ✗        |
| .1307           | 06/11/2010 | Bolton       | Tottenham    | H | .2519               | .2523 | .4958 | .3384             | .3358 | .3259 | ✓        |
| .1270           | 22/08/2010 | Newcastle    | Aston Villa  | H | .2693               | .3161 | .4146 | .3828             | .3514 | .2658 | ✓        |
| .1228           | 25/01/2011 | Wigan        | Aston Villa  | A | .3436               | .3431 | .3133 | .2058             | .3433 | .4508 | ✓        |
| .1219           | 29/12/2010 | Liverpool    | Wolves       | A | .7162               | .1717 | .1121 | .8058             | .1406 | .0536 | ✗        |
| .1156           | 23/04/2011 | Sunderland   | Wigan        | H | .4138               | .3310 | .2552 | .2848             | .3568 | .3584 | ✗        |
| .1150           | 01/02/2011 | Sunderland   | Chelsea      | A | .2661               | .3861 | .3478 | .1556             | .3363 | .5082 | ✓        |
| .1104           | 27/12/2010 | Arsenal      | Chelsea      | H | .4034               | .3383 | .2583 | .2828             | .3578 | .3594 | ✗        |
| .1102           | 28/12/2010 | Sunderland   | Blackpool    | A | .5200               | .2791 | .2009 | .3929             | .3380 | .2692 | ✓        |
| .1063           | 25/09/2010 | Arsenal      | West Br.     | A | .8196               | .1499 | .0305 | .7063             | .2424 | .0512 | ✓        |
| .1023           | 22/01/2011 | Wolves       | Liverpool    | A | .3070               | .3465 | .3466 | .4038             | .3465 | .2497 | ✗        |



## APPENDIX E: Betting simulation given objective forecasts

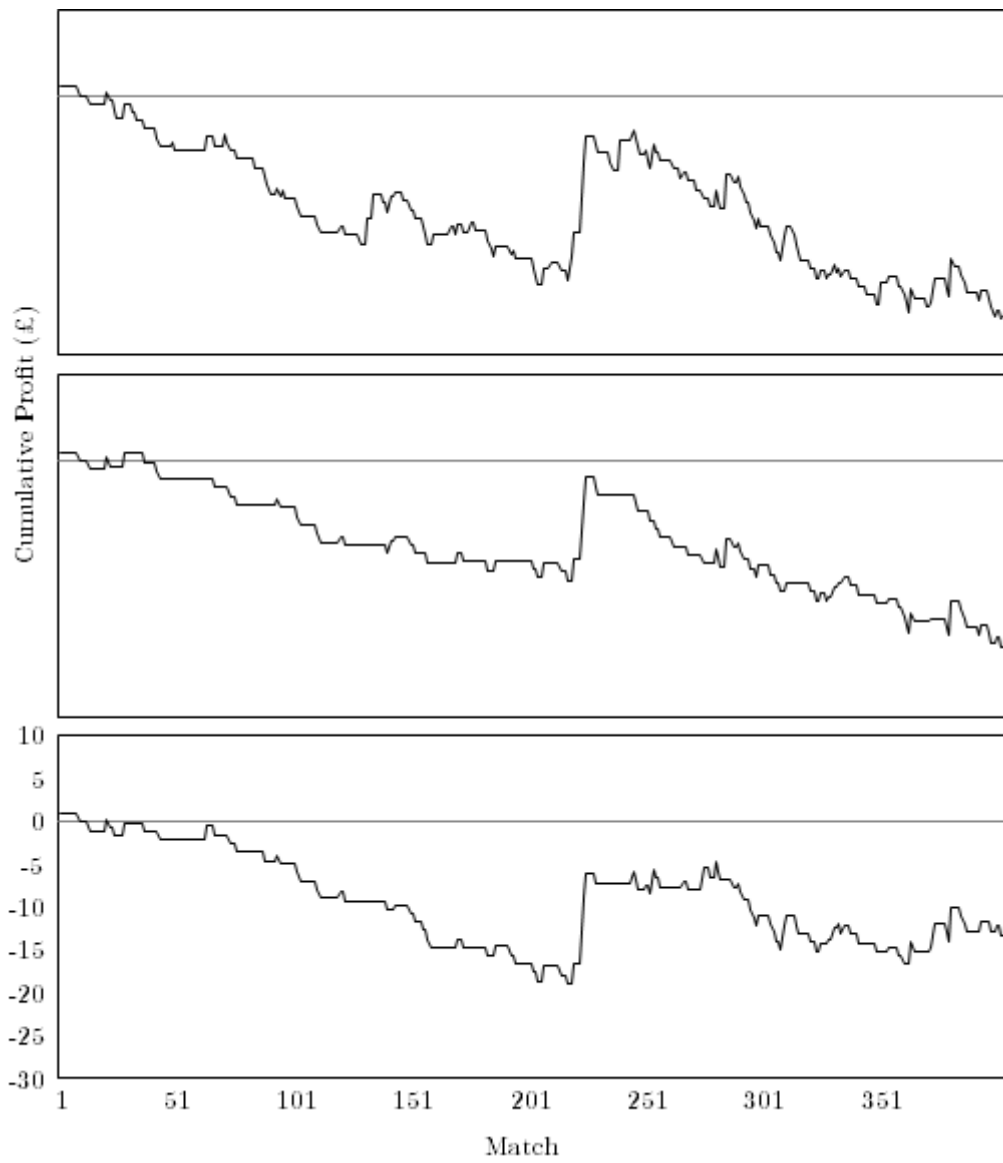


Figure E1. Cumulative profit/loss observed given  $f_o$  when simulating the standard betting strategy at discrepancy levels of  $\geq 5\%$  against a)  $f_{maxB}$ , b)  $f_{meanB}$  and c)  $f_{WH}$ .

**APPENDIX F: Betting simulation at different levels of discrepancy given  $f_S$**

Table F.1. Betting simulation stats given  $f_S$  against )  $f_{maxB}$ , b)  $f_{meanB}$  and c)  $f_{WH}$  at discrepancy levels from 1% to 20%

| Discrepancy | Maximum odds |             |                 | Mean odds   |             |                 | William Hill odds |             |                 |
|-------------|--------------|-------------|-----------------|-------------|-------------|-----------------|-------------------|-------------|-----------------|
|             | No. of bets  | Returns (£) | Profit/Loss (£) | No. of bets | Returns (£) | Profit/Loss (£) | No. of bets       | Returns (£) | Profit/Loss (£) |
| 1%          | 358          | 356.24      | -0.49%          | 280         | 266.25      | -4.91%          | 284               | 276.04      | -2.80%          |
| 2%          | 325          | 320.21      | -1.47%          | 240         | 225.93      | -5.86%          | 234               | 235.98      | 0.85%           |
| 3%          | 275          | 277.85      | 1.04%           | 189         | 187.07      | -1.02%          | 192               | 191.12      | -0.46%          |
| 4%          | 225          | 236.87      | 5.28%           | 136         | 144.85      | 6.51%           | 147               | 159.44      | 8.46%           |
| 5%          | 169          | 183.19      | 8.40%           | 109         | 112.13      | 2.87%           | 123               | 134.66      | 9.48%           |
| 6%          | 131          | 148.4       | 13.28%          | 85          | 84.96       | -0.05%          | 95                | 102.31      | 7.69%           |
| 7%          | 107          | 119.92      | 12.07%          | 68          | 64.86       | -4.62%          | 67                | 68.91       | 2.85%           |
| 8%          | 84           | 92.43       | 10.04%          | 53          | 54.79       | 3.38%           | 45                | 49.53       | 10.07%          |
| 9%          | 71           | 82.36       | 16.00%          | 36          | 39.19       | 8.86%           | 34                | 32.71       | -3.79%          |
| 10%         | 52           | 62.61       | 20.40%          | 26          | 16.97       | -34.73%         | 24                | 23.55       | -1.88%          |
| 11%         | 41           | 55.61       | 35.63%          | 15          | 7.82        | -47.87%         | 19                | 21.82       | 14.84%          |
| 12%         | 25           | 18.05       | -27.80%         | 12          | 7.82        | -34.83%         | 13                | 7.82        | -39.85%         |
| 13%         | 15           | 10.39       | -30.73%         | 10          | 7.82        | -21.80%         | 10                | 7.82        | -21.80%         |
| 14%         | 12           | 8.3         | -30.83%         | 8           | 7.82        | -2.25%          | 10                | 7.82        | -21.80%         |
| 15%         | 10           | 8.3         | -17.00%         | 7           | 7.82        | 11.71%          | 7                 | 7.82        | 11.71%          |
| 16%         | 7            | 8.3         | 18.57%          | 5           | 6.2         | 24.00%          | 6                 | 6.2         | 3.33%           |
| 17%         | 6            | 8.3         | 38.33%          | 2           | 0           | -100%           | 3                 | 2.4         | -20.00%         |
| 18%         | 5            | 5.9         | 18.00%          | 2           | 0           | -100%           | 2                 | 0           | -100%           |
| 19%         | 2            | 0           | -100%           | 1           | 0           | -100%           | 1                 | 0           | -100%           |
| 20%         | 2            | 0           | -100%           | 1           | 0           | -100%           | 1                 | 0           | -100%           |

## APPENDIX G: Example of forecasts generated by pi-football

Table G.1. Objective ( $f_o$ ) and subjective ( $f_s$ ) forecasts generated by pi-football, at the beginning of the EPL season 2010/11

| Date       | Home Team   | Away Team   | Result | Objective ( $f_o$ ) |        |        | Subjective ( $f_s$ ) |        |        |
|------------|-------------|-------------|--------|---------------------|--------|--------|----------------------|--------|--------|
|            |             |             |        | p(H)                | p(D)   | p(A)   | p(H)                 | p(D)   | p(A)   |
| 14/08/2010 | Aston Villa | West Ham    | H      | 60.92               | 23.971 | 15.109 | 61.735               | 23.67  | 14.596 |
| 14/08/2010 | Blackburn   | Everton     | H      | 34.382              | 29.314 | 36.304 | 36.338               | 29.781 | 33.881 |
| 14/08/2010 | Bolton      | Fulham      | D      | 46.863              | 29.199 | 23.938 | 46.863               | 29.199 | 23.938 |
| 14/08/2010 | Chelsea     | West Brom   | H      | 87.055              | 12.706 | 0.24   | 89.581               | 10.227 | 0.192  |
| 14/08/2010 | Sunderland  | Birmingham  | D      | 44.366              | 29.623 | 26.011 | 44.197               | 29.679 | 26.124 |
| 14/08/2010 | Tottenham   | Man City    | D      | 35.178              | 33.654 | 31.168 | 32.82                | 33.756 | 33.424 |
| 14/08/2010 | Wigan       | Blackpool   | A      | 53.939              | 30.156 | 15.905 | 53.939               | 30.156 | 15.905 |
| 14/08/2010 | Wolves      | Stoke       | H      | 38.763              | 31.563 | 29.674 | 37.778               | 31.746 | 30.477 |
| 15/08/2010 | Liverpool   | Arsenal     | D      | 51.705              | 27.305 | 20.99  | 54.007               | 26.773 | 19.22  |
| 16/08/2010 | Man United  | Newcastle   | H      | 81.665              | 16.058 | 2.277  | 83.853               | 14.18  | 1.966  |
| 21/08/2010 | Arsenal     | Blackpool   | H      | 85.569              | 12.668 | 1.763  | 85.695               | 12.56  | 1.746  |
| 21/08/2010 | Birmingham  | Blackburn   | H      | 44.269              | 29.088 | 26.643 | 49.695               | 28.632 | 21.673 |
| 21/08/2010 | Everton     | Wolves      | D      | 73.202              | 17.433 | 9.365  | 69.731               | 20.077 | 10.192 |
| 21/08/2010 | Stoke       | Tottenham   | A      | 27.657              | 29.283 | 43.059 | 28.289               | 29.58  | 42.13  |
| 21/08/2010 | West Brom   | Sunderland  | H      | 36.848              | 33.163 | 29.989 | 36.325               | 33.216 | 30.459 |
| 21/08/2010 | West Ham    | Bolton      | A      | 39.606              | 32.217 | 28.177 | 35.012               | 33.074 | 31.913 |
| 21/08/2010 | Wigan       | Chelsea     | A      | 9.945               | 16.713 | 73.342 | 6.465                | 14.345 | 79.19  |
| 22/08/2010 | Fulham      | Man United  | D      | 13.416              | 22.345 | 64.239 | 12.059               | 21.442 | 66.499 |
| 22/08/2010 | Newcastle   | Aston Villa | H      | 26.934              | 31.612 | 41.455 | 38.277               | 35.144 | 26.58  |
| 23/08/2010 | Man City    | Liverpool   | H      | 55.566              | 26.104 | 18.33  | 59.331               | 24.983 | 15.686 |
| 28/08/2010 | Blackburn   | Arsenal     | A      | 29.444              | 31.547 | 39.009 | 24.496               | 31.194 | 44.31  |
| 28/08/2010 | Blackpool   | Fulham      | D      | 28.052              | 31.672 | 40.276 | 28.272               | 31.732 | 39.996 |
| 28/08/2010 | Chelsea     | Stoke       | H      | 80.673              | 16.736 | 2.591  | 84.022               | 13.905 | 2.073  |
| 28/08/2010 | Man United  | West Ham    | H      | 82.525              | 15.553 | 1.922  | 84.627               | 13.711 | 1.662  |
| 28/08/2010 | Tottenham   | Wigan       | A      | 73.716              | 17.443 | 8.841  | 73.327               | 17.74  | 8.934  |
| 28/08/2010 | Wolves      | Newcastle   | D      | 40.609              | 32.837 | 26.554 | 37.192               | 33.491 | 29.318 |
| 29/08/2010 | Aston Villa | Everton     | H      | 45.276              | 31.446 | 23.277 | 44.676               | 31.63  | 23.695 |
| 29/08/2010 | Bolton      | Birmingham  | D      | 39.858              | 31.208 | 28.934 | 36.146               | 32.013 | 31.84  |
| 29/08/2010 | Liverpool   | West Brom   | H      | 80.318              | 15.187 | 4.495  | 77.822               | 17.212 | 4.967  |
| 29/08/2010 | Sunderland  | Man City    | H      | 21.155              | 20.44  | 58.405 | 21.584               | 21.237 | 57.179 |
| 11/09/2010 | Arsenal     | Bolton      | H      | 70.745              | 19.864 | 9.391  | 70.751               | 19.861 | 9.388  |
| 11/09/2010 | Everton     | Man United  | D      | 27.891              | 25.825 | 46.284 | 31.386               | 28.593 | 40.021 |
| 11/09/2010 | Fulham      | Wolves      | H      | 46.98               | 29.379 | 23.641 | 48.281               | 29.125 | 22.594 |
| 11/09/2010 | Man City    | Blackburn   | D      | 69.118              | 20.636 | 10.246 | 62.251               | 25.453 | 12.296 |
| 11/09/2010 | Newcastle   | Blackpool   | A      | 55.782              | 31.301 | 12.918 | 51.035               | 33.384 | 15.581 |
| 11/09/2010 | West Brom   | Tottenham   | D      | 22.674              | 28.013 | 49.314 | 25.911               | 30.475 | 43.614 |
| 11/09/2010 | West Ham    | Chelsea     | A      | 7.98                | 16.013 | 76.007 | 7.879                | 15.911 | 76.21  |
| 11/09/2010 | Wigan       | Sunderland  | D      | 40.77               | 32.102 | 27.128 | 41.178               | 32.039 | 26.784 |
| 12/09/2010 | Birmingham  | Liverpool   | D      | 30.374              | 29.364 | 40.262 | 35.557               | 31.287 | 33.155 |
| 13/09/2010 | Stoke       | Aston Villa | H      | 29.946              | 29.846 | 40.208 | 35.597               | 31.808 | 32.595 |
| 18/09/2010 | Aston Villa | Bolton      | D      | 67.813              | 20.418 | 11.768 | 66.943               | 21.027 | 12.03  |
| 18/09/2010 | Blackburn   | Fulham      | D      | 49.733              | 28.365 | 21.902 | 48.58                | 28.861 | 22.559 |
| 18/09/2010 | Everton     | Newcastle   | A      | 64.358              | 22.042 | 13.6   | 63.488               | 22.615 | 13.898 |
| 18/09/2010 | Stoke       | West Ham    | D      | 45.372              | 31.286 | 23.342 | 39.697               | 33.048 | 27.255 |
| 18/09/2010 | Sunderland  | Arsenal     | D      | 17.051              | 20.505 | 62.444 | 21.997               | 30.62  | 47.383 |
| 18/09/2010 | Tottenham   | Wolves      | H      | 74.223              | 17.506 | 8.271  | 43.964               | 40.629 | 15.407 |
| 18/09/2010 | West Brom   | Birmingham  | H      | 33.397              | 32.167 | 34.436 | 34.729               | 32.261 | 33.01  |
| 19/09/2010 | Chelsea     | Blackpool   | H      | 88.112              | 11.363 | 0.525  | 88.753               | 10.751 | 0.496  |
| 19/09/2010 | Man United  | Liverpool   | H      | 58.15               | 28.169 | 13.681 | 61.165               | 26.618 | 12.217 |
| 19/09/2010 | Wigan       | Man City    | A      | 23.721              | 26.167 | 50.113 | 25.023               | 27.358 | 47.619 |

## REFERENCES

- Agena. (2012). Agena: Bayesian Network and Simulation Software for Risk Analysis and Decision Support. Retrieved January 26, 2012, from <http://www.agenarisk.com>
- Agena. (2009). *AgenaRisk 5.0 User Manual*. Agena Ltd, p. 195.
- Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 37:2, 253-264.
- Buchner, A., Dubitzky, W., Schuster, A., Lopes, P., O'Doneghue, P., Hughes, J., et al. (1997). Corporate evidential decision making in performance prediction domains. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI'97)*. Brown University, Providence, Rhode Island, USA.
- Cain, M., Law, D., & Peel, D. (2000). The Favourite-Legshot Bias and Market Efficiency in UK Football Betting. *Scottish Journal of Political Economy*, 47, 25-36.
- Constantinou, A. C., & Fenton, N. E. (2012a). Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models. *Journal of Quantitative Analysis in Sports*, Vol. 8: Iss. 1, Article 1.
- Constantinou, A. C., & Fenton, N. E. (2012b). Evidence of an (intended) inefficient Association Football gambling market. Under Review, Draft available at: <http://www.constantinou.info/downloads/papers/evidenceOfInefficiency.pdf>
- Crowder, M., Dixon, M., Ledford, A., & Robinson, M. (2002). Dynamic modelling and prediction of English football league matches for betting. *The Statistician*, 51, 157-168.
- Dixon, M., & Coles, S. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics*, 46, 265-80.
- Dixon, M., & Pope, P. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting*, 20, 697-711.
- Dunning, E. G., & Joseph A. Maguire, R. E. (1993). *The Sports Process: A Comparative and Developmental Approach*. Champaign: Human Kinetics, p. 129.
- Dunning, E. (1999). *Sport Matters: Sociological Studies of Sport, Violence and Civilisation*. London: Routledge.
- Elo, A. E. (1978). *The rating of chess players, past and present*. New York: Arco Publishing.
- Epstein, E. (1969). A Scoring System for Probability Forecasts of Ranked Categories. *Journal of Applied Meteorology*, 8, 985-987.
- Falter, M., & Perignon, C. (2000). Demand for football and intra-match winning probability: an essay on the glorious uncertainty of sports. *Applied Economics*, 32, 1757-1765.
- Fenton, N. E., Neil, M. and Caballero, J. G. (2007). Using Ranked nodes to model qualitative judgments in Bayesian Networks. *IEEE TKDE* 19(10), 1420-1432.
- Football-Data. (n.d.). *Football-Data.co.uk*. Retrieved August 25, 2011, from *Football Results, Statistics & Soccer Betting Odds Data*: <http://www.football-data.co.uk/englandm.php>
- Forrest, D., & Simmons, R. (2000). Forecasting sport: the behaviour and performance of football tipsters. *International Journal of Forecasting*, 16, 317-331.

- Forrest, D., & Simmons, R. (2001). Globalisation and efficiency in the fixed-odds soccer betting market. University of Salford, Salford: Mimeo.
- Forrest, D., & Simmons, R. (2002). Outcome uncertainty and attendance demand in sport: the case of English soccer. *The Statistician*, 2, 241-291.
- Forrest, D., & Simmons, R. (2008). Sentiment in the betting market on Spanish football. *Applied Economics*, 40, 119-126.
- Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, 21, 551-564.
- Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21, 331-340.
- Goddard, J., & Asimakopoulous, I. (2004). Forecasting Football Results and the Efficiency of Fixed-odds Betting. *Journal of Forecasting*, 23, 51-66.
- Graham, I., & Stott, H. (2008). Predicting bookmaker odds and efficiency for UK football. *Applied Economics*, 40, 99-109.
- Halicioğlu, F. (2005a). Can we predict the outcome of the international football tournaments? : the case of Euro 2000. *Doğuş Üniversitesi Dergisi*, 6, 112-122.
- Halicioğlu, F. (2005b). Forecasting the Professional Team Sporting Events: Evidence from Euro 2000 and 2004 Football Tournaments. 5th International Conference on Sports and Culture: Economic, Management and Marketing Aspects. Athens, Greece, pp. 30-31.
- Hirotsu, N., & Wright, M. (2003). An evaluation of characteristics of teams in association football by using a Markov process model. *The Statistician*, 52: 4, 591-602.
- Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, 26, 460-470.
- Jolliffe, I. T., & Stephenson, D. B. (2003). *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. West Sussex, England: Wiley.
- Joseph, A., Fenton, N., & Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems*, 7, 544-553.
- Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *The Statistician*, 52: 3, 381-393.
- Knorr-Held, L. (2000). Dynamic rating of sports teams. *The Statistician*, 49, 261-276.
- Koning, R. (2000). Balance in competition in Dutch soccer. *The Statistician*, 49: 3, 419-431.
- Koning, R., Koolhaas, M., Renes, G., & Ridder, G. (2003). A simulation model for football championships. *European Journal of Operational Research*, 148, 268-276.
- Kuonen, D. (1996). *Statistical Models for Knock-Out Soccer Tournaments*. Technical Report, Department of Mathematics, École Polytechnique Federale de Lausanne.
- Kuyppers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics*, 32, 1353-1363.
- Lee, A. J. (1997) Modeling scores in the Premier League: is Manchester United really the best? *Chance*, 10, 15-19.
- Leitch, G., & Tanner, J. E. (1991). Economic Forecast Evaluation: Profits Versus The Conventional Error Measures. *American Economic Association*, 580-590.

- Leitner, C., Zeileis, A., & Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. *International Journal of Forecasting*, 26, 471-481.
- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36, 109-118.
- Min, B., Kim, J., Choe, C., Eom, H., & McKay, R. B. (2008). A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems*, 21, 551-562.
- Mueller, F. O., Cantu, R. C., & Camp, S. P. (1996). Catastrophic Injuries in High School and College Sports. Champaign: Human Kinetics, p. 57.
- Murphy, A. (1969). On the "ranked probability score". *Journal of Applied Meteorology*, 8, 988-989.
- Murphy, A. (1970). The ranked probability score and the probability score: A comparison. *Monthly Weather Review*, 98, 917-924.
- Peel, D. A., & Thomas, D. A. (1989). Outcome uncertainty and the demand for football. *Scottish Journal of Political Economy*, 35, 242-249.
- Peel, D. A., & Thomas, D. A. (1992). The demand for football: some evidence on outcome uncertainty. *Empirical Economics*, 4, 567-570.
- Peel, D. A., & Thomas, D. A. (1997). Handicaps, outcome uncertainty and attendance demand. *Applied Economic Letters*, 4, 567-570.
- pi-football (2010) Probabilistic Intelligence in Football, Retrieved May 25, 2012, from Forecasts for season 2010/11: [http://www.pi-football.com/past\\_predictions.aspx](http://www.pi-football.com/past_predictions.aspx)
- Pope, P., & Peel, D. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica*, 56, 323-341.
- Ridder, G., Cramer, J., & Hopstaken, P. (1994). Estimating the Effect of a Red Card in Soccer. *Journal of the American Statistical Association*, 89: 427, 1124-1127.
- Rotshtein, A., Posner, M., & Rakytyanska, A. (2005). Football predictions based on a fuzzy model with genetic and neural tuning. *Cybernetics and Systems Analysis*, 41: 4, 619-630.
- Rue, H., & Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *The Statistician*, 3, 339-418.
- Soccer Base. (2012). *Soccer Base*. Retrieved February 23, 2012, from Man City: Full Transfer History: [http://www.soccerbase.com/teams/team.sd?team\\_id=1718&teamTabs=transfers](http://www.soccerbase.com/teams/team.sd?team_id=1718&teamTabs=transfers)
- Tsakonas, A., Dounias, G., Shtovba, S., & Vivdyuk, V. (2002). Soft computing-based result prediction of football games. The First International Conference on Inductive Modelling (ICIM'2002). Lviv, Ukraine.
- Vecer, J., Kopriva, F., & Ichiba, T. (2009). Estimating the Effect of the Red Card in Soccer. *Journal of Quantitative Analysis in Sports*, 5: Iss. 1, Article 8.
- Wilks, D. (1995). *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series, Vol. 59, Academic Press, p. 467.
- William Hill PLC. (2012). Retrieved January 25, 2012, from About William Hill: <http://www.williamhillplc.com/wmh/about/>

Wing, C., Tan, K., & Yi, L. (2007). *The Use of Profits as Opposed to Conventional Forecast Evaluation Criteria to Determine the Quality of Economic Forecasts*. Singapore: Nanyang Business School, Nanyang Technological University, Nanyang Avenue.