Knowledge-Based Systems 36 (2012) 322–339

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

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

Anthony C. Constantinou *, Norman E. Fenton, Martin Neil

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

ARTICLE INFO

Article history:
Received 2 April 2012
Received in revised form 14 July 2012
Accepted 15 July 2012
Available online 1 August 2012

Keywords:
Bayesian reasoning
Football betting
Football predictions
Soccer predictions
Subjective information

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 $\geqslant$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'.

1. Introduction

Association Football (hereafter referred to simply as 'football') is the world's most popular sport [11,43,12], and constitutes the fastest growing gambling market [7]. 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 outcomes [36,4,35,26,27] or league positions [34], 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 [41,9,38,32]. A similar version is also reported in [10] 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 [53] in which the authors demonstrate profitability against Intertops (a bookmaker located in Antigua, West Indies), and refinements of this technique were later proposed in [8] 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, [37] considered team performance data as well as published bookmakers' odds, whereas [24,22] considered team quality, recent performance, match significance and geographical distance. Ref. [23] 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 [55], 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, [52] 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 [33] used a dynamic cumulative link model to generate ratings for top division football teams in Germany. The ELO rating

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

0950-7051/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2012.07.008

that was initially developed for assessing the strength of chess players [13] has been adopted to football [3]. In [29], 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. Ref. [40] 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 [28], ball possession [28], and red cards [51,56].[1]

Recently researchers have considered Bayesian networks and subjective information for football match predictions. In particular, [31] 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 [42] through simulating various predifined strategies along with subjective information, whereas in [2] 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 [49];
(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 [17]. The forecasts generated by the model were for season 2010/11, a total of 380 EPL matches.

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

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[2] that the particular team achieved during that particular season as shown in Table 1.

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 2006–2007 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 2009–2010 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 2010–2011. Because the historical performances of Manchester City and Aston Villa prior to season 2010–2011 were in no way representative of their strength in season 2010–2011, 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 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 pre-

[1] While this work falls within the scope of our interest, other empirical forecasting studies such as attendance demand [46–48,15,20], and the effectiveness of football tipsters [18] do not.

[2] 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.

dict 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 Fig. 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'.

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 [16], the Bayesian Network software we have used [1] automatically updates the forecast taking account of the subjective proximity as shows for different examples in Fig. 2. Fig. 3 illustrates how the four components are linked. We will refer to the revised (and final) forecasts as 'subjective forecasts'.

3.1. Component 1: Team strength

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

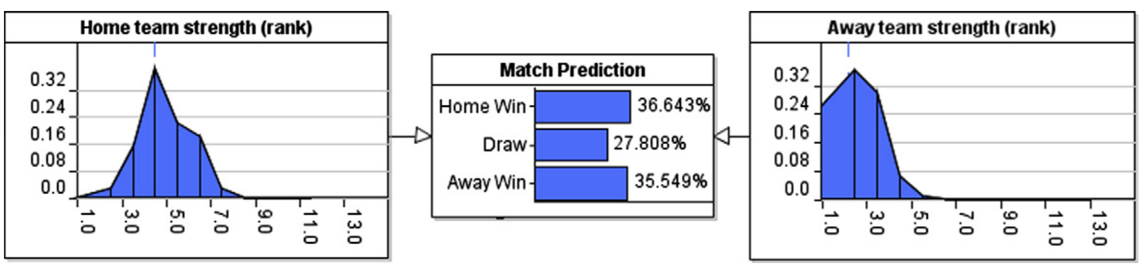

Fig. 1. An example of an objective forecast generated at component 1.

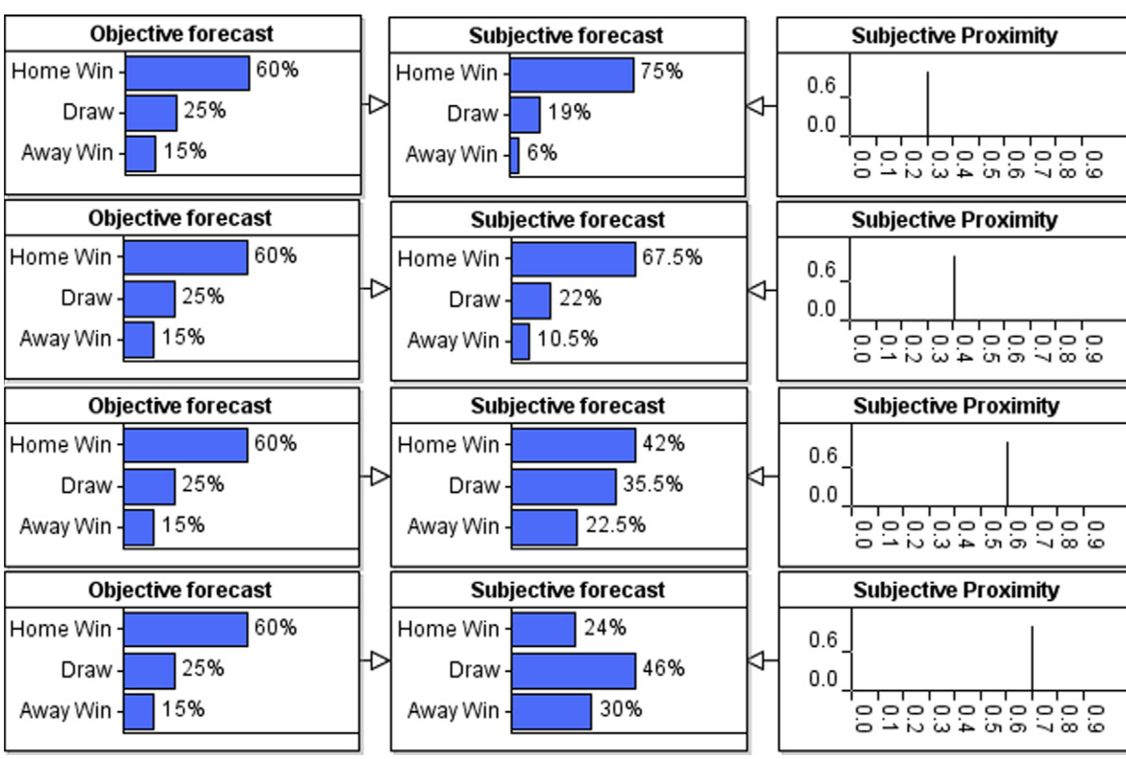

Fig. 2. Forecast revision given different indications of subjective proximity.

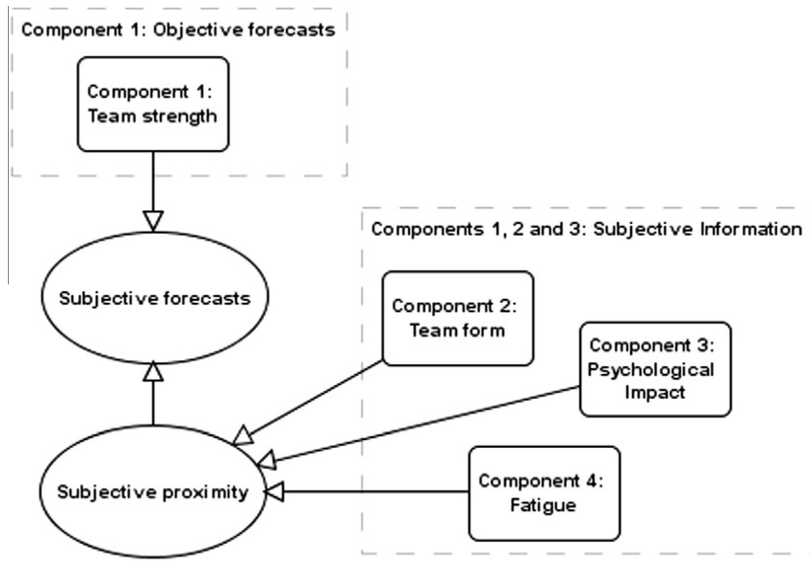

Fig. 3. How components 1, 2, 3 and 4 are linked.

(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[3] 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 [54].

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 Fig. 5a;
(b) *Current information*: this indication receives decreased rates of variance (and hence become more important) after each subsequent gameweek,[4] following the exponential decay illustrated in Fig. 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*[5] (Fig. 5a).

Further information regarding the variables and available scenarios of this process is provided in Table A.1. An example with observations from the actual match between Man City and Man United dated 10th of November 2010 is illustrated in Appendix B.

3.2. Component 2: Team form

This Bayesian network component is shown in Fig. 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[6] 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).[7] 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 A.2.

3.3. Component 3: Psychological impact

This Bayesian network component is shown in Fig. 7. The psychology of a team is determined by subjective indications

[3] 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.
[4] A complete EPL season consists of 38 gameweeks.
[5] 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 5th preceding season.
[6] Represented by what the model had initially forecasted.
[7] Form decreases if the team has new first-team injuries and increases when important players return back to action.

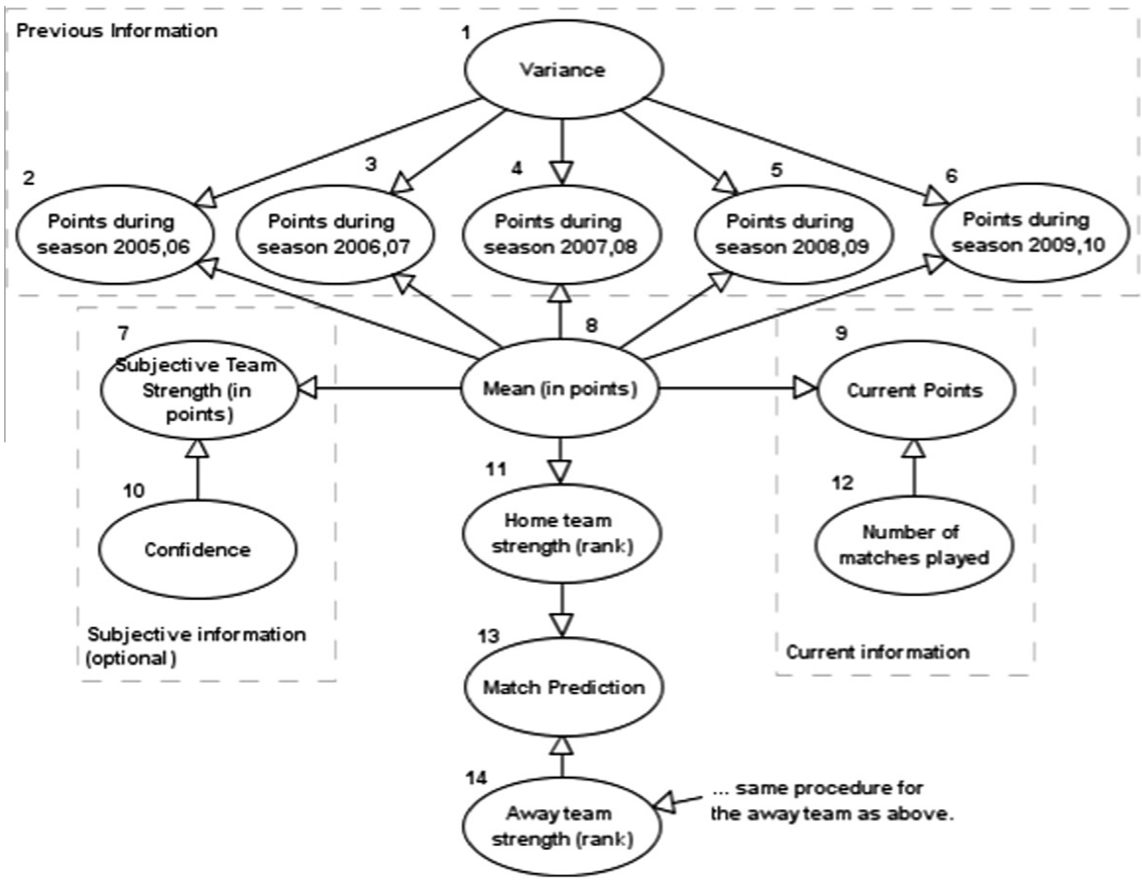

Fig. 4. Component 1: Non-symmetric Bayesian parameter learning network for measuring the strength of the two teams and generating objective match predictions.

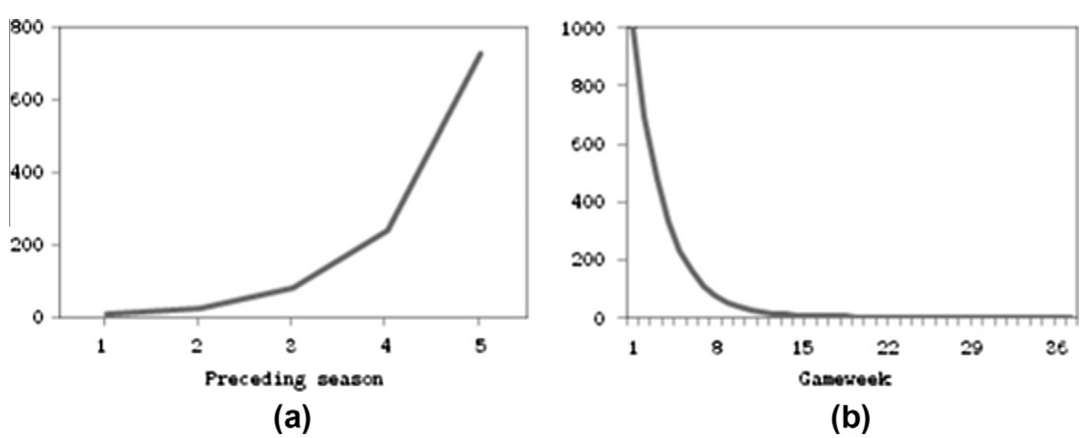

Fig. 5. Limited memory process achieved by exponential growth/decay rates of uncertainty for (a) the previous seasons and (b) the gameweeks played under the current season.

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 A.3.

3.4. Component 4: Fatigue

This Bayesian network component is shown in Fig. 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

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

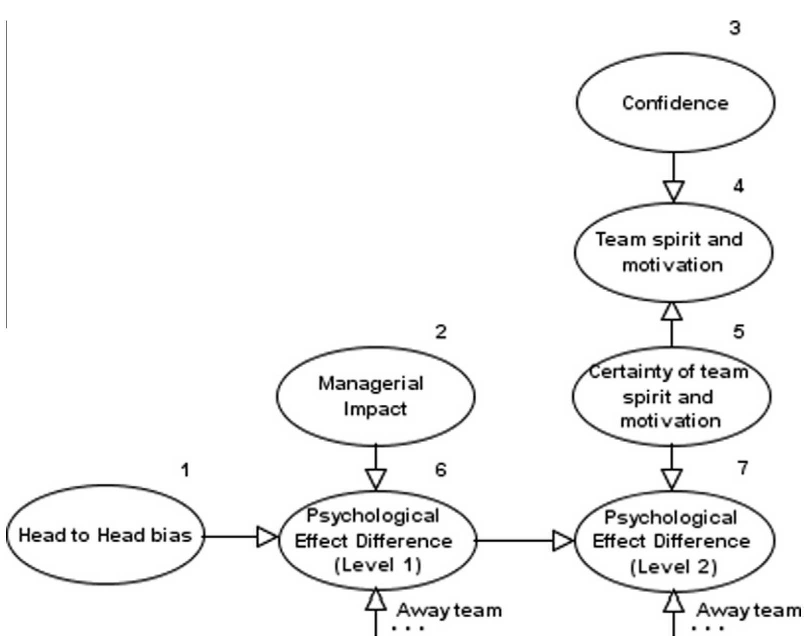

Fig. 7. Component 3: Expert constructed Bayesian network for estimating potential advantages in psychological impact between the two teams.

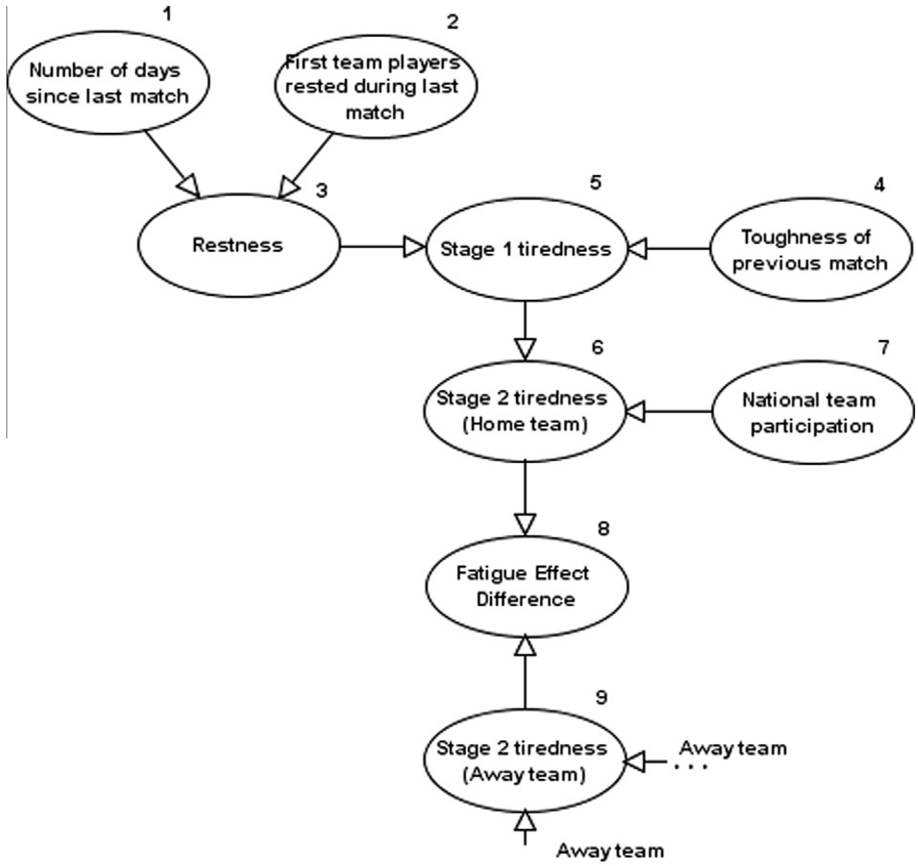

Fig. 8. Component 4: Expert constructed Bayesian network for estimating potential advantages in fatigue between the two teams.

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[8] (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[9] (nodes 6, 7). If there is no national team participation the fatigue will receive no increase. Further information regarding the variables and available scenarios of this process is provided in Table A.4.

4. Results and 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 [39] and that a combination of the two might be best [59]. 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 [14], and which has been described to be particularly appropriate in assessing both interval and ordinal scale probabilistic variables [45]. We explained why it was the most rational scoring rule of those that have been proposed and used for football outcomes in [6]. 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 [57], which is subject to a negative bias that is strongest for small ensemble size [30]. RPS is both strictly proper and sensitive to distance [44,45]. 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;

[8] Where (a) is defined to be twice as important to (b) when calculating ‘Restness’ (node 3).

[9] 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.

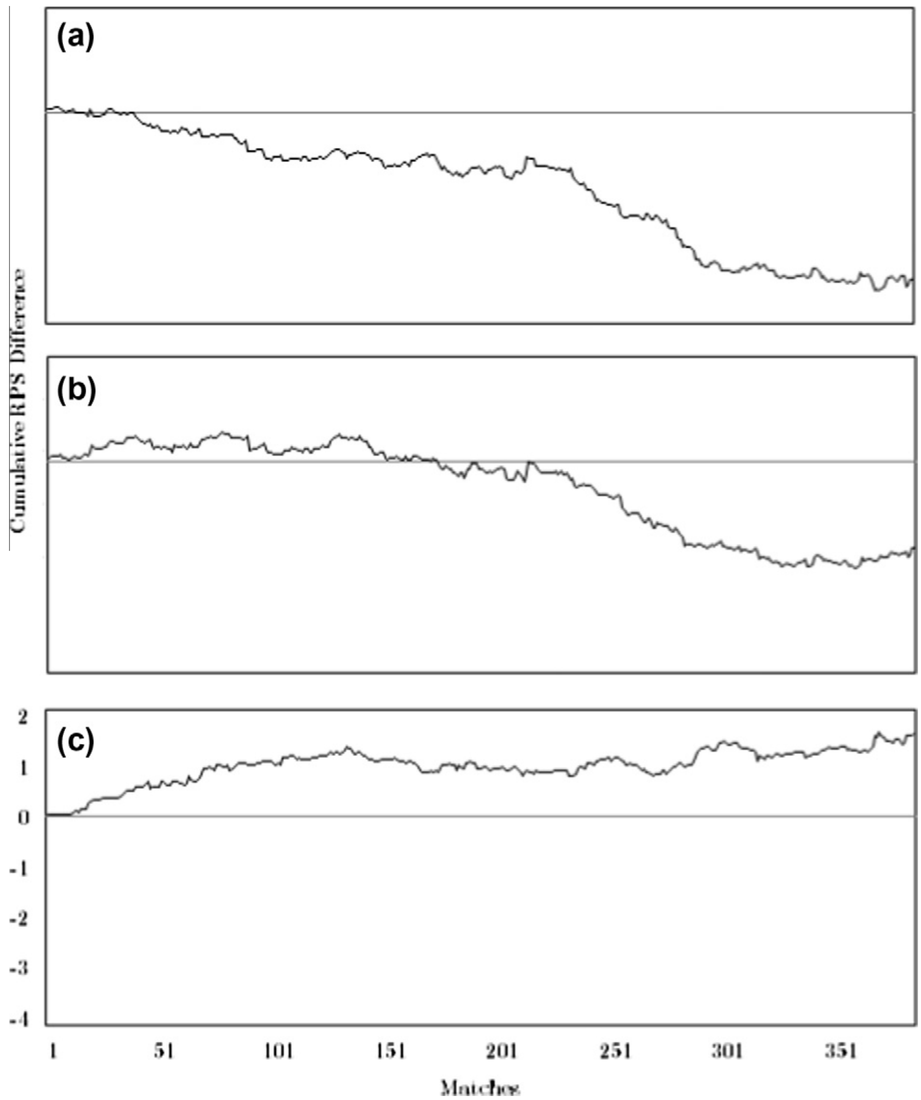

Fig. 9. Cumulative RPS difference when (a) $f_B - f_O$, (b) $f_B - f_S$, and (c) $f_O - f_S$.

(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[10] 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 [10,22,7]. However, instead of selecting a single bookmaker we make use of the mean[11] bookmakers' odds as provided by (Football-Data). Fig. C1 demonstrates the RPS generated per forecast under the three datasets.

Fig. 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;
(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.

[10] 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 [7].

[11] The mean odds are measured by considering a minimum of 28 and a maximum of 40 different bookmakers per match instance (Football-Data).

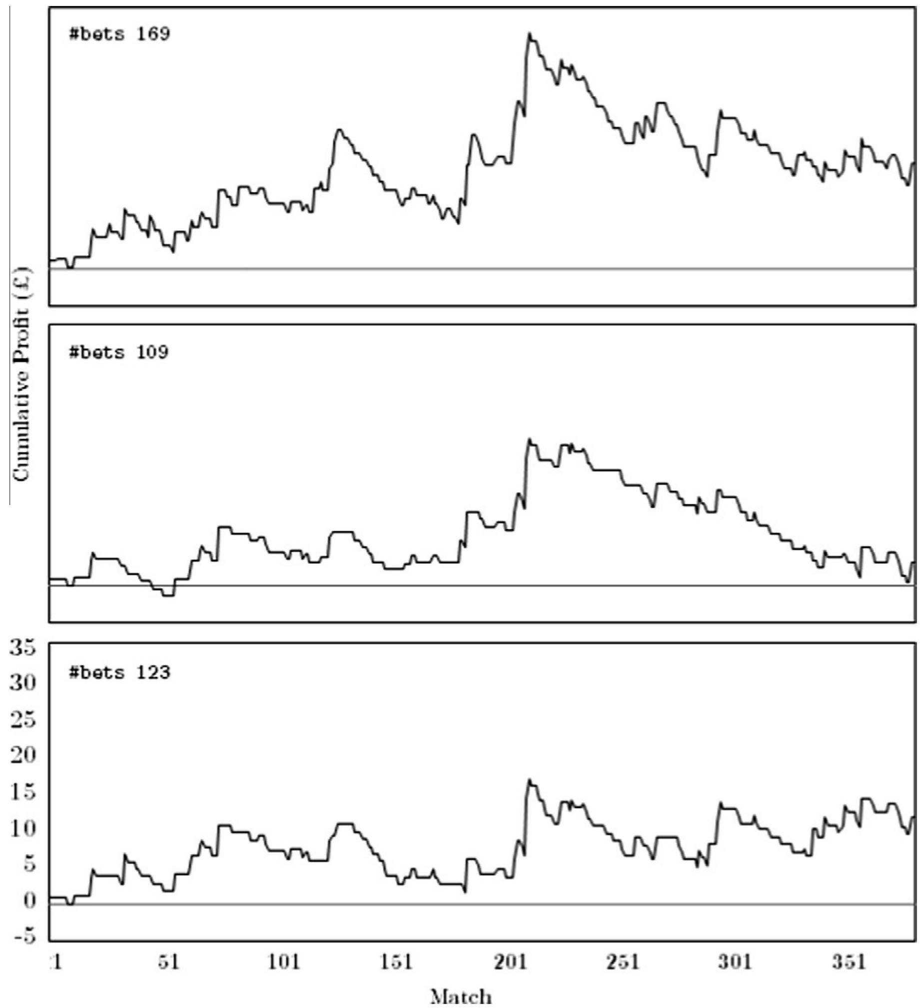

Fig. 10. Cumulative profit/loss observed given f_S when simulating the standard betting strategy at discrepancy levels of $\geqslant$5% against (a) f_{maxB}, (b) f_{meanB}, and (c) f_{WH}.

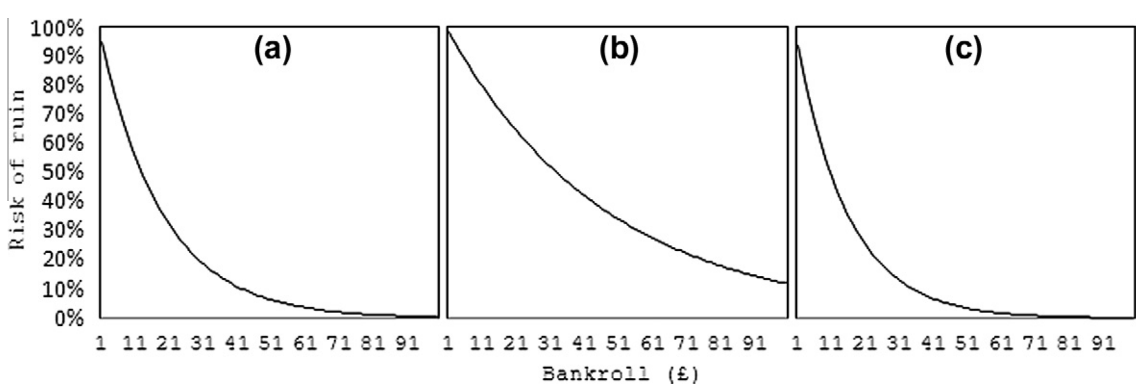

Fig. 11. Risk of Ruin given the specified betting simulation against (a) f_{maxB}, (b) f_{meanB}, and (c) f_{WH}.

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.[12] In contrast to previous papers [20,22], the work in [7] shows that the published odds of a single bookmaker are not representative of the overall market. Unlike the

[12] See also the following studies on the football gambling market: [50,9,37,53,19,10,24,21,25].

Table 2
Betting simulation stats given f_S against: (a) f_{maxB}, (b) f_{meanB}, and (c) f_{WH} at discrepancy levels of ⩾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

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[13]:

(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 [58].

Fig. 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.[14] Fig. 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 ⩽5% for f_{maxB} and f_{WH} respectively. In the case of f_{meanB} the profit rate is not high enough to ensure a risk of ruin ⩽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 [5,19,20,22,25,7]. 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%.

Table F.1 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 ⩾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 ⩾5% for all three components.

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[15] 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 and 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' underlying approach which generates predictions by only considering the strength of the two competing teams given results data and total

[13] The bookmakers' odds are also provided by (Football-Data).

[14] We have also performed the identical betting simulation given f_O. Fig. E.1 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.

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

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_O (cumulative profit increase/decrease)	Revised accuracy relative to f_O (cumulative profit increase/decrease)	Measures agree in the direction of the revision
1 (All components True)	All	40	22	+£16.56	−0.4823	True
2 (Exactly two components True)	Fatigue, psychology	31	13[a]	+£7.70[a]	−0.3394	True[a]
	Fatigue, form	9[a]	2[a]	−£2.60[a]	−0.2004[a]	False[a]
	Psychology, form	57	27	+£6.28	−0.4008	True
3 (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

[a] Sample size too small to contribute to conclusions.

Table A.1
Team strength (as presented in Fig. 2).

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

Table A.2
Team form (as presented in Fig. 3).

ID	Variable (node)	Description	Subjective scenarios
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]
II.	Secondary key-player availability	Expert indication regarding his confidence about the availability of the secondary key-player	[Very High, High, Medium, Low, Very Low]
III.	Tertiary key-player availability	Expert indication regarding his confidence about the availability of the tertiary key-player	[Very High, High, Medium, Low, Very Low]
IV.	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]
V.	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 Fig. 4).

ID	Variable (node)	Description	Subjective scenarios
I.	Team spirit and motivation	Expert indication regarding the team's level of motivation and team spirit	[Very High, High, Normal, Low, Very Low]
II.	Confidence	Expert indication regarding its confidence about his input in (I)	[Very High, High, Medium, Low, Very Low]
III.	Managerial impact	Expert indication regarding the impact of the current managerial situation	[Very High, High, Normal, Low, Very Low]
IV.	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 Fig. 5).

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

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 [9], 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 [25] 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 [7]. 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 [49] 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 (8–11%) 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 [7]. 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)

See Tables A.1, A.2, A.3 and A.4.

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

Fig. B.1 presents a real component 1 example between Manchester City (home team) and Manchester United, as prepared for the 11th 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,

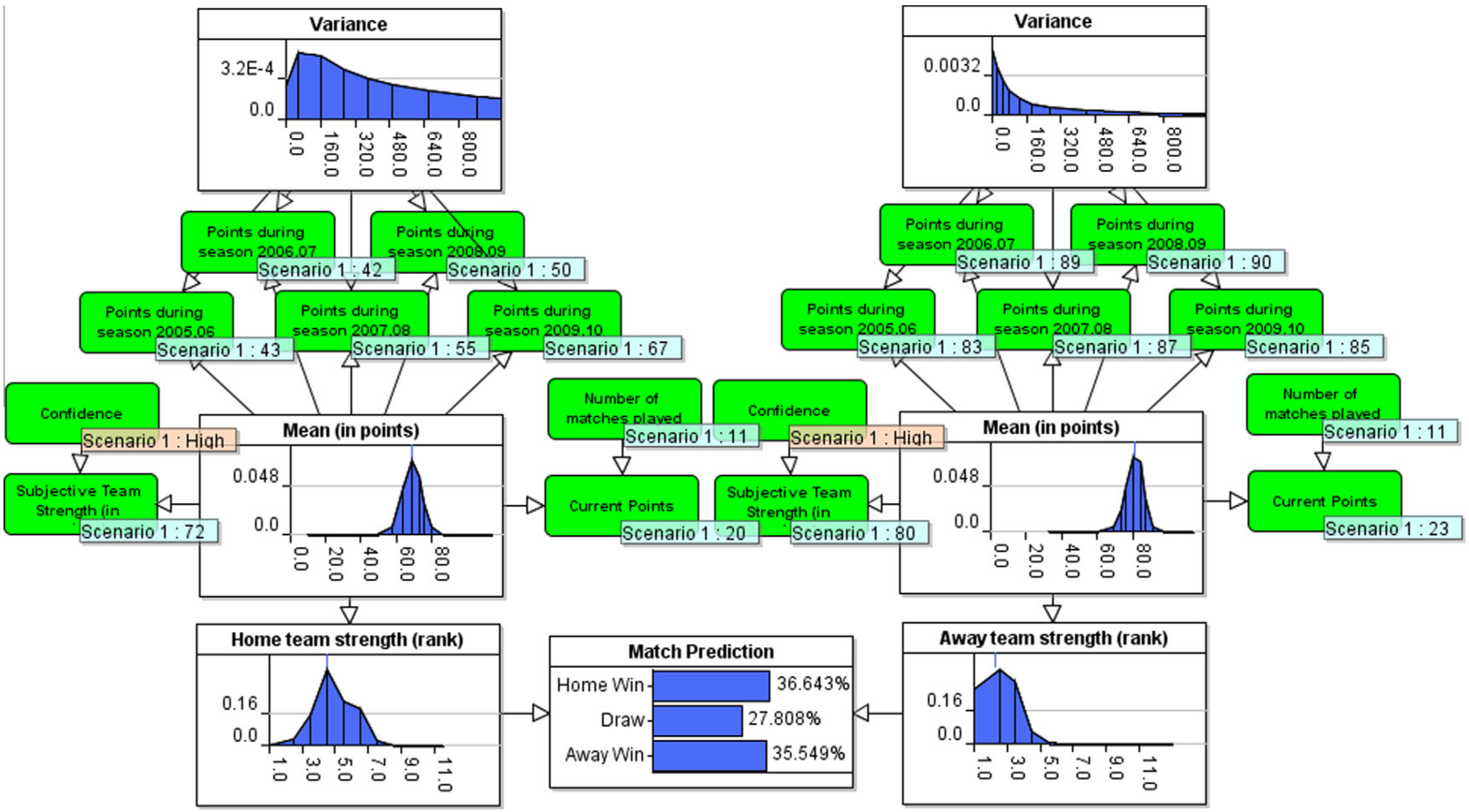

Fig. B.1. An actual example of the Bayesian network (from Fig. 2) at component 1. The parameters represent the actual observations provided from the Man City versus Man United match, 10th of November, 2010.

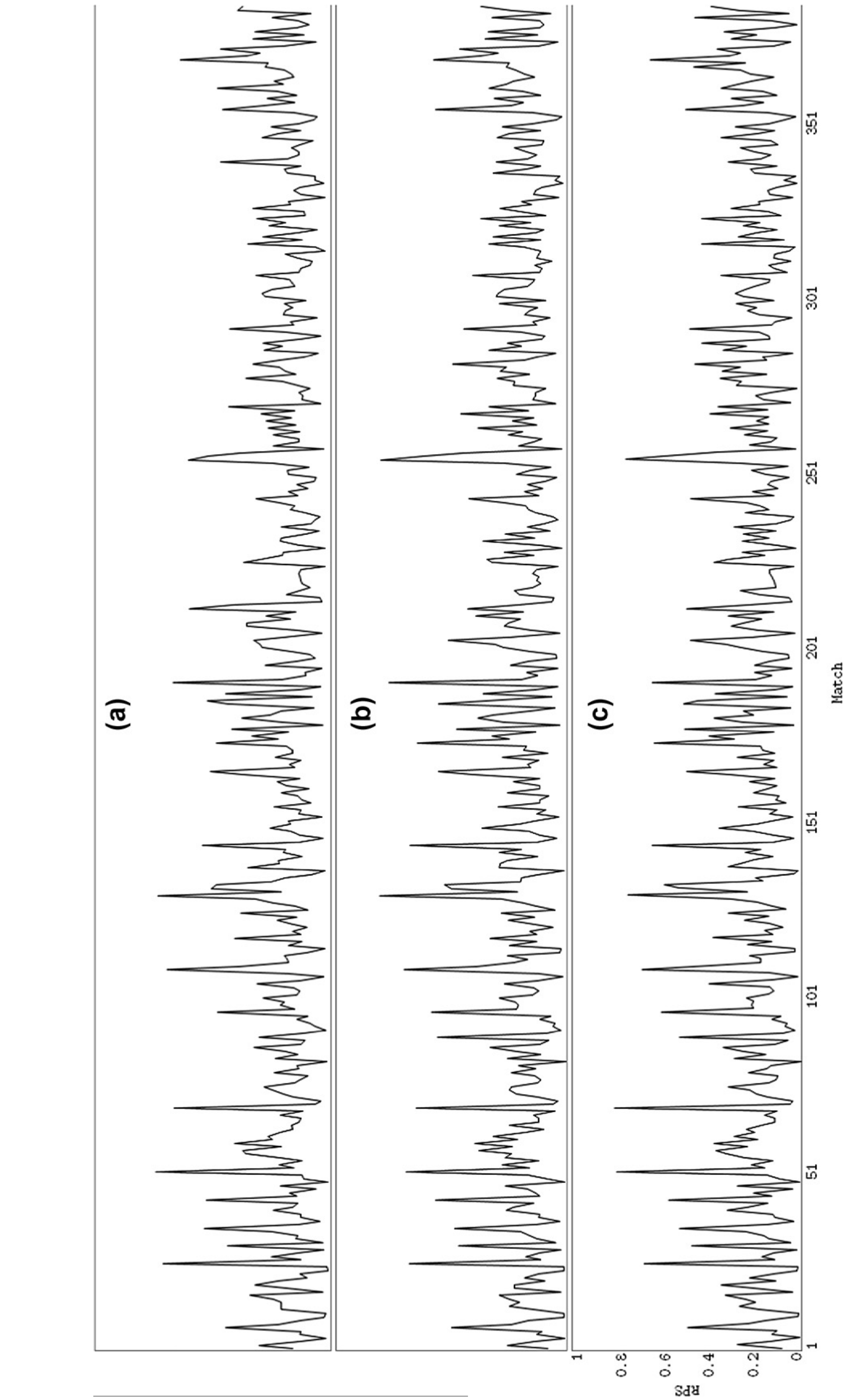

Fig. C.1. RPS per match for datasets f_O (a), f_S (b), and f_B (c) respectively.

Table D.1
RPS discrepancies $\geqslant$0.1 between objective (f_O) and revised (f_S); ranked by highest discrepancy.

RPS discrep.	Date	Home team	Away 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	×

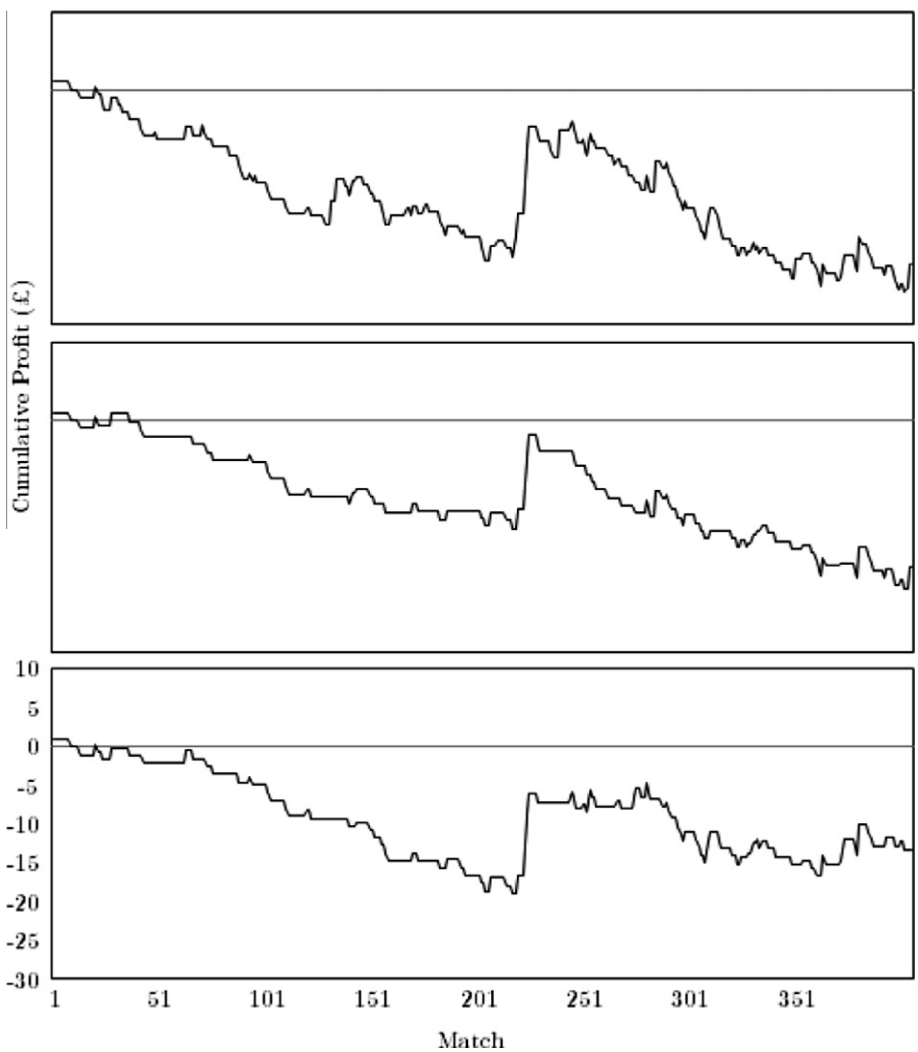

Fig. E.1. Cumulative profit/loss observed given f_O when simulating the standard betting strategy at discrepancy levels of $\geqslant$5% against: (a) f_{maxB}, (b) f_{meanB}, and (c) f_{WH}.

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").[16] On the other hand, we have introduced a 5-point decrease for Man United with "High" confidence.[17] Accordingly, the inputs were [72, 'High'] and [80, 'High'] for Man City and Man United respectively.

Table F.1
Betting simulation stats given f_S against (a) 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

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

(continued on next page)

Table G.1 (*continued*)

Date	Home team	Away team	Result	Objective (f_O)			Subjective (f_S)		
				p(H)	p(D)	p(A)	p(H)	p(D)	p(A)
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

(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.

Appendix C. Match RPS per dataset

See Fig. C.1.

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 D.1 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.

Appendix E. Betting simulation given objective forecasts

See Fig. E.1.

Appendix F. Betting simulation at different levels of discrepancy given f_S

See Table F.1.

Appendix G. Example of forecasts generated by pi-football

See Table G.1.

References

[1] Agena, Agena: Bayesian Network and Simulation Software for Risk Analysis and Decision Support, 2012. <http://www.agenarisk.com> (retrieved 26.01.12).
[2] G. Baio, M. Blangiardo, Bayesian hierarchical model for the prediction of football results, Journal of Applied Statistics 37 (2) (2010) 253–264.
[3] J. Buchdahl, (2003). Fixed Odds Sports Betting: Statistical Forecasting and Risk Management. High Stakes.
[4] A. Buchner, W. Dubitzky, A. Schuster, P. Lopes, P. O'Doneghue, J. Hughes, et al., 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, 1997.
[5] M. Cain, D. Law, D. Peel, The favourite-loghshot bias and market efficiency in UK football betting, Scottish Journal of Political Economy 47 (2000) 25–36.
[6] A.C. Constantinou, N.E. Fenton, Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models, Journal of Quantitative Analysis in Sports 8 (1) (2012) (article 1).
[7] A.C. Constantinou, N.E. Fenton, Evidence of an (intended) inefficient Association Football gambling market, 2012. <http://www.constantinou.info/downloads/papers/evidenceOfInefficiency.pdf> (Under Review).
[8] M. Crowder, M. Dixon, A. Ledford, M. Robinson, Dynamic modelling and prediction of English football league matches for betting, The Statistician 51 (2002) 157–168.
[9] M. Dixon, S. Coles, Modelling association football scores and inefficienies in the football betting market, Applied Statistics 46 (1997) 265–280.
[10] M. Dixon, P. Pope, The value of statistical forecasts in the UK association football betting market, International Journal of Forecasting 20 (2004) 697–711.
[11] E.G. Dunning, A. Joseph, R.E. Maguire, The Sports Process: A Comparative and Developmental Approach, Champaign: Human Kinetics, 1993, p. 129.
[12] E. Dunning, Sport Matters: Sociological Studies of Sport, Violence and Civilisation, Routledge, London, 1999.
[13] A.E. Elo, The Rating of Chess Players, Past and Present, Arco Publishing, New York, 1978.
[14] E. Epstein, A scoring system for probability forecasts of ranked categories, Journal of Applied Meteorology 8 (1969) 985–987.
[15] M. Falter, C. Perignnon, Demand for football and intra-match winning probability: an essay on the glorious uncertainty of sports, Applied Economics 32 (2000) 1757–1765.
[16] N.E. Fenton, M. Neil, J.G. Caballero, Using ranked nodes to model qualitative judgments in Bayesian networks, IEEE TKDE 19 (10) (2007) 1420–1432.
[17] Football-Data. (n.d.). Football-Data.co.uk. Football Results, Statistics & Soccer Betting Odds Data: <http://www.football-data.co.uk/englandm.php> (retrieved 25.08.11).
[18] D. Forrest, R. Simmons, Forecasting sport: the behaviour and performance of football tipsters, International Journal of Forecasting 16 (2000) 317–331.
[19] D. Forrest, R. Simmons, Globalisation and Efficiency in the Fixed-Odds Soccer Betting Market, University of Salford, Salford, Mimeo, 2001.
[20] D. Forrest, R. Simmons, Outcome uncertainty and attendance demand in sport: the case of English soccer, The Statistician 2 (2002) 241–291.
[21] D. Forrest, R. Simmons, Sentiment in the betting market on Spanish football, Applied Economics 40 (2008) 119–126.
[22] D. Forrest, J. Goddard, R. Simmons, Odds-setters as forecasters: the case of English football, International Journal of Forecasting 21 (2005) 551–564.
[23] J. Goddard, Regression models for forecasting goals and match results in association football, International Journal of Forecasting 21 (2005) 331–340.
[24] J. Goddard, I. Asimakopoulos, Forecasting football results and the efficiency of fixed-odds betting, Journal of Forecasting 23 (2004) 51–66.
[25] I. Graham, H. Stott, Predicting bookmaker odds and efficiency for UK football, Applied Economics 40 (2008) 99–109.
[26] F. Halicioglu, Can we predict the outcome of the international football tournaments?: the case of Euro 2000, Doğuş Üniversitesi Dergisi 6 (2005) 112–122.
[27] F. Halicioglu, Forecasting the professional team sporting events: evidence from Euro 2000 and 2004 football tournaments, in: 5th International Conference on Sports and Culture: Economic, Management and Marketing Aspects, Athens, Greece, 2005, pp. 30–31.
[28] N. Hirotsu, M. Wright, An evaluation of characteristics of teams in association football by using a Markov process model, The Statistician 52 (4) (2003) 591–602.

[29] L.M. Hvattum, H. Arntzen, Using ELO ratings for match result prediction in association football, International Journal of Forecasting 26 (2010) 460–470.
[30] I.T. Jolliffe, D.B. Stephenson, Forecast Verification. A Practitioner's Guide in Atmospheric Science, Wiley, West Sussex, England, 2003.
[31] A. Joseph, N. Fenton, M. Neil, Predicting football results using Bayesian nets and other machine learning techniques, Knowledge-Based Systems 7 (2006) 544–553.
[32] D. Karlis, I. Ntzoufras, Analysis of sports data by using bivariate Poisson models, The Statistician 52 (3) (2003) 381–393.
[33] L. Knorr-Held, Dynamic rating of sports teams, The Statistician 49 (2000) 261–276.
[34] R. Koning, Balance in competition in Dutch soccer, The Statistician 49 (3) (2000) 419–431.
[35] R. Koning, M. Koolhaas, G. Renes, G. Ridder, A simulation model for football championships, European Journal of Operational Research 148 (2003) 268–276.
[36] D. Kuonen, Statistical Models for Knock-Out Soccer Tournaments. Technical Report, Department of Mathematics, Ècole Polytechnique Federale de Lausanne, 1996.
[37] T. Kuypers, Information and efficiency: an empirical study of a fixed odds betting market, Applied Economics 32 (2000) 1353–1363.
[38] A.J. Lee, Modeling scores in the Premier League: is Manchester United really the best?, Chance 10 (1997) 15–19
[39] G. Leitch, J.E. Tanner, Economic forecast evaluation: profits versus the conventional error measures, American Economic Association (1991) 580–590.
[40] C. Leitner, A. Zeileis, K. Hornik, Forecasting sports tournaments by ratings of (prob)abilities: a comparison for the EURO 2008, International Journal of Forecasting 26 (2010) 471–481.
[41] M.J. Maher, Modelling association football scores, Statististica Neerlandica 36 (1982) 109–118.
[42] B. Min, J. Kim, C. Choe, H. Eom, R.B. McKay, A compound framework for sports results prediction: a football case study, Knowledge-Based Systems 21 (2008) 551–562.
[43] F.O. Mueller, R.C. Cantu, S.P. Camp, Catastrophic Injuries in High School and College Sports, Human Kinetics, Champaign, 1996. p. 57.
[44] A. Murphy, On the ranked probability score, Journal of Applied Meteorology 8 (1969) 988–989.
[45] A. Murphy, The ranked probability score and the probability score: a comparison, Monthly Weather Review 98 (1970) 917–924.
[46] D.A. Peel, D.A. Thomas, Outcome uncertainty and the demand for football, Scottish Journal of Political Economy 35 (1989) 242–249.
[47] D.A. Peel, D.A. Thomas, The demand for football: some evidence on outcome uncertainty, Empirical Economics 4 (1992) 567–570.
[48] D.A. Peel, D.A. Thomas, Handicaps, outcome uncertainty and attendance demand, Applied Economic Letters 4 (1997) 567–570.
[49] pi-football, Probabilistic Intelligence in Football, Forecasts for season 2010/11, 2010. http://www.pi-football.com/past_predictions.aspx> (retrieved 25.05.12).
[50] P. Pope, D. Peel, Information, prices and efficiency in a fixed-odds betting market, Economica 56 (1989) 323–341.
[51] G. Ridder, J. Cramer, P. Hopstaken, Estimating the effect of a red card in soccer, Journal of the American Statistical Association 89 (427) (1994) 1124–1127.
[52] A. Rotshtein, M. Posner, A. Rakytyanska, Football predictions based on a fuzzy model with genetic and neural tuning, Cybernetics and Systems Analysis 41 (4) (2005) 619–630.
[53] H. Rue, O. Salvesen, Prediction and retrospective analysis of soccer matches in a league, The Statistician 3 (2000) 339–418.
[54] Soccer Base, Soccer Base. From Man City: Full Transfer History, 2012. <http://www.soccerbase.com/teams/team.sd?team_id=1718&teamTabs=transfers> (retrieved 23.02.12).
[55] A. Tsakonas, G. Dounias, S. Shtovba, V. Vivdyuk, Soft computing-based result prediction of football games, in: The First International Conference on Inductive Modelling (ICIM'2002), Lviv, Ukraine, 2002.
[56] J. Vecer, F. Kopriva, T. Ichiba, Estimating the effect of the red card in soccer, Journal of Quantitative Analysis in Sports 5 (1) (2009) (article 8).
[57] D. Wilks, Statistical methods in the atmospheric sciences, International Geophysics Series, vol. 59, Academic Press, 1995. p. 467.
[58] William Hill PLC, From About William Hill, 2012. <http://www.williamhillplc.com/wmh/about/> (retrieved 25.01.12).
[59] C. Wing, K. Tan, L. Yi, The Use of Profits as Opposed to Conventional Forecast Evaluation Criteria to Determine the Quality of Economic Forecasts, Nanyang Business School, Nanyang Technological University, Nanyang Avenue, Singapore, 2007.