



A ratings based Poisson model for World Cup soccer simulation

D Dyte and SR Clarke*

Swinburne University, Hawthorn, Australia

In this paper a method is suggested for predicting the distribution of scores in international soccer matches, treating each team's goals scored as independent Poisson variables dependent on the Fédération Internationale de Football Association (FIFA) rating of each team, and the match venue. The results of a Poisson regression to estimate parameters for this model were used to simulate matches played during the 1998 World Cup tournament. For the model to be a more effective predictor, some manual adjustments must be made to the ratings data. The predictions of the model were placed on a web page to create interest in applications of mathematics, and proved popular with the general public.

Keywords: sports; stochastic processes; simulation; soccer

Introduction

The World Cup soccer tournament, held every four years, is the most watched sports event—bigger even than the Olympic Games. Some 192 countries participate in qualifying matches over a two year period, with the winners of group matches from around the world playing off in the final tournament. The 1998 World Cup tournament was held in France, and for the first time, an expanded tournament included 32 teams.

The teams were divided into eight groups of four. Teams were allocated to groups by random draw, one from each of four pools arranged according to FIFAs estimate of each team's standard of play. Each group played six round robin matches. The top two teams from each group advanced to the second round, being the first part of a simple 16 team knockout arrangement. A third place playoff was also held for the losers of the semi finals, so that 64 matches were played in total. Host nation France won its first World Cup, defeating Brazil, the highest ranked team in the official FIFA standings, 3–0 in the final. The most recent host nations to win the World Cup prior to 1998 were England in 1966, West Germany in 1974, and Argentina in 1978.

Just prior to the start of the event, we decided to provide forecasts on our web site of the individual matches and the tournament as a whole. One of the purposes of our Sports Statistics Web site is to demonstrate the applications of mathematics and statistics in an area of interest to the general population. An event such as the World Cup is the

perfect vehicle, and methods for generating team ratings suitable for forecasting match results have been outlined in the literature.^{1,2} While confident these could be adapted to our purpose, to do so would require regular collection of results and maintenance of the team ratings. The time and effort required could not be justified. How useful were the official ratings? In a similar manner to a simulation of the Wimbledon tennis tournament described in Clarke and Dyte,³ we attempted to add value to the official ratings for an international sport by using them to generate a model for the outcome of matches. In this case, we used the official Coca-Cola FIFA ratings to predict outcomes of matches in the 1998 World Cup tournament. Individual match predictions were then used to run a complete simulation, which provided predictions for the outcome of the tournament as a whole.

Stefani^{4,5} provides a summary and analysis of the official rating systems used by the governing bodies of skiing, tennis, soccer and golf. He finds that each of these ratings relies primarily on event outcomes rather than actual scores, although the soccer ratings do account for scores to a small degree. Often the rating systems reflect more than a simple attempt to list competitors in order of skills. Tennis, in particular, has rating systems designed as much to encourage high player participation as to identify the best players.

The FIFA ratings are calculated using the results of all soccer matches between senior national teams. Points are allocated for wins, draws and losses, based on the relative strength of the teams before the match. Bonus points are awarded for away wins and goals scored. Also, a multiplication factor is used to reflect the importance of the match (World Cup finals matches count highest) and for

*Correspondence: Dr SR Clarke, School of Mathematical Sciences, Swinburne University of Technology, PO Box 218, Hawthorn 3122, Australia. E-mail: sclarke@swin.edu.au

each of the six continental federations, to reflect their relative strength. Points attained in past matches are gradually reduced to zero over a period of eight years. This stands in contrast to the Association of Tennis Professionals (ATP) and Women's Tennis Association (WTA) tennis ratings, in which points retain their full value for exactly one year, and are then erased.

We wish to gauge, then, the worth of the FIFA ratings in a predictive sense. Using the 1998 World Cup tournament as a test, we derive a probability distribution for the goals scored by a team in each match and hence the overall score for the match, and use these to simulate the entire tournament. Results were updated daily, to account for matches already played, and published on the Swinburne University Sports Statistics world wide web site at <http://www.swin.edu.au/sport/>.

Goal predictions: the basic model

The model for this simulation rests on two assumptions:

- The number of goals scored by a team in a soccer match is *Poisson distributed*
- It is *independent* of the number of goals scored by the opposing team

These two assumptions are by no means taken for granted in previous work. Norman⁶ gives a good summary of previous attempts to model soccer scores. Pollard *et al*⁷ argue for a negative binomial distribution of goals based on no assumptions about team strengths. However Maher⁸ suggests that if we find an appropriate method of calculating estimated mean scores for each possible pairing of teams, a Poisson distribution gives a good fit to the observed data.

Dixon and Robinson⁹ study some 4000 matches in English club competitions, and model goals scored as two interacting birth processes using such parameters as attacking and defensive strength, match venue, current game score, and time to play in the match. The results model the data very well, including a holdout sample of some 600 matches of the 1995–96 season. Such detail is, however, beyond the scope of the model specified here. We merely seek to measure, in some sense, the predictive power of the official FIFA rating system, rather than uncover the deeper mechanics underlying a complete soccer match. In this instance, an independent Poisson model certainly has the advantage of simplicity. Lee¹⁰ uses such a model to obtain team scores in a simulation of a season's UK Association football. He found no evidence against the assumption of independence.

Having settled on the assumptions of Poisson distribution and independence, data were collected of the monthly FIFA ratings over the 12 months leading up to the 1998 World Cup tournament in France, and all A international soccer matches played in this time.

Therefore, for each team in each match, we were able to tabulate:

Current FIFA rating
Opponent's current FIFA rating
Goals scored
Venue status (home/away/neutral)

The final data set consisted of some 477 matches, or 954 observations.

Using the SAS/STAT package, a Poisson regression was performed on the number of goals scored by each team in each match, using as an expression for the mean:

$$\ln(m) = a + b \cdot TR + c \cdot OR + v$$

where, m is the expected number of goals scored

TR is the team's FIFA rating

OR is the opponent's FIFA rating

v is a parameter which changes according to venue (home, away or neutral)

The parameter estimates obtained from the data were:

$$\hat{a} = 0.1193$$

$$\hat{b} = 0.0218$$

$$\hat{c} = -0.0246$$

$$\hat{v}_{\text{home}} = 0.3462$$

$$\hat{v}_{\text{neutral}} = 0.2885$$

Note that the model was arranged so that v_{away} identically equal to zero. Obviously, in this tournament, the only team which benefited from home advantage was the eventual winner, France. Note that the parameter change for a team when moving from a neutral venue to a home venue (+0.06) is much smaller than the effect on their opponent who moves from a neutral venue to an away venue (−0.29). This indicates that home advantage in soccer, at least for the data set used, applies more to defence than attack. The negative effect on the away team's goal scoring rate is much greater than the positive effect on the home team's goal scoring rate.

It is also interesting to note that $\hat{b} \approx -\hat{c}$, which implies that a simplified formula for the mean number of goals scored may apply:

$$\ln(m) = a + b \cdot (TR - OR) + v$$

In this version, the difference between two teams' ratings is sufficient to specify the expected number of goals for either team. Clarke and Dye³ uses such a scheme for predicting the outcome of tennis sets using only the difference between players' ATP ratings. However, for the more complicated matter of predicting two teams' scores, thought to be independently distributed, the more complex model was chosen.

Match prediction

The opening match of the tournament, between Brazil and Scotland, provides a convenient example of how the method works.

$$R_{\text{Scotland}} = 48.02$$

$$R_{\text{Brazil}} = 71.75$$

$$\hat{m}_{\text{Scotland}} = e^{0.1193 + (0.0218 \times 48.02) - (0.0246 \times 71.75) + 0.2885} = 0.73$$

$$\hat{m}_{\text{Brazil}} = e^{0.1193 + (0.0218 \times 71.75) - (0.0246 \times 48.02) + 0.2885} = 2.20$$

So we expected Scotland to score 0.73 goals to Brazil's 2.20. Using these as means gives the marginal probabilities for each team's Poisson distribution of goals scored, and using independence they can be multiplied to give the probability of each individual match result as shown in Table 1 (for probabilities 0.001 or more only). Of course, such tables are difficult for some soccer fans to interpret. This being the case, when these analyses were published on the web, we also provided win and draw probabilities, and the most likely score. In this case:

| | |
|---------------|-----|
| Brazil win: | 71% |
| Scotland win: | 11% |
| Draw: | 18% |

Most likely exact score: Brazil 2—Scotland 0 (12.9%).

These probabilities were generated from Table 1 by summing the upper right, lower left, and diagonal elements of the table respectively. For the record, the actual result was a 2–1 win to Brazil, a result we had allocated a probability of 9.4%.

This table illustrates the basic problem of soccer prediction. There are approximately 20 plausible outcomes shown in the table. Although soccer fans might want one of these predicted with a high degree of certainty, natural variation in the scores makes such a task impossible. In practice, this caused a degree of complaint when the predictions were published on the Swinburne web site, from people who simply read the most likely score and took little notice of the comparatively small probability attached to it.

Tournament prediction

Using the latest FIFA ratings to calculate the expected number of goals estimated by the regression analysis, it was then possible to generate two Poisson random variables for every game, and run a simulation of the entire tournament. Under the World Cup regulations, extra time periods of up to 30 minutes were played after drawn matches in the knockout stages, with the first goal (or golden goal) deciding the match. In the simulation, these extra periods were treated as a similar Poisson process, but with reduced mean to cater for the reduced time. Any matches that passed the extra time stage were decided by a penalty shootout. In the simulation, these were simply treated as a coin toss.

The advantages of simulation, as opposed to direct calculation of probabilities, are readily apparent in this case. Merely calculating win probabilities within groups in the first round is insufficient, as goal difference and other tie breakers make the exact score a requirement. Extra data are easy to gather, such as the probability of making the semi final stage, or the probability of a particular pair of teams meeting in the final. All this information is of interest to both the casual observer and those who may wish to stake money on tournament outcomes.

The FIFA ratings are by no means a perfect measure of a team's ability—due to the fact that several nations from lesser known federations had played many games in their own area and few outside, and appear to have acquired ratings above their true ability. An arbitrary decision was made to scale these ratings down by a uniform factor before running the simulation. The countries were USA, Iran, Saudi Arabia, Japan, South Korea and Jamaica. Each rating was multiplied by 0.85. The figure of 0.85 was chosen prior to the tournament, on a purely subjective basis, testing a number of figures until the teams appeared to be ranked correctly in the overall list of 32 competing nations. These are the adjusted ratings shown in Table 2.

The decision to adjust the official FIFA ratings was made with some reluctance, and analyses based on both adjusted and unadjusted ratings are given later. Although an accurate simulation is a desirable outcome, the exercise was also intended to demonstrate the utility of the official system as it stands.

Table 1 Probability distribution of scores for Brazil vs Scotland

| | | <i>Brazil</i> | | | | | | | | | <i>Total</i> |
|-----------------|-------|---------------|-------|-------|-------|-------|-------|-------|-------|-------|--------------|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| <i>Scotland</i> | 0 | 0.053 | 0.117 | 0.129 | 0.095 | 0.052 | 0.023 | 0.008 | 0.003 | 0.001 | 0.480 |
| | 1 | 0.039 | 0.086 | 0.094 | 0.069 | 0.038 | 0.017 | 0.006 | 0.002 | 0.001 | 0.352 |
| | 2 | 0.014 | 0.031 | 0.035 | 0.025 | 0.014 | 0.006 | 0.002 | 0.001 | 0.000 | 0.129 |
| | 3 | 0.003 | 0.008 | 0.008 | 0.006 | 0.003 | 0.002 | 0.001 | 0.000 | 0.000 | 0.032 |
| | 4 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.006 |
| | 5 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 |
| | Total | 0.110 | 0.243 | 0.268 | 0.197 | 0.109 | 0.048 | 0.018 | 0.006 | 0.002 | |

Table 2 Simulation results during the tournament—number of tournament wins for each team in 10 000 runs

| <i>Team</i> | <i>Rating (adjusted)</i> | <i>Pre-Cup</i> | <i>One game played</i> | <i>Two games played</i> | <i>Rd 1 played</i> | <i>Rd 2 played</i> | <i>QF played</i> | <i>SF played</i> |
|--------------|--------------------------|----------------|------------------------|-------------------------|--------------------|--------------------|------------------|------------------|
| France | 56.1 | 892 | 1075 | 1169 | 1007 | 1563 | 2741 | 3811 |
| Brazil | 71.8 | 2215 | 2299 | 2665 | 2658 | 3553 | 5027 | 6189 |
| Croatia | 55.6 | 252 | 297 | 352 | 305 | 521 | 1239 | |
| Netherlands | 54.2 | 174 | 139 | 228 | 233 | 433 | 993 | |
| Germany | 64.8 | 1063 | 1186 | 1129 | 1025 | 1857 | | |
| Argentina | 60.0 | 488 | 585 | 603 | 584 | 1071 | | |
| Italy | 56.7 | 257 | 288 | 352 | 395 | 644 | | |
| Denmark | 52.9 | 158 | 200 | 268 | 321 | 358 | | |
| Mexico | 61.4 | 616 | 663 | 617 | 677 | | | |
| England | 60.1 | 458 | 580 | 392 | 605 | | | |
| Yugoslavia | 59.4 | 499 | 480 | 552 | 590 | | | |
| Norway | 59.7 | 404 | 362 | 216 | 588 | | | |
| Chile | 59.0 | 431 | 395 | 335 | 522 | | | |
| Romania | 55.0 | 168 | 262 | 322 | 304 | | | |
| Paraguay | 52.5 | 152 | 110 | 142 | 176 | | | |
| Nigeria | 36.9 | 0 | 5 | 8 | 10 | | | |
| Colombia | 58.2 | 323 | 148 | 152 | | | | |
| Spain | 56.4 | 291 | 140 | 97 | | | | |
| Morocco | 57.2 | 257 | 265 | 178 | | | | |
| South Africa | 54.8 | 210 | 101 | 52 | | | | |
| Tunisia | 55.1 | 203 | 75 | 0 | | | | |
| Austria | 51.6 | 100 | 61 | 61 | | | | |
| Bulgaria | 50.5 | 94 | 91 | 17 | | | | |
| Belgium | 50.2 | 62 | 58 | 55 | | | | |
| Japan | 57.7 (49.1) | 57 | 27 | 0 | | | | |
| USA | 58.1 (49.4) | 53 | 34 | 0 | | | | |
| South Korea | 55.4 (47.1) | 39 | 18 | 0 | | | | |
| Scotland | 48.0 | 23 | 23 | 30 | | | | |
| Cameroon | 45.5 | 21 | 18 | 2 | | | | |
| Jamaica | 52.3 (44.4) | 21 | 8 | 0 | | | | |
| Saudi Arabia | 50.8 (43.2) | 10 | 3 | 0 | | | | |
| Iran | 47.6 (40.5) | 9 | 4 | 6 | | | | |

It may also be true that Nigeria's very poor FIFA rating at the beginning of the tournament was misleading, but no adjustment was made for this. Although highly rated by media analysts, Nigeria had performed very badly in African matches for a year prior to the World Cup, and entered the tournament ranked only 74th, immediately below such soccer minnows as Togo and Iceland.

After each day of play, the simulation was run 10 000 times, and the results tabulated. Simulation results at various stages of the tournament are shown in Table 2.

Results

Summing the probabilities generated for each of the 64 matches played in the World Cup gives the expected numbers of wins, draws and losses for the favoured team. These are compared with the observed numbers in Table 3. Although not strictly appropriate to these data (trials are independent but do not have identical distributions), a chi squared test gives $\chi^2 = 2.24$ (2 d.f.), $P = 0.33$.

Similarly, the expected number of goals for the favourite and underdog in each game may be tallied. As these are

sums of Poisson random variables, they are also Poisson random variables. These are compared with the actual numbers of goals (in the standard 90 minutes—golden goals not included) in Table 4, and two sided significance probabilities calculated.

The most pleasing feature of these results is the total number of goals scored, 170 versus 168.2 expected. This is

Table 3 Expected vs observed matched outcomes

| | <i>Expected number</i> | <i>Observed number</i> |
|-----------------|------------------------|------------------------|
| Favourite wins | 31.8 | 32 |
| Draw | 15.9 | 20 |
| Favourite loses | 16.3 | 12 |

Table 4 Expected vs actual goals scored

| | <i>Expected number</i> | <i>Observed number</i> | <i>P</i> |
|-----------|------------------------|------------------------|----------|
| Favourite | 101.3 | 110 | 0.37 |
| Underdog | 66.9 | 60 | 0.43 |
| Total | 168.2 | 170 | 0.88 |

Table 5 Expected vs observed match outcomes for 'minnow' matches

| | <i>Expected number</i> | | <i>Observed number</i> |
|--------------|------------------------|-----------------|------------------------|
| | <i>Raw</i> | <i>Adjusted</i> | |
| Minnow wins | 4.3 | 3.0 | 0 |
| Draw | 3.6 | 3.3 | 2 |
| Minnow loses | 6.1 | 7.8 | 12 |

Table 6 Expected vs actual goals scored for 'minnow' matches

| | <i>Expected number</i> | | <i>Observed number</i> |
|----------|------------------------|-----------------|------------------------|
| | <i>Raw</i> | <i>Adjusted</i> | |
| Minnow | 15.7 | 13.6 | 5 |
| Opponent | 20.6 | 24.5 | 32 |

a good indication that the form of Poisson model used for this simulation is plausible at the very least.

There remains the question of the subjective rating adjustments applied to this model prior to the tournament. The fact that these adjustments were considered necessary at all implies a degree of mistrust in the FIFA ratings, and their ability to correctly rank teams. The expected results in matches played by these 'minnow' teams have been calculated for both the true and adjusted ratings, and compared to the observed scores. Two matches where both teams were in this adjusted group are not included. The results are shown in Tables 5 and 6.

These tables show the raw FIFA ratings to be a slightly poorer predictor than the manually adjusted ones, although the manually adjusted ratings are still far from perfect. The tables suggest the manual adjustment should have been greater. To some extent, this reflects the intention of the FIFA ratings to reward participation, and not to be seen to be reinforcing the stereotype that all good soccer is played in Europe and South America.

By summing the probability of each team in each game scoring a particular number of goals, we may calculate the expected number of times each individual team score occurred, with both the adjusted and raw ratings in place. The two sets of expected values shown in Table 7 are similar. Both are a good fit to the actual observations, with a chi squared test giving $P > 0.85$ in both cases.

Conclusions

The official FIFA ratings, along with information on match venue, have been shown to provide useful predictive power under the assumption of goals scored being Poisson distributed. Home advantage, under this model, was found to be of greater effect on defence than attack, the decrease in goals scored against the home team outweighed the increase in goals scored by the home team.

When applied to the 1998 World Cup competition, the model performed plausibly well, with upset results occurring in similar numbers to those predicted, and a total number of goals very close to that predicted.

Part of this plausible performance was, however, due to a subjective modification of the raw FIFA ratings for several teams from non-traditional soccer playing regions. Results suggest that the adjustment could have been greater. Had these teams retained their original rating, the model results would have been worse.

As other soccer federations improve, and approach, in playing standard, the established regions of Europe and South America, so the ability of the FIFA ratings to predict match outcomes should improve.

While the simulation was here used to supply fans with predictions of individual matches and the chance teams would ultimately win the tournament, there are many other questions of interest. A simulation gives outcomes at any level of detail, and these may be of interest to fans or have implications for gambling. What is the chance two teams will meet in the final? What is the probability a particular team or set of teams will score more than a given number of goals? How many draws can we expect in the tournament? While the answers to all these questions could not be supplied on a web site, we are investigating the possibility of writing the results of the simulation to a database, which web users can then interrogate. In this way we do not restrict the questions fans may ask.

A possible improvement is to update team ratings as the tournament progresses. Since the FIFA rating is based on eight years performance, merely recalculating the official rating would produce little change. Stefani² describes predicting the 1978 World Cup, by applying least squares adjustments to equal ratings at the start of the tournament. Such methods could be adapted to alter the original FIFA ratings as the tournament progressed. Therefore the rating a team brings to the tournament as a reflection of its overall

Table 7 Expected vs actual number of observations of team scores

| | <i>Number of goals</i> | | | | | | | |
|--------------------------------|------------------------|----------|----------|----------|----------|----------|----------|------------|
| | <i>0</i> | <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> | <i>5</i> | <i>6</i> | <i>7 +</i> |
| Observed frequency | 34 | 45 | 32 | 11 | 3 | 2 | 1 | 0 |
| Expected frequency (raw model) | 36.26 | 44.41 | 28.45 | 12.71 | 4.45 | 1.30 | 0.33 | 0.09 |
| Expected frequency (adj model) | 36.22 | 43.97 | 28.33 | 12.90 | 4.67 | 1.42 | 0.38 | 0.12 |

ability, would be gradually replaced with one that gives a greater weight to current form. In this way the official body undertakes the onerous task of collecting results and updating ratings between major tournaments, while match predictions can still depend on current form (at least by the final stages). While this was not attempted for this World Cup, it is being investigated with regard to tennis.

During the tournament, nearly 4000 visits were made to our soccer page, more than double the number that visit our school home page in a whole year. While the general population is showing less interest in studying mathematical based courses, they can be enticed to view the results of applying mathematics and statistics. In this respect we rated the project a success.

References

- 1 Clarke SR (1993). Computer forecasting of Australian Rules football for a daily newspaper. *J Opl Res Soc* **44**: 753–759.
- 2 Stefani RT (1980). Improved least squares football, basketball and soccer predictions. *IEEE Trans Systems, Man and Cybernetics* **10**: 116–123.
- 3 Clarke SR. and Dyte DS (1999). Using official tennis ratings to estimate tournament chances. In: Massoud M (ed). *Proceedings of the 28th Annual Meeting of the Western Decision Sciences Institute*, WDSI: Mexico, pp 777–779.
- 4 Stefani RT (1997). Survey of the major world sports rating systems. *J App Stats* **24**: 635–646.
- 5 Stefani RT (1998). Predicting Outcomes. In: Bennett J (ed). *Statistics in Sport*. Arnold: London, pp 249–273.
- 6 Norman JM (1998). Soccer. In: Bennett J (ed). *Statistics in Sport*. Arnold: London, pp 105–118.
- 7 Pollard R, Benjamin B and Reep C (1977). Sport and the negative binomial distribution. In: Ladany SP and Machol RE (eds). *Optimal Strategies in Sports*. North Holland: Amsterdam, pp. 188–195.
- 8 Maher MJ (1982). Modelling association football scores. *Statistica Neerlandica* **36**: 109–118.
- 9 Dixon MJ and Robinson ME (1998). A birth process model for association football matches. *The Statistician* **47**: 523–538.
- 10 Lee AJ (1997). Modelling scores in the Premier League: Is Manchester United really the best? *Chance* **10**: 15–19.

*Received April 1999;
accepted April 2000 after two revisions*