SOCCER/WORLD FOOTBALL

DAVID R. BRILLINGER
Statistics Department,
University of California,
Berkeley, California

Soccer/world football, also known as *association*, is a sport involving two teams of players kicking and heading a medium-sized ball about on a large rectangular field. Each team has a goal that is quite wide and high and seeks to shoot the ball into the other team's goal. The basic objective is to see which team of the two can score the most goals and thereby win the game. The sport is subject to laws set down by the Fédération Internationale de Football Association (FIFA). Typically, there is a group of teams playing against each other in a league or tournament, with a formal schedule of games. A team's intention then is to win a tournament or championship by winning the final game or scoring the most points, where 3 points are awarded for a win, 1 for a tie, and 0 for a loss.

The sport of soccer has a long and full history possibly going back to 2500 BC [1]. In the present time, the teams each have at most 11 players. One, the goalie, may handle the ball but only within a restricted area in front of the goal. There is a neutral referee whose decisions are final. The grandest tournament, the World Cup, moves amongst different countries and occurs nominally every 4 years. The past world champions are Uruguay, Argentina, Brazil, England, France, Germany, and Italy. The sport has become a multibillion dollar business.

Tactics are basic to the games, with many styles and plating formations available. A team's choice has been based, in part, on studies of match statistics. The pioneer in the field of match performance analysis is Charles Reep [2–4].

DATA COLLECTION AND DESCRIPTIVE ANALYSES

Data collection and analysis have been basic to soccer for many years. The most common statistics are a game's result and the number of goals scored by each team. Outputs of the data analyses include reports, tables, graphs, talks, images, and videos.

Many things go on during a game. A basic question is what is to be recorded for the analysis and how to do so. Charles Reep [2,3] developed solutions starting in the early 1950s. The study of basic events in a game has been called both *match performance analysis* and *match analysis* [2,5–7].

A recent addition to the type of information available for individual games is near-continuous high frequency digital spatial-temporal data. Companies come to a stadium and set up an array of video cameras. By signal processing, they then develop the spatial-temporal coordinates of the changing locations of the players on the field, the ball, the linesmen and the referee. Companies doing this include Match Analysis, Prozone Holdings, and Sport-Universal SA. Di Salvo *et al.* [8] have validated one such system. The basic quantities that may now be tabled directly include counts of tackles, crosses, distances covered, key moments and actions, possessions, passes, interceptions, runs with the ball, incorrect referee decisions, fouls, penalty kicks, challenges, entries into the opponent's area, ball touches, blocks, forward passes, long balls, high intensity running, ball velocity and accuracy, and balls received. Individual player's performances may be evaluated directly. The data can lead to changes in tactics during a game [9,10]. Substantial soccer databases have been available for many years. Websites include www.uefa.com, www.fifa.com [11,12], www.soccerbase.com, www.soccerway.com, and www.soccerpunter.com.

Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran

One early investigation of basic data is that of Moroney ([13], pp. 101–103). He studied the numbers of goals scored in 480 games and prepared a histogram of the counts. In another early study, the number of successful passes in passing movements were studied for the English First Division during the years 1957–1958 and 1961–1962 [14]. The results were presented in tabular form. Reep and Benjamin found a stable correspondence between shots on goal and goals scored of about 10 to 1. Distances traveled by individual players during a game are graphed by player position in Ref. 5. Hughes and Franks [15] present scatter and time series plots. One early discovery was the existence of a home advantage. Schedules are often set up so that the two teams meet both, at home and away to deal with this advantage. Jochems [16,17], while addressing the question of whether the Dutch football pools were breaking the gambling laws, found the percentages to be 46.6 for a home win, 31.0 for a visitor win, and 22.4 for a draw/tie. Managers make use of such data in their decision-making, more so every year. Stochastic models are important in this connection.

STOCHASTIC MODELING

Stochastic models are pertinent to soccer statistics, because in part there is much uncertainty in what happens and what may happen. During the preceding 50 years, stochastic models have been constructed to address soccer questions.

Between Game Modeling

Some models may be distinguished as to concerning final goals, win–tie–loss, or points. Specific distributions and models that have been employed include bivariate Poisson, exponential, extreme value, GARCH, generalized linear, logistic, Markov, negative binomial, ordinal, regression, prior, point process, and state space. Many analyses are based on the Poisson distribution whose probability function is given by

$$\text{Prob}\{Y = y\} = \mu^y e^{-y}/y! \text{ for } y = 0, 1, 2, \ldots.$$

Here Y might represent the number of goals a team scores in a future game. In an early work, Moroney [13], using the data mentioned above, compared the number of goals scored by a team in a game to their estimated expected frequency assuming that a Poisson distribution held. On examining the result, he was led to seek to improve it and fit a negative binomial distribution. This is a generalization of the Poisson. This fit was found satisfactory. However, Greenhough *et al.* [18] found that when data from 169 countries were pooled, extreme value distributions were needed.

Reep *et al.* [14,19] work on the number of passes in successful passing movements. They fit the negative binomial and found it was good when 0-length cases were excluded.

Suppose that team i at home is playing against team j visiting. Denote the final score by (X_{ij}, Y_{ij}) with X refering to the home team. Various authors have assumed that X_{ij} and Y_{ij} are independent Poissons with respective means,

$\alpha_i\beta_j, \gamma_i\delta_j$	Ref. 20
$\exp\{\alpha + \eta + \gamma_i + \delta_j\}$, $\exp\{\alpha + \gamma_j + \delta_i\}$	Refs 21 and 22
$\alpha_i\beta_j\eta, \alpha_j\beta_i$	Ref. 23
$\exp\{\alpha + \eta + \gamma_i - \delta_j - \phi X(\varepsilon_i - \varepsilon_j)\}, \exp\{\alpha + \gamma_j - \delta_i + \phi(\varepsilon_i - \varepsilon_j)\}$	Ref. 24

here η is the home advantage, ϕ a psychological effect, and $\varepsilon_i = \gamma_i + \delta_i$. In their fitting, Dixon and Coles modify the low score model probabilities. The probability that team i wins the game may be estimated having estimated the model parameters.

Maher [20] estimates $\alpha, \beta, \gamma, \delta$ by maximum likelihood and mentions fitting the bivariate Poisson. Karlis and Ntzoufras [25] give details of fitting a bivariate Poisson, studying the data for 24 leagues. They find that the assumption of independence is not rejected in 15 out of the 24 cases. They also consider the negative binomial distribution and the inclusion of interaction terms.

The goal difference $X_{ij} - Y_{ij}$ is also important. It determines whether a game result is win, tie or loss (W, T, or L) and is used to resolve conflicts when the points are equal. Stefani [26] considers the model

$$X_{ij} - Y_{ij} = \eta + \rho_i - \rho_j + \text{random error}$$

with η the home advantage and ρ_i the rankings. Karlis and Ntzoufras [27] fit the Poisson difference distribution to observed goal differences including a home effect, attacking, and defensive parameters.

Fahrmeir and Tutz [28] employ a logistic latent variable-cutpoint setup to model an ordinal-valued variable. Their model is

$$\begin{aligned}\text{Prob}\{Y_{ij} = r\} &= F(\theta_r + \alpha_i - \alpha_j) \\ &\quad - F(\theta_{r-1} + \alpha_i - \alpha_j) \\ &\text{for } r = W, T, L,\end{aligned}$$

with F the logistic and α_i the ability of team i. In contrast, Brillinger [29–33] employs an extreme value variable-cutpoint approach. It leads to the model

$$\begin{aligned}&\text{Prob}\{\, i \text{ wins at home playing } j\} \\ &\quad = 1 - \exp\{-\exp\{\beta_i + \gamma_j + \theta_2\}\}, \\ &\text{Prob}\{\, i \text{ ties at home against } j\} \\ &\quad = \exp\{-\exp\{\beta_i + \gamma_j + \theta_2\}\} \\ &\qquad \times - \exp\{-\exp\{\beta_i + \gamma_j + \theta_1\}\}, \\ &\text{Prob}\{\, i \text{ loses at home against } j\} \\ &\quad = \exp\{-\exp\{\beta_i + \gamma_j + \theta_1\}\},\end{aligned}$$

with $\theta_2 > \theta_1$ cutpoints and the standardizations $\sum \beta_i, \sum \gamma_i = 0$. The parameters β_i and γ_i represent home and away effects of team i. The extreme value distribution employed is longer tailed to the right. The result of the preceding game and the distances between the cities involved were considered as explanatories. The $\beta_i - \gamma_i$ can be interpreted as the advantage of team i when playing at home. These values were also considered in Ref. 34. These authors model the goal differences as

$$X_{ij} - Y_{ij} = \eta + \alpha_i + \beta_j + \text{random error},$$

calling the α_i and β_j "forces." Least squares estimation is employed. Including past results did not improve things. Goodard [35] fits an ordered probit model, that is, F above relates to the normal, with explanatories.

In a linear regression analysis, Panaretos [36] studies the number of points collected in the course of a double round robin tournament and the effect of the explanatories goals scored, goals conceded, and time of ball possession. The regression model

$$\begin{aligned}&\text{points} = \alpha \\ &\quad + \beta \log(\text{goals scored}/\text{goals conceded}) \\ &\quad + \gamma \,\text{ball possession} + \text{random noise}\end{aligned}$$

is fitted and a proportion of variance explained of 0.971 is found.

Barnett and Hilditch [37] study whether the field being artificial turf affects the game result. They find that the home team's advantage was increased. The use of artificial turf was abandoned.

Attention now turns to the case where time or round, t, plays an essential role in the modeling. Jochems [16,17] defines, as a measure of strength of team i at home and j visiting, $\lambda_{ij} = W_i - W_j$, with W_i the average points of team i in its preceding games. The prediction of the result for team i is

$$\begin{aligned}&i \text{ wins if } \lambda_{ij} > 0, \\ &i \text{ wins if } \lambda_{ij} = 0 \quad \text{home advantage}, \\ &i \text{ loses if } \lambda_{ij} < 0.\end{aligned}$$

Jochem's study had been commissioned by the Netherlands Lottery Commission to see if any skill was involved on the part of tipsters.

The models listed above can be turned into forecasting procedures by simply adding a t to the subscripts of the parameters and estimating them by using the data up to time t. One then uses the schedule of remaining games to develop forecasts. For example, Stefani [26] considers the model

$$X_{ijt} - Y_{ijt} = \eta_{ijt} + \rho_i - \rho_j + \text{random error},$$

with ρ_i a rating based on i's past games and $\eta_{ijt} = \pm\eta$ represents a home advantage. The quantity ρ_i represents team i's ability. The fit is by conditional least squares.

Dixon and Coles [23] employ independent Poissons and introduce an exponentially decaying time function to damp out the effects of previous rounds. The Poisson means are $\alpha_{i(t)}\beta_{j(t)}\eta$ and $\alpha_{j(t)}\beta_{i(t)}$, respectively. The variable t is time or round number and η stands for home advantage. Estimation is by maximizing the likelihood.

Fahrmeir and Tutz [28] work with a latent variable-motivated ordinal model,

$$\text{Prob}\,\{R_{ij}^t = r\} = F(\theta_{tr} - \alpha_{ti} - \alpha_{tj}) \\ -F(\theta_{tr-1} - \alpha_{ti} - \alpha_{tj}),$$

with R the result, and $r = 0,\ 1,\ 2$ a win–tie–loss indicator. Further F is the logistic, and α_{ti} are random walks in t. The computations involve a Kalman filter and simulation. Rue and Salvesen [24] employ a Bayesian dynamic generalised linear model with independent normals having conditional means given the past

$$\exp\{\alpha + \eta + \gamma_i^t - \delta_j^t - \phi(\varepsilon_i^t - \varepsilon_j^t)\},$$
$$\exp\{\alpha + \gamma_j^t - \delta_i^t + \phi(\varepsilon_i^t - \varepsilon_j^t)\},$$

to predict next weekend's results. They employ a directed graph to describe the causal structure of the model as a function of time.

Brillinger [32,33], for each round, t, uses previous round results. He fits a latent variable-based model and then uses simulation to make projections of the final points and standings of the teams using the knowledge of the remaining games in the schedule. Estimates are then available for future rounds.

Harvey and Fernandes [38] study the series of goals scored by England against Scotland in a biyearly match. In a state space approach, they assume a negative binomial model with mean an exponentially decaying function of past values.

Within Game Modeling

Next, consider modeling the progress of a single game. One approach breaks the course of a game into states corresponding to zones in the field. Pollard and Reep [39] carry out a logistic regression analysis to model the probability of scoring from various positions on the field. In a series of papers [40–42], Markov chains with a finite number of states are employed. Hirotsu and Wright [40] use a four-state model to determine the optimal timing of substitution and tactical decisions. The expected number of points to be gained from a change in tactics is considered an objective function, and dynamic programming is employed. The model includes transition rates of scoring and conceding, and rates of gaining and losing possession. Hirotsu and Wright found evidence for replacing a defender with an attacker when losing.

A common problem is to identify change taking place. Croucher [43] studied the effect of changing the points awarded. The object of this change is to generate more attacking play. Ridder *et al.* [44] study the effect of a red card in game. They infer that the scoring rate does change after the ejection, and has a negative effect on the team with fewer players. Other studies that deal with the effect are given in Refs 45–47. The latter take a point process approach. When team i is playing team j, the times of goals are modeled by Poisson processes with log rates

$$\nu_{ij}(t) = \lambda_i(t)\exp\{\alpha_j + \beta z_{ij}(t)\} \text{ and}$$
$$\nu_{ji}(t) = \lambda_j(t)\exp\{\alpha_i + \beta z_{ji}(t)\},$$

respectively, t being time into the game. The function, $\lambda_i(t)$, is referred to as the *attack intensity* for team i and α_i as the defense parameter. The z_{ij} are the explanatories

$$z_{ij1}(t) = 1 \text{ when rival player off with red card, 0 otherwise;}$$
$$z_{ij2}(t) = 1 \text{ if actual score positive;}$$
$$z_{ij3}(t) = 1 \text{ if actual score negative;}$$

where t is time, β and z_{ij} are the 3 vectors. The parameters are estimated using data from the 2006 World Cup. The fitted model may be employed to generate simulations. Dixon and Robinson [48] also take a point process approach, employing a birth process, to model the scoring.

The match video data being collected these days may be processed to determine trajectories of players and the ball. Then models can be developed. Xie *et al.* [49] take a video of a game, break it into segments, for example, "play" or "break." They deal with camera pans and zooms using the green grass ratio and motion intensity. Their analysis uses dynamic programming and spatial-temporal hidden Markov processes. Min *et al.* [50] work on the same problem making use of rule-based reasoning, Bayesian inference, simulation, and expert systems. Palavi *et al.* [51] develops a graph-based multiplayer detection and tracking system. In the simplest case of a noteable goal scoring movement in one game, Brillinger [31] develops a potential function approach to obtain a stochastic model.

Berument *et al.* [52] look for evidence of soccer results affecting the Turkish stock market. They do find an effect for one team as well as a day of the week effect.

Garvia [53] asks, "Is soccer dying?" He uses a Markov switching model to look for a break in the series of yearly average goals per game. He finds a drop for England, Italy, and Spain around 1965.

RANKING

Rankings/ratings/seedings are numerical values meant to describe the relative performances of teams in a league. They are used for a variety of specific purposes. One is the preparation of schedules for knock-out tournaments. Another is for the maximization of the probability that the best competitors meet in a tournament's later stages. If there are N teams the collection of ranking values might be the sequence of integers 1 to N. A famous example is the FIFA/Coca-Cola world ranking [www.fifa.com.worldfootball/ranking/procedure/men.html [54]], where the values derive from a formula involving major games during the previous 4 years, strength of opponent, W–T–L points, and several other items. It orders the teams of all the countries that are members of the FIFA and is updated steadily as more game results become available.

Jochems [16,17] employed the differences of average game scores and compared this to tipsters' predictions. Hill [55] uses Kendall's tau to compare tipster results with the end of season standings and finds association. Stefani [26] suggests the model

$$X_{ij} - Y_{ij} = \eta + \rho_i - \rho_j + \text{random error},$$

with the ρ_i's rankings. Stefani [56] surveyed the major world sports rating systems. Bassett [57] suggests improving the estimates by employing a robust estimation method to handle long tails. Forrest and Simmons [58,59] investigate the predictive quality of newspaper tipsters match results. Stefani and Pollard [60] provide a critical survey of ranking procedures. Gelade and Dobson [61] relate the international standing of a country's team to various social factors including number playing regularly, wealth, and climate. McHale and Davies [62] focus on the FIFA rankings.

There are papers on ratings for other sports; for example, Refs 63–65.

TOURNAMENTS AND SCHEDULING

Scheduling is a basic step involved when games need to be organized amongst the members of a group of teams. Two basic types of tournament design are the round robin and the knock-out. In the round robin, each pair of teams plays against each other the same number of times, equally at home and away. The champion of the group is the team with the most points at the end of all the games. In the knock-out case, there is a seeding order and the first round is based on it. The tournament progresses with two teams playing in pairs setup by the seeding order with the winner going on to the following round. This continues until only one team is left, the champion.

Sometimes there are numerical-valued objective functions that can be employed in developing schedules. Their forms can include distance traveled, total expenses, numbers of consecutive home or away games, and referees' time. There are often constraints that need to be taken note of in the scheduling. For example, there may be multiple venues, television schedules, desired dates for two teams from the same city. Analytic tools employed in an optimization include integer programming, maximization, knapsack algorithms, simulation and stochastic models.

Kuonen [21,66] employs a logistic regression model and shows how output can be used to estimate probabilities of interest for a knock-out tournament. Unlike a round robin, future opponents for the following rounds are only known probabilistically. Kuonen works with European Cup Winners Cup data. He investigates three methods for calculating seeding coefficients based on the data for the 3 preceding years.

Urban and Russell [67] consider scheduling competitions on multiple venues, venues not associated with any of the participants. Della Croce and Oliveri [68] present an integer linear programming approach for scheduling the Italian League. They had to deal with cable TV and with games between teams in the same city.

Scheduling the Chilean League has also posed special problems. As described in Refs 69 and 70, these were dealt with via using mathematical programming, recognized existing weaknesses, stadiums availability, international requirements, and reducing travel distance.

Objectives may be described probabilistically and then stochastic models are involved. When the goal is to determine a champion, it is natural to ask questions like: what is the probability that the "best" team actually wins the tournament, or how much better is the first team than the second? Appleton [71] provides definitions and reviews many tournament structures. He employs simulations.

Scarf *et al.* [72] provide an extensive review including discussion of: tournament metrics, how the design influences outcome uncertainty, the use of simulation, robustness (to teams dropping out), effects of rule changes, and probability model employed. He defines the parameter

$$P_{qR} = \text{Prob}\{\text{team in top } 100q \text{ pretournament rank percentile progresses advances to round } R+1\};$$

here R is round achieved and $0 < q < 1$. To understand the P_{qR} values, he graphs them against q.

It is necessary to schedule the referees also. Yavuz *et al.* [73] provide an extensive review. The topics covered include leagues in different countries requiring different objective functions and constraints, tension between referees and clubs caused by past incidents, reducing frequent assignment of the same referee to a team's games. In amateur leagues, there may be several games the same day for the same referee, perhaps at different places even.

GAME THEORY

The expression "game theory" has both a general and a technical meaning. The "technical" comes from the classic Von Neumann–Morgenstern [74] setup. The "general" refers to other things, particularly tactics. Both meanings provide general frameworks with which to study soccer. A difficulty arising in the theories' applications to soccer is that the game is highly dynamic. Jones and Trantner's [75] book is unusual in that, it starts by emphasizing the defensive formations. Wilkinson's [76] book is devoted to the topic of tactics.

The formation selected would be controlled by the coach. The 2-3-5 was very common for many years. In it, there are two defensive specialists, five attacking, and three midfielders in between. Newer formations include 3-5-2, 5-4-1, 4-5-1, 4-3-3, and even 4-1-4-1 with an extra row. The formation 4-4-2 is favored by various famous teams. The two attackers can drop back to assist in obtaining the ball or be on the sides up front. When the team has possession of the ball, the two outside midfielders can move forward to increase pressure on the opponent. Wilson [77] provides a review and insightful analysis of many of the formations employed in games dating from 1878 to 2008.

Pollard and Reep [39] seek to quantify the effectiveness of different playing strategies.

They break the action of a game down into discrete events, for example, pass, center, shoot. They define the "yield" as the probability a goal is scored minus the probability one is conceded. This quantity is used to evaluate the expected outcome of a team possession.

The book, The Science of Soccer, has a chapter on "game theory" [78]. It discusses how the strength and deployment of the team moderates the apparent random motion and discusses the fact that low scoring benefits the weaker team.

One of the momentous occasions in a soccer game is the awarding of a penalty shot. There have been a number of formal studies. Greenless *et al.* [79] study how the penalty takers' uniform color and prepenalty gaze affect the impressions formed by the goal-keeper. When penalties take place to break a tie, McGarry and Franks [80] identify an optimal order for a team to take the shots in. Their conclusion is to begin by ranking the players 1 (best) to 11 (worst). Then use the order: 5, 4, 3, 2, 1, and if necessary after that 6, 7, 8, 9, 10, and 11. Jordet *et al.* [81] considers the roles of stress, skill, and fatigue for answering why the English are poor at penalties. He analyzed 200 shots from World Cup and European Championship. In an analysis of many penalties, Franks *et al.* [82] learned that in 80% of the cases studied, if the nonkicking foot points to the left, the ball will be shot to the left. They developed a training program to test this discovery.

Larsen [2] describes the arrival of Reep's ideas in Norway and the success that teams adopting it enjoyed.

The theory of von Neumann and Morgenstern provides general definitions and leads to random strategies. Haigh [83] provides the following intuitive example. The table shows estimated percentages of successful penalty shots based on 1876 attempts.

Goalie	Kicker		
	Left	Center	Right
Left	60	90	93
Center	100	30	100
Right	94	85	60

These data concern the goalie's direction of motion and the direton of the Kicker's shot.

The situation can be considered is a two-person zero-sum game. If the kicker wins by scoring, the goalie loses. If the goalie wins by saving, the kicker loses. Haigh finds that Nash equilibrium theory leads to the goalie's chosing to move left, center or right with the respective probabilities of 0.44, 0.13, and 0.43. For the kicker, they are 0.37, 0.29, and 0.34, respectively. If either player uses their strategy, 80% of the shots would be scored in the long run. This turns out to be close to the actual percentage.

Other references to applications of the formal theory include Refs 84–86. Hirotsu and Wright [85] consider a zero-sum game focusing on who wins the game, while Hirotsu *et al.* [87] focus on points gained. Here one has a nonzero-sum game. Bennett *et al.* [88] use hypergame analysis, involving the ranking of various forms of hooligans' actions and authorities' reactions followed by hooligan's reactions and so on. They explore the implications of such scenarios.

ECONOMICS AND MANAGEMENT

Reilly and Williams [89] write, "In the 1980s it became apparent that the football (soccer) industry and professionals in the game could no longer rely on the traditional methods of previous decades. Methods of management science were applied to organizing the big soccer clubs and the training of players could be formulated on a systematic basis." Desbordes [90, Chapter 10] advocates the introduction of modern management tools into soccer. This makes sense if for no other reason than the fact that billions of dollars are now involved in the sport. One does need to remember that at the youth and amateur levels, the amount of money involved is little. Really, all that is needed for a game is a "ball" and a field or a beach. This is one charm of the game.

Management refers to a sport's structure, owners, sponsors, marketers, financiers, schedulers, and others with some form of control. Management provides the superstructure. The fans expect both their team's players and its management to be successful in acquiring players, providing facilities, and hiring the coach. Management assists by supporting the enterprise generally. It provides the infrastructure. Other things that management is responsible for include cash inflows–outflows, advertising, manpower planning and hiring, planning new and keeping up old stadiums, and signing players. The stadium owners provide facilities for security, ticket collection, food, souvenirs, and parking. Management studies the performance of their players and others [10,91].

Sometimes, there is a numerical-valued objective function for the management to work with. Then optimization methods may be employed to good advantage. Other concerns are loss functions, earnings/revenue, prestige/honor, recruitment, salary negotiation, scouting, referees, merchandise, commercialism, productivity, profitability, travel, and television rights.

There is a common wish to improve individual player's and a team's performance. An accompanying question is how to assess performance and efficiency. One paper on the topic for English soccer is Ref. 92. Others are Refs 93 and 94. The latter concerns the Spanish Soccer League. There have been analyses. Audas *et al.* [95] study the impact of managerial change on team performance using the data of English football match results since the 1970s, while Partovi and Corredoira [96] investigate models for prioritizing and designing rule changes.

In a study of competitive balance in the Dutch League, Koning [97] employs the values C_{ij}, corresponding to W, T, L, generated via

$$\begin{aligned} D_{ij} &= \varphi_i - \varphi_j + \varepsilon_{ij} \text{ with the } \varepsilon\text{'s mean } 0, \\ &\quad \text{common variance normals,} \\ C_{ij} &= 1 \text{ if } D_{ij} > c_2 \\ &= 0 \text{ if } D_{ij} = 0 \\ &= -1 \text{ if } D_{ij} < c_1. \end{aligned}$$

It is used to study changes in competitive balance in Holland.

A novel application of programming theory is presented in Ref. 98. It applies an integer program to the generalized knapsack problem to find a "Dream Team." Player values were established by a public poll with the participants given a budget of 2.5 million dollars each. A knapsack algorithm was then employed to establish the best team at the lowest price.

Calster *et al.* [99] study an unwanted happening in soccer, the scoreless draw. The occurrence is related to indices of a team's offensive performance including total goals and earned points per game.

SOME REMAINING TOPICS

The book, The Inner Game of Soccer, provides details of the game from a referee's standpoint [100]. It discusses the laws and the mechanics and talks of the characteristics of the best referees and the ones who can turn benign games into ugly unsportsmanlike confrontations.

Reference 78 lays out some of the physics of soccer. For example, the chapter titled "The ball in flight" analyzes the movement of a ball spinning in flight and how players can make it bend. Hall [101] elaborates on that describing the physical forces at work in a double banana shot, which curves in one direction then swerves in another. Such a shot is very difficult for the goalie to handle.

Sports biomechanics concerns physical actions that occur in sport that can be observed and improved, for example, to reduce injuries. The Refs 102 and 103 are on the topic of biomechanics in soccer. Grimpampi *et al.* [104] concern computational biomechanics. Bradley *et al.*'s study [105] is a human biology study. It refers to results on high-intensity running using data collected via an array of television cameras.

There are many articles on employing stochastic models in evaluating bets. Most of the articles referred to above go on to study the efficiency of various gambling strategies.

What is ahead for the game? Wright [106] writes on 50 years of operations research in sport to provide a background. The future is discussed in Ref. 107 and by the sports writer Garner [108] in World Soccer. He suggests the following changes: bigger goals, smaller penalty area, a revised offside rule, fouls denoted contact or technical, red-carded player punished but replacement allowed. Reviewing the literature and practice suggest that one can expect to see much more in the way of the collection and analysis of the spatial-temporal data collected in a game.

REFERENCES

1. Henshaw R. The encyclopedia of world soccer. Washington, DC: New Republic; 1979.
2. Larsen O. Charles Reep: a major influence on British and Norwegian football. Soccer Soc 2001;2:55–78.
3. Pollard R. Charles Reep (1904–2002): pioneer of notational and performance analysis in football. J Sports Sci 2002;20:853–855.
4. Pollard R. Home advantage in soccer: variations in its magnitude and a literature review of the inter-related factors associated with its existence. J Sport Behav 2006;29: 169–189.
5. Reilly T, Thomas V. A motion analysis of work-rate in differential positional roles in professional football match-play. J Hum Mov Stud 1976;2:87–97.
6. McGarry T, Franks IM. The science of match analysis. In: Reilly T, Williams AM, editors. Science and soccer, 2nd ed. New York: Routledge; 2003. pp. 265–275.
7. Hughes M. Notational analysis. In: Reilly T, Williams AM, editors. Science and soccer, 2nd ed. London: Routledge; 2003. pp. 245–264.
8. Di Salvo V, Collins A, McNeill B, Cardinale M. Validation of Prozone: a new video-based performance analysis system. Int J Perform Anal Sport 2006;6:108–119.
9. Carling C, Reilly T, Williams AM. Handbook of soccer match analysis. London: Routledge; 2005.
10. Carling C, Reilly T, Williams AM. Performance assessment for field sports. London: Routledge; 2009.
11. FIFA. Laws of the game. (2008/2009). www.fifa.com.
12. FIFA. Disciplinary code. 2009. www.fifa.com/mm/50/02/75/disco_2009_en.pdf.
13. Moroney M. Facts from Figures. London: Pelican; 1951.
14. Reep C, Benjamin B. Skill and chance in association football. J R Stat Soc A 1968;131:581–585.
15. Hughes M, Franks I. Analysis of passing sequence, shots and goals in soccer. J Sports Sci 2005;23:509–514.
16. Jochems DB. Voetbalprognoses. Stat Neerl 1958;12:17–31.
17. Jochems DB. Forecasting the outcomes of soccer matches: a statistical appraisal of the Dutch experience. Metrika 1962;5:194–207.
18. Greenhough J, Birch PC, Chapman SC, Rowlands G. Football goal distributions and extremal statistics. Phys A 2002;316: 615–624.
19. Reep C, Pollard R, Benjamin B. Skill and chance in ball games. J R Stat Soc A 1971;134:623–629.
20. Maher MJ. Modelling association football scores. Stat Neerl 1982;36:109–118.
21. Kuonen D. Modelling the success of soccer teams in European championships. Technical Report 96.1, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne; 1996.
22. Lee AJ. Modelling soccer in the Premier League: is Manchester United really the best? Chance 1997;15–19.
23. Dixon MJ, Coles SC. Modelling association football scores and inefficiencies in the football betting market. Appl Stat 1997;46:265–280.
24. Rue H, Salvesen O. Prediction and retrospective analysis of soccer matches in a league. J R Stat Soc D 2000;49:399–418.
25. Karlis D, and Ntzoufras I. Analysis of sports data using bivariate Poisson models. Statistician 2003;52:381–393.
26. Stefani RT. Improved least squares football, basketball, and soccer predictions. IEEE Trans Syst, Man, Cybern 1980;10:116–123.
27. Karlis D, Ntzoufras I. Bayesian modelling of football outcomes: using the Skellam's distribution for goal difference. IMA J Manag Math 2009;20:133–145.
28. Fahrmeir L, Tutz G. Dynamic stochastic models for time-dependent ordered paired comparison systems. J Am Stat Assoc 1994;89:1438–1449.
29. Brillinger DR An analysis of an ordinal-valued time series. In: Athens conference on applied probability and time series analysis. Volume 2, Lecture Notes in Statistics Volume 115. Springer; 1996. pp. 73–87.
30. Brillinger DR Modelling some Norwegian soccer data. In: Nair V, editor. Advances

in statistical modelling and inference. Singapore: World Scientific; 2006. pp. 3–20.

31. Brillinger DR A potential function approach to the flow of play in soccer. J Quant Anal Sports 2007;3(1):3.
32. Brillinger DR. Modelling game outcomes of the Brazilian 2006 Series A Championship as ordinal-valued. Brazilian J Probab Stat 2008;22:89–104.
33. Brillinger DR. An analysis of Chinese Super League partial results. Sci China A: Math 2009;52:1139–1156.
34. Leeflang PSH, van Praag BMS. A procedure to estimate relative powers in binary contacts and an application to Dutch Football League results. Stat Neerl 1971;25:203–208.
35. Goodard J. Regression models for forecasting goals and match results in association football. Int J Forecast 2005;21:331–340.
36. Panaretos VM (2002). A statistical analysis of the European Soccer Champions League. ASA Proceedings of the Joint Statistics Meetings. pp. 2600–2602.
37. Barnett V, Hilditch S. The effect of an artificial pitch surface on home team performance in football (soccer). J R Stat Soc A 1993;156:39–50.
38. Harvey AC, Fernandes C. Time series models for count data of qualitative observations. J Bus Econ Stat 1989;7:407–422.
39. Pollard R, Reep C. Measuring the effectiveness of playing strategies at soccer. J R Stat Soc D 1997;46:541–550.
40. Hirotsu N, Wright M. Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. J Oper Res Soc 2002;53:88–96.
41. Hirotsu N, Wright M. An evaluation of characteristics of teams in association football by using a Markov process model. Statistician 2003;52:591–602.
42. Hirotsu N, Wright M. Determining the best strategy for changing the configuration of a football team. J Oper Res Soc 2003;54:878–887.
43. Croucher JS. The effect of changing competition points in the English football league. Teach Stat 1984;6:39–42.
44. Ridder G, Cramer JS, Hopstaken P. Down to ten: estimating the effect of a red card in soccer. J Am Stat Assoc 1994;89:1124–1127. [see also Albert J, Bennett J, Cochran JJ. Anthology of statistics in sports. Phildelphia: SIAM; 2005. pp. 299–302.]
45. Wright M, Hirotsu N. The professional foul in football: tactics and deterrents. J Oper Res Soc 2003;54:213–221.
46. Vecer J, Kopriva F, Ichiba T. Estimating the effect of a red card in soccer: when to commit an offense in exchange for preventing a goal opportunity. J Quant Anal Sports 2009;5(1):8.
47. Volf P. A random point process model for score in matches. IMA J Manag Math 2009;20:121–131.
48. Dixon MJ, Robinson ME. A birth process model for association football matches. Statistician 1998;47:523–538.
49. Xie L, Xu P, Chang S-F, Divarkanm A, Sun H. Structure analysis of soccer video with domain knowledge and hidden Markov models. Pattern Recognit Lett 2004;25:76–775.
50. Min B, Kim J, Choe C, Eom H, McKay RI. A compound framework for sports prediction: a football case study. Knowledge-Based Syst 2008;21:551–562.
51. Palavi V, Mukherjee J, Majumdar AK, Sural S. Graph-based multiplayer detection and tracking in broadcast soccer videos. IEEE Trans Multimedia 2008;10:794–805.
52. Berument H, Ceylan NB, Gozpinar E. Performance of soccer on the stock market: evidence from Turkey. Social Sci J 2006;43: 695–699.
53. Garvia A. Is soccer dying? A time series approach. Appl Econ Lett 2007;7:275–278.
54. FIFA. Ranking. 2006. www.fifa.com.world football/ranking/procedure/men.html.
55. Hill ID. Association football and statistical inference. Appl Stat 1974;23:203–208.
56. Stefani RT. Survey of the major world sports rating systems. J Appl Stat 1997;24: 635–646.
57. Bassett GW. Robust sports ratings based on least absolute errors. Am Stat 1997;51(2).
58. Forrest D, Simmons R. Making up the results: the work of the Football Pools Panel, 1963–1997. Statistician 2000;49:253–260.
59. Forrest D, Simmons R. Forecasting sport: the behaviour and performance of football tipsters. Int J Forecast 2000;16:317–331.
60. Stefani R, Pollard R. Football rating systems for top-level competition: a critical survey. J Quant Anal Sports 2007;3(3):3.
61. Gelade GA, Dobson P. Predicting the comparative strengths of national football teams. Social Sci Q 2007;88:244–258.

62. McHale I, Davies S. Statistical analysis of the effectiveness of the FIFA world rankings. In: Albert J, Koning RH, editors. Statistical thinking in sports. New York: Chapman and Hall; 2008. pp. 77–90.
63. Leake RJ. A method for ranking teams with an application to 1974 college football. North Holland: Management Science in Sports; 1976.
64. Harville D. The use of linear model methodology to rate high school or college football teams. J Am Stat Assoc 1977;72:278–289.
65. Stern H. On the probability of winning a football game. Am Stat 1992;45:179–183. [see also Albert J, Bennett J, Cochran JJ. Anthology of statistics in sports. Phildelphia: SIAM; 2005. pp. 53–57].
66. Kuonen D. Statistical models for knock-out soccer tournaments. Technical Report 97.3. Lausanne: Department of Mathematics, Swiss Federal Institute of Technology; 1997.
67. Urban TL, Russell RA. Scheduling sports competitions on multiple venues. Eur J Oper Res 2003;148:302–311.
68. Della Croce F, Oliveri D. Scheduling the Italian Football League: an ILP-based approach. Comput Oper Res 2006;33:1963–1974.
69. Noronha TF, Ribeiro CC, Duran G, Souyris S, Weintraub A. A branch-and-cut algorithm for scheduling the highly-constrained Chilean soccer tournament. Lect Notes Comput Sci 2007;3867:174–186.
70. Duran G, Guajardo M, Weintraub A, Wolf R. Scheduling the Chilean League using mathematical programming. OR/MS Today 2009.
71. Appleton DR. May the best man win? Statistician 1995;44:529–538.
72. Scarf P, Yusof MMN, Bilbao M. A numerical study of designs for sporting contests. Eur J Oper Res 2009;198:190–198.
73. Yavuz M, Inan UH, Figlah A. Fair referee assignments for professional football leagues. Comput Oper Res 2008;35: 2937–2951.
74. Von Neumann J, Morgenstern O. Theory of games and economic behavior. Princeton: Princeton University Press; 1944.
75. Jones R, Trantner T. Soccer Strategies: defensive and attacking tactics. Reedswain: Spring City; 1979.
76. Wilkinson WHG. Soccer tactics: top team strategies explained. Crowood: Marlborough; 1988.
77. Wilson J. Inverting the pyramid: a history of football tactics. London: Orion; 2008.
78. Wesson J. The science of soccer. New York: Taylor and Francis; 2002.
79. Greenless I, Leyland A, Thelwell R, Filby W. Soccer penalty takers' uniform colour and pre-penalty gaze affects the impressions formed by opposing goal-keepers. J Sports Sci 2008;26:569–576.
80. McGarry T, Franks IM. On winning the penalty shoot-out in soccer. J Sports Sci 2000;18:401–409.
81. Jordet G, Hartman E, Vischer C, Lemmink KAPM. Kicks from the penalty mark in soccer: the roles of stress, skill, and fatigue for kick outcomes. J Sport Sci 2007; 25:121–129.
82. Franks I, McGarry T, Hanvey T. From notation to training: analysis of the penalty kick. Insight 1999;2:24–26.
83. Haigh J. Uses and limitations of mathematics in sport. IMA J Manag Math 2009;20:97–108.
84. Pacacios-Huerta I. Professionals play minimax. Rev Econ Studies 2003;70:395–415.
85. Hirotsu N, Wright M. Modelling tactical changes of formation in association football as a zero-sum game. J Quant Anal Sports 2006;2(2):4.
86. Coloma G. Penalty kicks in soccer: an alternative methodology for testing mixed-strategy equilibria. J Sports Econ 2007;8:530–545.
87. Hirotsu N, Ito M, Miyaji C, Hamano K, Taguchi A. Modelling tactical changes of formation in association football as a non-zero-sum game. J. Quant Anal Sports 2009; 5(3):2.
88. Bennett PG, Dando MR, Sharp RG. Using hypergames to model difficult social issues: an approach to the case of soccer hooliganism. J Oper Res Soc 1980;31: 621–635.
89. Peiser B, Minten J. Soccer violence. In: Reilly T, Williams AM, editors. Science and soccer. 2nd ed. Abingon: Routledge; 2003. pp. 230–242.
90. Desbordes M. Marketing and football: an international perspective. Portsmouth: Butterworth-Heinemann; 2007.
91. Covell D, Walker S, Siciliano J, Hess PW. Managing sports organizations: responsibility for performance, 2nd ed. Portsmouth: Butterworth-Heinemann; 2007.

92. Dawson P, Dobson S, Gerrard B. Stochastic frontiers and the temporal structure of managerial efficiency in English soccer. J Sports Econ 2000;1–4:341–362.
93. Haas DJ. Technical efficiency of Major League Soccer. J Sports Econ 2003;04–03: 203–215.
94. Sala-Garrido R, Carrion VL, Esteve AM, Bosca JE. Analysis and evolution of efficiency in the Spanish Soccer League. J Quant Anal Sports 2009;5(1):3.
95. Audas R, Dobson S, Goddard J. The impact of managerial change on team performance in professional sports. J Econ Bus 2002;54:633–650.
96. Partovi FY, Corredoira RA. Quality deployment for the good of soccer. Eur J Oper Res 2002;137:642–656.
97. Koning RH. Balance in competition in Dutch soccer. Statistician 2000;49:419–431.
98. Mosheiov G. The solution of the soccer 'Dream League' game. Math Comput Model 1998;27:79–83.
99. Calster BV, Smits T, Van Huffel S. The curse of scoreless draws in soccer: the relationship with a team's offensive, defensive, and overall performance. J Quant Anal Sports 2008;4(1):4.
100. Sellin E. The inner game of soccer. Mountain View: World Publications; 1976.
101. Hall LE. Laws of motion. New York: Rosen; 2005.
102. Lees A, Nolan L. The biomechanics of soccer: a review. J Sports Sci 1998;16:211–234.
103. Lees A. Biomechanics applied to soccer skills. In: Science and soccer, 2nd ed, Reilly T, Williams AM, editors. New York: Routledge; 2003. pp. 109–119.
104. Grimpampi E, Pasculli A, Sacripanti A. Computational biomechanics, stochastic motion and team sports. 2008. arxiv.org/pdf/0811.3659
105. Bradley PS, Sheldon W, Wooster B, Olsen P, Boanas P, Krustrup P. High-intensity running in English FA Premier League soccer matches. J Sports Sci 2009;27:159–168.
106. Wright MB. 50 years of OR in Sport. J Oper Res Soc 2009;60:S161–S168.
107. The Orange Future of Football Report. 2008. http://www.orange.co.uk/sport/football/pics/3395_1.htm
108. Garner P. The future of football. World Soccer; London; 2009.

FURTHER READING

Archontakis F, Osborne E. Playing it safe? A Fibonacci strategy for soccer betting. J Sports Econ 2007;8:295–308.

Baxter M, Stevenson R. Discriminating between the Poisson and negative binomial distributions: an application to goal scoring in association football. J Appl Stat 1988;15:347–354.

Bellos A. Futebol, the Brazilian Way of Life. London: Bloomsbury; 2002.

Boeri T, Severgnini B. The Italian job: match rigging, career concerns and media concentration in 'Series A'. IZA Discussion Paper No. 3745. 2008.

Brimson D. March of the hooligans: soccer's bloody fraternity. London: Virgin; 2007.

Brocas I, Carrillo JD. Do the three-point victory and golden goal rules make soccer more exciting? J Sports Econ 2004;5:169–185.

Cantor A, Arcucci D. Goooal: a celebration of soccer. New York: Fireside; 1997.

Carmichael F, Thomas D, Ward R. Team performance: the case of English Premiership Football. Manag Dec Econ 2000;21:31–45.

Comeron M. The prevention of violence in sport. Strasbourg: Council of Europe; 2002.

Cordeiro SE. El futbol y la corrupcion. Volume 5. In: Carrion F, editor. Biblioteca del Futbol Ecuatoriano, 5 Volumes. Quito: Flacso; 2006. pp. 233–250.

Crowder M, Dixon M, Ledford A, Robinson M. Dynamic modelling and prediction of English Football League matches for betting. Statistician 2002;51:157–168.

David FN. Games, gods and gambling, London: Griffin; 1962.

Dobson S, Goddard J. The economics of football. Cambridge: Cambridge University Press; 2001.

Dobson S, Goddard J. Forecasting scores and results and testing the efficiency of the fixed odds betting market in Scotish League football. In: Albert J, Koning RH, editors. Statistical thinking in sports. New York: Chapman and Hall; 2008. pp. 91–109.

Dyte D, Clarke SR. A ratings based Poisson model for World Cup soccer simulation. J Oper Res Soc 2000;51:993–998.

Eastaway R, Haigh J. Beating the odds: the hidden mathematics of sport, London: Robson; 2007.

Emonet B. Revisiting statistical aspects of soccer, STS Report. Lausanne: Swiss Technical University; 2000.

England C. No more budha, only football, London: Sceptre; 2003.

Espita-Escuer M, Garcia-Cebrianm LI. Measuring the efficiency of Spanish First-Division soccer teams. J Sports Econ 2004;04–03: 203–215.

Fernandez-Cantelli E, Meeden G. An improved award system for soccer. Chance 2002;7: 275–278.

Foer F. How soccer explains the world: an unlikely theory of globalization, New York: Harper Collins; 2004.

Giulianotti R. Football: a sociology of the global game, Polity: Bodmin; 1999.

Giulianotti R, Bonney N, Hepworth M. Football, violence and social identity, London: Routledge; 1994.

Goddard J, Asimakopoulos I. Forecasting football results and the efficiency of fixed-odds betting. J Forecast 2004;23:51–66.

Holey AJ, Johnson BS. Menaces to management: a developmental view of British Soccer Hooligans, 1961–1986. Sports J 1998;1(1).

Hornby N. Fever pitch. Riverhead trade; London: 1992.

Hughes M, Bartlett R. The use of performance indicators in performance analysis. J Sports Sci 2002;20:739–754.

Hurley WJ. Special issue on operations research in sport. Comput Oper Res 2006;33:1871–1873.

Jackson DA. Index betting on sports. Statistician 1994;43:309–315.

Karlis D, Ntzoufras I. On modelling soccer data. Student 2000;3:229–244.

Kern M, Sussmuth B. Managerial efficiency in German top league soccer: an econometric analysis of club performances on and off the pitch. German Econ Rev 2005;6:485–506.

Knorr-Held L. Dynamic rating of sports teams. Statistician 2000;49:261–276.

Kristan M, Pers J, Perse M, Kovacic S. Towards fast and efficient methods for tracking players in sports. In: Pers J, Magee DR, editors. Proceedings of ECCV Workshop on Computer Vision Based Analysis in Sport Environments. 2006. pp. 14–25.

Kuonen D, Roehrl ASA. Was Frances's World Cup win pure chance? Student 2000;3:153–166.

Kuper S. Szymanski S. Soccernomics, New York: Nation; 2009.

Kuypers T. Information and efficiency: an empirical study of a fixed odds betting market. Appl Econ 2000;32:1353–1363.

Ladany SP, Machol RE, editors. Optimal strategies in sports. Amsterdam: North-Holland; 1977.

Lewis MM. Moneyball: the art of winning an unfair game. New York: Norton; 2003.

Lucena CJP, Noronha S, Urrutia S, Ribeiro CC. A multi-agent framework to retrieve and publish information on qualification and elimination data in sports tournaments. Rio de Janeiro, Brazil: Monografias em Ciência da Computação; 2006.

Machol RE, Ladany SP, Morrison DG. Management science in sports. Amsterdam: North-Holland; 1976.

Magnum. Soccer. London: Phaidon; 2002.

Mason T. Association football and english society 1863–1915. Brighton: Harvester; 1980.

McCullough P, Nelder JA. Generalized linear models, 2nd ed. London: Chapman and Hall; 1989.

McGuire EJ, Courneya KS, Widmeyer WN, Carron AV. Aggression as a potential mediator of the home advantage in professional hockey. J Sport Exerc Physiol 1992;14:148–158.

Morris D. The soccer tribe. Jonathan Cape; London: 1981.

Needham CJ. Tracking and modelling of team game interactions. Ph.D. Thesis, University of Leeds; 2003.

Osborne MJ. An introduction to game theory. Oxford: Oxford University Press; 2004.

Ribeiro CC, Urrutia S. OR on the ball: applications in sports scheduling and management. OR/MS Today 2004; (June).

Sadovskii AL, Sadovskii LE. Mathematics and sports. Providence: American Mathematical Society; 1993.

Simons R. Bamboo goalposts. London: MacMillan; 2008.

Sloane PJ. The economics of professional football: the football club as a utility maximiser. Scottish J. Polit Econ 1971;8:121–146.

Sloane PJ editor. The economics of sport. Econ Affairs 1997;17:Special issue.

Stefani R. Measurement and interpretation of home advantage. In: Albert J, Koning RH, editors. Statistical thinking in sports. New York: Chapman and Hall; 2008. pp. 203–216.

Stefani RT. Predicting score difference versus score total in rugby and soccer. IMA J Manag Math 2009;20:147–158.

SUP AMISCO PRO. http://213.30.139.108/sport-universal/uk/amiscopro.htm; 2010.

Turnbull J, Raab A, Satterlee T. The global game: writers on soccer. Bison; London: 2008.

Wright DB. Football standings and measurement levels. Statistician 1997;46:105–110.

Yue Z, Broich H, Seifritz F, Mester J. Mathematical analysis of a soccer game. Part 1: individual and collective behaviors. Studies Appl Math 2008;121:223–243.

Yue Z, Broich H, Seifritz F, Mester J. Part 2: energy, spectral and correlation analyses. Studies Appl Math 2008;121:245–261.