## Guest Editorial 'Statistical Modelling for Sports Analytics'

The current issue is the second and final part of the special issues on 'Statistical Modelling for Sports Analytics', which altogether is a collection of ten contributions that focus on baseball, basketball and football and cover a wide range of topics, such as technical issues concerning the game and the way of playing, performance analysis, prediction, sports preferences, match-fixing and betting. All of the articles have met our challenging aim of disseminating knowledge that represents important developments, extensions and applications on statistical modelling in sports by appropriately combining methodology and practice.

In this issue, one contribution is devoted to testing and ranking on round-robin designs for sports data with an application to basketball, while the remaining three focus on different aspects in (association) football.

Corain et al. (2019) model the results of sport matches as a set of paired fixed effect linear models, based on the concept of multivariate stochastic dominance, and propose a reference framework for modelling, testing and ranking on multivariate scoring sport data. In particular, they show that traditional scoring outputs can be used to do inference on parameters related to the net relative strength or weakness of teams within a league. The novelty of their approach refers to two main aspects: first, the team ranking is obtained by hypothesis testing and not simply by ordering any point estimate of suitable parameters; second, the proposed testing and ranking approach is multivariate in nature, so a (possibly large) set of outcomes/performances can be considered. The authors prove the validity of the proposed methodology by extensive Monte Carlo simulations and describe an interesting application to basketball, and are able to highlight that the proposed methodology can be effective in facing some real problems in sport performance analytics.

Diquigiovanni and Scarpa (2019) propose an innovative hierarchical clustering method to divide a sample of undirected weighted networks into groups. The proposed methodology is applied in football, in order to detect the main tactics shown by the Italian Serie A teams and to verify the effects of the styles of play on the number of goals scored. The application to the analysis of playing styles shows interesting results: for example, the building up of the offensive manoeuvre from the lateral zones of the field has a positive effect on the number of goals scored by a team.

Ley et al. (2019) aim at building a ranking reflecting a football teams' current strengths. The idea is to produce a new ranking based on ten different strength-based statistical models used to model soccer match outcomes. The models are of four main types: Thurstone-Mosteller, Bradley-Terry, Independent Poisson and Bivariate Poisson, and their common aspect is that the parameters are estimated via weighted

## 4 Guest Editorial

maximum likelihood, the weights being a match importance factor and a time depreciation factor giving less weight to matches that were played a long time ago. The models' predictive performance is then compared via the Rank Probability Score at the level of both domestic leagues and national teams. The usefulness of the newly-obtained rankings is discussed by means of some examples where the existing rankings fail to provide enough information.

Carpita et al. (2019) propose a study exploring a large open database of soccer leagues in ten European countries, retrievable from the online platform Kaggle. Role-based indicators of teams' performance have been built and used to model the winning probability of the home team, by means of a Binomial Logistic Regression model, in its base version and also extended to include the ELO rating predictor and two random effects, due to the hierarchical structure of the dataset. Comparison of those models with some alternative models shows interesting results, suggesting that role-based indicators substantially improve the performance of all the models. Defensive performance is a very important aspect, especially in the most recent seasons, while inclusion of both the ELO rating predictor and the random effects doesn't substantially improve prediction.

This special issue has been promoted by BDsports (Big Data Analytics in sports), a project developed as part of the activities of the Big & Open Data Innovation Laboratory of the University of Brescia, born in 2016 thanks to the financial support of Fondazione Cariplo and Regione Lombardia and designed to set up a unique collaboration of experts interested in sport analytics both from a scientific and a practical point of view. At the very end, we want to express our thanks to all the authors and reviewers that have made this special issue possible; it really was a pleasure for us to have the opportunity to read so many interesting contributions! We also wish to thank the Editor Jeff Simonoff for his strong support throughout the whole publication process.

> Andreas Groll Marica Manisera Gunther Schauberger Paola Zuccolotto Guest Editors