

THEMATIC ISSUE ON "STATISTICS FOR PERFORMANCE AND MATCH ANALYSIS IN SPORTS - EDITORIAL"

Maurizio Carpita

Department of Economics and Management, University of Brescia, Contrada Santa Chiara, 25122, Brescia, Italy

Rodolfo Metulini

Department of Economics, University of Bergamo, Via Caniana, 2, 24127, Bergamo, Italy

Hans Van Eetvelde

Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281-S9, 9000, Ghent, Belgium

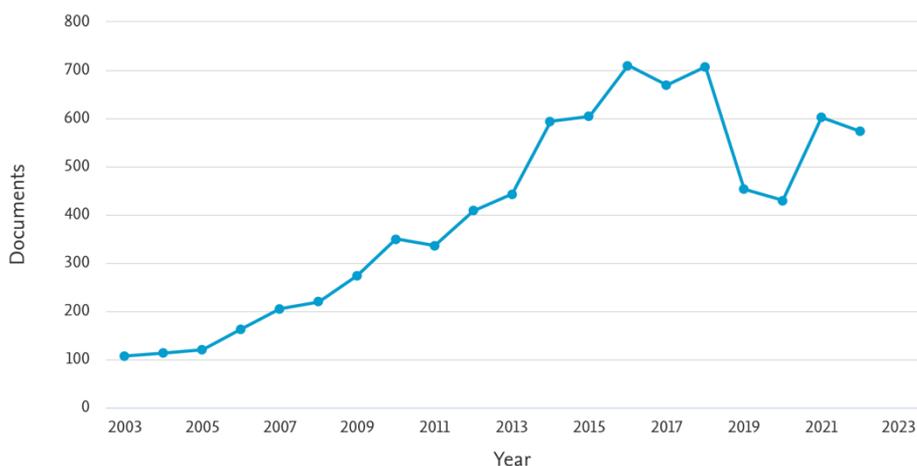
Statistics is more and more adopted in all sports with a variety of aims, ranging from predicting the outcome of a match of competition [3] and the analysis of performance [6, 7], to the prediction and prevention of injuries [9], amongst many others.

Statistical research and application in sports are fostered by the joint force of i) the increased availability of large amount of data of different types (e.g., play-by-play, trajectories, images) and from several sources (e.g., manual annotation, sensors, and tracking systems) and ii) the advances in information technologies and the computer storing capabilities [5, 8].

The interest in the topic is proved by the appearance of dedicated special issues [1, 4, 10], workshops, proceedings [2] and spontaneous contributions, also thanks to the birth of many research projects, such as the *BDsports* (<https://bodai.unibs.it/bdsports/>), which supports this thematic issue with the *ISI Special Interest Group on Sports Statistics* (<https://www.isi-web.org/isi-community/committees/sports-statistics>).

The invasiveness of this topic in the scientific research is evident: In the last twenty years (2003-2022) a total of 8,080 articles with both the words “statistics” and “sport” in the title, in the abstract or in the keywords have been published by journals indexed in Scopus. As showed in Figure 1, in the early 2000s, the number of publications satisfying such a criterion was around 100 per year. These numbers increased to more than 700 in year 2016 (with an average annual growth rate of 15.7%), they experimented a decrease to about 400 in 2019 and 2020 and they finally increased again reaching 600 articles per year in the last two years (2021 and 2022).

Figure 1. *Number of articles published in journals indexed in Scopus with both “statistics” and “sport” in the title, abstract or keywords.*



This thematic issue follows up to methodological developments in the field, by collecting original contributions that focus on the application of up-to-date statistics and machine learning methods and techniques on sport-specific problems, as are the prediction of game outcomes, the evaluation of player/athlete’s performances and traits, the search for the optimal strategy and tactics to be adopted.

This thematic issue collects ten works about individual as well as team sports. Specifically, five of them are related to basketball, three refer to soccer and two to tennis. The issue is divided in two volumes.

The first paper of the first volume of the thematic issue, by Bonnini, Corain, Pesarin, and Salmaso investigates the application of the multivariate McNemar's test for evaluating the effect of the field factor on the performance of basketball players. The proposed method is based on the nonparametric combination of permutation tests.

Gjøen, Hvattum, Moltubak, and Hvattum show through simulations that in basketball, there are game situations where a strategy of taking fewer three-point attempts at the expense of more two-point attempts will improve the probability of winning the game.

The paper of van der Wurp and Groll compares classical univariate regression approaches with copula models explicitly accounting for the dependency structure as well as with modern machine learning techniques in the context of modelling and predicting of football results in the major European leagues.

A description of the characteristics of the R-package "welo" is given by Candila. The package is dedicated to calculating the weighted and unweighted Elo rates for tennis players. It allows the user to obtain the WElo and Elo rates easily and quickly, as well as the predicted probabilities of winning.

Dona and Swartz introduce two quantitative definitions of pace in soccer, whose calculations are facilitated through the availability of player tracking data. Their study investigates the influence of playing pace on the number of shots taken by a team.

The second volume of the thematic issue begin with the paper by Biancalani, Gnecco, and Metulini studies whether, for a basketball player, obtaining a large salary can be explained by its average marginal contribution to the team performance, measured using generalized Shapley values. The study is applied to players in the NBA.

Wu and Swartz have developed automatic methods that analyze the activities of players that are "off the-ball" in soccer. They introduced a metric which

measures defensive anticipation, based on the velocity of a defensive player in a given situation. The analysis is facilitated through player tracking data.

The work by Macis, Manisera, Sandri, and Zuccolotto studies which skills are associated to the probability for a basketball player of scoring a certain number of points during a NBA season segment, by applying a Stepwise Cox regression model and a Lasso Cox regression.

Tracking data systems gain a lot of interest in football, but they are still expensive. Broadcasting videos provide an alternative for tracking data, but they are of less quality and these data are censored. Therefore, the study by Karlis and Kontos explores interpolation methods for retrieving the missing information of players' and ball's positions and rectifies the effect of censoring.

The thematic issue concludes with the study by Milekhina, Breznik, and Restaino, that aims to investigate the existence of professional tennis players' psychological traits. For this purpose, datasets on tennis matches of professional male and female tennis players were collected and a dynamical network analysis was applied using the RSiena program.

Finally, many thanks to all the reviewers that make this special issue possible.

The Guest Editors

Maurizio Carpita, University of Brescia, Italy
Rodolfo Metulini, University of Bergamo, Italy
Hans Van Eetvelde, Ghent University, Belgium

REFERENCES

1. Brefeld, U., Zimmermann, A. (2017). Guest editorial: Special issue on sports analytics. *Data Mining and Knowledge Discovery*, 31(6), 1577–1579. doi: 10.1007/s10618-017-0530-1
2. Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (2018). Machine

- learning and data mining for sports analytics. Cham: Springer. doi: 10.1007/978-3-030-17274-9
3. Bunker, R., Susnjak, T. (2022). The Application of Machine Learning Techniques for Predicting Match Results in Team Sport: A Review. *Journal of Artificial Intelligence Research* 73:1285–1322. doi: 10.1613/jair.1.13509
 4. Groll, A., Liebl, D. (2022). Editorial special issue: Statistics in sports. *AStA Advances in Statistical Analysis*, 1-7. doi: 10.1007/s10182-022-00453-9
 5. Gudmundsson, J., Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys (CSUR)*, 50(2), 1-34. doi: 10.1145/3054132
 6. Lord, F., Pyne, D.B., Welvaert, M., Mara, J.K. (2020). Methods of performance analysis in team invasion sports: A systematic review. *Journal of Sports Sciences* 38:2338–2349. doi: 10.1080/02640414.2020.1785185
 7. Sarlis, V., Tjortjis, C. (2020). Sports analytics — Evaluation of basketball players and team performance. *Information Systems* 93:101562. doi: 10.1016/j.is.2020.101562
 8. Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., Kosub, S., Schreck, T., Keim, D.A., Grossniklaus, M. (2017). How to make sense of team sport data: From acquisition to data modeling and research aspects. *Data*, 2(1), 2. doi: 10.3390/data2010002
 9. Van Eetvelde, H., Mendonça, L.D., Ley, C., Seil, R., Tischer, T. (2021). Machine learning methods in sport injury prediction and prevention: a systematic review. *J EXP ORTOP* 8:27. doi: 10.1186/s40634-021-00346-x
 10. Zuccolotto, P., Manisera, M., Kenett, R. (2017). Statistics in Sports: Project BDSports (Big Data Analytics in Sports) bodai. unibs.it/BDSports. *Electronic Journal of Applied Statistical Analysis*, 10(3). ISSN: 2070-5948