
The combination of non-linear analyses and machine learning for the quantification of competition and match workload in men and women soccer players

Nicolas Babault*¹, Carole Cometti¹, Joao Luiz Quagliotti Durigan², Christos Paizis¹,
and Tom Douchet^{1,3}

¹Center for performance expertise, CAPS, U1093 INSERM, University of Bourgogne Europe, Sport Science faculty, 3 allée des stades universitaires, BP 27877, 21078 Dijon Cedex – INSERM U1093, Cognition Action et Plasticité Sensorimotrice – France

²Universidade de Brasilia = University of Brasilia [Brasília] – Brésil

³Racing Club de Lens – RCLens – France

Résumé

Introduction

The demands of team sport competitions and training are regularly monitored. It is easily achieved thanks to the democratization of outdoor and indoor tracking systems. Studies have extensively explored the physical demand of soccer games for different playing positions, playing standards or age groups (Harkness-Armstrong et al., 2022). It has been quantified using various parameters such as the total distance, distance within velocity thresholds or number of accelerations/decelerations. A large inter-individual variability is often revealed. Players also exhibit a huge intra-individual variability that could be quantified using the most intense periods during a match; the so-called worst-case scenario (Douchet et al., 2025). However, the different methods generally address the magnitude of the variability and not the structure of this variability (i.e., regularity or periodicity of the displacements variations). Non-linear analyses based on entropy or fractality have been shown to be an alternative signal computing that can witness various magnitudes and temporal dynamics of the player displacement. For instance, using non-linear analyses, studies have shown an increase in the displacements' variability during small-sided games with increasing the number of players with a dominance of small as compared to large velocity fluctuations (Cruz & Sampaio, 2020). In addition, offensive players demonstrated greater velocity regularity than defensive players and an increased regularity and predictability was also recorded with the progression of the game (Babault et al., 2024). Therefore, non-linear analyses provide complementary outcomes to those frequently used to explore soccer workload. Accordingly, such signal computing could be of interest to improve machine learning classifications. The aim of this study was to explore the combination of linear and non-linear analyses to machine learning to classify playing positions during official games in elite soccer players.

Methods

The present study was based on GPS signals recorded during official games in a professional soccer team. 159 GPS files were analyzed using linear and non-linear methods. Linear

*Intervenant

analyses were based on total distance, moderate, high-speed and sprint distance and number of accelerations and decelerations. Non-linear analyses were based on the sample entropy (an index of irregularity) and multifractal spectrum (range of variability and velocity fluctuations). Different models of machine learning were then applied and included Linear regression, Naïve Bayes, Support Vector Machine, Random Forest or eXtreme Gradient Boosting.

Results

Random Forest and eXtreme Gradient Boosting appeared the most efficient machine learning methods to classify playing positions with accuracy ranging between 0.63-0.72. Accuracy and F1 scores revealed these two methods were efficient to classify central and lateral defenders and less efficient for midfielders and forwards. Considering linear and non-linear analyses altered machine learning results. Including sample entropy calculated from velocity and acceleration traces enhanced machine learning classifications with a slight increase in most machine learning indexes. Including multifractal results had no effect on machine learning prediction.

Discussion

Non-linear analyses revealed complementary outcomes to explore soccer workload with intra-individual variability dynamics. While this additional information could help the training staff to optimize training sessions, it appeared it had slight effect on machine learning prediction. Amongst the different non-linear parameters that could be quantified, sample entropy seemed the most important.

Conclusions/Perspectives

Non-linear analyses, and more particularly sample entropy, could be interesting parameters to improve the analysis of soccer workload and to improve machine learning efficiency.

References

- Babault, N., Rodot, G., Cometti, C. & Vieira, D. C. L. (2024). Elite women's soccer match demand can be described using complexity-based analyses and multifractals. *Chaos, Solitons & Fractals*, 189, 115612. <https://doi.org/10.1016/J.CHAOS.2024.115612>
- Cruz, I. F. & Sampaio, J. (2020). Multifractal Analysis of Movement Behavior in Association Football. *Symmetry* 2020, Vol. 12, Page 1287, 12(8), 1287. <https://doi.org/10.3390/SYM12081287>
- Douchet, T., Michel, A., Verdier, J., Babault, N., Gosset, M. & Delaval, B. (2025). Intensity vs. Volume in Professional Soccer: Comparing Congested and Non-Congested Periods in Competitive and Training Contexts Using Worst-Case Scenarios. *Sports* 2025, Vol. 13, Page 70, 13(3), 70. <https://doi.org/10.3390/SPORTS13030070>
- Harkness-Armstrong, A., Till, K., Datson, N., Myhill, N. & Emmonds, S. (2022). A systematic review of match-play characteristics in women's soccer. *PLoS ONE*, 17(6 June), e0268334. <https://doi.org/10.1371/journal.pone.0268334>