Does Training Affect Match Performance? A Study Using Data Mining And Tracking Devices

Javier Fernández[1,2], Daniel Medina[1], Antonio Gómez[1], Marta Arias[2], and Ricard Gavaldà[2]

[1] Fútbol Club Barcelona, Ciudad Deportiva Joan Gamper, Barcelona, Spain
[2] Universitat Politècnica de Catalunya, Campus Nord, Barcelona, Spain

Abstract. FIFA has recently allowed the use of electronic performance and tracking systems (*EPTS*) in professional football competition, providing teams with novel and more accurate data. Physical performance has not yet taken much attention from the research community, due to the difficulty of accessing this information with the same devices during training and competition. This study provides a methodology based on machine learning and statistical methods to relate the physical performance variation of players during time-framed training sessions, and their performance in the following matches. The analysis is carried out over F.C. Barcelona B, season 2015-2016 data, and makes emphasis on exploiting the design characteristics of the *structured training* methodology implemented within the club. The use of summarized physical variation data has provided a remarkable relation between higher magnitudes of variation in 3-week time frames during training, and higher physical values in the following matches. With increased data availability this and new approaches could provide a new frontier in physical performance analysis. This is, up to our knowledge, the first study to relate training and matches performance through the same *EPTS* devices in professional football.

Keywords: GPS, tracking devices, football physical performance, sports analytics, dtw, cluster analysis

1 Introduction

Professional football has attracted the attention of the data science community in the last decade due to the increasing availability of quantitative data. The latest technology has provided the possibility of gathering different kinds of specific metrics, from team statistics to in-game detailed events, contributing to the improvement of typical and critical tasks such as team tactics evaluation, opponent analysis, player scouting and training design. The idea that exploiting data-related analysis can become a competitive advantage within professional sports is increasingly supported [1]. However, it should be noted that few of the current studies are devoted to the analysis of physical information of the players [2]. This has to do mainly with the difficulty of having access to this data through training and competition, which is considered highly valued by football clubs [2]. Typically, such information is gathered through the use of electronic performance and tracking systems (EPTS) which include GPS and microsensor technology such as accelerometeres, gyroscopes and magnetometers. Such is the case of professional sections at F.C. Barcelona where these tools are used for monitoring load and many other physical variables. Despite the existing concerns regarding its reliability, they have increasingly being adapted and accepted in sports such as Rugby, Australian football, Cricket and Hockey [3]. Recently, the Football Association Board (IFAB) has authorized the use of these devices during official football competition for the 2015-2016 season [4], opening the doors for novel research regarding physical performance of players during the season.

At F.C. Barcelona, EPTS devices have been recently used to aid the evaluation of the applied training methodology, the *structured training*, a system that sets the baselines for the planning and adaptation of the training activities along the season, providing the novelty of incorporating competition activities in this design. This involves the idea of providing a schema in which the player is promoted to adapt to the training demands and evolve in each of its structures, beyond the strictly physical conditions [5]. A *player optimization* is sought through the application of training situations that cause imbalance in one of the subject's structures in order to promote its adaptation, so forcing a continuous auto-organization process in sets of 3 weeks periodization [5]. This methodology considers not only training as a stimulus to induce adaptation but also competition as the most relevant stimulus to optimize the athlete capabilities. This implies that physical demands for players during training are structured within consecutive cycles but are not strictly defined, so the measured physical player values can provide uncertainty and richness in its analysis. Also, given the idea of *deterministic chaos* present in biological systems [6], players are expected to evidence different adaptational behaviors along the season trainings. Based on this, it is plausible to think that periodical variation of physical values could provide valuable information regarding the adaptability and fitness of the player.

The main objective of this study is to find whether there exist significant relations between physical performance of players during training and the measured performance in subsequent matches, for F.C. Barcelona B data from season 2015-2016. Machine learning algorithms are used in order to exploit the contribution of the high amount of measured variables as a whole, all of which are expected to contribute explaining the player's dynamic up to some extent. The study is structured in three main stages. A data preparation stage in which data is pre-processed and normalized, and two datasets are created. An exploration stage where dynamic time warping and cluster analysis is applied in order to obtain representative natural groups from data. And finally, a validation stage, where the matches associated with clustered series are extracted and statistical tests are performed to determine the existence of significant differences. Final conclusions and future work suggestions are detailed, regarding the usefulness of this approach and the finding of moderate standardized differences between groups presenting high and low variations of physical values from week to week.

2 Methodology

2.1 Data Collection

F.C. Barcelona B has collected both training and matches physical performance measurements, for season 2015-2016, using the *StatsSports GPS Viper Pod* devices. The resulting tracking information is manually segmented by physical coaches, and further visualized through a software integrated with the devices which outputs several variables. From this set of variables, we have selected 15 along physical coaches, described in Table 1, which summarize the considered most relevant performance information. Variables are structured in three main groups: locomotor, metabolic and mechanical. Locomotor variables refer to simple direct measurements of travelled distance and speed, that are obtained solely through GPS. Metabolic variables are associated with energy expenditure and exertion, while mechanical variables relate with intensity changes and impacts [7]. For these last two groups variables are calculated by a combination of GPS and accelerometers. The data consists of 153 training sessions and 34 matches, which adds up to 2478 training rows and 473 match rows among all the 42 different players throughout the season 2015-2016. The season information is queried from the central database containing the total 2951 rows, where each one contains the measured variables for a single player in a specific session and additional variables that contextualize the information such as player id, position, name, total session time, the session id and session type.

2.2 Data Processing

The dataset is initially processed, adding additional contextualization variables and performing several types of normalizations. Within F.C. Barcelona training structure, training days are labelled in strict relation with the following match day, where match is labelled as MD, the following two days MD+1 and MD+2, and the previous days MD-1 up to MD-4. Each day-type follows specific design rules for training drills. For simplicity of the study, only day MD-3 sessions are used, due to they similarities to match days in terms of number of players, playing spaces and opposition level. Additionally, day MD-3 involves the highest differences between physical values. Goalkeepers are deleted from the database since they face considerably different physical challenges than field players. A new variable, load percentage (PER) is added in order to reflect the session load, which is calculated as a ratio of the average AMP from matches. All the measured values are normalized by dividing by the total time of duration of the session. Variables that already represent averages or maximums are kept

Table 1: Description of selected physical variables splitted in three groups: locomotor, metabolic and mechanical.

Locomotor Variables	
Name and Acronym	**Description**
Travelled Distance (DIS) [8]	Total distance travelled during session drills or matches
Sprints (SPR) [8]	Number of times over $5.5m/s$ during > 1
High Speed Running (HSR) [8]	Travelled meters when speed $> 5.8m/s$
Max Speed (MAX) [8]	Maximum speed reached by the player

Metabolic Variables	
Name and Acronym	**Description**
Average Metabolic Power (AMP) [8]	Energy expended by the player per second per kg, measured in W/Kg
High Metabolic Load Distance (HML) [8]	Distance travelled by a player when the metabolic power is $> 25.5W/Kg$
High Metabolic Efforts (HEF) [9]	The number of separate movements/efforts undertaken in producing HML distance
Load Percentage (PER)	Proportion of AMP with respect to an average 9.5 AMP in matches

Mechanical Variables	
Name and Acronym	**Description**
Fatigue Index (FAI) [8]	Accumulated DSL from the total session volume, in terms of speed. (DSL/SPI)
Dynamic Stress Load (DSL) [8]	Total of the weighted impacts, based on accelerometer values over 2g
Lower Speed Loading (LSL) [8]	Load associated with the low speed activity alone
Total Loading (TLO) [8]	The total of the forces on the player over the entire session based on accelerometer data alone
Accelerations (ACC) [8]	Number of increases of speed during at least 0.5 s ($> 3m/s^2$)
Decelerations (DEC)[8]	Number of decreases of speed during at least 0.5 s ($< 3m/s^2$)
Step Balance (STE) [8]	Ratio of left step impact to the sum of the left step impact and right step impact

as originally measured, such as AMP, FI, PER, STE and MAX. Additionally, summarized information is added to matches data such as the average training minutes, average fatigue and total (training plus match) load in the previous three weeks. An additional normalization is applied where absolute values are transformed into the number of standard deviations of each particular player in the given day label type. This transformation is performed in order to avoid differences that arise due to player physical characteristics instead of a response to training. Finally, a last transformation performed over training data seeks to quantify the degree of variability from week to week on each physical value. The idea is to measure the difference between registered values from two consecutive weeks, as presented in Figure 1.

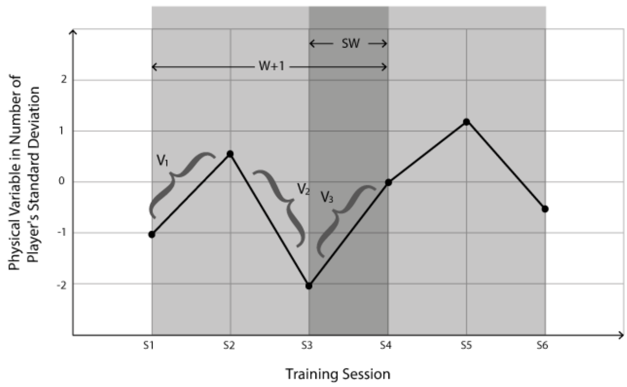

Fig. 1: Representation of a series of measured values of a particular variable during weekly training sessions (x axis). V_i values refer to the difference of values registered at sessions S_{i+1} and S_i. W is the size of the sliding window, used to build time-series and summarized datasets. SW refers to the amount of weeks to slide each time.

Each value V_i represents the absolute difference between a value registered at sessions S_{i+1} and S_i. Two datasets were built: the first one consists of time-series of W window size. A sliding window approach is followed by using a fix-sized (W) window of consecutive weeks. The time-series dataset is conformed by groups of W rows containing the 15 physical variables, corresponding to a player in a specific period of the season. Selected windows sizes during experiments are 3 and 6 in order to match the methodology of the club. Windows are moved SW steps each time, so to control the degree of coincidence of values between windows. The value of SW was selected following Equation (1) to avoid an excessive overlap between windows and to avoid a too strict separation that would reduce significantly the amount of data. Another dataset is built which summarizes each group of W rows in each variable, by calculating the average of absolute differences. Equation (2) describes the performed calculations, where

P_{jvd} corresponds to the absolute average of window differences of a variable v of a player j, measured in the window frame d, substracted by the mean of P_{ivd} for every other player i. P corresponds to the set of all possible players.

$$SW = W - (W/3) \tag{1}$$

$$P_{jvd} = \frac{\sum_{i=2}^{W+1} \|S_i - S_{i-1}\|}{W} - \frac{\sum_{i \neq j}^{|P|} P_{ivd}}{|P|} \tag{2}$$

2.3 Data Exploration

Visual Exploration. Specific differences of physical variables where assessed visually through boxplots and analytically through one-way *ANOVA* and Post Hoc tests observing the differences between type-days (i.e MD-4, MD-3, MD-2, etc.). A *PCA* analysis was also performed, and the two principal components where plotted accounting for 69% of variance and observing the acknowledged differences. On the other hand, different plots over the time-series and summarized datasets allowed to visualize oscillatory patterns along the season that respond to cycles design. Also, it is observed how players tend to oscillate in similar patterns due to the training design. There exist, however, several cases in which certain players magnitude of variations starts differing considerably from the mean variation. The results of these observations coincided with the understanding of physical responses in training from the club's physical coaches. For space restriction reasons, the graphical results are omitted from this section.

Calculating Series Similarities through Dynamic Time Warping. Dynamic time warping (DTW) is a highly used method that allows to measure the similarity between two temporal series, while being less sensitive to signal transformations such as shifting, uniform amplitude scaling or uniform time scaling [12]. DTW was applied over the time-series dataset in order to calculate similarity between windowed variations along the season on different players. The idea is to find variation patterns that are more similar to each other, independently from the specific player or position. A distance or dissimilarity matrix is found for each pair of series in the dataset. Euclidean distance was used, in order to prioritize vectors magnitude over angles since the degree of variation is believed to be more informative than the actual followed pattern, in order to approximate the physiological response. Once the dissimilarity matrix is found, the *k-mediods* algorithm is applied for finding a natural clustering of the time series.

Cluster Analysis For both datasets cluster analysis is applied to find natural groupings of variation. It is critical to observe that the clustering procedure is applied to multidimensional data, aiming to incorporate the relation between each of the variables. For the time-series dataset the k-mediods algorithm is used, since its capability of being applied to distance matrices and the flexibility of controlling the number of clusters. For the summarized dataset, *k-means* is used instead. The selection of number of clusters is performed by calculating five internal indices and selecting the number of clusters picked by the majority. These indices are: C-index, C-H index, DB index, Silhouette index and the Ratkowsky-Lance index [13]. Also, the dimensionality reduction technique T-Stochastic Neighbor Embedding *t-SNE* [14] was applied to visually asses the quality of clusters. Once the training sessions information is clustered, each of the window-frames is associated with next upcoming match, generating a cluster-labelled dataset containing the absolute values of matches physical variables.

3 Results

Results for both the time-series and the summarized datasets are presented together since they follow an identical approach in its evaluation. For both cases, the selected number of clusters was 2 by four of the five different indices, the sample size of the training sessions dataset is 112, and the sample size of associated matches dataset is 82. Only the results for 3-week window are presented, since no statistically significant relation was found with 6-week window frames. For each of the variables conforming the two groups (in each dataset) the standardized difference of means was calculated to describe the effect size. The limits of the effect sizes are those suggested by Hopkins [16] which are recommended in sports related data and for practical applications (trivial effect: < 0.2, small effect: $0.2 - 0.6$, moderate effect: $0.6 - 1.2$, large effect $1.2 - 2.0$ and, very large: > 2.0), with a confidence interval of 90%.

Detailed results are presented in Table 2. It can be clearly observed that for the summarized dataset almost every variable in training registered a moderate to large effect size when comparing groups. So, we are observing the detection of two groups: one where the average magnitude of variations of each variable is higher (*high variation group*), and one where is lower (*low variation group*). It is critical to observe that separation among groups is not absolute, and there exist ranges of values which overlap. This has to do with multivariate nature of the clustering procedure, and coincides with the original expectation of this study. It can also be observed that for the timeseries dataset few variables where able to stand out just with a small size effect. Even with the selection of Euclidean distance to favor magnitudes, the cluster analysis over the DTW procedure was not able to found a clear separation between groups. The procedure over the summarized dataset, instead, did find a considerably separation between training groups so the analysis over associated matches is easier to interpret and translate to practice. Figure 2 presents the effect sizes for the associated matches in both datasets. It can be observed for both cases that variables registering high intensity efforts, energy consumption and distance travelled appear with higher magnitude in the *high variation group* consistently, while the total load percentage and training minutes in the previous three weeks are considerably low in this same group. HML, AMP and DIS present moderate effect size in the summarized dataset, variables belonging to metabolic group (the first two) and locomotor group. For the timeseries dataset only HML presents a moderate effect size, toward the same tendency. A small effect size is also observed in other locomotor (MAX), metabolic (PER and HSR) and mechanical variables (DSL, DEC) toward the same tendency. Three-weekly PER and training minutes show also a moderate effect in differences, towards lower values. Sample size for associated matches allows to conclude with certainty about moderate size effects. Small effects should be taken into account, but must be further validated with the future increase of availability of data.

Table 2: Mean and standard deviation for each physical variable in each of the clustered groups. For both training data and the associated matches, values obtained in both summarized and timeseries datasets are presented. The standardized difference of means *SDM* is presented for each case. Training results refer to the absolute average of variation while matches results refer to the actual measured physical values.

	Training (mean ± SD)						Matches (mean ± SD)					
	Summarized			Timeseries			Summarized			Timeseries		
Variable	Cluster 1	Cluster 2	SDM	Cluster 1	Cluster 2	SDM	Cluster 1	Cluster 2	SDM	Cluster 1	Cluster 2	SDM
DSL p/m	0.89 ± 0.29	0.58 ± 0.28	*Moderate*	0.41 ± 0.198	0.41 ± 0.18	*Trivial*	3.60 ± 1.30	3.26 ± 1.37	*Small*	3.37 ± 1.17	4.29 ± 1.84	*Small*
ACC p/m	1.011 ± 0.45	0.67 ± 0.31	*Moderate*	0.45 ± 0.26	0.49 ± 0.23	*Trivial*	0.56 ± 0.22	0.52 ± 0.24	*Trivial*	0.65 ± 0.23	0.561 ± 0.21	*Small*
DEC p/m	0.94 ± 0.38	0.51 ± 0.23	*Large*	0.43 ± 0.25	0.38 ± 0.19	*Trivial*	0.80 ± 0.27	0.69 ± 0.26	*Small*	0.82 ± 0.29	0.785 ± 0.307	*Trivial*
SPR p/m	0.74 ± 0.35	0.59 ± 0.28	*Small*	0.36 ± 0.17	0.40 ± 0.20	*Trivial*	0.37 ± 0.11	0.311 ± 0.09	*Small*	0.37 ± 0.12	0.38 ± 0.20	*Trivial*
HSR p/m	0.98 ± 0.46	0.53 ± 0.26	*Large*	0.39 ± 0.17	0.43 ± 0.28	*Trivial*	13.30 ± 4.73	10.78 ± 4.31	*Small*	12.58 ± 5.27	10.686 ± 3.837	*Small*
AMP	0.62 ± 0.36	0.29 ± 0.14	*Large*	0.28 ± 0.10	0.25 ± 0.13	*Small*	10.35 ± 1.08	9.65 ± 1.17	*Moderate*	10.27 ± 1.09	10.26 ± 1.50	*Trivial*
HML	0.58 ± 0.24	0.40 ± 0.20	*Moderate*	0.33 ± 0.14	0.25 ± 0.13	*Small*	39.12 ± 10.05	32.05 ± 10.01	*Moderate*	36.09 ± 9.11	30.892 ± 7.95	*Moderate*
HEF p/m	0.74 ± 0.31	0.40 ± 0.22	*Large*	0.34 ± 0.17	0.29 ± 0.19	*Small*	2.23 ± 0.52	1.94 ± 0.5	*Small*	2.31 ± 0.5	2.20 ± 0.64	*Trivial*
FAI	0.93 ± 0.39	0.71 ± 0.34	*Moderate*	0.47 ± 0.24	0.48 ± 0.21	*Trivial*	0.62 ± 0.19	0.64 ± 0.25	*Trivial*	0.602 ± 0.18	0.75 ± 0.29	*Moderate*
DIS p/m	0.58 ± 0.36	0.27 ± 0.15	*Large*	0.25 ± 0.10	0.21 ± 0.20	*Small*	111.5 ± 10.98	104 ± 11.3	*Moderate*	109.5 ± 10.72	110.71 ± 15.41	*Trivial*
TLO p/m	0.83 ± 0.25	0.39 ± 0.21	*Large*	0.35 ± 0.16	0.32 ± 01.9	*Small*	1.59 ± 0.23	1.48 ± 0.32	*Small*	1.56 ± 0.22	1.67 ± 0.38	*Small*
MAX	0.80 ± 0.39	0.63 ± 0.32	*Trivial*	0.40 ± 0.198	0.42 ± 0.22	*Trivial*	29.6 ± 2.11	28.79 ± 2.32	*Small*	29.612 ± 2.57	29.28 ± 1.61	*Trivial*
STE	1.05 ± 0.57	0.97 ± 0.53	*Trivial*	0.61 ± 0.35	0.596 ± 0.317	*Trivial*	0.006 ± 0.03	0.008 ± 0.026	*Trivial*	0.012 ± 0.022	0.002 ± 0.029	*Small*
PER	0.57 ± 0.35	0.27 ± 0.15	*Large*	0.23 ± 0.12	0.22 ± 0.19	*Trivial*	0.96 ± 0.25	0.85 ± 0.29	*Small*	0.86 ± 0.31	0.73 ± 0.34	*Small*
3W Training PER	–	–	–	–	–	–	5.26 ± 1.24	8.29 ± 1.68	*Moderate*	7.04 ± 1.09	7.45 ± 1.79	*Moderate*
3W Training Minutes	–	–	–	–	–	–	796 ± 171	964 ± 201	*Moderate*	725 ± 186.50	873.30 ± 202.87	*Moderate*
3W Total PER	–	–	–	–	–	–	7.71 ± 1.08	8.29 ± 1.68	*Small*	7.04 ± 1.09	7.45 ± 1.79	*Small*
3W Average FAI	–	–	–	–	–	–	0.65 ± 0.14	0.67 ± 0.18	*Trivial*	0.629 ± 0.14	0.77 ± 0.22	*Moderate*

4 Conclusions and Future Work

The presented approach allowed to observe considerable relation between training variations and match performance. The players presenting higher variations during training reflected in higher values in 11 of the 15 analyzed variables for locomotor (4/4), metabolic (4/4) and mechanical (3/7) groups in the next matches, and also lower training minutes and accumulated load during training. This approach might provide a way for analyzing the adaptation of players to training dynamics, and even to evaluate training design. The procedure follows a series of simplifications such as the selection of day-type MD-3 which might incur in loss of information. However, this type of calculations can be easily integrated to daily routine performance analysis carried out by physical coaches, without the need of additional systems or requiring high processing times. The findings provide sufficient evidence to suggest the incorporation of this calculation in daily analysis and track its evolution in order to further measure is

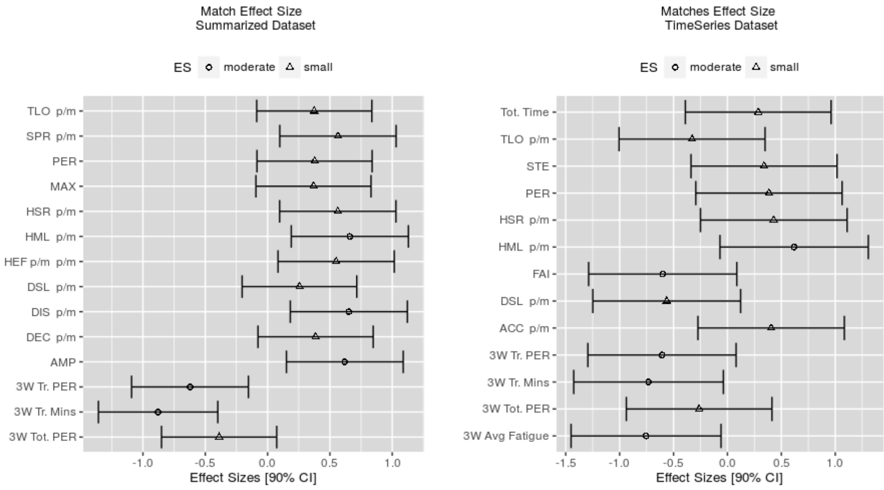

Fig. 2: Effect size differences in group mean values in standardize units for matches groups found through the summarized dataset (left) and the timeseries dataset (right). Trivial effect sizes are not shown.

effectiveness on relating with match performance.

The summarized dataset allowed a more representative grouping and more conclusive results. In practice, high and low variations can be found directly by using the ranges found by the clustering procedure for each variable. Also, time-window aggregated information is showing to add value for performance analysis and should be considered in future research. On the other hand, DTW could not provide sufficiently clear results in this study, most probably due the the short-size characteristics of the analyzed time series and that exact match of variation patterns might be too strict for the few data available. Also, the player normalization seems to favor a cleaner comparison between players, instead of using absolute values which could lead to differences that are more related to physical characteristics than actual adaptation patterns.

This is the first study, up to our knowledge, to relate training and match physical values directly registered from player using EPTS devices during training and matches for a whole season. In the following years, with higher availability of data these remarks must be further validated. Future work should incorporate new day-types in the analysis and factors beyond the physical such as tactical information and variables related with psychological information such as the rate of perceived exertion (RPE). The yearly knowledge of physical evolution of training dynamics and even specific players might provide new insights about the physical preparation of teams and the performance during competition.

References

1. McCall A., Davison M., Carling C., Buckthorpe M., Coutts A.J., Dupont G., Can off-field 'brains' provide a competitive advantage in professional football?. Br J Sports Med, vol. 50, pp. 710-712 (2016)
2. Gyarmati L., Hefeeda M., Estimating the Maximal Speed of Soccer Players on Scale, In Proc. of Machine Learning and Data Mining for Sports Analytics Workshop, Porto, Portugal, (2015).
3. Cummins C., Orr R., OConnor H., West C. Global Positioning Systems (GPS) and Microtechnology Sensors in Team Sports: A Systematic Review. Sports Med vol. 43, pp. 10251042 (2013)
4. IFAB. 129th Annual General Meeting The Football Association, http://resources.fifa.com/mm/document/affederation/ifab/02/60/90/85/2015agm_minutes_v10_neutral.pdf (2015)
5. Arjol J., La planificación actual del entrenamiento en fútbol. Análisis comparado del enfoque estructurado y la periodizacin táctica. Acciónmotriz, vol. 8, pp. 27-37 (2012)
6. Lesne A., Chaos in Biology. Biology Forum / Rivista di Biologia. vol. 99, issue 3, pp. 467-481 (2006)
7. Gaudino P., Alberti G., and Iaia M. Estimated metabolic and mechanical demands during different small-sided games in elite soccer players. Elsevier, vol. 36, pp. 123-133. (2014)
8. STATSports Technologies Ltd. STATSSports Viper Metrics. version 1.2 (2012)
9. Sandbakk Ø., Cunningham D., Shearer D., Drawer S., Eager R., Taylor N., Cook C., Kilduff L. Movement Demands of Elite U20 International Rugby Union Players. Plos One vol. 11, issue 4, pp. e0153275 (2016)
10. Talukder, Hisham and Vincent, Thomas and Foster, Geoff and Hu, Camden and Huerta, Juan and Kumar, Aparna and Malazarte, Mark and Saldana, Diego and Simpson, Shawn Simpson. Preventing in-game injuries for NBA players. MIT Sloan Sports Analytics Conference 2016. http://www.sloansportsconference.com/wp-content/uploads/2016/02/1590-Preventing-in-game-injuries-for-NBA-players.pdf
11. Bangsbo J, Mohr M., Krustrup P. Physical and metabolic demands of training and match-play in the elite football player. Journal of Sports Sciences, vol. 24 , pp. 665-674 (2006)
12. Ratanamahatana, Chotirat A., and Eamon k. Everything you know about dynamic time warping is wrong Third Workshop on Mining Temporal and Sequential Data. (2004)
13. Desgraupes B., University Paris Ou. (2013) https://cran.r-project.org/web/packages/clusterCrit/vignettes/clusterCrit.pdf
14. Van der Maaten L., Hinton G. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research, vol. 9. pp. 2579-2605 (2008)
15. Cohen, J. A power primer. Psychological Bulletin, vol. 112, pp. 155-159 (1992)
16. Hopkins WG, Marshall SW, Batterham AM, and Hanin J. Progressive statistics for studies in sports medicine and exercise science. Medicine and science in sports and exercise vol. 41, pp. 3-13. (2009)