Applied Mathematical Modelling 45 (2017) 642–652

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Multidimensional scaling analysis of soccer dynamics

J.A. Tenreiro Machado[a], António M. Lopes[b,*]

[a] *Institute of Engineering, Polytechnic of Porto, Dept. of Electrical Engineering, R. Dr. António Bernardino de Almeida, 431, 4249 – 015 Porto, Portugal*
[b] *UISPA–LAETA/INEGI, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200 – 465 Porto, Portugal*

ARTICLE INFO

Article history:
Received 17 March 2016
Revised 8 November 2016
Accepted 5 January 2017
Available online 12 January 2017

Keywords:
Multidimensional scaling
Visualization
Long-term performance
Sport dynamics

ABSTRACT

This paper studies the behavior of teams competing within soccer national leagues. The dissimilarities between teams are measured using the match results at each round and that information feeds a multidimensional scaling (MDS) algorithm for visualizing teams' performance. Data characterizing four European leagues during season 2014–2015 is adopted and processed using three distinct approaches. In the first, one dissimilarity matrix and one MDS map per round are generated. After, Procrustes analysis is applied to linearly transform the MDS charts for maximum superposition and to build one global MDS representation for the whole season. In the second approach, all data is combined into one dissimilarity matrix leading to a single global MDS chart. In the third approach, the results per round are used to generate time series for all teams. Then, the time series are compared, generating a dissimilarity matrix and the corresponding MDS map. In all cases, the points on the maps represent teams state up to a given round. The set of points corresponding to each team forms a locus representative of its performance versus time.

1. Introduction

In the last decades the mathematical tools of dynamical systems have been used to provide a framework for modeling sport dynamics [1–6]. More than statistics of single aspects of the match [7,8], the main focus became the quantitative analysis of space-time patterns that emerge during the game [2,9–11], and the collective behavior of the teams [12].

Along with the match dynamics, the long-term competitiveness of the teams has been investigated by many authors [13–16]. Traditional approaches measure competitiveness focused on simple statistics, such as the win ratios, points accumulation, or the dominance of subsets of teams over a number of seasons [17,18]. However, many rankings do not give the proper relevance to the match results, the match importance, or the opponent strength, among others, and may yield dubious conclusions. Competitiveness resulting from the behavior of teams competing within the same league has been studied using data envelopment analysis [16], complex networks [19], graphs and centrality measures [20], among others [21].

This paper studies the behavior of teams competing within a soccer league season by means of multidimensional scaling (MDS). Soccer (also known as association football, or football) is one of the most popular sports around the world, involving over than 250 million players in about 200 countries [1,22]. The game is played by 2 teams, with 11 players each, on a rectangular field with a goal placed at each end. The match has 2 periods of 45 minutes. The objective of the game is to score by getting the ball into the opposing goal. The team that scores more goals by the end of the match wins. The 10

* Corresponding author.
E-mail address: aml@fe.up.pt (A.M. Lopes).

http://dx.doi.org/10.1016/j.apm.2017.01.029

field players can maneuver the ball using any part of their bodies except hands and arms, while the goalkeeper is allowed to touch the ball with the whole body, as long as he/she stays in his/her penalty area.

In our methodology, we consider the results of the matches at each round of a soccer league for quantifying dissimilarities between teams. We adopt data from four national leagues, namely the Spanish “La Liga”, the English “Premiership”, the German “Bundesliga” and the Portuguese “Primeira Liga”, for the season 2014–2015. We process data by means of the MDS technique, generating visualization maps that unveil patterns and relationships between teams. We test three distinct approaches. In the first strategy, we construct one dissimilarity matrix and we generate the corresponding MDS maps per round. We then apply Procrustes analysis (PA) to linearly transform the MDS charts for maximum superposition, leading to one global MDS map. In the second approach, we combine all data into a single dissimilarity matrix that yields one MDS global chart. In the third scheme, we use the results per round to generate time series for describing the performance of each team, where the discrete time unit corresponds to one round. We then compare all time series and we construct one dissimilarity matrix to generate one global MDS map. For all cases, a point in the MDS global map represents one team at a given round. The set of points corresponding to one team forms one locus that captures the dynamic performance of the team along the season.

Bearing these ideas in mind, this paper is organized as follows. Sections 2 and 3 describe the experimental dataset the mathematical tools adopted, respectively. Section 4 analyses the teams behavior along the 2014–2015 soccer season by means of MDS. Finally, Section 5 outlines the main conclusions.

2. Description of the dataset

Data for worldwide soccer is available at http://www.worldfootball.net/. The database contains information about national leagues and international competitions. For the national leagues, the matches results are available in a per season basis. For each match we have the names of the home and away teams, the goals scored, the points won, and the date of the match, among other data.

We consider data of four top national European leagues:

- “La Liga” was established in 1929 as the top professional division of the Spanish soccer league system. It has been considered by UEFA the strongest league in Europe for the last five years. Since 1997 “La Liga” is contested by 20 teams. Every season the three lowest placed teams are relegated to the “Second Division”, being replaced by the top two teams of this league plus the winner of a play-off competition. A total of 60 teams have competed in “La Liga” since 1929. Nine teams have been crowned champions, with Real Madrid winning the title for 32 and FC Barcelona for 23 times. The other champions were Atlético Madrid, Athletic Bilbao, Valencia CF, Real Sociedad, Real Betis, Sevilla FC and Deportivo La Coruña.
- The “Premier League”, or “Premiership”, is the most important league of the English association football. It was established in 1992, following the decision made by the top 22 clubs to break away from the “Football League”. The “Premiership” is now contested by 20 teams. Similarly to “La Liga” it uses a system of promotion and relegation with the “Football League”. The “Premier League” is now the most watched football league in the world, while registering the higher stadium occupancy among all soccer leagues in Europe. A total of 47 clubs have competed in the “Premiership” since 1992. Five teams have won the title, namely Manchester United, Chelsea FC, Arsenal FC, Manchester City and Blackburn Rovers. The Premier League rank 2nd in the UEFA coefficients of European leagues.
- The “Bundesliga” is contested by 18 teams and as most European leagues operates on a system of promotion and relegation with the so-called “2. Bundesliga”. The “Bundesliga” was created in 1962 and is the main soccer competition in Germany. A total of 53 clubs have competed since the league creation. Bayern München has dominated in terms of the number of titles won. The other champions were Borussia Dortmund, Hamburger SV, Werder Bremen, Bor. Mönchengladbach, VfB Stuttgart, 1. FC Kaiserslautern, 1. FC Köln, 1. FC Nürnberg, Eintracht Braunschweig, TSV 1860 München and VfL Wolfsburg. The “Bundesliga” is one of the top national leagues in Europe, currently ranked in the 3rd place, according to UEFA’s league coefficient ranking.
- The Portuguese “Primeira Liga” occupies the 5th place of the UEFA’s league ranking. It was officialised in 1938 and was named “Primeira Divisão” (First Division) until 1999. “Primeira Liga” is the top professional association football division of the Portuguese football league system. A total of 70 teams have competed since the league creation. Among them, the so-called “big three” (SL Benfica, FC Porto and Sporting CP) won every season with the exception of seasons 1945–1946 and 2000–2001, where Os Belenenses and Boavista were crowned champions, respectively. The “Primeira Liga” is now contested by 18 teams. The two lowest placed are relegated to the “Segunda Liga” and replaced by the top-two teams from this division.

3. Mathematical tools

The first tool is the MDS method. MDS is a computational technique for visualizing complex information that explores dissimilarities between the numerical data [23–29]. For N items in space $\mathbb{R}^m$, the MDS requires a dissimilarity index, δ_{ij}, between the pair of elements $(i, j) = \{1, 2, \ldots, N\}$, and the calculation of matrix $\boldsymbol{\Delta} = [\delta_{ij}]$, $N \times N$ dimensional, of item-to-item dissimilarities. In classical MDS, $\delta_{ij} = \delta_{ji}$, $\delta_{ij} \geq 0$ and $\delta_{ii} = 0$. MDS assigns a “point” to each item in $\mathbb{R}^m$ and tries to estimate numerically its coordinates, yielding a configuration in $\mathbb{R}^d$ ($d \leq m$) that approximates the observed dissimilarities. MDS uses an iterative algorithm that evaluates configurations in $\mathbb{R}^d$ for optimizing the goodness-of-fit with the dissimilarities in $\mathbb{R}^m$. To access how well a configuration reproduces $\mathbf{\Delta}$, the raw stress, σ, is often adopted:

$$\sigma = \left[d_{ij} - f(\delta_{ij})\right]^2, \tag{1}$$

where d_{ij} and δ_{ij} stand for the reproduced and observed distances between items i and j, respectively, and $f(\cdot)$ represents some monotonic transformation. The smaller the value of σ, the better is the fit between d_{ij} (for points in $\mathbb{R}^d$) and δ_{ij} (for points in $\mathbb{R}^m$).

Low dimensional spaces (e.g., $d = 2$, or $d = 3$) allow the items (points) to be visualized in a "map". Instead of dissimilarities MDS can also accept similarities, or proximities, between items, and in that case, $\delta_{ij} \leq 1$ and $\delta_{ii} = 1$.

Since we depart from item-to-item measurements, we can rotate, translate, or zoom the MDS map, and the distances between points remain identical.

The quality of the MDS can be assessed by means of the stress and Shepard plots. The stress (or scree) plot represents the stress σ versus the number of dimensions d of the MDS map. The plot has a monotonic decreasing behavior and we choose d as a compromise low σ and high d. The Shepard diagram compares d_{ij} versus δ_{ij} for a particular value of d. Therefore, a narrow scatter around the 45 degree line indicates a good fit between d_{ij} and δ_{ij}.

The second tool is PA [30–34]. PA is a statistical method that takes a collection of shapes and transforms them for maximum superposition. Shape means all the geometrical information that remains after location, scale and rotational effects are removed from an object [32]. A mathematical representation of a p-point shape, $\mathbf{x}$, in m dimensional space is a pm-vector concatenating all dimensions. For planar shapes ($m = 2$) we have $\mathbf{x}_k = [x_{1k}, x_{2k}, \ldots, x_{pk}, y_{1k}, y_{2k}, \ldots, y_{pk}]^T$. PA performs linear transformations, namely translation, reflection, orthogonal rotation and scaling, while minimizing the Procrustes distance, P_d, given by the sum of squared distances between corresponding points in each shape. For shapes $\mathbf{x}_1$ and $\mathbf{x}_2$, we have:

$$P_d = \sqrt{\sum_{i=1}^{p} [(x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2]}. \tag{2}$$

The shape alignment involves the steps:

- computing the centroid of each shape;
- re-scaling each shape for having equal size;
- aligning the shapes both in position and orientation.

4. Data analysis and visualization

We consider N teams competing within a soccer league that comprises $R = 2(N - 1)$ rounds. Along the season, each team plays $R/2$ matches at home and $R/2$ matches away from home. In this section three distinct schemes are tested.

In Section 4.1 we use PA to process R individual MDS maps (one per round) yielding one global chart that represents the long-term performance versus time dynamics of the teams. In Section 4.2 we combine data for the R rounds into one dissimilarity matrix that leads to a single global MDS chart. In Section 4.3 we use the match results for generating one time series per team and round. We compare the time series to obtain a dissimilarity matrix and we generate the corresponding MDS map. The 3 approaches will be denoted by strategies 1, 2, or 3, and the visualizing charts obtained are interpreted based on the locus of points and on the emerging clusters.

4.1. MDS visualization based on Procrustes analysis

For the ith team, $i \in [1, N]$, we define the variables:

- h_i - goals scored at home;
- a_i - goals scored away from home;
- sh_i - goals suffered at home;
- sa_i - goals suffered away from home.

At the end of kth round, $k \leq R$, we construct a $N \times N$ dimensional matrix $\mathbf{M}(k) = [m_{ij}(k) : i, j = 1, \ldots, N,\ i \neq j]$. The elements $m_{ij}(k)$ ($i \neq j$) are given by:

$$m_{ij}(k) = \begin{cases} h_i(\xi) - a_j(\xi), \text{ if team } i \text{ played at home with } j, \text{ at } \xi \leq k \\ \langle h_i(k)\rangle - \langle sh_i(k)\rangle + \left\langle sa_j(k)\right\rangle - \left\langle a_j(k)\right\rangle, \text{ otherwise,} \end{cases} \tag{3}$$

where $\langle \phi_i(k)\rangle$ represents the average value of $\phi_i(k)$, that is:

$$\langle \phi_i(k)\rangle = \frac{1}{k}\sum_{\xi=1}^{k} \phi_i(\xi). \tag{4}$$

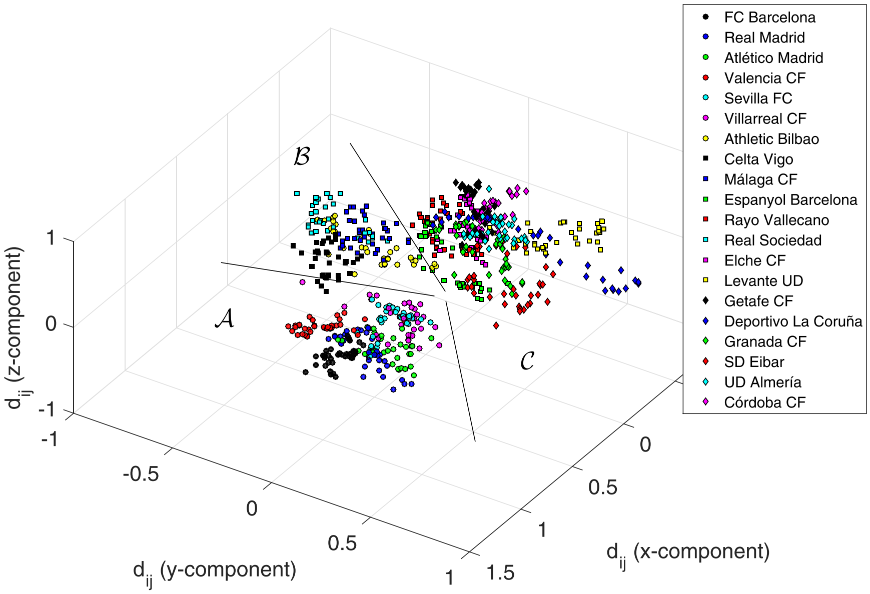

Fig. 1. MDS global map obtained after PA for "La Liga", season 2014–2015, using strategy 1.

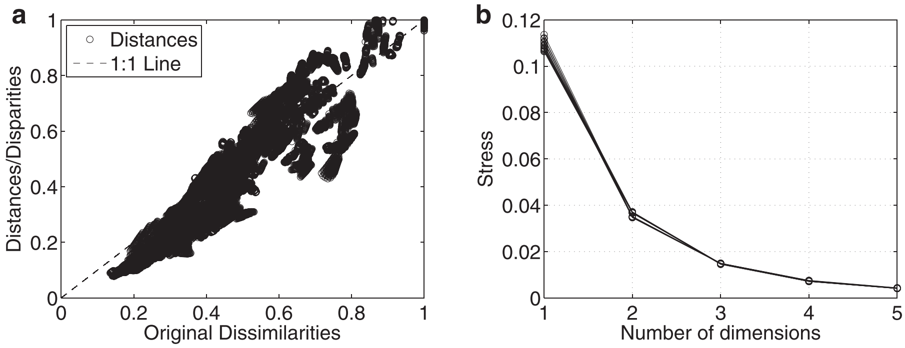

Fig. 2. Standard charts for assessing the MDS results: (a) Shepard plot for the 3D map; (b) stress plot.

Eq. (3) means that at the end of kth round we fill down $kN/2$ cells of $\mathbf{M}(k)$ based on results of matches played until then. The remaining $N \cdot (N - k/2 - 1)$ off-diagonal elements, corresponding to matches not yet played, are updated with "noise" calculated based on past average scores. The elements $m_{ii}(k)$ are left undefined ($m_{ii}(k) = \text{NaN}$).

We then construct the $1 \times 2(N-1)$ dimensional vectors, $\mathbf{v}_i(k)$, as:

$$\mathbf{v}_i(k) = \left[\mathbf{M}_{i,\tau}(k)\ \mathbf{M}^T_{\tau,i}(k)\right],\ \tau = 1, \ldots, N,\ \tau \neq i. \tag{5}$$

Vector $\mathbf{v}_i(k)$ represents the set of features that characterize team i at round k, where $\mathbf{M}_{i,\tau}(k)$ and $\mathbf{M}_{\tau,i}(k)$ correspond to the team performance when playing at home and away from home, respectively.

Vectors $\mathbf{v}_i(k)$ are objects to be compared by means of MDS. Therefore, we calculate the $N \times N$ dimensional matrices $\boldsymbol{\Delta}(k) = [\delta_{ij}(k)]$, where $\delta_{ij}(k)$ represents the arc-cosine correlation between the pair of items i and j ($i, j = 1, \ldots, N$) given by:

$$\delta_{ij}(k) = \arccos\left(\frac{\mathbf{v}_i(k) \cdot \mathbf{v}_j^T(k)}{||\mathbf{v}_i(k)|| \cdot ||\mathbf{v}_j(k)||}\right). \tag{6}$$

For each $k = 1, \ldots, R$, we generate a single MDS map. These individual maps are then combined by means of PA. Therefore, we (i) choose the first map in the set for initial reference, (ii) use PA to superimpose the next MDS map into the current reference, (iii) make the current set of superimposed maps the new reference, (iv) continue to step (ii) until all maps have been conformed.

Fig. 1 depicts the 3D MDS global map obtained for the Spanish national soccer league ("La Liga") and season 2014–2015. In this season the league was competed by $N = 20$ teams along $R = 38$ rounds. The final standings were FC Barcelona, Real Madrid, Atlético Madrid, Valencia CF, Sevilla FC, Villarreal CF, Athletic Bilbao, Celta Vigo, Málaga CF, Espanyol Barcelona,

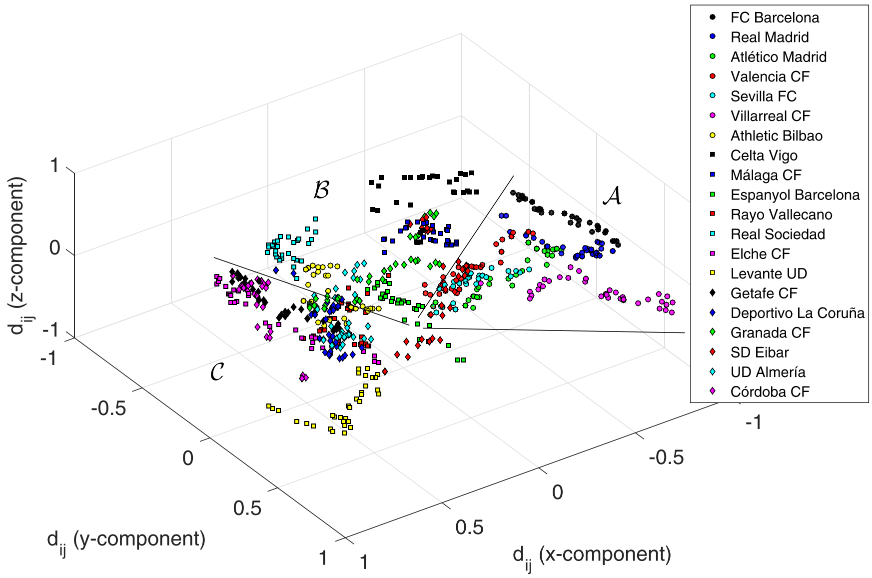

Fig. 3. MDS global map for “La Liga”, season 2014–2015, using strategy 2.

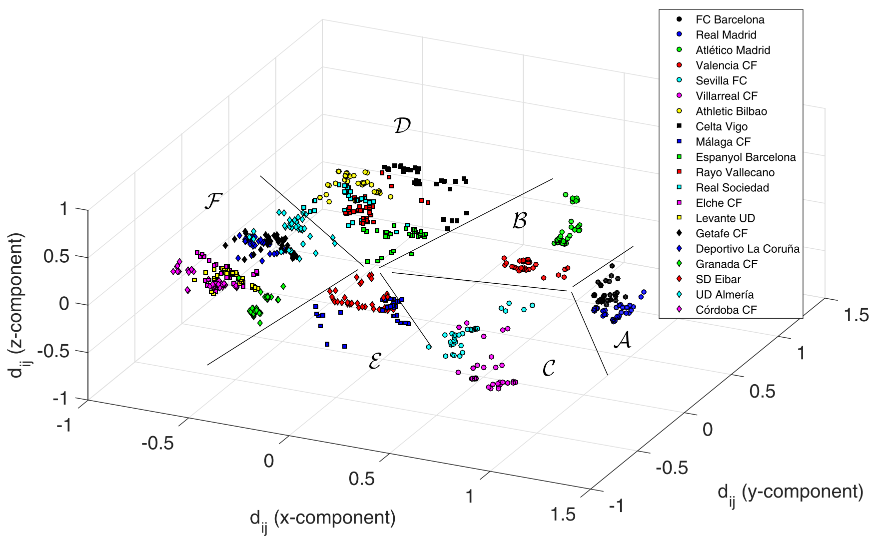

Fig. 4. MDS map for time series data of “La Liga”, season 2014–2015, using strategy 3.

Rayo Vallecano, Real Sociedad, Elche CF, Levante UD, Getafe CF, Deportivo La Coruña, Granada CF, SD Eibar, UD Almería and Córdoba CF. From the chart we observe the emergence of a locus of points for each team. The locus length is related to the regularity of the team along the season. A shorter (longer) locus corresponds to a more regular (irregular) behavior. Moreover, we can see the emergence of 3 main clusters, revealing similarities (or dissimilarities) between teams. Cluster $\mathcal{A} =$ {FC Barcelona, Real Madrid, Atlético Madrid, Valencia CF, Sevilla FC, Villarreal CF} includes the top six teams in season 2014–2015. Cluster $\mathcal{B} =$ {Athletic Bilbao, Celta Vigo, Málaga CF, Real Sociedad} comprises almost up mid-table teams. Cluster $\mathcal{C} =$ {Espanyol Barcelona, Rayo Vallecano, Elche CF, Levante UD, Getafe CF, Deportivo La Coruña, Granada CF, SD Eibar, UD Almería, Córdoba CF} includes all teams standing bellow the 11th place. Therefore, the methodology adopted is able to identify patterns and relationships between the data that are dependent on the teams performance and time dynamics, but with limited ability to discriminate between objects. The results obtained for the “Premiership”, “Bundesliga” and “Primeira Liga” are similar and therefore are not depicted.

Fig. 2 depicts the Shepard and stress plots for the superimposed R individual maps. The Shepard diagram shows a good distribution of points around the 45 degree line, particularly for the 3D representation, which means a good fit of the distances to the dissimilarities. The 2D map is not represented for the sake of parsimony. The stress plot reveals that a three dimensional space describes well the locus of the N objects. The stress diminishes strongly until the dimensionality is two,

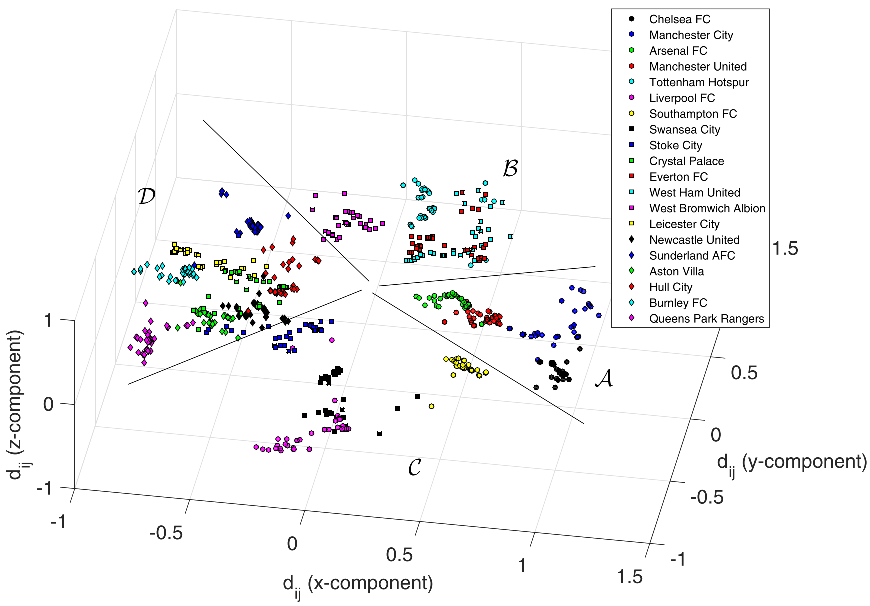

Fig. 5. MDS map for time series data of “Premiership”, season 2014–2015, using strategy 3.

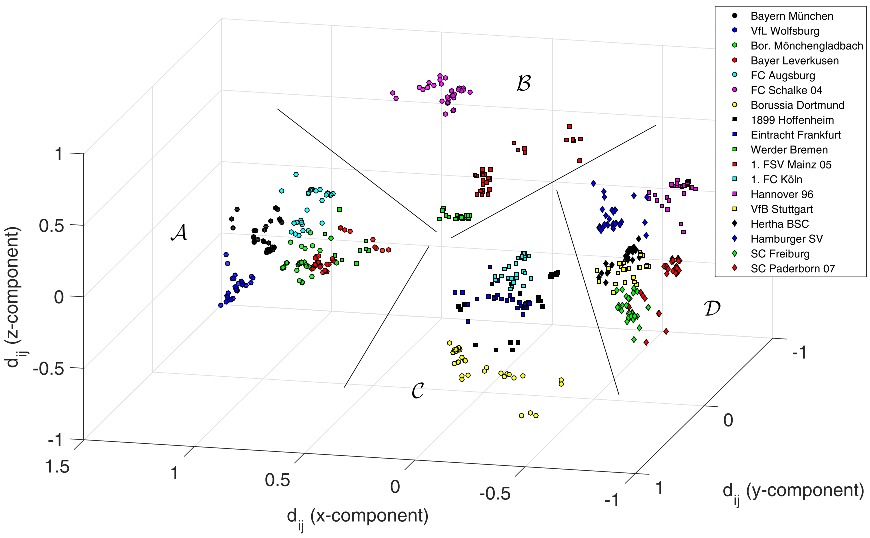

Fig. 6. MDS map for time series data of “Bundesliga”, season 2014–2015, using strategy 3.

moderately towards dimensionality three and weakly onward. This means that, although four or more dimensions would represent the data more accurately, 3D maps are a good compromise between accuracy and easiness of visualization.

4.2. MDS visualization based on direct data analysis

We now compare objects $\mathbf{v}_i(k)$ ($i = 1, \ldots, N$, $k = 1, \ldots, R$) at once. Therefore, we define the enumeration set $\mathcal{Z} = \{\mathbf{z}_p : p = 1, \ldots, NR\} = \{\mathbf{v}_1(1), \ldots, \mathbf{v}_1(R), \ldots, \mathbf{v}_N(1), \ldots, \mathbf{v}_N(R)\}$, and we calculate the $NR \times NR$ dimensional matrix $\mathbf{\Delta} = [\delta_{pq}]$, $p, q = 1, \ldots, NR$, where δ_{pq} is the arc-cosine between vectors $\mathbf{z}_p$ and $\mathbf{z}_q$ (6).

Fig. 3 depicts the 3D MDS global map for “La Liga”, season 2014–2015. The results have similarities with those obtained in Section 4.1.

4.3. MDS visualization based on time series scores

In the sequel, for every team $i = 1, \ldots, N$, we construct R vectors, $\mathbf{v}_i(k)$, $k = 1, \ldots, R$. Each $\mathbf{v}_i(k)$, $1 \times R$ dimensional, represents team i up to the kth round. As such, we have:

$$\mathbf{v}_i(k) = [\epsilon_i(1) \cdots \epsilon_i(k) \ \ \mathbf{0}], \ \ k = 1, \ldots, R, \tag{7}$$

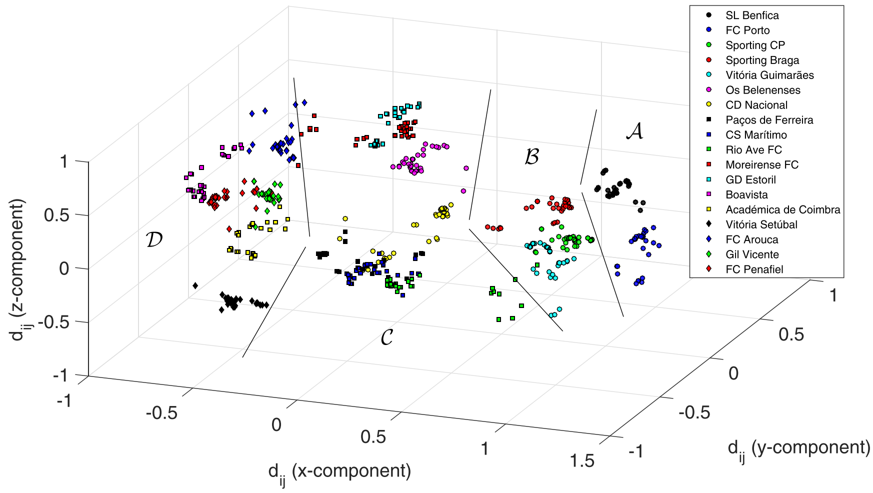

Fig. 7. MDS map for time series data of "Primeira Liga", season 2014–2015, using strategy 3.

$$\epsilon_i(\xi) = |h_i(\xi) + a_i(\xi) - sh_i(\xi) - sa_i(\xi)| \cdot (-1)^u, \ \xi = 1, \ldots, k, \tag{8}$$

where $u = 1$ ($u = 2$) if team i was defeated (won) the match.

Similarly to the strategy 2 formulated in Section 4.2, we now define the enumeration set $\mathcal{Z} = \{\mathbf{z}_p : p = 1, \ldots, NR\} = \{\mathbf{v}_1(1), \ldots, \mathbf{v}_1(R), \ldots, \mathbf{v}_N(1), \ldots, \mathbf{v}_N(R)\}$, and we calculate the $NR \times NR$ dimensional matrix $\mathbf{\Delta} = [\delta_{pq}],\ p, q = 1, \ldots, NR$.

Figs. 4–7 depict the resulting 3D MDS maps for "La Liga", "Premiership", "Bundesliga" and "Primeira Liga", respectively, for season 2014–2015.

We now observe the emergence of new clusters that reveal the superior ability of the time series approach (strategy 3) for unveiling dissimilarities between objects. The results for the four leagues yield the following comments:

- "La Liga" - Six clusters emerge from the MDS map (Fig. 4). Cluster $\mathcal{A}$ = {FC Barcelona, Real Madrid} includes the champion and vice-champion, respectively. These teams are the top Spanish teams. Every season they are the most serious candidates for the title. Clusters $\mathcal{B}$ = {Atlético Madrid, Valencia CF} and $\mathcal{C}$ = {Sevilla FC, Villarreal CF} comprise teams that occupy the places from 3rd to 6th. In some seasons they can contest for the title and, usually, they classify for the European UEFA leagues. Cluster $\mathcal{D}$ = {Athletic Bilbao, Celta Vigo, Espanyol Barcelona, Rayo Vallecano, Real Sociedad} includes mid-table teams. Cluster $\mathcal{E}$ = {Málaga CF, SD Eibar} is composed of two teams that are distant apart in the standings table and have an irregular behavior along the season. Cluster $\mathcal{F}$ = {Elche CF, Levante UD, Getafe CF, Deportivo La Coruña, Granada CF, UD Almería, Córdoba CF} comprises bottom-table teams.
- "Premiership" - In season 2014–2015 the standings table was Chelsea FC, Manchester City, Arsenal FC, Manchester United, Tottenham Hotspur, Liverpool FC, Southampton FC, Swansea City, Stoke City, Crystal Palace, Everton FC, West Ham United, West Bromwich Albion, Leicester City, Newcastle United, Sunderland AFC, Aston Villa, Hull City, Burnley FC, Queens Park Rangers. From Fig. 5 we observe the emergence of four main clusters. Cluster $\mathcal{A}$ = {Chelsea FC, Manchester City, Arsenal FC, Manchester United} is made of the top classified in the season. Clusters $\mathcal{B}$ = {Tottenham Hotspur, Everton FC, West Ham United, West Bromwich Albion} and $\mathcal{C}$ = {Liverpool FC, Southampton FC, Swansea City, Stoke City} include mid-table teams, even though some are European level teams, as Liverpool FC and Southampton FC. Cluster $\mathcal{D}$ = {Crystal Palace, Leicester City, Newcastle United, Sunderland AFC, Aston Villa, Hull City, Burnley FC, Queens Park Rangers} comprises the worst classified in the league.
- "Bundesliga" - The standings table in 2014–2015 was Bayern München, VfL Wolfsburg, Bor. Mönchengladbach, Bayer Leverkusen, FC Augsburg, FC Schalke 04, Borussia Dortmund, 1899 Hoffenheim, Eintracht Frankfurt, Werder Bremen, 1. FSV Mainz 05, 1. FC Köln, Hannover 96, VfB Stuttgart, Hertha BSC, Hamburger SV, SC Freiburg and SC Paderborn 07. Four clusters emerge from the MDS of Fig. 6. Cluster $\mathcal{A}$ = {Bayern München, VfL Wolfsburg, Bor. Mönchengladbach, Bayer Leverkusen, FC Augsburg} comprises the top five classified in the season, including the champion, Bayern München, one of the best world teams. Clusters $\mathcal{B}$ = {FC Schalke 04, Werder Bremen, 1. FSV Mainz 05}, that is, includes the 6th and two mid-table teams. $\mathcal{C}$ = {Borussia Dortmund, 1899 Hoffenheim, Eintracht Frankfurt, 1. FC Köln} comprises mid-table teams. Cluster $\mathcal{D}$ = {Hannover 96, VfB Stuttgart, Hertha BSC, Hamburger SV, SC Freiburg and SC Paderborn 07} comprises the bottom-table teams.
- "Primeira Liga" - In season 2014–2015 the standings table was SL Benfica, FC Porto, Sporting CP, Sporting Braga, Vitória Guimarães, Os Belenenses, CD Nacional, Paços de Ferreira, CS Marítimo, Rio Ave FC, Moreirense FC, GD Estoril, Boavista, Académica de Coimbra, Vitória Setúbal, FC Arouca, Gil Vicente and FC Penafiel. Four clusters emerge from the MDS of Fig. 7. Cluster $\mathcal{A}$ = {SL Benfica, FC Porto} includes the champion and vice-champion, respectively. In the last decade

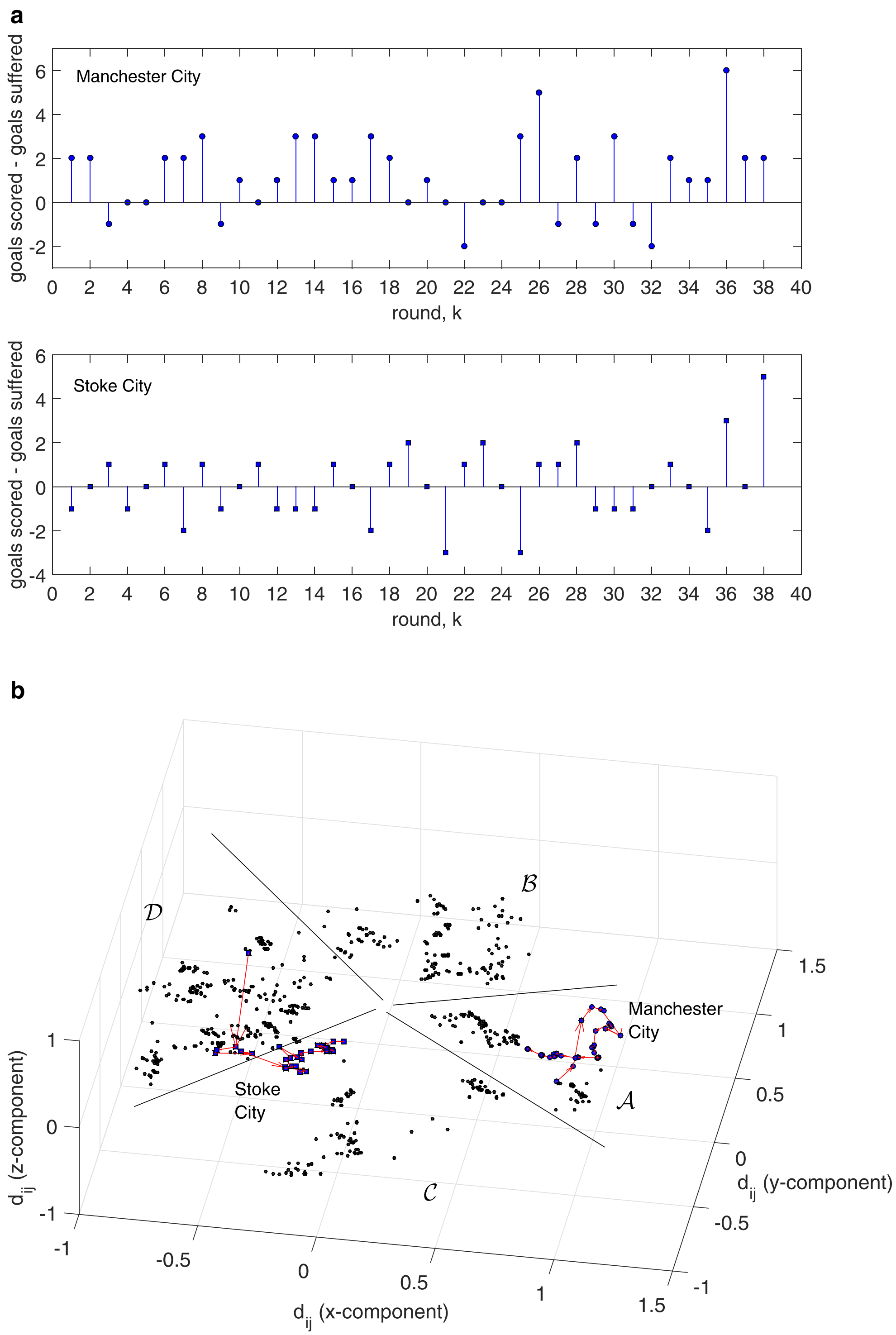

Fig. 8. Dynamics of Manchester City and Stoke City: (a) time series; (b) trajectories within the MDS map for season 2014–2015, using strategy 3.

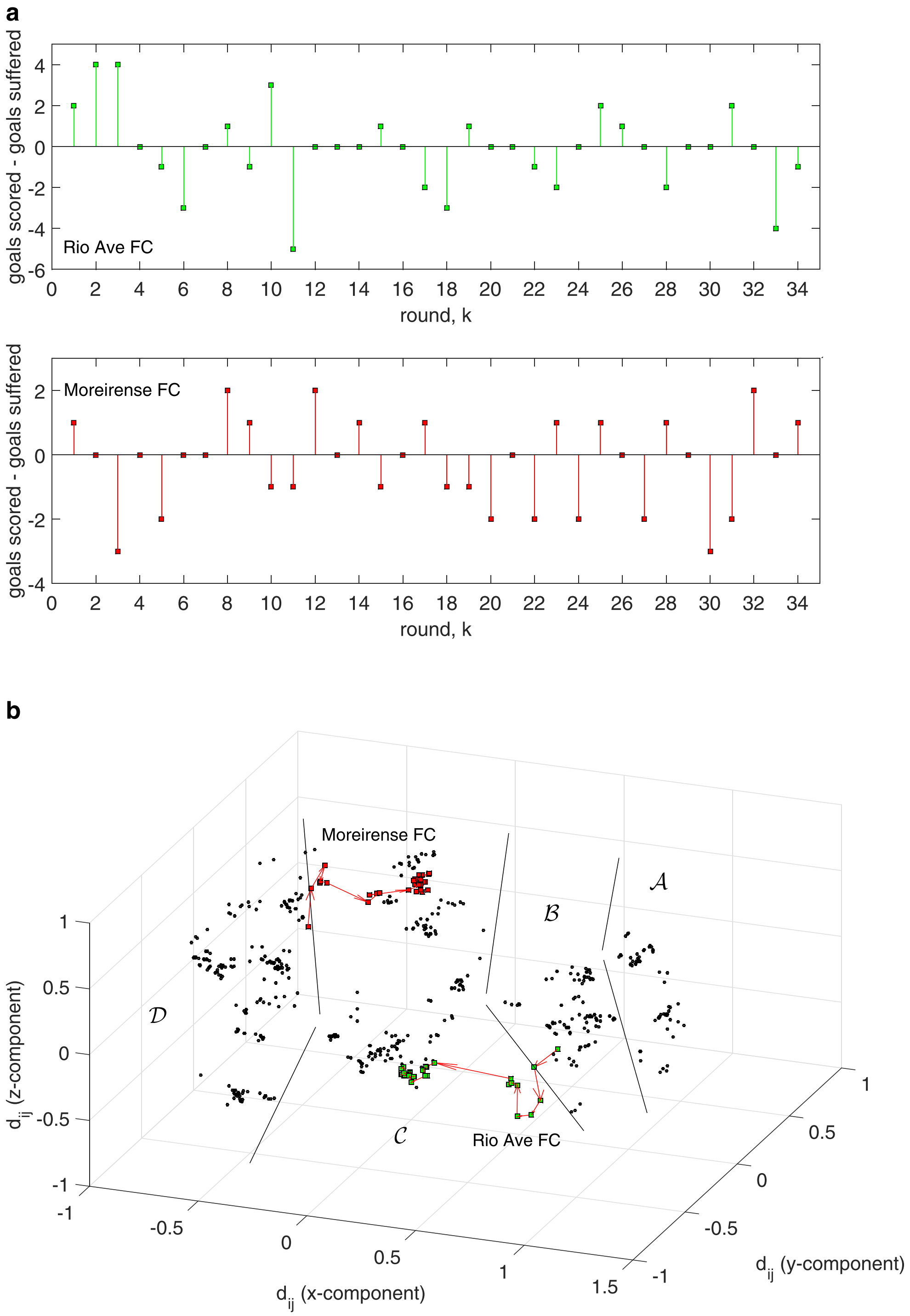

Fig. 9. Dynamics of Rio Ave FC and Moreirense FC: (a) time series; (b) trajectories within the MDS map for season 2014–2015, using strategy 3.

these teams have been on the top of the Portuguese league, being the so-called “chronic” candidates to the title. Cluster $\mathcal{B}=$ {Sporting CP, Sporting Braga, Vitória Guimarães} comprises Sporting CP, the 3rd of the “big three”, as well as the rivals Sporting Braga and Vitória Guimarães. These three usually classify for the UEFA leagues. Cluster $\mathcal{C}=$ {Os Belenenses, CD Nacional, Paços de Ferreira, CS Marítimo, Rio Ave FC, Moreirense FC, GD Estoril} includes mid-table teams. Cluster $\mathcal{D}=$ {Boavista, Académica de Coimbra, Vitória Setúbal, FC Arouca, Gil Vicente and FC Penafiel} comprises bottom-table teams that contest to stay in the “Primeira Liga”, but alternate between the top division and “Segunda Liga” by means of the system of promotion and relegation that rules the professional Portuguese soccer leagues. Moreover, it is interesting to note the locus of some teams, namely CD Nacional, Rio Ave FC and Moreirense FC, that spread in space, revealing irregularity along the season.

With this strategy, the loci in the MDS yield trajectories that represent the teams' time dynamics. Curly trajectories between distant points are associated to irregular behavior during some period of time. Trajectories that connect points close to each other correspond to consistent behavior. Figs. 8 and 9 illustrate the time dynamics of Manchester City and Stoke City "Premiership" teams, and Rio Ave FC and Moreirense FC "Primeira Liga" teams, respectively. The following comments can be drawn:

- Manchester City classified in 2nd in the 2014–2015 edition of "Premiership". As shown in Fig. 8a, representing the time series of the results, the team had a good performance in the first half of the championship, suffering 2 defeats only. The second half was more irregular, namely in the period between journeys 26 and 33, alternating wins and defeats, and ending with 7 defeats in total. Such behavior is shown in the MDS of Fig. 8b, with a trajectory spanning over a considerable region within cluster $\mathcal{A}$. Stoke City classified in 9th place. It started in the worst cluster, $\mathcal{D}$, since until journey 15 it won just 33% of the matches. From journey 16 onward the team recovered from the bottom of the table, increasing the win ratio to 43%. This behavior led to a trajectory in the MDS from cluster $\mathcal{D}$ towards $\mathcal{C}$.
- Rio Ave FC and Moreirense FC classified in 10th and 11th in the 2014–2015 edition of "Primeira Liga", respectively. As shown in the time series of results in Fig. 9a, Rio Ave FC had a good beginning, with 3 consecutive wins, followed by some irregularity until round 10. At round 11 the team suffered a heavy defeat and from then onward alternated wins, defeats and draws, but with a slight larger number of defeats than wins. Such behavior reflects in the MDS of Fig. 9b. In fact, we observe Rio Ave FC starting in the cluster of up mid-table teams, $\mathcal{B}$. After the first rounds, it then moved to cluster $\mathcal{C}$, and at round 10 occurs a large jump. From round 11 onward the trajectory stays almost constant, with a small tendency towards cluster $\mathcal{D}$, that corresponds to the low mid-table teams. Moreirense FC had an opposite trajectory, starting at cluster $\mathcal{D}$, and assuming a consistent place in $\mathcal{C}$ after journey 5.

The analysis adopted is based on real data, namely the results of past matches. It is not a "mathematical model", in the sense that it does not use an integro-differential formulation, but it can be considered a "computational model", since it makes use of algorithmic visualization techniques. The performance of teams depends on distinct factors, such as the physical and psychological condition of the players, the ability of the coach to keep a high level of motivation, the budget to replace long term injured top players, the skills of the coach, and many others. The proposed MDS approach assumes that all relevant factors are implicitly embedded in the results of played matches. Therefore, the modeling assumes a slow evolution of all parameters influential to the outcome, without considering such factors apart from the results of previous matches. The underlying hypotheses limit the usefulness of the proposed modeling technique for forecasting and analysis of long term performance. Nonetheless, the computational modeling based on MDS reveals promising perspectives and future algorithms may extend its utility toward some of the aforementioned directions.

5. Conclusions

We used MDS techniques for visualizing long-term performance of teams competing within a national soccer league. We studied five top European leagues, namely "La Liga", "Premiership", "Bundesliga" and "Primeira Liga", for season 2014–2015. In all cases we got an assertive representation of the team dynamics. With the proposed approach we verify the emergence of trajectories in a abstract MDS space. This methodology avoids the limitations of classical system modeling. In fact, the requirement of models inspired in the laws of physics are now substituted by a computational modeling based on the real data. The resulting description represents a novel paradigm for analyzing system dynamics pointing towards quantitative conclusions in the line of thought of human-based intuitive experience.

References

[1] J.L. Wallace, K.I. Norton, Evolution of world cup soccer final games 1966–2010: Game structure, speed and play patterns, J. Sci. Med. Sport 17 (2) (2014) 223–228.
[2] T. McGarry, D.I. Anderson, S.A. Wallace, M.D. Hughes, I.M. Franks, Sport competition as a dynamical self-organizing system, J. Sports Sci. 20 (10) (2002) 771–781.
[3] L. Vilar, D. Araújo, K. Davids, Y. Bar-Yam, Science of winning soccer: emergent pattern-forming dynamics in association football, J. Syst. Sci. Complex. 26 (1) (2013) 73–84.
[4] P. Passos, K. Davids, D. Araujo, N. Paz, J. Minguéns, J. Mendes, Networks as a novel tool for studying team ball sports as complex social systems, J. Sci. Med. Sport 14 (2) (2011) 170–176.
[5] Y. Zengyuan, H. Broich, F. Seifriz, J. Mester, Kinetic energy analysis for soccer players and soccer matches, Prog. Appl. Math. 1 (1) (2011) 98–105.
[6] M.S. Couceiro, F.M. Clemente, F.M. Martins, J.A.T. Machado, Dynamical stability and predictability of football players: the study of one match, Entropy 16 (2) (2014) 645–674.
[7] L. Malacarne, R. Mendes, Regularities in football goal distributions, Phys. A Stat. Mech. Appl. 286 (1) (2000) 391–395.
[8] C. Lago, R. Martín, Determinants of possession of the ball in soccer, J. Sports Sci. 25 (9) (2007) 969–974.
[9] J.-F. Grehaigne, D. Bouthier, B. David, Dynamic-system analysis of opponent relationships in collective actions in soccer, J. Sports Sci. 15 (2) (1997) 137–149.
[10] O.F. Camerino, J. Chaverri, M.T. Anguera, G.K. Jonsson, Dynamics of the game in soccer: detection of T-patterns, Eur. J. Sport Sci. 12 (3) (2012) 216–224.
[11] S.H.M.G. Mack, K.E. Dutler, J.K. Mintah, Chaos theory: A new science for sport behavior? Athl. Insight 2 (2) (2000) 8–16.
[12] B. Travassos, D. Araújo, V. Correia, P. Esteves, Eco-dynamics approach to the study of team sports performance, Open Sports Sci. J. 3 (2010) 56–57.

[13] Y.H. Lee, R. Fort, Competitive balance: Time series lessons from the English premier league, Scottish J. Polit. Econ. 59 (3) (2012) 266–282.
[14] A. Jessop, A measure of competitiveness in leagues: a network approach, J. Oper. Res. Soc. 57 (12) (2006) 1425–1434.
[15] E. Ben-Naim, F. Vazquez, S. Redner, Parity and predictability of competitions, J. Quant. Anal. Sports 2(4), ISSN (Online) 1559-0410, doi:https://doi.org/10.2202/1559-0410.1034.
[16] C.P. Barros, S. Leach, Performance evaluation of the English premier football league with data envelopment analysis, Appl. Econ. 38 (12) (2006) 1449–1458.
[17] R. Fort, J. Maxcy, Competitive balance in sports leagues: an introduction, J. Sports Econ. 4 (2) (2003) 154–160.
[18] P.D. Owen, M. Ryan, C.R. Weatherston, Measuring competitive balance in professional team sports using the Herfindahl–Hirschman index, Rev. Ind. Organ. 31 (4) (2007) 289–302.
[19] R. Criado, E. García, F. Pedroche, M. Romance, A new method for comparing rankings through complex networks: model and analysis of competitiveness of major European soccer leagues, Chaos Interdiscip. J. Nonlinear Sci. 23 (4) (2013) 043114.
[20] J. Park, M.E. Newman, A network-based ranking system for US college football, J. Stat. Mech. Theory Exp. 2005 (10) (2005) P10014.
[21] T. Callaghan, P.J. Mucha, M.A. Porter, Random walker ranking for NCAA division IA football, Am. Math. Monthly 114 (9) (2007) 761–777.
[22] E. Dunning, Sport Matters: Sociological Studies of Sport, Violence and Civilisation, Routledge, London and New York, 1999.
[23] R.N. Shepard, The analysis of proximities: Multidimensional scaling with an unknown distance function. I, Psychometrika 27 (2) (1962) 125–140.
[24] J.B. Kruskal, M. Wish, Multidimensional Scaling, Sage, Beverly Hills, 1978.
[25] I. Borg, P.J. Groenen, Modern Multidimensional Scaling: Theory and Applications, Springer Science & Business Media, 2005.
[26] A.M. Lopes, J.T. Machado, C. Pinto, A. Galhano, Multidimensional scaling visualization of earthquake phenomena, J. Seismol. 18 (1) (2014) 163–179.
[27] A.M. Lopes, J.T. Machado, C.M. Pinto, A.M. Galhano, Fractional dynamics and MDS visualization of earthquake phenomena, Comput. Math. Appl. 66 (5) (2013) 647–658.
[28] A.M. Lopes, J.T. Machado, Analysis of temperature time-series: Embedding dynamics into the MDS method, Commun. Nonlinear Sci. Numer. Simul. 19 (4) (2014) 851–871.
[29] J.A.T. Machado, A.M. Lopes, Analysis and visualization of seismic data using mutual information, Entropy 15 (9) (2013) 3892–3909.
[30] F.L. Bookstein, Landmark methods for forms without landmarks: morphometrics of group differences in outline shape, Med. Image Anal. 1 (3) (1997) 225–243.
[31] J.C. Gower, G.B. Dijksterhuis, Procrustes Problems, vol. 3, Oxford University Press, Oxford, 2004.
[32] M.B. Stegmann, D.D. Gomez, A brief introduction to statistical shape analysis, Inf. Math. Modell. Tech. Univ. Den. 15 (2002) 11.
[33] A.M. Lopes, J. Tenreiro Machado, A.M. Galhano, Multidimensional scaling visualization using parametric entropy, Int. J. Bifurc. Chaos 25 (14) (2015) 1540017.
[34] J. Tenreiro Machado, A.M. Lopes, A.M. Galhano, Multidimensional scaling visualization using parametric similarity indices, Entropy 17 (4) (2015) 1775–1794.