Case Study

Institution of MECHANICAL ENGINEERS

Journal of SPORTS ENGINEERING AND TECHNOLOGY

Proc IMechE Part P:
J Sports Engineering and Technology
2016, Vol. 230(2) 124–132

DOI: 10.1177/1754337115583198
pip.sagepub.com
SAGE

Developing a tactical metric to estimate the defensive area of soccer teams: The defensive play area

Filipe M Clemente[1,2], Fernando ML Martins[1,3], Micael S Couceiro[4,5], Rui S Mendes[1] and António J Figueiredo[2,6]

Abstract
This study proposes a computational method to inspect the tactical position of players during the match and a new metric to analyse the defensive pressure made by a soccer team. These metrics only require Cartesian information about the players' positions on the field. As a case study, three matches played by the same professional soccer team were considered, including variables computed for the half of the match (first half vs second half) and the final score of the game for an analysis of variance of tactical performance, trying to identify the influence of such variables on the collective organisation. The data were collected at 1 Hz and from this process, 9218 instances of useful time were collected. The results revealed that the different kinds of final scores had significant effects on the tactical performance. The comparison between two halves of the match revealed significant differences with a small effect size on tactical performance. In summary, this study showed that these new tactical metrics can be a computational option to increase a coaches' knowledge about the defensive organisation of soccer teams, giving them the possibility to augment their own perception with metrics that can provide specific information.

Keywords
Match analysis, tactical metrics, defensive play area, tactics, soccer

Date received: 5 February 2015; accepted: 28 March 2015

Introduction

In the last few years, match analysts have introduced a new approach to the analysis of collective behaviour. Beyond traditional notational analysis and manual observational procedures,[1] it is now possible to use technological advances to inspect the performance of soccer players.[2] Such technological advances can be called metrics (parameters or measures of quantitative assessment used for measurement, comparison or to track performance)[3] that use the spatio-temporal relationship among players throughout the match to estimate collective synchronisation.[4] Such metrics benefit from the Cartesian data about a soccer player's position during the match to compute a set of metrics that characterised the collective organisation of a given team.[5]

One of the first metrics was proposed by Schollhorn[6] which focuses on the concept of the centroid in a soccer game. This metric is a geometric mean reference of different points (players) in the field and acts as the centre of gravity. Later, Frencken et al.[7] used the same concept to apply new metrics for small-sided soccer games and Bourbousson et al.[4] for basketball games. An adaptation of the centroid was also proposed by Clemente et al.[8] for soccer games that additionally considers the ball position to attribute weight to the players' positions,

[1]Department of Education, Coimbra College of Education (ESEC), Polytechnic Institute of Coimbra (Instituto Politécnico de Coimbra), Coimbra, Portugal
[2]Faculty of Sport Sciences and Physical Education, University of Coimbra, Coimbra, Portugal
[3]Instituto de Telecomunicações – Delegação da Covilhã, Coimbra, Portugal
[4]RoboCorp, Department of Electrical Engineering (DEE), Engineering Institute of Coimbra (ISEC), Polytechnic Institute of Coimbra (Instituto Politécnico de Coimbra), Coimbra, Portugal
[5]Ingeniarius, Lda., Mealhada, Portugal
[6]CIDAF, Faculty of Sport Sciences and Physical Education, University of Coimbra, Coimbra, Portugal

Corresponding author:
Filipe M Clemente, Department of Education, Coimbra College of Education (ESEC), Polytechnic Institute of Coimbra (Instituto Politécnico de Coimbra), Rua Dom João III Solum, 3030-329 Coimbra, Portugal.
Email: Filipe.clemente5@gmail.com

denoting such metric as ‘weighted centroid’. This metric provides a different weight to each player that contributes for the final centroid position. Thus players with closer proximity to the ball have greater weights than players who are far away from the ball. This metric allowed integrating the goalkeeper and reducing the impact of outliers’ positions.[8]

Complementarily to the centroid, Bourbousson et al.[4] applied the concept of a stretch index that measures the dispersion of players from the team’s centroid. A similar method was proposed by Clemente et al.[8] that benefits from the same weighted methodology adopted for the ‘weighted centroid’. Both metrics use the position of players to inspect the centre-of-team and the players’ dispersion.

Other metrics were developed to estimate the area covered by a team. The first methods were the coverage area[9] and the surface area,[7] both developed for soccer games. These metrics allow estimating the area covered by a team, using the information about the minimum number of triangulations performed by all players. The polygon formed by all players within one team is used to determine the coverage area. Another concept was proposed by Clemente et al.[8] called the ‘effective area of play’. This metric determines the defensive and offensive triangulations and benefits from information about both teams to estimate the triangulations with higher or lower effectiveness. This metric is also different from another metric called the ‘effective play space’ that only determines the polygon of both teams and does not consider the effective triangulations of players in defensive and attacking moments.[10]

In summary, all of these metrics allow to computationally characterise the collective organisation of teams. Nevertheless, such metrics are not specific enough to a given sport modality such as soccer.

In a soccer game, it is very important to determine the strategic positions of all players, but this strategic definition is not static. The tactical position (TP) of a player changes throughout the match. One of these cases of position changes is the lateral defender (the defender that acts in the wing side of the field), which contributes to the attacking process. Therefore, a lateral midfielder, or a lateral forward, assumes a more advanced position in the field when compared to the lateral defender during attacking phases. Thus, their positions vary based on the moments with or without ball possession. Nevertheless, no method had been developed to determine the position of each player throughout a match so far. Another important issue in a soccer game is to determine the specific areas of pressing during the defensive process. In fact, during defensive moments, the majority of players act in synchronisation within a collective plan/strategy that aims to reduce the possibilities of success for the opposing team. Nevertheless, such collective organisation changes due to the contested regions of the field. This happens most often close to the goal and in the central area of the field during pressure provided by the opposed team. Despite the usefulness of quantifying such phenomena, no metric has been proposed to measure the different kinds of pressure within the team.

Bearing these ideas in mind, this study proposes a method and one metric to inspect both these issues. The method will allow determining the TP throughout the match and the metric, herein denoted as defensive area of play, will be used to analyse the defensive pressure made by soccer teams. This research aims to provide new computational methods to apply in match analysis on soccer. As a proof of concept, three matches will be considered. Regarding the defensive play area (DPA) of play, only the moments without ball possession will be considered. The variance between the first and second halves of the matches, as well as with the final scores, will be inspected to understand whether different halves and scores induce changes in the tactical performance of soccer teams.

Methods

Sample

Three official home matches of the same professional soccer team from the *Portuguese Professional Premier League* were used as a dataset. There were three different final scores in these three matches: win, lose and draw. All matches were analysed for matches at home during the 2012/2013 season. The initial home-team distribution was 1-4-3-3 for the three matches. From these three matches, all useful times (the time that the ball is not stopped or out of the field) were recorded and then processed in a computational system that recorded the players’ Cartesian positional data in the field. The data were collected at 1 Hz and from this process, 9218 moments of useful data were collected. The entire process complies with the American Psychological Association (APA) ethical standards for the treatment of human or animal subjects.

Data collecting

The data from the three official matches were obtained using a digital camera (*GoPro Hero* with 1280×960 resolution) with the capacity to process images at 30 Hz. To capture the whole field, the camera was placed on an elevated surface above the ground. The dimensions of the soccer field were 104×68 m.

Before processing the images, the soccer field was calibrated using 19 markers throughout the official lines of field. Such markers allowed identifying the points of calibration in the virtual space, thus matching the virtual space (pixels) with real physical space (metres) by benefiting from the direct linear transformation (DLT) method.[11] The reliability of such a conversion was assessed through experimental trials with ground truth information.

The soccer players and the ball were tracked using a manual tracking method. Such a procedure was done after calibration, thus returning the position of players and the ball over time in Cartesian coordinates. A graphical user interface was developed to visualise the match by controlling the frame rate with a sampling rate of 1 Hz. In each frame, the locations of all players and the ball were recorded, following a typical point-and-click approach. Such identification corresponded to one point at the centre of the feet of each player and the centre of the ball, respectively. The whole process associated with this approach (i.e. detection and identification of player trajectories, space transformation and metric computation) was performed using the high-level calculation tool MATLAB.[12]

Considering the proposed metrics, only the useful playing time of each match was considered, excluding all moments in which the ball was not in the field (e.g. ball out-of-bounds). Because the proposed methodology is computationally complex, a downsampling was considered to analyse the data at 1 Hz.

Developing the DPA

The DPA proposed in this project emerges from the observational analysis performed by Seabra,[13] which introduced the concept of effective space covered. In Seabra's concept, the effective space covered is determined by the successive position of players located within the perimeter formed by the group of players of both teams at a given moment *T*, generating a polygonal surface figure (excluding the goalkeeper). The effective space covered enables locating actions in the periphery – beside, behind and opposite – or in their interior. Based on that method, a new computational metric was developed which additionally includes the spatio-temporal information of players and computes the TP of each player throughout the match. In fact, the strategic mission of a given player is relatively permanent during the match (e.g. for a lateral defender). Yet, throughout the match, the lateral defender moves forward during the attacking phase. At this moment, their position is different than the one they have during the defensive phase. Bearing this idea in mind, a metric that computes the position of each player over time is then proposed.

Computing the momentary TP. To develop the DPA, the first task is to compute the surface area of players constituted by the sum of triangulations between teammates. This effective area of play can be computed as proposed by Clemente et al.[8] and as illustrated in Figure 1. The polygon of the overall team, denoted as the effective area of play, is computed based on the sum of triangulations among teammates. Each triangulation is generated based on the minimum distance between players, in such a way that the surface area represents the minimum number of triangulations generated with the overall polygon of the team. In brief, the algorithm starts by generating every possible triangle formed by the 11 players of a team, resulting in a total of 165 triangles. After generating all triangles of each team, the next step of the algorithm is to consider non-overlapping triangles within the team. This step is immediately followed by computing all triangles of each team without relevant interception. In the presence of interceptions between opposing triangles, and based on the supposition that effective defensive triangles (i.e. triangles with perimeter under 36 m) can overlap the offensive triangles,[14] the effective area to be considered is the one of the defensive triangles, thus reducing the effective area of the offensive team.

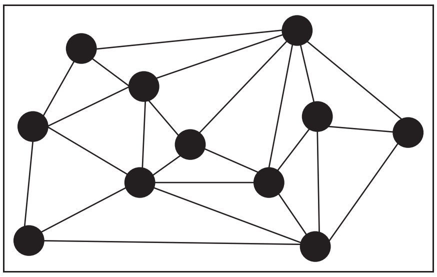

Figure 1. Surface area of one team in a given moment.

After computing the surface area, the second step is to define the criteria to classify the TP of soccer players. Therefore, one can define a positional relationship, at each instant, between the most backward player (typically the goalkeeper) and the most advanced player. The most adequate way to represent this relationship is by considering the distance between two players from a given team δ, herein denoted as d_δ^{max}. The variable d_δ^{max} can also be described as the longitudinal range of the team from end to end, being computed as

$$d_\delta^{max}[t] = \max_{N_\delta} \left\| X_\delta^1[t] \right\| \tag{1}$$

where $X_\delta[t]$ is the positioning matrix in which row n represents the planar position $\mathbb{R}^2$ of player n of team δ at time t as

$$X_\delta[t] = \begin{bmatrix} x_1[t] & y_1[t] \\ \vdots & \vdots \\ x_{N_\delta}[t] & y_{N_\delta}[t] \end{bmatrix}, \quad x_n[t], y_1[t] \in \mathbb{R}^1 \tag{2}$$

Note that equation (1) only considers the first dimension, that is, longitudinal x-axis, of players position, identifiable as $X_\delta^1[t]$. Such amplitude is then used to normalise the longitudinal (x-axis) coordinates of each player with respect to the most backward player (the goalkeeper), thus resulting in a normalised positioning x-vector

$$\hat{x}_\delta[t] = \begin{bmatrix} \frac{x_1[t]}{d_\delta^{max}[t]} \\ \vdots \\ \frac{x_{N_\delta}[t]}{d_\delta^{max}[t]} \end{bmatrix}, \quad \hat{x}_\delta[t] \in [0, 1] \tag{3}$$

Table 1. Possible combinations between the tactical positions on a field and the corresponding defensive area.

	G	D	D+D	D+M	D+F	M	M+M	M+F	F	F+F
G			1	2	2		2	3		3
D			2	2	3		3	3		3
D+D						2			3	
D+M						3			3	
D+F						3			3	
M							3	3		4
M+M									3	
M+F									4	
F										4
F+F										

1	Defensive backward region (DBR)	4	Defensive forward Region (DFR)
2	Defensive first half of the middle Region (D1HMR)	3	Defensive second half of the middle Region (D2HMR)
	Not possible to compute		The same process as in the example

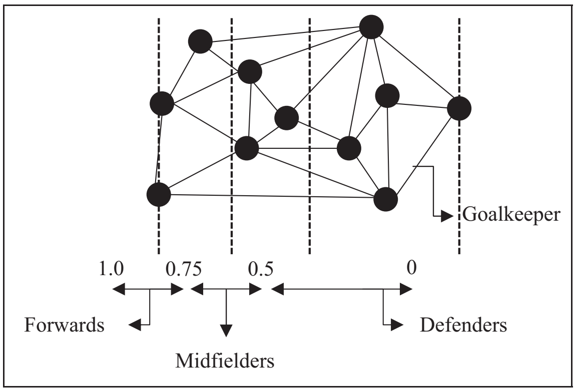

Figure 2. Momentary tactical position in a given moment.

After computing the normalised longitudinal position of players, a simple set of thresholds was used to classify them as goalkeeper, defenders, midfielders and forwards, as depicted in Figure 2.

The cut-off values proposed in this work were tested using 5% of overall data. The procedure was performed after observation of two expert soccer coaches (with more than 6 years of experience and 3 years working as soccer analysts). The analysis procedure was performed after the observation of a sequence of plays where the coaches identified the TP of each player in each frame. Based on coaches' identification and the algorithm cut-off, a test of comparison was carried out. The variance between the coaches' description and the computational cut-offs method was 7%. The method of cut-offs is independent from the players' locations. Thus, as example in the case that all players are within the penalty-box, the split is performed. The same case happens for situations where all players are behind the line of ball. Such decision-making was justified by the need to classify the attacking players and retreated players, even in small spaces.

From this method, it was possible to analyse the TP at each second (1 Hz), thus representing the number of defenders, midfielders and forwards. For statistical analysis purposes, each variable was codified as follows: (1) TP_Defensive, (2) TP_Middle and (3) TP_Attacking.

Computing the relative DPA. After computing the momentary TP of each player, it is possible to develop the concept of the relative DPA. The region between each position (goalkeeper, defender, midfielder and forward) can be classified during the defensive phase (in this case, during the moments without ball possession). From the different positions inherent to triangulations, it is possible to define a set of defensive regions, defined as follows: (1) defensive backward region – DBR (space between the defensive players and the goalkeeper), (2) defensive first half of the middle region – D1HMR (region between the defender and the midfielder), (3) defensive second half of the middle region – D2HMR (region between two midfielders and one attacking player) and (4) defensive attacking region – DFR (region between attacking players and one midfielder). Such classifications are generic, but they were defined from the multiple combinations that may occur throughout a match, as well as all the possibilities and the respective relative defensive area, as can be seen in Table 1.

Under these conditions (depicted in Table 1), it is possible to compute the relative area of each DPA and to compute the number of triangulations within a given DPA. Such information can be computed throughout all defensive moments, at a sampling rate of 1 Hz. Figure 3 illustrates a one-frame example in which the DPA can be observed.

The DPA only considers players close to each region and do not identify players performing real coverage. In fact, such qualitative analysis can only be identified by an observational method. This computational metric

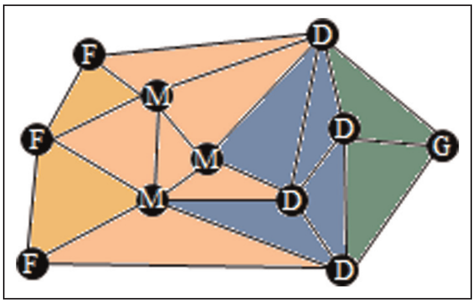

Figure 3. Example of one-frame and the respective relative defensive play area.

only identifies the potential covering regions as missions for each player. During the match, each player has specific instructions to cover a given region. Although sometimes the player is too far away to recover the ball, he can be the closest from such regions, thus acting as the next obstacle to the opponent.

From the example depicted in Figure 3, it is possible to observe two triangulations in DBR, four triangulations in D1HMR, six triangulations in D2HMR and two triangulations in DFR. The area of each triangle is estimated and computed based on the distance between the players that constitute such a triangle. Once again, such variables (the number of triangulations at each DPA and the DPA) are computed at 1 Hz, varying based on the conditions previously defined. From this output, it is possible to obtain the following variables inherent to the DPA: (1) DPA_DBR (partial area in m^2), (2) DPA_D1HMR (partial area in m^2), (3) DPA_D2HMR (partial area in m^2), (4) DPA_DFR (partial area in m^2), (5) Triang_DBR (number of triangulations in that region), (6) Triang_D1HMR (number of triangulations in that region), (7) Triang_D2HMR (number of triangulations in that region) and (8) Triang_DFR (number of triangulations in that region).

Statistical procedures

Descriptive statistics (mean, standard deviation and coefficient of variation) were considered to inspect the results from the positions, defensive area of play and the number of triangulations in a given region of a defensive area of play. The three case study matches were organised into two halves of each game, thus resulting in six variables: M1H1 (match 1 and half 1), M1H2, M2H1, M2H2, M3H1 and M3H2.

The respective influences of 'final score' and 'half of the match' factors on the dependent variables of 'TP_Defensive', 'TP_Middle', 'TP_Attacking', 'DPA_DBR', 'DPA_D1stHMR', 'DPA_D2ndHMR', 'DPA_DFR', 'Triang_DBR', 'Triang_D1stHMR', 'Triang_D2ndHMR' and 'Triang_DFR' were analysed using two-way multivariate analysis of variance (MANOVA) after validation of normality and homogeneity assumptions. This statistical test was based on the two factors being studied, as well as by the multiple dependent variables. The MANOVA was used afterwards to analyse the requirements about the normality and the homogeneity of the sample.

The assumption of normality for each of the univariate dependent variables was examined using the univariate tests of Kolmogorov–Smirnov (*p value* < 0.05). Although the univariate normality of each dependent variable was not verified, the large sample of $n \geqslant 30$ allows the use of MANOVA by considering the central limit theorem (CLT).[15] Consequently, the assumption of multivariate normality was validated.[16] The assumption of the homogeneity of variance/covariance matrix in each group was examined with the Box's M test. Such tests showed no regular homogeneity. The samples per match were $n = 3121$ in the first match, $n = 3300$ in the second match and $n = 2797$ in the third match. Despite the non-homogeneity and unequal samples, the MANOVA is robust to this situation due to the Pillai's Trace test, even for unequal samples.[15]

When MANOVA detected significant statistical differences between the two factors, we proceeded to the commonly used two-way analysis of variance (ANOVA) for each dependent variable[5] followed by Tukey's honestly significant difference (HSD) post-hoc test. When the factor interactions were verified using the two-way MANOVA, a new variable was developed by crossing the two independent variables (half of the match and ball possession status) for each dependent variable. The one-way ANOVA was then performed for each new factor (crossing both initial independent variables).[15] Next, Tukey's HSD Post-Hoc test was performed to analyse the multiple comparisons.

The classification of the effect size (i.e. the measure of the proportion of total variation in the dependent variable explained by the independent variable) and the power of the test were made according to Hopkins et al.[17] In this study, the independent variables were the final score and the half of the match. All the statistical analyses were performed using *IBM SPSS Statistics* (version 21) for a significance level of 5%.

Results

Two-way MANOVA results revealed that the final score had significant main effects and a small effect size (*Pillai's Trace* = 0.157; $F_{(20,\ 8924)} = 37.910$; *p value* = 0.001; $\eta_p^2 = 0.078$; *power* = 1.000) on the composite of TP, DPA and triangulations. The match half had a significant main effect and a small effect size (*Pillai's Trace* = 0.040; $F_{(10,\ 4461)} = 18.383$; *p value* = 0.001; $\eta_p^2 = 0.040$; *power* = 1.000) on the composite of TP, DPA and triangulations. Finally, significant interaction effects between the two factors of the variables' composite were observed (*Pillai's Trace* = 0.083; $F_{(20,\ 8924)} = 19{,}429$; *p value* = 0.001; $\eta_p^2 = 0.042$; *power* = 1.000; small effect size).

After observing the significance of the final score and the match half in MANOVA, a univariate ANOVA analysis and relevant post-hoc comparisons

Table 2. Comparison of tactical metrics between final scores.

	Loss			Draw			Win		
	Mean	SD	%CV	Mean	SD	%CV	Mean	SD	%CV
TP_Defenders	3.05[a,b]	1.97	64.59	3.51[c]	1.59	45.30	3.44[c]	1.93	56.10
TP_Midfielders	3.39[a,b]	1.42	41.89	2.94[b,c]	1.17	39.80	3.27[a,c]	1.35	41.28
TP_Forwards	3.56[b]	1.43	40.17	3.55[b]	1.42	40.00	3.30[a,c]	1.43	43.33
DPA_DBR	1744.17[a,b]	1331.82	76.36	2590.69[b,c]	1499.05	57.86	2038.50[a,c]	1410.14	69.18
DPA_D1stHMR	2668.71[a]	1478.38	55.40	2891.67[b,c]	1668.17	57.69	2523.63[a]	1336.69	52.97
DPA_D2ndHMR	3000.96[a]	1582.18	52.72	3787.28[b,c]	2236.88	59.06	2930.95[a]	1576.47	53.79
DPA_DFR	1333.18[a,b]	1238.43	92.89	1635.94[b,c]	1649.42	100.82	1084.61[a,c]	1127.18	103.92
Triang_DBR	1.42[a,b]	1.05	73.94	1.81[b,c]	0.98	54.14	1.64[a,c]	1.06	64.63
Triang_D1stHMR	3.81[b]	1.95	51.18	3.77[b]	1.77	46.95	4.12[a,c]	1.96	47.57
Triang_D2ndHMR	5.49[a]	1.84	33.52	5.22[b,c]	1.70	32.57	5.58[a]	1.78	31.90
Triang_DFR	3.39[b]	2.29	67.55	3.24[b]	2.26	69.75	2.96[a,c]	2.22	75.00

SD: standard deviation; CV: coefficient of variation.
Values are the average of the first and second halves.
[a]Significantly different compared to draw at *p value* < 0.05.
[b]Significantly different compared to win at *p value* < 0.05.
[c]Significantly different compared to loss at *p value* < 0.05.

Table 3. Comparison of tactical metrics between first and second halves.

	First half			Second half		
	Mean	SD	%CV	Mean	SD	%CV
TP_Defenders	3.33	1.94	58.26	3.29	1.79	54.41
TP_Midfielders	3.22	1.34	41.61	3.23	1.34	41.49
TP_Forwards	3.46	1.40	40.46	3.48	1.47	42.24
DPA_DBR	2213.45[a]	1491.21	67.37	1946.19[b]	1389.38	71.39
DPA_D1stHMR	2816.53[a]	1492.61	52.99	2548.80[b]	1486.65	58.33
DPA_D2ndHMR	3334.99[a]	1720.57	51.59	3058.56[b]	1910.08	62.45
DPA_DFR	1399.72[a]	1311.02	93.66	1267.80[b]	1382.58	109.05
Triang_DBR	1.62	1.06	65.43	1.59	1.03	64.78
Triang_D1stHMR	3.89	1.96	50.39	3.92	1.87	47.70
Triang_D2ndHMR	5.38[a]	1.76	32.71	5.51[b]	1.82	33.03
Triang_DFR	3.15[a]	2.20	69.84	3.26[b]	2.33	71.47

SD: standard deviation; CV: coefficient of variation.
Values are the average of the three final scores.
[a]Significantly different compared to second half at *p value* < 0.05.
[b]Significantly different compared to first half at *p value* < 0.05.

were performed for each dependent variable. The descriptive statistics and the corresponding post-hoc results in the final score variable can be found in Table 2. A similar procedure can be seen in Table 3 for the match half variable.

With respect to the final score, it was established as an alternative hypothesis that there are statistical significant differences on the dependent variables between the three possible results (win, lose and draw). Significant differences were found among the three final scores for TP_defensive ($F_{(2,\ 4470)} = 28.199$; p $value = 0.001$; $\eta^2 = 0.012$; $power = 1.000$; small effect size), TP_middle ($F_{(2,\ 4470)} = 39.998$; p $value = 0.001$; $\eta^2 = 0.018$; $power = 1.000$; small effect size), TP_attacking ($F_{(2,\ 4470)} = 16.310$; p $value = 0.001$; $\eta^2 = 0.007$; $power = 1.000$; very small effect size), DPA_DBR ($F_{(2,\ 4470)} = 138.692$; p $value = 0.001$; $\eta^2 = 0.058$; $power = 1.000$; small effect size), DPA_D1stHMR ($F_{(2,\ 4470)} = 21.383$; p $value = 0.001$; $\eta^2 = 0.009$; $power = 1.000$; very small effect size), DPA_D2ndHMR ($F_{(2,\ 4470)} = 94.884$; p $value = 0.001$; $\eta^2 = 0.041$; $power = 1.000$; small effect size), DPA_DFR ($F_{(2,\ 4470)} = 56.810$; p $value = 0.001$; $\eta^2 = 0.025$; $power = 1.000$; small effect size), Triang_DBR ($F_{(2,\ 4470)} = 54.004$; p $value = 0.001$; $\eta^2 = 0.024$; $power = 1.000$; small effect size), Triang_D1stHMR ($F_{(2,\ 4470)} = 13.213$; p $value = 0.001$; $\eta^2 = 0.006$; $power = 0.998$; very small effect size), Triang_D2ndHMR ($F_{(2,\ 4470)} = 13.011$; p $value = 0.001$; $\eta^2 = 0.006$; $power = 0.997$; very small effect size) and Triang_DFR ($F_{(2,\ 4470)} = 16.869$; p $value = 0.001$; $\eta^2 = 0.007$; $power = 1.000$; very small effect size).

In the case of the half of the match, it was established as an alternative hypothesis that there are statistical significant differences on the dependent variables between the two possible halves of the match (first half and second half). Significant differences were found among the three final scores for TP_defensive ($F_{(2, 4470)} = 1{,}737$; p value = 0.188; $\eta^2 = 0.000$; power = 0.261; very small effect size), TP_middle ($F_{(2, 4470)} = 0.466$; p value = 0.495; $\eta^2 = 0.000$; power = 0.105; very small effect size), TP_attacking ($F_{(2, 4470)} = 1.161$; p value = 0.281; $\eta^2 = 0.000$; power = 0.190; very small effect size), DPA_DBR ($F_{(2, 4470)} = 44.980$; p value = 0.001; $\eta^2 = 0.010$; power = 1.000; small effect size), DPA_D1stHMR ($F_{(2, 4470)} = 19.471$; p value = 0.001; $\eta^2 = 0.004$; power = 0.993; very small effect size), DPA_D2ndHMR ($F_{(2, 4470)} = 17.476$; p value = 0.001; $\eta^2 = 0.004$; power = 0.987; very small effect size), DPA_DFR ($F_{(2, 4470)} = 4.735$; p value = 0.030; $\eta^2 = 0.001$; power = 0.585; very small effect size), Triang_DBR ($F_{(2, 4470)} = 2.925$; p value = 0.087; $\eta^2 = 0.001$; power = 0.401; very small effect size), Triang_D1stHMR ($F_{(2, 4470)} = 0.019$; p value = 0.891; $\eta^2 = 0.000$; power = 0.052; very small effect size), Triang_D2ndHMR ($F_{(2, 4470)} = 5.559$; p value = 0.018; $\eta^2 = 0.001$; power = 0.654; very small effect size) and Triang_DFR ($F_{(2, 4470)} = 4.328$; p value = 0.038; $\eta^2 = 0.001$; power = 0.548; very small effect size).

Discussion

This study proposed a method to identify TPs in soccer by only considering the Cartesian positional data of players during the match. Such information was used to introduce a new concept denoted as the DPA, which estimates the regions of the defensive pressure and analyse the defensive pressure made by soccer teams. This approach was validated over three official matches as a case study. Moreover, using the outcomes from the proposed metrics, it was possible to perform an ANOVA between different final scores and halves of each match.

The first method developed in this study defined the TP of all players in a given moment. Such a method was computed based on the Cartesian positional data of all players throughout the match. From such information, it was possible to estimate the number of defensive, midfielder and attacking players at a sampling rate of once per second. It is important to highlight that although one player can be a lateral defender (as a strategic position), he may still have the position of a midfielder during the same match.[18] This method, therefore, considers the spatial position instead of the strategic position.[19] The outcomes suggested that on average, the highest number of defensive players (3.51), and consequently, the lowest number of midfielders, may be observed when the game ends in a draw. On one hand, the smallest number of defensive players is identified when the team loses a match (3.05), in which one can also observe the highest number of midfielders (3.39) and attacking players (3.56). On the other hand, the smallest number of attacking players (3.30) may be found in a winning score. In that situation, the defensive organisation had more defensive and less attacking players than during losing scores and more midfielders than drawing scores.

The influence of each half of the match in the variation of position was also inspected. No statistical differences were found between the first and second halves. Nevertheless, it was possible to identify a small increasing tendency to decrease the number of defensive players and to increase the number of midfielders and attacking players from the first to the second half. Such a result can be justified by the attempt to exploit an opponent during the second half, and try to perform counter-attacks, thus increasing the defensive pressing by the highest average of attacking players.[3]

To inspect the area covered by teammates during defensive moments in soccer, the metric of DPA was developed. Such a metric was proposed based on the interactions between positions. The DPA was organised in four relative regions. The results showed that the defensive pressing in the second half of the midfield had the highest area coverage in all scores in comparison with the other defensive regions. The second highest area was the first half of the midfield, thus suggesting that a great amount of defensive pressing occurs on the whole midfield. The attacking and backward regions of pressing correspond to the smallest regions of pressing. In the four defensive regions, the greatest amount of area was performed in final score of draw, being statistically different from the remaining results (win and loss). This result can be justified by the team's strategy in gaining disadvantage or advantage. In fact, if a team needs to reverse their losing status, then it will increase the pressing in block, thus reducing the area of action and increasing the intensity of tactical actions to quickly recover the ball.[14] Such analysis can be very important to identify some patterns of teammates' interactions in ball recovery used by observational methods.[20]

The comparison between first and second halves showed that the greatest area in all regions of defensive pressing was higher in the first half. This observation is in line with a previous study that measured the weighted stretch index and effective area of play in soccer matches.[3] In this study, a large amount of effective area of play and weighted stretch index in first half was found.

Regarding the number of defensive triangulations for each region, it was found that in winning status, a team had the highest statistical average of triangulations in the first and second midfield regions. This result is truly interesting because it suggests that higher pressing in the midfield region can be associated with successful results. Nevertheless, such a condition must be inspected in depth through further studies with a larger number of matches. Another interesting result involves the differences between the first and second halves. The number of triangulations was higher in the second half of the midfield and attacking pressing regions. Such a result is interesting because the greater coverage area was covered in the first half. Such results may thus suggest that cooperation is higher in the second half (due to the number of triangulations), but it is performed in a closest way (reducing the total of coverage area per region).

In summary, this study showed that the greater defensive pressing occurs more often in the midfield region than in attacking or backward regions. Moreover, it was found that the specific team under analysis used a balanced distribution between defensive, middle and attacking positions throughout defensive moments. The greatest amount of defensive coverage was found in the first half, although the greatest number of defensive triangulations was in the second half. It is now expected that this metric can augment the coaches' knowledge during the match about the kind of defensive pressure performed by their team having access to new technological devices such as augmented reality glasses.

Despite these findings, this study had some limitations mainly in the sample size and the computation of momentary TP. The main goal of this study was to propose two new metrics of match analysis for tactical behaviour, using three case study matches as a proof-of-concept.

Further studies must be done by considering a larger number of matches, as well as simultaneously using other tactical metrics, such as effective area of play or weighted stretch index, to correlate all metrics and better understand the collective organisation. It would be interesting to identify how teams vary their positions and defensive area of play throughout a match in accordance with different momentary scores.

Conclusion

Three tactical metrics were proposed in this study, based on Cartesian position of soccer players. The momentary TP allowed an understanding of the variation of TPs and players distribution throughout a match. The DPA allowed identifying the defensive pressing coverage per region. The triangulations of defensive pressing showed how teammates interact to generate the pressing area. It was possible to propose a set of new computational metrics that can help coaches to use information provided by tracking methods, such as video-tracking or global positioning systems, to increase the available information about the collective behaviour of soccer teams. Using the information generated by these metrics, one may characterise the defensive processes of teams and augment the knowledge of soccer coaches. Now it possibly increases the coaches' knowledge during the match. To do that, devices such as augmented reality glasses can be a good solution to increase the information for the coach during the match or daily training sessions.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Funding

This study was carried out in the scope of R&D Unit 50008, financed by UID/EEA/50008/2013.

References

1. Hughes M and Franks M. *Notational analysis of sport: systems for better coaching and performance in sport*. New York: Routledge, 2004.
2. Duarte R, Araújo D, Correia V, et al. Sports teams as superorganisms: implications of sociobiological models of behaviour for research and practice in team sports performance analysis. *Sports Med* 2012; 42: 633–642.
3. Clemente FM, Couceiro MS, Martins FML, et al. Measuring collective behaviour in football teams: inspecting the impact of half of the match on ball possession. *Int J Perform Anal Sport* 2013; 13: 678–689.
4. Bourbousson J, Sève C and McGarry T. Space-time coordination dynamics in basketball. Part 2: the interaction between the two teams. *J Sport Sci* 2010; 28: 349–358.
5. Bartlett R, Button C, Robins M, et al. Analysing team coordination patterns from player movement trajectories in football: methodological considerations. *Int J Perform Anal Sport* 2012; 12: 398–424.
6. Schollhorn W. Coordination dynamics and its consequences on sports. *Int J Computer Sci Sport* 2003; 2: 40–46.
7. Frencken W, Lemmink K, Delleman N, et al. Oscillations of centroid position and surface area of football teams in small-sided games. *Eur J Sport Sci* 2011; 11: 215–223.
8. Clemente FM, Couceiro MS, Martins FML, et al. Measuring tactical behaviour using technological metrics: case study of a football game. *Int J Sport Sci Coach* 2013; 8: 723–739.
9. Okihara K, Kan A, Shiokawa M, et al. Compactness as a strategy in a football match in relation to a change in offense and defence. *J Sport Sci* 2004; 22: 515.
10. Gréhaigne JF and Fernandez A. Qualitative observation tools to analyze soccer. *Int J Perform Anal Sport* 2001; 1: 52–61.
11. Abdel-Aziz Y and Karara H. *Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry*. Falls Church, VA: American Society of Photogrammetry, 1971 (ASP symposium on close-range photogrammetry, pp.1–18).
12. Couceiro MS, Clemente FM and Martins FML. Analysis of football player's motion in view of fractional calculus. *Cent Eur J Phys* 2013; 11: 714–723.
13. Seabra F. *Identificação e análise de padrões de circulação da bola no futebol*. Unpublished Thesis, Escola de Educação Física e Esporte, Universidade de São Paulo, São Paulo, 2010.
14. Trapattoni G. *Coaching high performance football*. Spring City, PA: Reedswain Inc., 1999.
15. Maroco J. *Análise Estatística com ultização do SPSS*. Lisboa: Edições Silabo, 2010, p.199.

16. Pallant J. *SPSS survival manual: a step by step guide to data analysis using the SPSS program*. Crows Nest, NSW, Australia: Allen & Unwin, 2011.
17. Hopkins KD, Hopkins BR and Glass GV. *Basic statistics for the behavioral sciences*. Boston, MA: Allyn & Bacon, 1996.
18. Couceiro MS, Clemente FM, Martins FML, et al. Dynamical stability and predictability of football players: the study of one match. *Entropy* 2014; 16: 645–674.
19. Gréhaigne JF, Bouthier D and David B. Dynamic-system analysis of opponent relationship in collective actions in football. *J Sport Sci* 1997; 15: 137–149.
20. Barreira D, Garganta J, Guimarães P, et al. Ball recovery patterns as a performance indicator in elite soccer. *Proc IMechE, Part P: J Sports Engineering and Technology* 2014; 228(1): 61–72.