Chaos, Solitons and Fractals 138 (2020) 109831

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Collective movement analysis reveals coordination tactics of team players in football matches

Rui Marcelino[a,b,*], Jaime Sampaio[a,c], Guy Amichay[d,e,f], Bruno Gonçalves[g,h,i], Iain D. Couzin[d,e,f], Máté Nagy[d,e,f,j,k,*]

[a] Research Center in Sports Sciences, Health Sciences and Human Development, CIDESD, CreativeLab Research Community, Vila Real, Portugal
[b] University Institute of Maia, Maia, Portugal
[c] Sport Sciences Department, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal
[d] Max Planck Institute of Animal Behavior, Department of Collective Behaviour, Konstanz, Germany
[e] Department of Biology, University of Konstanz, D-78457 Konstanz, Germany
[f] Centre for the Advanced Study of Collective Behaviour, University of Konstanz, 78464, Konstanz, Germany
[g] Comprehensive Health Research Center (CHRC), Évora, Portugal
[h] Departamento de Desporto e Saúde, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal
[i] Portugal Football School, Portuguese Football Federation, Oeiras, Portugal
[j] MTA-ELTE Statistical and Biological Physics Research Group, Budapest, Hungary
[k] MTA-ELTE 'Lendület' Collective Behaviour Research Group, Hungarian Academy of Sciences, Eötvös Loránd University, Budapest, Hungary

ARTICLE INFO

Article history:
Received 27 January 2020
Revised 3 April 2020
Accepted 17 April 2020

Keywords:
Collective behaviour
Team sports
Cooperation and competition
Directional correlation analysis

ABSTRACT

Collective behavior is a hallmark of every living system and utilizing methods from statistical physics (such as correlation functions) could aid in our understanding of their underlying rules. We analyzed five football (soccer) matches as this game provides a unique but yet mostly unexplored example to study a system of collective cooperation and competition. The aim of our study was to analyze the collective motion patterns exhibited by football players to unfold the underlying coordination among them in order to understand collective strategies associated with team performance. By analyzing pairwise relationships among all the players using spatio-temporal correlation functions we reveal that there exist identifiable collective dynamics that characterize winning and losing teams. Using our metric we find clear and robust differences between the players, indicating a difference in their behavior and their interactions. And this enables us to assign a unique behavioral pattern - a 'fingerprint' - for each individual and for each team. Furthermore, we reveal there exists a relationship between the market value of the players and the metrics introduced here, suggesting that these metrics could potentially serve as valuable performance indicators in the future, with applications ranging from talent identification to player scouting. In a broader context team sports could open up new directions for quantitative analyses of human collective behavior.

1. Introduction

Collective behavior characterizes the way in which an individual's activity is influenced by other members of its group, and in turn how they influence it [1]. In a system consisting of many similar 'units' [2] (such as fish in a shoal, birds in a flock, or players of a team sport [3]) the interactions between them can be simple (attraction/repulsion) or more complex, and can occur between neighbours in space [4] or through an underlying social network [5]. In human behavior joint action is ubiquitous, such as during pedestrian traffic, teamwork (*e.g.* bucket brigade or human chain), team sports, or group dance, where movements are adjusted according to the ongoing actions of others to fit the demands of the tasks at hand [6–9]. Here, we investigate joint actions of football players (European association football; soccer) to understand how players perform as a collective and how this relates to team success.

Football is played between two opposing teams where players interact directly, indirectly and concurrently to achieve an objective that involves scoring goals while simultaneously preventing the opposition from scoring [10,11]. In this sense, football is composed of individuals who coordinate their movements and develop

* Corresponding author.
E-mail addresses: rmarcelino@ismai.pt (R. Marcelino), nagymate@hal.elte.hu (M. Nagy).

https://doi.org/10.1016/j.chaos.2020.109831

cooperative relations to outcompete their opponents [12]. Having this theoretical knowledge as a base ground, a set of metrics have already been developed to incorporate a wide range of complex behaviors and explore how these impact and be impacted by group structures. For instance, the teammates' displacement synchronization, computed through the relative phase [11], have been used to assess the dynamic players' interpersonal coordination [13,14]. Preliminary results have shown that, by being more synchronized in positioning with teammates, players decreased their physical and physiological responses during the game [13]. Also, advances in knowledge have been made through the increased variety of processing techniques to assess regularities between whole teams and between individual players with their team as a function of time [15,16]. Entropy computations have been used to identify if individuals' positioning dynamics express a predictable pattern which may, in turn, provide insights about the local information sources that are being used in movement decisions [17].

Historically, the analysis of football has focused on several descriptive indicators (with a preponderance to player movement characterization) and in the comparison of different single players' performance according to different contexts [18]. In the behavioral sciences, by contrast, a paradigm shift has occurred in the study of group behavior, shifting focus from the study of single 'focal' individuals to the quantitative study of simultaneous interactions among organisms, and the resulting collective properties exhibited by groups. As a result of advances that allow automatic player identification and tracking [19–23], some recent works started to explore group behaviour analyses [24–26] (for a systematic review see [27]). These analyses mostly relied on metrics that gave overall collective scores (for example: "team centroid" - geometric mean position from all team players, or "dispersion" - an indicator of how "compact" the team is, *i.e.* large dispersion means players are more distant from each other) and less detailed insight to players' individual contributions to the collective behaviour (i.e. how players coordinate with each other). Thus, and since the literature has already explored some collective indicators that result from the spatial relationships of the players, it only makes sense to move towards more fine-tuned analyses to understand the individual role of each 'unit' (player) on the behaviour of the remaining 'units' of the same system (teammates and opponents).

Analyzing performance with the aim of revealing the dynamics and complexity of team sports is largely an unexplored area and will likely lead to new avenues with theoretical and practical implications [28–30]. The integrated use of positional data, for example, may make it possible to develop performance indicators capable of describing and understanding the mechanisms that underlie players' collective movements [31,32] and to describe their coordination when performing different tasks together, or how their individual and collective behavior adapts to that of their opponents. The practical implications of applying this type of analysis may have an impact, for example, in the assessment of talent in a sports context, with the immediate consequences in the transfer of players between teams. It is public knowledge that each year players are being transferred between clubs for record-breaking fees. Nevertheless, it is especially surprising that the evaluation of talented players relies on very few performance indicators (such as the number of goals scored) and circumstantial, subjective impressions [33], and there is no indication of the existence of performance indicators that consider the way in which players cooperate and/or compete, considering their spatial movements during competitions.

The aim of our study was to analyze collective patterns in the motion of football players to reveal the underlying coordination among players within teams as well as between players at different teams in order to reveal collective strategies associated with team performance. To do so we used directional correlation with time delay analysis [34] - which detects similarity/difference between concurrent motion patterns among players. Using high-spatiotemporal resolution positional data of the players (and of the ball), we demonstrated how football players dynamically adapt their behavior depending on different contexts, such as changing their movements relative to their teammates and opponents according to their playing position (goalkeepers, defenders, midfielders or forwards) and according to the match phase (offense vs. defense). This work opens up new possibilities for quantifying the performance of sport teams which can serve as an innovative method for applications from design and development of team strategies to the screening of players with tactical talent.

2. Methods

2.1. Study design

This study followed a cross-sectional design in which the spatial coordinates of all players were collected, during the entire duration of the analyzed matches. For each player from his high resolution trajectory, spatio-temporal correlation based metrics were calculated with the other players (teammates and opponents) and the ball, in order to identify the *highly correlated segments* (HCS). After calculating the HCS for each pair of players, individual scores and team scores were produced and analyzed according to different time scales (minute by minute, 1st half *versus* 2nd half, and full matches), location in the field (heatmaps), ball possession (in offense *versus* in defense), player position (defenders, midfielders, attackers), time delay of HCS (ahead in time *versus* behind in time), game status (winning *versus* losing). The aggregated metrics were compared to each other, to the match result, to the team success and to other performance scores available for the players, in order to put our results into perspective.

2.2. Participants and data collection

Five full matches from the German Bundesliga 2015-2016 season were analyzed, played by nine different teams, during one week period. The data set were provided by an external company, and no a priori criteria were set in choosing the actual matches. Two of the matches were balanced (finished with a tie) and three were unbalanced matches (finished with one team winning with two or more goals). All the matches ended with goals scored (number of goals: $n = 20$). The X and Y positional coordinates (over 26 million locations) of all players ($n = 139$ in total; Goalkeeper: $n = 10$; Defenders: $n = 44$; Midfielders: $n = 49$; Forwards: $n = 36$) and of the ball ($n = 5$) were collected with TRACAB® Optical Image Tracking System, Gen5 (ChyronHego) at 25 Hz. TRACAB® has been successfully validated as technology to collect spatiotemporal measurements in a football-specific test conditions with high accuracy [35]. We included in the analysis only those pairs that played together at least 10 minutes (to have enough data on the players behaviour/performance but the actual value for the cut off had no impact on the results within $\pm 20\%$ relative variation). We collected the players' market value data from Transfermarket (www.transfermarkt.com) website.

The study protocol was approved and followed the guidelines stated by the Ethics Committee of the Research Centre for Sport Sciences, Health and Human Development, based at Vila Real (Portugal) and conformed to the recommendations of the Declaration of Helsinki. Under this protocol, all data were collected for 'real world use', without informed consent, and shared with the research team. As such, these data conform to U.S. Department of Health and Human Services, "Regulatory considerations regarding classification of projects involving real world data" [36].

2.3. Highly correlated segments calculation

A highly correlated segment (HCS) was defined for each pair of players ($i \neq j$) for a time t when their directional correlation, $C_{ij}(t)$—the time-averaged dot product of their time delayed directions of motion (normalized velocities) for a time window—was higher than a threshold ($C^{MIN} = 0.99$). In more detail: $C_{ij}(t) = max_{\tau \in [-4s:4s]}(\langle v_i(t^*) \cdot v_j(t^* + \tau)\rangle_{t^* \in [t-\frac{T}{2}:t+\frac{T}{2}]}) \geq C^{MIN}$, where $v_i(t)$ denotes i's normalized velocity, $\langle \ldots \rangle$ denotes time averaging for the $T = 6s$ long interval centred at t, and a maximum value was searched for τ in the interval $[-4s : 4s]$ as a conservatively large range. We pooled data with multiple bin sizes ranging from 1 minute to the duration of a full match. To get a value for highly correlated segments that could be compared between different binning we calculated the *proportion of HCS* (S_{ij}) as the total number frames when $C_{ij}(t) \geq C^{MIN}$ and divided that by the total number of frames the two players were together on the field during that time bin. To characterize the performance of players we calculated the average over all possible other players ($S_i = \sum S_{ij}$), all other teammates or all opponents. Data was analyzed using custom-written dedicated scripts in Perl (v5.22.1) and Cuda 500 (v7.5.17).

2.4. Statistical procedures

Magnitude-based inferences [37] techniques were used to evaluate the differences of HCS for all the multiple filters used in this work (see Fig. 2, and for more details the Supplementary Methods). Calculations were performed with the r package "mbir". The association of several facets of the HCS was assessed via Linear Regression Trend Models (r^2) with 95% confidence intervals. Pearson's product moment correlation (r) was used to evaluate the correlation of the HCS with the players market values as well to assess the robustness of the HCS metrics. When comparing the data of the two halves, Spearman's rank correlation was used as the data didn't follow a normal distribution. We reported two-tailed *p-values*, and the exact values for r and p are presented, unless those where p were lower than 0.001. We suggest considering $p \leq 0.05$ as a threshold for significance, and $r^2 \geq 0.5$ for magnitude, but let the reader decide on their interpretation of the statistical tests.

2.5. Data availability

The positional data of the players and the ball are owned by the company Opta and the authors did not get authorisation to share the raw positional data to a third party. However, we provide a sample data with open access from an official match (available in https://github.com/nagymate80/HCS_football), with the full trajectories of all field players during the first half (45 min). The sample data differ from the data analyzed here in their resolution and they do not contain positional data of the goalkeepers and of the ball, nor information about the inactive match periods (ball out of bounds, injured player, celebration of a goal, etc.) (S5 Table). Despite these differences, we show that our analysis is robust and the sample dataset results in similar patterns as compared to the original dataset when resampled to accommodate for these differences (S6 Fig). All the custom-written dedicated scripts in Perl and Cuda 500 used in this paper are freely available in the previously mentioned github repository.

3. Results

We analyzed high spatio-temporal resolution positional data of players and the balls from 5 official football matches between 9 different top level teams (hereafter referred to as *tA* to *tI* to preserve their anonymity), including one team that played twice (Table 1). In order to quantify the joint actions among players in these matches, we calculated correlations between path segments of pairs of players to detect co-occurring similar motion patterns. We defined *highly correlated segments* (HCS) when the difference in the movement direction of two players was less than a threshold angle (8 degrees on average). These trajectories are typically spatially displaced. This metric can give us a first approximation of how the players are cooperating and competing, by moving in similar or dissimilar directions in different tasks.

We calculated the HCS by cross-correlating two trajectories (with possible time delays) and searching for peak values. A detected HCS characterizes an event (with a duration of at least a few seconds) of highly correlated motion between two players at a given time point (Fig. 1). Different players spent different amounts of time on the field (and we analyzed the data on multiple timescales), so to be able to compare these data, we used the proportion of HCS which characterize in what portion of the relevant time were HCS events detected.

From this pairwise event-based dataset we pool data using multiple 'filters' (or a combination of them): We (*i*) use different time bins ranging from one minute, a match half (45 minutes) up to the whole match length; (*ii*) pool data for a player with all of his teammates/opponents to get individual scores; (*iii*) pool data for the whole team; (*iv*) pool data when a focal individual (or team) is being compared only either to the same or to the opposite team's members; (*v*) pool data based on ball possession (whether the focal individual (or focal team) is in offense/defense) (*vi*) pool data from multiple individuals based on the players' 'positions' (*i.e.*, role: attacker/midfielder/defender/goalkeeper); (*vii*) separate data based on the time delay between the players during the HCS; and (*viii*) use the location of the players in the field while the HCS occurs (Fig. 2). We stress the fact that the ball has different motion characteristics compared to the players - between passes it moves mostly in a straight line in the (x, y) horizontal plane and it can move much faster than the running speed of the players. This could result in the absence of HCS between the players and the ball, although the players' motions are obviously affected by the ball.

3.1. Highly correlated segments (HCS) with teammates and opponents

We found that 82.0% of the field players had greater proportion of highly correlated segments (HCS) with their opponents than they did with their teammates (Effect size: 0.48, 99% Confidence Interval (CI) [0.35, 0.61]; difference in mean: 36.8%, 99% CI [28.7, 44.9], for more details see Supplementary Methods), with the percentages ranging from 60.9% to 92.3% from match to match. The goalkeepers behave in the opposite way and 8 out of the 10 goalkeepers had more HCS with their teammates than they did with their opponents (S1 Table). It is important to note that the distances that goalkeepers travel per game are very low and, therefore, high correlations would not be expected.

Interestingly, in Match #2, which ended with a draw but was played between unbalanced teams (team tD had a final ranking twelve places above team tC), the players of the lower ranked team had much more HCS among themselves than they did with opponents (Fig. 3A). This difference is predominantly due to defenders of the lower ranked team which exhibited a lot of HCS while their team was in offense (Fig. 3B–E). As the motion of the attacking players (and their movement relative to each other) should be relatively unexpected, moving in unison (with high HCS) may be a sign of poor offensive threats. Thus, one possible interpretation of this result could be that these defenders performed poorly in supporting their team's attack by following very similar motion patterns, and not 'surprising' the opponent with unique movements and positions.

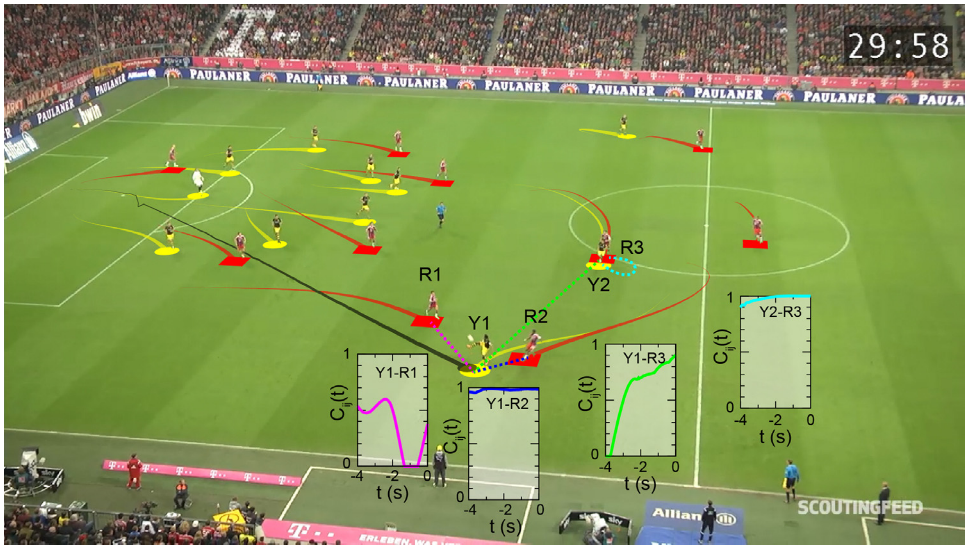

Fig. 1. Example of tracking and the method for finding highly correlated segments (HCS) by directional correlation functions between players overlaid on top of a video frame. Players' (with red and yellow for the two teams) and the ball's trajectories (black) of the last 4 seconds were projected in perspective on the plane of the field. Red squares and yellow circles show the position of the players at the moment corresponding to the displayed frame ($t = 0$). Insets show examples of pairwise directional correlation functions for four selected pairs (indicated by dashed links between the selected players). For each timepoint, the highest correlation is given from all the possible correlations with different time delays. The correlation for a pair is symmetrical. Highly correlated segments can be seen between players Y1-R2 (blue) and Y2-R3 (cyan). The pair Y1-R3 (green) at the beginning of the plotted 4-second-long period move in approximately perpendicular directions (thus their correlation is close to 0), then as they move more and more towards the same direction their correlation increases. To keep the anonymity of the players this figure shows a match that was not analysed in this paper.

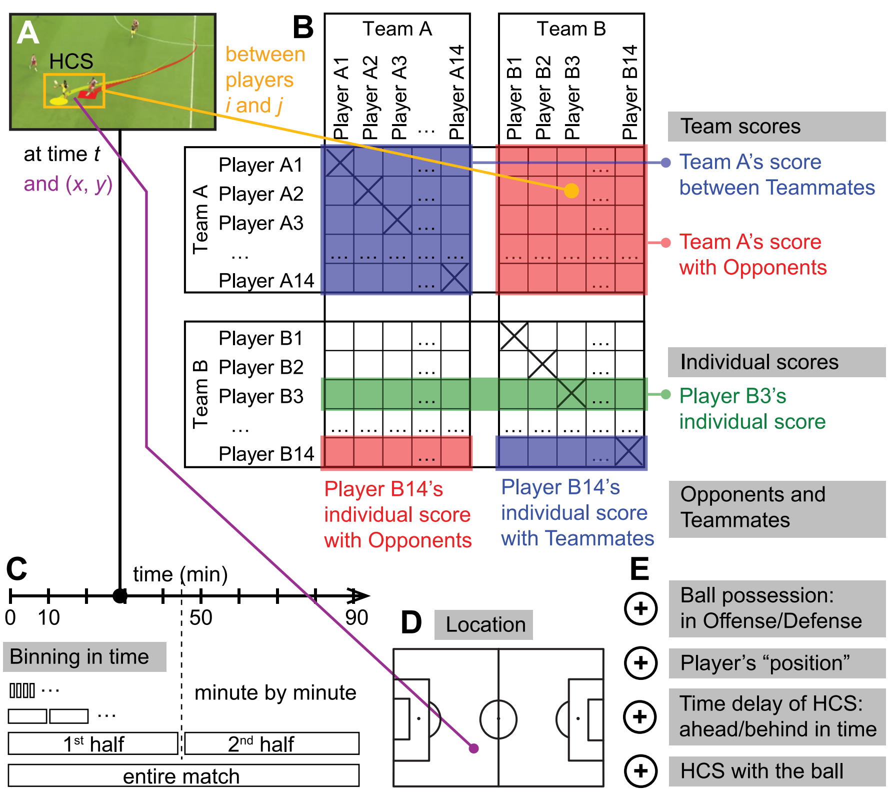

Fig. 2. A schematic explanation of the pooling the pairwise and event-based HCS into different individual and team scores as well as the other metrics used. A) When a HCS is detected between two players at a given time and location, it gives contributions to several metrics (B–E). B) We show here a simplified version of the full matrix of all possible pairings consists of the 14 players (11 regulars + 3 substitutes) from both teams (S2 Fig). We averaged all pairwise values in a row for individual scores, and all values in a submatrix of the same (or the opposite) team for team scores. C) We pooled data into time bins of various length. D) Heatmaps are generated from the location of the HCS. E) Other filters that were used in this paper to analyze the data (but not shown in details on the figure).

Table 1
Details of the matches and the teams analyzed.

Match No.	Played	Team ID	Goals	Weekly Rank	Final Rank	Marker color
	Home	*tA*	1	11	[16–18]	Blue empty
1	Away	tB	1	6	[7–9]	Blue full
	Home	tC	2	2	[1–3]	Orange full
2	Away	tD	2	9	[13–15]	Orange empty
	Home	**tE**	**5**	**1**	[1–3]	**Green full**
3	Away	tF	1	3	[7–9]	Green empty
	Home	**tG**	**2**	**3**	[4–6]	**Red full**
4	Away	*tA*	0	11	[16–18]	Red empty
	Home	**tH**	**4**	**14**	[4–6]	**Purple full**
5	Away	tI	2	16	[10–12]	Purple empty

Team ranking is according to the teams' league position, where 1 stands for 1st place, and so forth (the worst team has the highest number). Weekly Rank denotes the temporal ranking during the week the matches were played. The Final Rank shows the results at the end of the season but to ensure the anonymity of the teams only a range is given. In case of unbalanced matches, the winner team is shown highlighted by bold font type. Team tA played twice in matches #1 and #4.

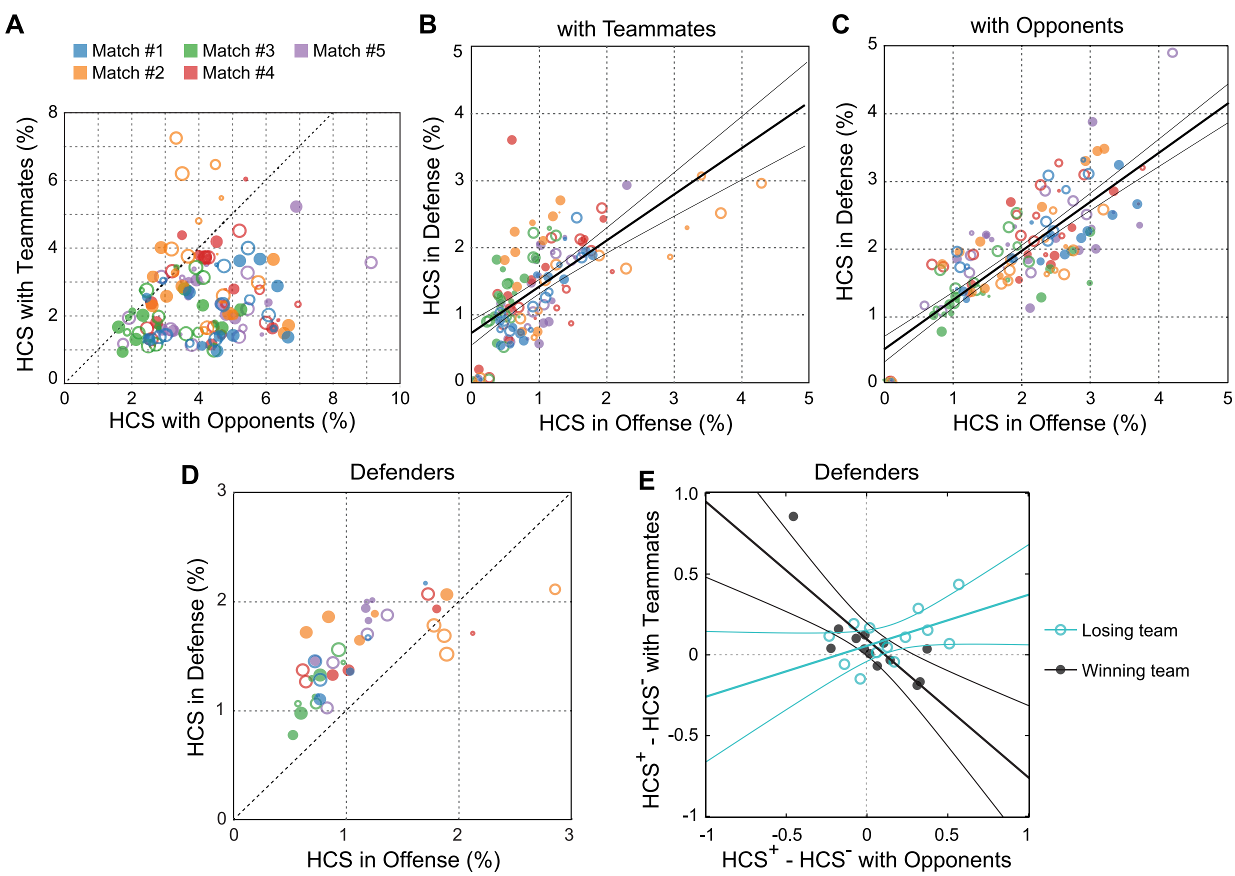

Fig. 3. The average proportion of highly correlated segments for the full match using different filtering: with teammates/opponents (A), in offense/defense (B,C), by player positions (defenders; D) and being ahead/behind in time (E). A–D) Each dot represents one player color-coded by the match they played in and sized proportional to the time spent on the field. A filled marker shows the team with higher final ranking (which coincided with winning the match in unbalanced matches) as opposed to empty markers for their opponents. B-C) Black lines represent the Linear Trend Model (with confidence bands showing upper and lower 95% confidence lines). Both models are significant at $p < 0.0001$ (S4 Table). D) Individual HCS (correlated with all players including teammates and opponents) for defenders ($n = 41$). Other positions are shown on S2 Fig. E) For defenders, the difference between HCS when ahead (HCS^+) and when behind (HCS^-) in time as compared to teammates (x axis) and opponents (y axis) is shown color-coded by the final result of the match. Trend lines for winning teams (black; slope $ß = -0.86$, $R^2 = 0.62$, $p = 0.001$) and for losing teams (cyan; slope $ß = 0.32$, $R^2 = 0.27$, $p = 0.07$).

We also found that the higher proportion of HCS among players within teams was dominated by that among defenders when in defense (Fig. 3B,D). These results highlight the potentially vital role that defenders have in a team, both for defensive and offensive roles.

We also investigated the precise timing of motion path between the players by calculating the time delay of the maximal correlations. In other words, if two players were highly correlated but one of them was making similar moves two seconds later, he was *behind* in this event. By comparing the relative timing in HCS metrics of winning and losing teams, we find that the winning defenders that were more often *ahead in time* compared to their teammates were more often *behind in time* when compared to their opponents (R-square $= 0.62$; $n = 13$, $p = 0.001$; Fig. 3E). Various parameters affect the decisions and motions of players during a match, so from the observed time lag in their motion we cannot conclude whether a player was directly leading or following another player. That said, one potential explanation for this finding is that sometimes defenders 'team-up' against an attacker, typically when he is a very talented forward player. If two defenders decide to move toward an attacker, for example, we would expect to see this relationship.

On the other hand, the winning defenders that were more often *behind in time* compared to their teammates were more often *ahead in time* when compared to their opponents. A possible scenario where this phenomenon could happen is the so-called 'offside trap', where the final line of defense (the defenders that are currently positioned closest to their own goal) tries to coordinate relative to the position of the opposing attacker (if the player is without the ball), to cause an offside ruling which will stop the attack.

By contrast, the defenders from losing teams did not follow this trend (R-square = 0.27, n = 13, p = 0.07). Typically, they were either more *ahead* or more *behind in time* relative to other players, but the trend found was the same when compared to both teammates or opponents. It is worth mentioning that for a player to be able to simultaneously track the motion of multiple players and the ball is challenging and may demand high spatial awareness. The aforementioned examples of successful defenders perhaps require such skills.

3.2. HCS in offense and in defense

Previous studies have tended to analyse matches as a whole, usually without dividing the match according to its phases (offense and defense) [15,18,24]. Here, by implementing this distinction, we reveal differences in the player's behavior when in offense and in defense (S3 Table). If a player has the possession of the ball, his teammates will move in order to facilitate the movement of the ball into the goal. To achieve that, they do not necessarily need to move with similar motion patterns either with respect to the player possessing the ball or with other teammates. Rather, players coordinate their movement by fulfilling different tasks, eventually moving in a less correlated way in order to surprise the opponents' defenders and to facilitate scoring opportunities. On the other hand, players from the defending team need to prevent the goal scoring, and, consequently, their movements will tend to be more correlated - they need to adopt similar moving trajectories as their opponents (more HCS) as well as to move together with their teammates in order to keep their defense organized.

Generally, players show significantly more HCS in defense than in offense (S3 Table). However, when comparing the similarity of motion with respect to opponents, we find no difference in HCS in defense and HCS in offense.

Players that have more HCS in offense are the ones that have more HCS in defense. There is a linear relationship both for the HCS with teammates (HCS in defense = 0.69 * HCS in offense + 0.73, R^2 = 0.38, p < 0.001; Fig. 3B) and with opponents (HCS in defense = 0.72 * HCS in offense + 0.52, R^2 = 0.67, p < 0.001; Fig. 3C). Independently of the match phase (being in offense or in defense) and of whom the other players are (whether they are teammates or opponents), some players are more 'collectively connected' (with more HCS) than are other players. This suggests that football players have an individual hidden 'fingerprint' with respect to the movements they make relative to others on the pitch.

3.3. Temporal and spatial dynamics

When analyzing the HCS of competing teams throughout matches in a minute-by-minute time resolution, we detect some changes in the HCS metrics, but overall the metric is robust (S3 fig). We also found that players played similarly in the first and the second halves (Spearman's r = 0.712, n = 70, p <0.001; S4 Fig). This confirms that no extreme outliers were affecting our data.

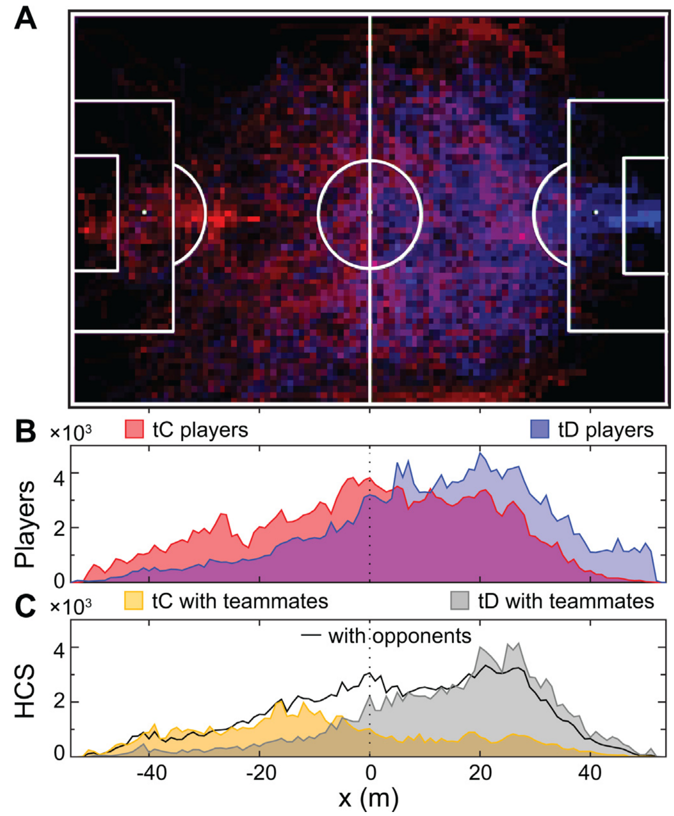

Fig. 4. Spatial distribution of the players (A,B) and the detected HCS events (C) between teammate and between opponent players from match #2. A) Two-dimensional heatmap showing the positions of the two teams for both halves of the match. Two color channels are used (red for team tC and blue for team tD) and where the two distributions overlap it results in purple. The positions were mirrored for the second half when the two teams switched sides. B) Distribution of the players' positions projected to the x axis - the main axis of the pitch connecting the two goals. C) Distribution of HCS detected on the same x axis between teammates and the opponents.

Comparing the spatial distributions of player positions and HCS locations across the pitch (see Fig. 4 for an example from one match), we found that the players of *Team tD* were predominantly in defensive positions, and correspondingly also had peak values of HCS there. This is what we would expect for a team that is occupied with defence. *Team tC* had lower HCS values in attacking positions (Fig. 4B, C), in line with the comparison of HCS according to ball possession (offense and defense).

3.4. Robustness of the HCS metrics

Are HCS values stable between different matches? We can address this question as one of the teams (tA) played against two different opponents. The HCS ratio of team tA's players showed highly significant correlations between the two matches both with their teammates (in offense: r = 0.97, p < 0.001; and in defense: r = 0.99, p < 0.001) and with the opponents (r = 0.99, p < 0.001; r = 0.97, p < 0.001, respectively; for players who played at least 15 minutes in both matches; n = 10; S5 Fig.).

These results, together with the similarities of the HCS metrics between the first half and the second half (S3, S4 Figs.), demonstrates the ability of our method to detect and reveal the patterns of the players' collective movements. Despite the different contextual situations (one match was played at home and ended in a draw and the other match was played away where the focal team lost), there is a regularity of the collective movements of the different players.

3.5. Comparison of the HCS metric to the market value of the players as a proxy for player quality

Players' market values are estimates of the transfer fees most likely to be paid for them [38]. Each successive year players are transferred between elite clubs for record-breaking fees. Although there are conceptual differences, market values and transfer fees are comparable [39]. As it is a very complex task to determine the quality of a player, we use the players' market value as a proxy for their quality [40,41]. We examined whether we could find correlations between the metrics defined in this paper and the market value of specific players/teams. As expected, the most successful teams (smallest number in team rank) indeed had higher average market values (Pearson's r = -0.704, n = 9, p = 0.034; Fig. 5A). Notably, players of the strongest team (overall in the season) showed distinct behavior when compared to the least successful team (HCS in offense vs. in defense; using the difference values ($HCS_{offense}$-$HCS_{defense}$) in a Welch-test for those two teams' field players: df = 21, p = 0.002; Fig. 5B). Higher market value players had a lower number of HCS with their teammates while attacking (attackers' mean for a team: Pearson's r = -0.719, n = 9, p = 0.029; all players: Pearson's r = -0.245, n = 125, p = 0.005; midfielders: Pearson's r = -0.317, n = 44, p = 0.036; and attackers: Pearson's r = -0.427, n = 30, p = 0.019; Fig. 5C–F). The lower market value players' moved more often in similar directions to other players, being more predictable and, consequently, likely to be more easily defeated. Previous research highlighted attributes such as positioning and decision-making as tactical skills that best predict adult performance levels [42,43]. We now offer the HCS metric as a key performance indicator in the talent identification process of young football players.

4. Discussion

This work is one of the first studies in a new interdisciplinary research field in which team sports are considered from the perspective of collective behavior. Previous studies are scarce, but in one nice recent study, Yokoyama and Yamamoto [9] analyzed spatiotemporal coupling and synchronization of humans playing a 3 vs. 1 football possession game. Our current study goes beyond previous attempts and the coordination of all the players (and also of the ball) are analyzed in formal football matches from high spatiotemporal data. The motivation for this work is twofold, being both application-oriented and basic-science research. High precision spatiotemporal data is now abundant in various biological systems, and over various scales. Currently, sports-data companies like Opta (www.optasports.com), STATSports® (https://statsports.com), Sportsradar (https://www.sportradar.com), among others, collects prodigious amounts of detailed performance data in football, which could be used for player and team evaluation. These data, together with the positional coordinates of each subject (the players, ball, referees, etc.) in a football match provided by advanced optical tracking and image-processing technologies, such as the TRACAB® system (www.chyronhego.com), enables unprecedented opportunities for data-analysis and inquiry. Here, we utilize data of sport players and analyze their movements employing an approach that is already common practice in studies of collective behavior in non-human animals [2], but was originally inspired by methods from statistical physics [34].

Data of this type could be analyzed to quantify many different patterns in team sports, such as the dynamics of the structure of the team and how individual and team behavior adapts to the motion of their opponents, or how synchronized the players are when performing different tasks together. Also, there is large scope for an increased variety of techniques that could be used to analyse team sports, such as methods from information theory, network science or supervised/unsupervised machine learning (including deep learning) [44].

So far, performance analysis studies in sports have mostly considered personal measurements, such as various physiological conditions - e.g. cardio-vascular or acceleration data [20,27,31,42]. This is indeed important, especially in sports that put emphasis on physical ability. However, team sports add several layers of complexity, and thus could benefit greatly from new approaches of analysis that consider collective action. The inherent limitations of the traditional individual-focused approach might be the reason that football is considered by many to be mostly random [45], and presently it is hard to predict the outcome of matches [46,47]. Thus, there is an urgent need for a holistic approach to sport performance, which will open up new ways of analysing and assessing team and individual performance. The produced knowledge, together with the increasingly known individual and collective performance profiles (*e.g.* travelled distances at different speeds, or different surface/stretch index areas; for more details see literature review from Cummins and colleagues [48]), could improve performance in a broad range of applications from talent identification to injury prevention or performance optimization.

It is known that in invasion team sports, in this particular case in football, team movements are guided by a collective tactic/strategy that aims, above all, to achieve a territorial advantage that enables teams to achieve their purposes: to score goals and not to concede goals. The method that was used in this study, when considering the position on the field, and the associated movements, of all players during the effective time duration of the game, allows to assess the instantaneous suitability of the movement of each player (in his relationship with teammates and opponents). Being a good player in an invasion sport implies that, in addition to being technically evolved when contacting the ball, there is a need to be in the right place at the right time. Here we have developed a metric that enables us to quantify the similarities and differences in the motion path segments between teammates and between opponents. We have demonstrated this to be a meaningful metric that could reveal qualities in players which may be often overlooked. Moreover, we find this to be a robust measurement, stable both within and between matches.

In the future it could also be interesting to use our methodology to compare club teams (as we considered in this work) to national teams. It is well known that national teams usually don't have the same amount of time for training (as a team) compared to clubs, despite competing in very important matches such as the World Championship. Results from these analyses could help in understanding the influence of training [49] as a team on the collective performance of football players. Studying sport games can also benefit the scientific research of collective animal behavior. When addressing the complexity of fish schools and bird flocks, scientists try to reveal the interaction rules between the conspecifics that enable their cohesive movement. In sports, we are aware (to some extent) of the rules governing their behavior. For instance, as we know about the offside rule, we can better understand behaviors that might seem puzzling and intriguing when analyzing these situations without this knowledge.

For football's sake alone, as the world's most popular game, insights from quantitative approaches, such as ours, could help coaches, scouts and staff in making wiser decisions, both during matches and in the transfer market. As we demonstrate, highly coordinated defenders can be crucial, and this requires skills (e.g. spatial awareness) that aren't typically tested for in youth systems [18, 50]. In addition, comparing between different styles of

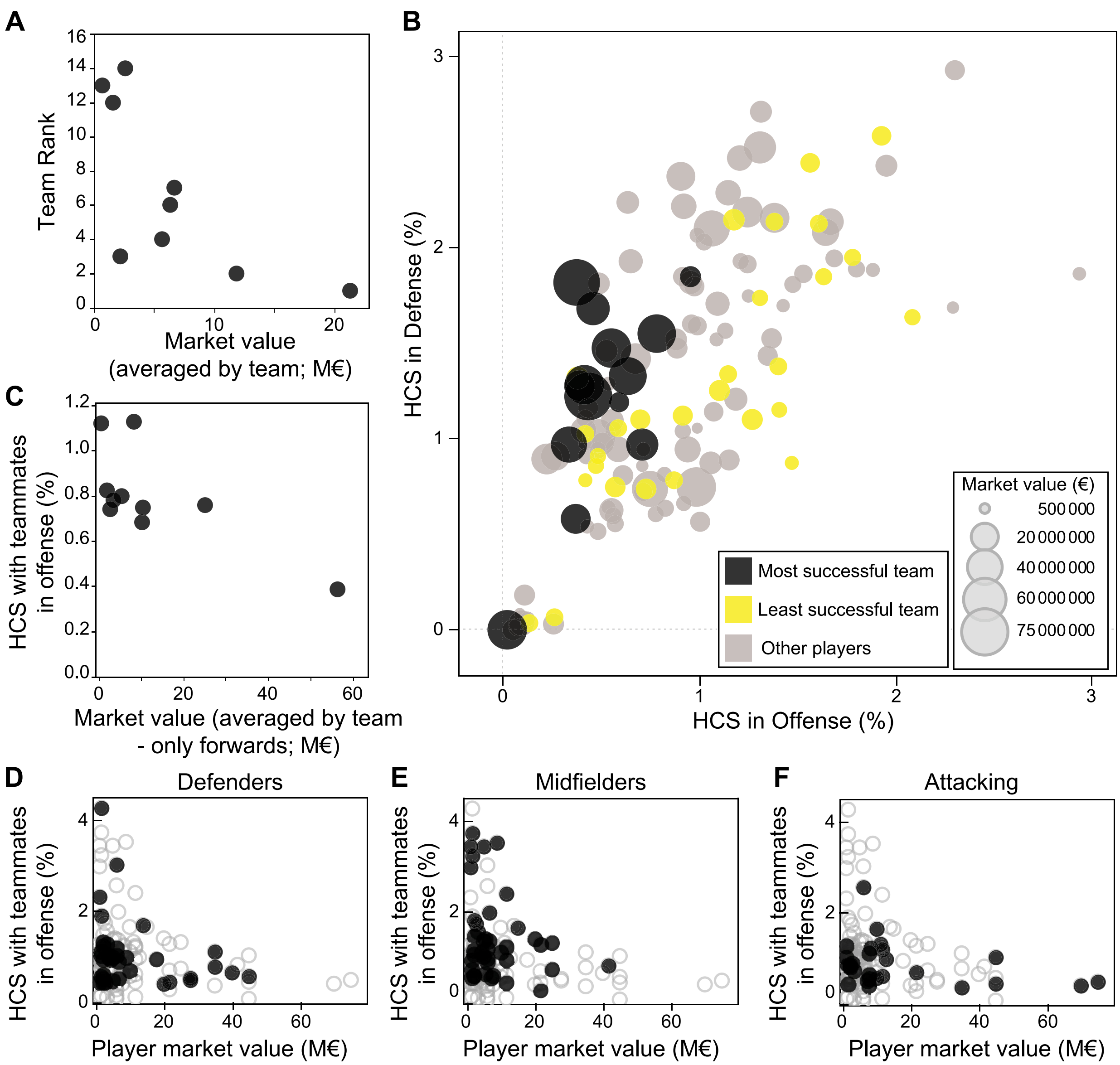

Fig. 5. Market value, team ranking and HCS metrics of players/teams. A) Final team ranking versus the average market value of all the players of the teams analysed to confirm the usage of the market value as a proxy for players' quality. Teams with players valued higher by the market ended the season at a higher rank (smaller number in ranking). B) HCS with teammates in defense vs. in offense is shown by circles for all players analyzed. The size of each circle is proportional to the market value of the player. The two teams with extreme ranking from our data set are highlighted by different colors (black is the most and yellow is the least successful team). C) Average HCS with teammates in offense for the teams' forward players versus their average market value. D–F) HCS with teammates in offense as a function of players' market value are shown for different positions: defenders (D), midfielders (E) and attackers (F) as full circles. Empty circles represents data from all players for reference.

play (from different leagues in different countries, known to be more technical, or more physical) could shed light on what makes them unique. Finally, the comparison between youth vs professional teams may provide useful insight into how experience and years of formal training changes players' abilities. In a broader context, the analysis of the collective dynamics in football (or team sports/games in general) offers a peek into human interactions (coordination/cooperation) in a 'controlled' setting, in which the participants are restricted to a certain range of behaviors, allowing for a more rigorous investigation than is possible in most social scenarios.

The methodology proposed in this work has practical applications that apply to the context of sports in particular, but which may also have an impact on other areas of human behavior. In the sports context, the use of this new individual metric can be applied in processes as diverse as: (1) identifying and recruiting talent, (2) optimize training processes in order to maximize the teams' tactical/strategic behaviors, (3) be used as an evaluation indicator for safe return to practice after sports injury, (4) when applied in real time, it can help coaches to make decisions regarding players who are lacking in their performance, or (5) to identify opponents in situations of greater fragility.

As for practical applications in other domains (with the necessary adaptations for each specific context), it can be used for instance to monitor groups of firefighters - identifying anomalies in movement that may be suggestive of exposure to danger; or, children in a schoolyard context - identifying students with a tendency to be isolated from other students, signaling them as likely to be victims of social isolation and, eventually, maybe more exposed to bullying. In summary, the methodology presented in this article has potential for the exploration and monitoring of many human collective movement activities.

It is important to acknowledge the limitations of this present work. This is one of the first public domain, open access studies that carried out an analysis on high resolution positions in football which includes information of all of the 22 individuals and of the ball, at all times (not only goals/assists, clean-sheets, etc). That said, we were limited in the amount of games available to us, and therefore there could be an inherent bias here. Future work should look at more games from different teams and leagues, and also from the same teams, but at different times throughout the season to see developments over time.

Another limitation is in our methodology - which served as a first assessment of how players interact. The use of directional correlation enabled us to reveal some coordination patterns; but, attacking teams could be highly coordinated without moving on similar paths - without being correlated. Future work could focus on developing new ways to quantify the behavior of very offensively-talented teams [25], to try to reveal the secrets behind their success in finding space within the opposition's defense.

5. Conclusions

In this work we aimed to advance the emerging field of analysis of football on a collective scale. By understanding what makes some teams better than others, we can inspire both coaches and football staff on their decision-making, and hopefully will also push football as a 'model-system' of collective action in general. With our pairwise analysis - focusing on the interactions between the players, we were able to find clear differences between the different roles on the pitch. In a nutshell, defenders need to be more organized as a unit, whereas attackers aim for the opposite, that is, to somehow move in unpredictable patterns (at least for their opponents). Also, within the same role, we show that our metric is a good proxy for the quality of the players and thus can help in predicting team success.

Declaration of Competing Interest

The authors declare no conflict of interest

CRediT authorship contribution statement

Rui Marcelino: Conceptualization, Resources, Data curation, Formal analysis, Funding acquisition, Validation, Investigation, Methodology, Writing - original draft, Project administration, Writing - review & editing. **Jaime Sampaio:** Supervision, Writing - review & editing. **Guy Amichay:** Validation, Writing - original draft, Writing - review & editing. **Bruno Gonçalves:** Data curation, Investigation. **Iain D. Couzin:** Supervision, Funding acquisition, Writing - review & editing. **Máté Nagy:** Conceptualization, Resources, Data curation, Software, Formal analysis, Supervision, Validation, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review & editing.

Acknowledgments

We thank Raphael Reiners (head of data collection, PERFORM, Opta Sports) for support with the data acquisition, Fritz Francisco and Michael Smith for commenting on our manuscript, and Tamás Vicsek and David Sumpter for valuable discussion. This work was supported by the Portuguese Foundation for Science and Technology (FCT) and European Social Fund (ESF), through a Post Doctoral grant endorsed to R.M. (SFRH/BPD/89806/2012), under the Human Potential Operating Program (POPH); and through the project Symbiotic technology for societal efficiency gains: Deus ex Machina (DEM), NORTE-01-0145-FEDER-000026, co-financed by Fundo Europeu de Desenvolvimento Regional (FEDER) by NORTE 2020. This work is supported by national funding through the Portuguese Foundation for Science and Technology, I.P., under projects UID/DTP/04045/2019 and UID04045/2020. G.A., I.D.C. and M.N. were founded by DFG Centre of Excellence 2117 "Centre for the Advanced Study of Collective Behaviour" (ID: 422037984). I.D.C was founded by Office of Naval Research grant N00014-19-1-2556. This research was partially supported by the Hungarian Academy of Sciences (a grant to the MTA-ELTE 'Lendület' Collective Behaviour Research Group).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chaos.2020.109831.

References

[1] Vicsek T. A question of scale. Nature 2001;411:421.
[2] Vicsek T, Zafeiris A. Collective motion. Phys Rep 2012;517:71–140.
[3] Zaccaro S, Rittman A, Marks M. Team leadership. Leadership Q 2001;12:451–83.
[4] Couzin I, Krause J, Franks N, Levin S. Effective leadership and decision-making in animal groups on the move. Nature 2005;433:513–16.
[5] Bode N, Wood A, Franks D. The impact of social networks on animal collective motion. Anim Behav 2011;82:29–38.
[6] Konvalinka I, Vuust P, Roepstorff A, Frith CD. Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. Q J Exp Psychol 2010;63:2220–30.
[7] Noy L, Dekel E, Alon U. The mirror game as a paradigm for studying the dynamics of two people improvising motion together. Proc Natl Acad Sci 2011;108:20947–52.
[8] Zhou G, Bourguignon M, Parkkonen L, Hari R. Neural signatures of hand kinematics in leaders vs. followers: a dual-MEG study. Neuroimage 2016;125:731–8.
[9] Yokoyama K, Yamamoto Y. Three people can synchronize as coupled oscillators during sports activities. PLoS Comput Biol 2011;7.
[10] Gréhaigne J-F, Godbout P, Bouthier D. Performance assessment in team sports. J Teach Phys Educ 1997;16:500–16.
[11] Kelso S. Dynamic patterns: the self-organization of brain and behavior. MIT Press; 1997.
[12] Duarte R, Araújo D, Correia V, Davids K. Sports teams as superorganisms: Implications of sociobiological models of behaviour for research and practice in team sports performance analysis. Sports Med 2012;42:633–42.
[13] Folgado H, Goncalves B, Sampaio J. Positional synchronization affects physical and physiological responses to preseason in professional football (soccer). Res Sports Med 2018;26:51–63.
[14] Goncalves B, Esteves P, Folgado H, Ric A, Torrents C, Sampaio J. Effects of pitch area-restrictions on tactical behavior, physical, and physiological performances in soccer large-sided games. J Strength Cond Res 2017;31:2398–408.
[15] Goncalves B, Coutinho D, Travassos B, Folgado H, Caixinha P, Sampaio J. Speed synchronization, physical workload and match-to-match performance variation of elite football players. Plos One 2018;13:e0200019.
[16] Goncalves BV, Figueira BE, Macas V, Sampaio J. Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. J Sports Sci 2014;32:191–9.
[17] Goncalves B, Marcelino R, Torres-Ronda L, Torrents C, Sampaio J. Effects of emphasising opposition and cooperation on collective movement behaviour during football small-sided games. J Sports Sci 2016;34:1346–54.
[18] Sarmento H, Marcelino R, Anguera M, Campaniço J, Matos N, LeitAo J. Match analysis in football: a systematic review. J Sports Sci 2014;32:1831–43.
[19] Abt G, Bray J, Benson A. Measuring moderate-intensity exercise with the apple watch: validation study. JMIR Cardio 2018;2:e6.
[20] Figueroa P, Leite N, Barros R. Tracking soccer players aiming their kinematical motion analysis. Comput Vision Image Understand 2006;101:122–35.
[21] Khatoonabadi S, Rahmati M. Automatic soccer players tracking in goal scenes by camera motion elimination. Image Vision Comput 2009;27:469–79.
[22] Stein M, Haussler J, Jackle D, Janetzko H, Schreck T, Keim DA. Visual soccer analytics: understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. Isprs Int J Geo-Inf 2015;4:2159–84.
[23] Gomez M-A, Reus M, Parmar N, Travassos B. Exploring elite soccer teams' performances during different match-status periods of close matches' comebacks. Chaos Solitons Fractals 2020;132:109566.
[24] Clemente FM, Sarmento H, Aquino R. Player position relationships with centrality in the passing network of world cup soccer teams: win/loss match comparisons. Chaos Solitons Fractals 2020;133:109625.
[25] Buldu JM, Busquets J, Echegoyen I, Seirullo F. Defining a historic football team: using network science to analyze Guardiola's FC Barcelona. Scient Rep 2019;9.
[26] Buldu JM, Busquets J, Martinez JH, Herrera-Diestra JL, Echegoyen I, Galeano J, et al. Using network science to analyse football passing networks: dynamics, space, time, and the multilayer nature of the game. Front Psychol 2018;9:1900.

[27] Sarmento H, Clemente FM, Araujo D, Davids K, McRobert A, Figueiredo A. What performance analysts need to know about research trends in association football (2012-2016): a systematic review. Sports Med 2018;48:799–836.
[28] Memmert D, Lemmink K, Sampaio J. Current approaches to tactical performance analyses in soccer using position data. Sports Med 2017;47:1–10.
[29] Dutt-Mazumder A, Button C, Robins A, Bartlett R. Neural network modelling and dynamical system theory: are they relevant to study the governing dynamics of association football players? Sports Med 2011;41:1003–17.
[30] Roesch D, Hodgson R, Peterson L, Graf-Baumann T, Junge A, Chomiak J, et al. Assessment and evaluation of football performance. Am J Sports Med 2000;28:S29–39.
[31] Folgado H, Duarte R, Marques P, Gonçalves B, Sampaio J. Exploring how movement synchronization is related to match outcome in elite professional football. Sci Med Football 2018:1–7.
[32] Sampaio J, Maças V. Measuring tactical behaviour in football. Int J Sports Med 2012;33:395–401.
[33] Pastor-Vicedo JC, Contreras-JordÁN O, Prieto-Ayuso A. Performance indicators as a resource for the selection of talented football players. J Hum Sport Exer 2017;12:S797–806.
[34] Nagy M, Akos Z, Biro D, Vicsek T. Hierarchical group dynamics in pigeon flocks. Nature 2010;464:890–3.
[35] Linke D, Link D, Lames M. Football-specific validity of TRACAB's optical video tracking systems. Plos One 2020;15:e0230179.
[36] HHS.gov. Attachment A: human subjects research implications of "big data" studies. Office for Human Research Protections. Secretary's Advisory Committee on Human Research Protections: U.S. Department of Health & Human Services; 2015. https://www.hhs.gov/ohrp/sachrp-committee/recommendations/2015-april-24-attachment-a/index.html.
[37] Cumming G, Calin-Jageman R. Introduction to the new statistics: estimation, open science, and beyond. New York: Routledge; 2016.
[38] Herm S, Callsen-Bracker H, Kreis H. When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. Sport Manag Rev 2014;17:484–92.
[39] He M, Cachucho R, Knobbe A. Football player's performance and market value. In: Proceedings of the 2nd workshop of sports analytics, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD). Porto: ECML PKDD; 2015. p. 1–10.
[40] Gerhards J, Mutz M. Who wins the championship? Market value and team composition as predictors of success in the top European football leagues. Eur Soc 2017;19:223–42.
[41] Müller O, Simons A, Weinmann M. Beyond crowd judgments: data–driven estimation of market value in association football. Eur J Oper Res 2017;263:611–24.
[42] Gonçalves B, Figueira B, Maças V, Sampaio J. Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. J Sports Sci 2014;32:191–9.
[43] Kannekens R, Elferink-Gemser MT, Visscher C. Positioning and deciding: key factors for talent development in soccer. Scand J Med Sci Sports 2011;21:846–52.
[44] Seidl T, Cherukumudi A, Hartnett A, Carr P, Lucey P. Bhostgusters: realtime interactive play sketching with synthesized NBA defenses. In: MIT SLOAN sports analytic conference. Boston: 42 Analytics; 2018. p. 1–13.
[45] Sumpter D. Soccermatics: Mathematical Adventures in the Beautiful Game. London: Bloomsbury Publishing; 2016.
[46] Harrop K, Nevill A. Performance indicators that predict success in an English professional League One soccer team. Int J Perform Anal Sport 2014;14:907–20.
[47] Kite C, Nevill A. The predictors and determinants of inter-seasonal success in a professional soccer team. J Hum Kinetics 2017;58:157–67.
[48] Cummins C, Orr R, O'Connor H, West C. Global positioning systems (GPS) and microtechnology sensors in team sports: a systematic review. Sports Med 2013;43:1025–42.
[49] Wong D, Hjelde G, Cheng C, Ngo J. Use of the Rsa/Rcod index to identify training priority in soccer players. J Strength Condition Res 2015;29:2787–2793.
[50] Mackenzie R, Cushion C. Performance analysis in football: a critical review and implications for future research. J Sports Sci 2013;31:639–76.