Ball possession dynamics in the game of Football: From data analysis to modeling

A. Chacoma,1 N. Almeira,1,2 J.I. Perotti,1 and O.V. Billoni1,2

1Instituto de Física Enrique Gaviola (IFEG-CONICET).
2Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba.

Collective dynamics studies of team sports such as football is a hard task involving large amounts of data and analysis. In this contribution we aim to address a particular subject of this broad field such as the dynamics of the ball possession intervals. To this end we have analyzed a novel football database comprising one season of the five major leagues of Europe. First we obtained the key elements of ball possession dynamics from a statistical analysis of the database. Using this input we developed a simple stochastic model based on three agents, two teammates and one defender. This model includes four parameters and can capture the main emergent statistical observables of the ball possession intervals in the database: the distribution of (i) possession time, (ii) number of passes performed, and (iii) the ball distance traveled on passes. In the last part, we show that the dynamics of the model, can be mapped into a Wiener process with a drift and an absorbing barrier.

I. INTRODUCTION

The statistical analysis of competing games based on data gathered from professional competitions is currently a growing area of research [1-8]. In the case of team sport games, these studies have a potentially high impact; which is boosted by commercial interests but also by its intrinsic complexity that caught the attention of basic research [9-12]. During the course of team sports games, complex behaviors are often observed which arise from individual or collective strategies. The complexity of these emerging behaviors is due to the interplay of dynamical processes governed at several well–differentiated spatio–temporal scales. In fact, it is known that these scales are important for both individual interactions among athletes and collective strategies [9].

Particularly interesting is the game of Football, where data analytics have been successfully tackled in the recent years [10-12]. For instance, in the field of complex systems, J. Buldú et al. used network tools in Ref. [13] to analyze Guardiola’s F.C. Barcelona performance. In that work, they consider a team as an organized social system and—in the frame of network theory—players are considered as nodes linked during the game through coordinated interactions.

Despite these recent and interesting contributions, Football analytics seems to be relegated as compared to other major team sports, for instance, Basketball or Baseball. That is why Football’s team management and strategy is far from being recognised as analytics driven. The specific problem with Football is concerned with strategy and player evaluation than off–ball dynamics in Football. The state of art in this kind of analysis is that—in spite of the plethora of articles present in the literature on this subject—the underlying mechanisms behind the statistical observables in the game of Football remain still undiscovered. Moreover, the modeling at microscopic level of the dynamics of Football’s games have been rarely proposed and there are few reports in this sense [13-17].

In the present contribution we have surveyed a novel database [18], collecting and analysing its most relevant information. Then using the input of these results we have devised an innovative football model that captures several of the most characteristic features of this game dynamics. In particular, our goals are focus on: (i) model the dynamics in the frame of both on–ball and off–ball actions, (ii) study the dynamic of the Ball Possession Intervals (BPI), defined as the consecutive series of actions carried out by a team. The last point is considered a key feature to understand team’s collective performances [19, 20].

The paper is organized as follow, in section Methods, we firstly introduce the database. Particularly, we describe the dataset Events, as well as other information regarding relevant fields. Secondly, we give a formal definition of the model and discuss in detail the key elements, the assumptions and the dynamical parameters. Thirdly, we present a method to systematically seek a suitable set of parameters for the model.

In section Results, we firstly analyse the empirical data extracted from the database, obtaining what we considered the most relevant features of the game. Secondly, we evaluate the outcomes of the model by performing numerical simulation and comparing its predictions with some of the results observed in the database. Thirdly, we place our model in a theoretical framework. This allows, under certain approximations, an interpretation of the emergent spatio–temporal dynamics of the model. Our results are briefly summarized in the last section.

*aachacoma@famaf.unc.edu.ar
II. MATERIAL AND METHODS

A. The dataset

In 2019, L. Pappalardo et al. published one of the largest football–soccer database ever released\cite{18}. Among the information provided in this astounding work, there is a dataset (Events) that gathers all the relevant spatio–temporal events taking place in football matches. This dataset contains the events for each match played in one season (2017/2018) of the following five professional football leagues in Europe: Spain, Italy, England, Germany and France. A typical event registered in this dataset (DS) bears information on:

- **Type of event.** Namely, pass, duels, free kicks, fouls, etc, subdivided into other useful subcategories. This field allows to evaluate in detail the correlation between particular actions and the consequences in the dynamics.

- **Spatio–temporal data.** Each event is tagged with temporal information referred to the match period and the time in seconds elapsed. Spatial data, in turn, is referred to the dimension of the stadium, as a percentage of the field length from the view of the attacking team.

- **Unique identifications.** Each event in the DS is linked to an individual player in a particular team. This allows us to accurately determine the BPIs, and moreover to perform a statistical analysis of the players involved.

In the light of this information, we define a BPI as the set of consecutive events generated by the same team. We gathered 3071395 events and 625195 BPIs from the DS; totalizing 1826 matches, involving 98 teams, and with the participation of 2569 different players. Since we aim to study a dynamical evolution, only BPIs with two or more events were collected. On the other hand, regarding spatial data, matches are played in different stadiums having in turn different sizes. Then in order to compare distances of different matches we have normalized all the distances in a match to the average distance calculated using the whole set of measures in that match. By doing this, we were able to compare and analyze distances from matches played on stadiums with different sizes.

B. The model

We aim to build a minimal model for the dynamics of the game of football, such that in every realization it emulates the dynamical evolution of a single BPI.

Let us think in a system with three agents (the players), two in the same team (the teammates), and one in the other (the defender). The players in this system are able to move in two dimension, and in particular the teammates are able to pass the ball to each other. In this minimal simulated game, the match evolves until the defender reaches the player with the ball and, emulating a Duel, ends the BPI. Bearing these ideas in mind, in the following we propose the rules that govern the motion of the agents, and consequently define the model’s dynamics.

Let $\mathbf{r}_i(t)$ be a 2D position vector for an agent $i$ ($i = 1, 2, 3$) at time $t$. Considering discrete time steps $\Delta t = 1$, at $t+1$ the agents will move as $\mathbf{r}_i(t+1) = \mathbf{r}_i(t) + \delta \mathbf{r}_i(t)$. In our model, we propose $\delta \mathbf{r}_i(t) = (R \cos \Theta, R \sin \Theta)$, where $R$ and $\Theta$ are two variables taken as follow,

1. **The displacement $R$**

   The three agents randomly draw a displacement from an exponential distribution $P_a(r) = \frac{1}{a} e^{-r/a}$, where $a$, the scale of the distribution, is the agent’s action radius (see Fig. 1A), i.e. the surroundings that each player controls.

2. **The direction $\Theta$**

   (a) **For the teammates.** They randomly draw an angle in $[0, 2\pi)$ from an uniform distribution.

   (b) **For the defender.** This agent takes the direction of the action line between itself and the agent with the ball.

   Then, according to the roles in the game, the players decide to accept the changes proposed as follows,

3. **The player with the ball** evaluates if the proposed displacement moves it away from the defender. If it does, the player changes the position; otherwise, it remains in the current position.

4. **The free player and the defender** always accept the change.

   As we mentioned before, in this model we consider the possibility that the teammates perform passes to each other. This decision is taken as follow,

5. If the defender’s action radius does not intercept the imaginary line joining the teammates; then the player with the ball plays a pass to the other teammate with probability $p$.

   Since in real football games the player’s movements are confined, for instance, by the field limits, in the model we introduce two parameters for these confinements: The inner and external radius, $R_1$ and $R_2$, respectively, (see Fig. 1B).

6. The inner radius $R_1$ is used to set the initial conditions. At $t = 0$, each one of the three agents is put at a distance $R_1$ from the center of the field, spaced with an angular separation of 120 degrees (maximum possible distance between each other).
7. The external radius $R_2$ defines the size of the field, setting the bounds for the simulation. If an agent proposes a new position $\vec{x}(t+1)$ out of the area $||\vec{x}(t+1)|| < R_2$, the change is forbidden, and the agent keeps its current position – note this overrules the decision taken from (3) and (4).

Lastly, a single realization of the model in the frame of the rules proposed above ends when

8. The defender invades the action radius of the agent with the ball; thus, if the distance $d$ between the player with the ball and defender satisfies $d < a$ the realization ends.

Let us provide a justification for the chosen rules and the different elements of the model. The explanation is broken down into four main points.

Firstly, it is well known that the game of football is complex. Fig. 2 (A) shows that many events are possible in the context of a BPI. However, we can see that the events *Pass* and *Duels* domain in the frequency of common events, and the events triggering a BPC, respectively. Therefore, a reasonable simplification is to propose a model with only two possible events, this is also in agreement with the data shown in the inset of Fig. 2 (B), i.e. regarding the number of different type of event observed during a BPI.

Secondly, while the game of football is played by twenty two players, considering only three player for the model could be criticized as an over simplification. However, as we show in the main plot of Fig. 2 (B), we can see that the number of players by BPI is mostly two. Therefore, a system with two teammates and a single defender triggering the BPCs seems to be a good approach.

Thirdly, let us discuss the players movement rules. In item (1) (see listing above), we propose the displacements are taken from an exponential distribution, with an action radius $a$ as the scale. The idea behind this, is to set a memoryless distribution, in the light that the players’ displacement are commonly related with both evasion and distraction maneuvers, which are more effective without a clear pattern [21]. The direction and the adoption of the new movement, on the other hand, is proposed as role-dependent. The player with the ball takes a random direction and adopts the movement if the displacement moves it away from the defender, otherwise it stays on the current position. The idea here is to slow down the player movement, since it is well know that the players on ball control, are slower than free players. The free player, on the other hand, follows a random walk. In this regards, we are trying to put into the model the possibility of performing passes of different lengths. The defender’s main role, in turn, is to capture the player with the ball, therefore rule (1.b) is straightforward.

Lastly, the incorporation of the confinement boundaries $R_1$ and $R_2$, is because the development of football matches takes place inside confined spaces. In particular $R_1$, brings into the model the possibility of capture short BPI, emulating plays into very reduced spaces, as, for instance, fast attacks. The incorporation of $R_2$, on the other hand, is straightforward, since the real football fields are not limitless. The main difference between the real and the model field’s bounds is the shape. In this regards, we neglect any possible contribution of the geometry of the field.

For further details on the model’s outcomes as a function of both different sets of parameters and alternative-testing rules, see Appendixes A and B. To summarize, the main parameters that rule the model dynamics are (i) the action radius $a$, (ii) the probability of performing a free pass $p$, and (iii) the confinement radius $R_1$ and $R_2$.

In the following section, we propose a simple method for tuning these parameters.

### C. On setting the model’s parameters

The model allows a certain degree of flexibility in the outcomes by using different combination of the parameters $a$, $p$, $R_1$ and $R_2$. For the sake of simplicity, we decided first to fix $a$, and refer the other radius to this scale, $R_1 \rightarrow \frac{R_1}{a}$ and $R_2 \rightarrow \frac{R_2}{a}$. For the other parameters, we designed a method for tuning the model based on the comparison between the real data and the model results. The method follows the algorithm below,

1. Propose a set of parameters $\rho = (p, R_1, R_2)$;
2. Perform $10^5$ realization, calculate $P(T), P(N)$ and $P(\Delta r, X)$
3. Compare the distributions obtained in step 2 with the real data, using the Jensen–Shannon divergence ($D_{JS}$, see definition in Ref. [22]), which is a metric to measure the similarity between two probability distribution.
4. Propose a new set of parameters $\rho$, seeking to lower $D_{JS}$.
5. Back to step 2, and repeat until find the lower similarity.

Notice, this method was not devised to perform a non-linear fit of the data, but to systematize the search for realistic parameters, avoiding the unrealistic ones. For further details on the model’s outcomes as a function of different sets of parameters, see Appendix A, Fig. 8

### III. RESULTS

#### A. Data analysis

Here we analyze and describe the most relevant statistical observables in the DS. The idea is to define and support the key elements of our football model for the BPIs' dynamic. Firstly, let us describe in Fig. 3 the statistic...
of the events found in the DS. In panel A, we plot the frequency of events by type (blue bars) and also the frequency of events that trigger a Ball Possession Change (BPC) (see red bars). By looking at the blue bars, we can see that the most common event is the Pass, with 1.56 million entries. Note that passes, almost duplicate the second most frequent type of event which is Duels. In particular, Duels is the most frequent event triggering a BPC (see red bars). Moreover, by comparing the two bars on Duels, we can see that \( \approx 75\% \) of the Duels ends in a BPC, showing that this type of event is very effective to produce a change of possession. On the other hand, the main plot in panel B shows the number of different players involved per BPI. We can see that, the case two players is the most common, with 0.27 million of observations, duplicating the case of three players, the second most common observed. The inset in that panel, shows the number of different type of events per BPI. Here the most frequent case is to find only two different type of events, with 0.4 million of cases recorded.

Secondly, let us inspect relevant spatio–temporal statistics found in the DS summarizing these analysis in Fig.3. Considering a single BPI, we define \( T \) as the total possession time i.e. the duration of the BPI, and \( N \) as the number of passes performed in the BPI. In addition, \( \Delta t \) and \( \Delta r \) as the time interval and the distance travelled between two (any) consecutive events of the BPI, respectively. The main plot in panel A, shows the distribution \( P(T) \) of possession times. In this case, we have measured the mean value in \( \langle T \rangle = 13.72 \text{ s} \). Moreover, a power–law behaviour can be seen in the tail, where \( P(T) \propto T^{-\gamma} \). By performing a non–linear fit, we found \( \gamma = 5.1 \pm 0.1 \), consequently this exponent is somehow large to ensure a power law tail. The inset in that panel, in turn, shows the distribution \( P(\Delta t) \) of time intervals. The same power–law behaviour is observed, which seems to indicate that in both plots, extreme events might not be linked to large values of \( T \), but to large \( \Delta t \), probably due to events such as interruptions in the match or similar. On the other hand, in panel B, we show the distribution \( P(\Delta r) \) of distances travelled between events. In this case, we divided the dataset in order to clearly see the contribution of the event tagged as Passes, since as we show in Fig.2A, these are the most recurrent events in the DS. Let us split \( P(\Delta r) \) as follows, \( P(\Delta r) = P(\Delta r, Y = \text{pass}) + P(\Delta r, Y = \text{other}) \), where \( Y \) stands for the type of event, the first term is the contribution coming from passes, and the second one from any other type of event. Moreover, we are able to divide the event pass, into two sub–types \( P(\Delta r, Y = \text{pass}) = P(\Delta r, Y = \text{simple pass}) + P(\Delta r, Y = \text{other pass}) \), where the first term is the contribution of the sub–type "Simple" Pass and the second is the contribution of any other sub–type (for example High pass, Cross, Launch, etc. c.f. [S] for further details). For the sake of simplicity, hereafter we refer to the type of events pass, simple pass and other pass as \( X, X_2 \) and \( X_3 \), respectively. Notably, we can see a highly significant contribution of the event Pass, and moreover, that the sub–type simple pass captures the features of the distribution having both, the peak at \( \Delta r = 1 \) (the mean value), and the hump around \( \Delta r \approx 3 \). This multimodal behaviour, likewise, might evidence the presence of two preferential distances, from where teammates are more likely to perform passes. Panel C shows the distribution \( P(N) \) of the number of passes per BPI. We observe the presence of a heavy tail at the right, but in this case we are not able to conclude whether this is power law. The mean value, \( \langle N \rangle = 3.1 \), indicates that in average we observe \( \approx 3 \) passes per BPI. In relation to this point, in panel D, we show the relation between the number of passes and the possession time. Interestingly, we observe a linear relation for values within \( 0 < T < 60 \text{ s} \) (see solid blue line in the panel). From our best linear fit in this region, we obtain \( \langle N \rangle(T) = \omega_p T \) with \( \omega_p = 0.19 \pm 0.03 (R^2 = 0.99) \). This parameter can be thought in overall terms as the rate of passes per unit of time. Therefore, we are able to conclude that, within the BPIs, \( \approx 0.2 \) passes per second are performed.

B. Modelling Ball Possession Intervals

In this section, we compare the results of our model with the observed in the dataset. The results are shown in Fig.3 (black solid lines). Panels A, B, C and D show the comparison between the results obtained from the real dataset (DS) (discussed in the previous section) and the model’s outcomes (MO). For the simulations, we used the set of parameters \( (p, a, R_1, R_2) = (0.3, 1, 2.5, 16) \). Although at first glance there is a good agreement between the simulated results and real data, there are points and features that deserve comments.

In order to search for the parameters’ values, and systematically compare the model’s outcomes and the observed in the real dataset, we used the Jensen–Shannon distance, which is a metric to measure the similarity between two probability distributions (see section Methods, On setting the model’s parameters, for further details). If \( D(P,Q)_{JS} \approx 0 \) we say that \( P \) and \( Q \) are similar. Then, for the distribution \( P(T) \) shown in panel A, we obtain \( D_{JS} = 0.017 \), which indicates a good similarity between the DS and the model. However, we observe a shift in the mean of this distribution of \( \approx -20\% \), and a problem to capture *the hump* of the curve around \( T \approx 30 \text{ s} \). The distribution of the traveled distance in passes \( P(\Delta r, X) \) showed in panel B, return a very good similarity \( D_{JS} = 0.008 \). Moreover, we can see the model succeeds in capturing the bi–modality of the distribution, which seems to indicate that the proposed rules are very effective for both nearby and distant passes. In panel C, we show the distribution of the number of passes \( P(N) \). The calculation for the Jensen–Shannon distance gives the value \( D_{JS} = 0.0007 \), which indicates a very good similarity between the curves. For this case, the value of parameter \( p \) seems to be critical. In this regards, note that the chosen value for \( p \) is near to the
rate $\omega_p = 0.19$ passes per second, reported in the previous section. In the relation $\langle N \rangle$ vs. $T$ in panel D, the DS shows that, in average, the number of passes cannot grow indefinitely with the possession time, which is likely a finite-size effect. Our simple model, in turn, allows the unrealistic unbounded growth of $\langle N \rangle$.

Lastly, let us put the parameter values in the context of real football dimensions. In regard to the action radius $a$, the literature includes reported estimations from kinetic and coordination variables [23, 24], where speed measurements [25, 26] show that professional players are able to move in a wide range within $1.1 - 4.8$ m/s. Thus, it would be easy for a professional player to control a radius of $a \approx 2$ m. If we set this value for $a$, we proportionately obtain for the internal and the external radius, the values $R_1 \approx 5$ m and $R_2 \approx 32$ m, respectively. Consequently, in the frame of our model, the BPIs take place into areas within a range of $78$ m$^2$ (approximately a goal area), and $3200$ m$^2$ ($\approx 47\%$ of the Wimbledon Greyhound Stadium). Therefore, the proposed parameters are indeed in the order of magnitude of real football field dimension, and we are able to confirm that the dynamic of the model, is ruled upon a realistic set of values.

C. Mapping the model into a Wiener process with drift and an absorbing barrier

We propose a theoretical framework to understand the distribution $P(T)$, observed from the model’s outcomes. Every realization can be thought as a process where the defender must capture the ball. A ball that, due to the movements and passes performed by the teammates, may follow a complicated path in the plane. However, since the defender always takes the direction towards the ball, the process can be reduced to a series of movements in one dimension. To visualize this mapping we fix the origin of our 1D coordinate system at the ball position and define the coordinate $x$ of the defender as the radial distance $d$ between the ball and the defender. In this frame the defender takes steps back and forth depending whether the radial distance between the ball and defender is increasing or decreasing, respectively. The size of the steps $\Delta d$ of this random walk is variable and the process ends when the coordinate $x$ of the defender is in the interval $(-a, a)$ (c.f. Section Methods, The model, rule 8).

In this process, the length step distribution characterizes the random walk.

Let us define $\delta = \frac{\Delta d}{a}$ as the size of the steps normalized to the initial distance between the players. Then, in Fig. 4A, we plot the distribution $P(\delta)$ analyzing two possible contributions for the steps, (i) the steps taken when the defender follows the player with the ball ($S_1$), (ii) those generated when a pass between teammates occurs ($S_2$). In order to visualize these contributions, we have plotted $P(\delta)$, and the joint probabilities $P(\delta, S_1)$ and $P(\delta, S_2)$, fulfilling $P(\delta) = P(\delta, S_1) + P(\delta, S_2)$. From this perspective, we can clearly see that the steps arising in passes ($S_2$) explain the extreme events, whereas simple steps ($S_1$) explain the peak. On the other hand, if we measure the mean value of both contributions we obtain $\langle \delta \rangle_{P(\delta, S_1)} = -0.14$, $\langle \delta \rangle_{P(\delta, S_2)} = 0.22$, which means that in average, the first contribution brings the defender towards the ball, and the second takes it away. However, notice that the full contribution is negative, $\langle \delta \rangle_{P(\delta)} = -0.07$, which indicates the presence of a drift leading up the defender towards the ball.

It turns, from the previous analysis, that the dynamics of the system can be mapped to a random walk with a drift, in the presence of an absorbing barrier. Moreover, in the approximation where $\delta$ is constant, the process described above is governed by the following Focker–Plank equation,

$$\frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} - \mu \frac{\partial p}{\partial x} = \frac{\partial p}{\partial t}$$

subject to the boundary conditions,

$$p(d_0, x; 0) = \delta(x),$$
$$p(d_0, x_b; t) = 0,$$

where $p(d_0, x, t)$ is the probability of find a walker that starts in $d_0$, in the position $x$ at time $t$. The coefficients $\mu$ and $\sigma$ are the drift and the diffusion, and $x_b$ indicates the position where the absorbing barrier is placed. Additionally, it can be proved that the probability distribution of the first passage time $\tau$, for a walker reaching the barrier, is given by (c.f. ref. [27]),

$$g(\tau) = \frac{x_b}{\sigma \sqrt{2\pi \tau^3}} \exp\left(-\frac{(x_b - \mu \tau)^2}{2\sigma^2 \tau}\right),$$

which can be straightforward linked to distribution $P(T)$.

In this theoretical framework, we used eq. (2) to perform a non-linear fit of $P(T)$, via the parameters $\mu$ and $\sigma$. We set $x_b = a$, the action radius can be thought as the barrier’s position. The result presented in Fig. 4B, shows the fitting is statistically highly significant, yielding a correlation coefficient $r^2 = 0.97$, with $\mu = 0.09 \pm 0.02$ and $\sigma = 0.39 \pm 0.03$. Moreover, notice that we achieve a very good agreement between the drift value and $\langle \delta \rangle_{P(\delta)}$, in magnitude. Therefore, one can conclude that, in the context of the model, a random walk with constant $\delta$ and drift $\mu$, is a good approximation for a walker drawing steps from $\langle \delta \rangle_{P(\delta)}$. Furthermore, this approximation explains the long tail observed in $P(T)$ for both, the outcomes of the model and the empirical observations.

IV. DISCUSSION

In this section, we briefly comment on the results that in our view are the most relevant contributions of this work.

We performed a data-driven statistical study of football matches. In particular, we have focused in the analysis and modeling of the intervals in which each team
has possession of the ball, defining the so-called Ball Possession Intervals (BPI). In the first part we worked on the characterization of BPIs, studying its main elements from an extensive dataset that compiles most of the events during the course of the matches. In this regards, we were able to identify four key elements in BPI that we later use to build a football model. Namely, we extracted the most frequent type of (i) events and (ii) events leading to a possession change, then we identified during a BPI (iii) the number of players participating and (iv) number of different types of events. We complete the analysis of the dataset measuring several important distributions regarding BPI, such as the time duration distribution of BPI, \( P(T) \), the distribution of the number of events in BPI, \( P(N) \), and the distribution of the length of the passes \( P(\Delta r, X) \), among others.

Using these findings, we proposed a minimal model made up of two teammates and a single defender that, following simple motion rules, emulates both on-ball and off-ball actions. We showed that the model can be tuned by setting four parameters \( a, p, R_1 \) and \( R_2 \), which control the action radius, the probability for making a pass, and the internal and external confinement radius, respectively. In order to set the parameters, we have introduced a simple method, based on the evaluation of the Jensen–Shannon distances, which allows us to fit the proper set of dynamical rules.

In the last part of our work, we map the model’s dynamics to a one-dimensional random walk in which the ball is fixed at the origin and the defender moves taking non-uniform steps of length \( \delta \). We showed that since on average \( \langle \delta \rangle_{P(\delta)} < 0 \) holds, then the defender moves following a preferential direction towards the ball. Then we are able to describe the dynamics of our model using the theoretical framework of a Wiener process with a drift and an absorbing barrier. In particular, we used the first passage time expression of this process to fit the distribution \( P(T) \) of the model, finding a very good agreement.

We conclude that BPI’s process in football matches is the result of a reduced set of players. The statistic of this process can be well reproduced using a dynamical toy model with three agents. Furthermore, the model can be mapped into a random walk with a drift and a trapping barrier. Finally, although the remarkable performance of the model, we point out that it would be interesting to study the dynamics, not for the internal ball possession intervals, but for the temporal interaction among them. If any correlation is there, the statistical emergent observables from this would give useful insights to propose a new model from where to study possible different strategies in the context of competition between teams. In this regards, we let the door open to future projects.

Appendix A: Varying the values of parameters \( p, R_1 \) and \( R_2 \)

In this section we assess the effect of systematically change the values of the dynamical parameters. As we have previously stated, the model’s outcomes depend on the set \( \{a, p, R_1, R_2\} \). Moreover, for the sake of simplicity, we fix \( a = 1 \), and refer the other radius to this scale, such that \( R_1 \rightarrow \frac{R_1}{a} \) and \( R_2 \rightarrow \frac{R_2}{a} \). In order to show the results in a straightforward manner, figures 6 and 7 follow the same pattern: (i) panels A, B and C show the effect on distributions \( P(T), P(\Delta r, X), \) and \( P(N) \), respectively, and (ii) coloured dots and black solid line in the background, show the results presented in Fig. 3 helping the reader to identify and compare among cases.

In Fig. 6 we show the effect on change \( p \). The idea is analyse the limits \( p \approx 0 \) and \( p \approx 1 \). As we can see, small values of \( p \) produce the reduction of the number of passes per ball possession interval (BPI); as well as the reduction of both, long possession times and long distance passes. In the latter case, the "hump" at \( \Delta r = 3 \) vanishes. In Fig. 7 we show the result on change \( R_1 \). As we can see, when this parameter is increased, the model becomes not able to reproduce short BPIs, see panel A at \( T \approx 0 \). Lastly, in Fig. 7 we show the result of change \( R_2 \). In our frame, large values means larger external bounds for the system. Clearly the number of large BPIs increases. Consequently, the model is not able to capture the behaviour at the distribution tails.

On the other hand, in the last part of section Methods, in the main text, we describe a simple algorithm to search for a set of parameters \( \rho \) such that the model's outcomes (MO) "fit" the observed in the dataset (DS). As we have said, this method is not a no-liner fit but a recipe to find a suitable set of parameters, in order to frame the result of the model in the context of real observables. To do so, we use the Jensen–Shannon metric \( D_{JS}(P,Q) \) to define a distance between the probability distributions observed in the DS and the results obtained from the model.

Let be \( P(\Delta r, X)_{DS}, P(N)_{DS}, P(\Delta r, X)_{MO} \) and \( P(\Delta r, X)_{MO} \); the probability distribution of \( T, N, \Delta r \) (Passes), obtained from the DS and MO, respectively. We define the distances \( D_{JS}^T = D_{JS}(P(T)_{MO}, P(T)_{DS}), D_{JS}^N = D_{JS}(P(N)_{DS}, P(N)_{MO}), \) and \( D_{JS}^{\Delta r} = D_{JS}(P(\Delta r, X)_{DS}, P(\Delta r, X)_{MO}) \). In this frame, we set \( R_2 = 16 \) and carried out simulations varying the values of \( p \in [0.1, 0.9] \) and \( R_1 \in [1, 5] \). Based on this outcomes, we calculated \( D_{JS}^T, D_{JS}^N, D_{JS}^{\Delta r} \) for the whole set of cases. The results are shown in the maps of Fig. 8 A, B and C. Panel D, in turn, show the sum of the three contribution, \( D_{JS}^{\rho} = D_{JS}^T + D_{JS}^N + D_{JS}^{\Delta r} \). Notably, from panel D, we can clearly see a region where the \( D_{JS}^{\rho} \) is minimum, and consequently a set of parameters suitable for our propose. Finally, is important to note that these values work for \( R_2 = 16 \), for alternative values of \( R_2 \) we can follow the same procedure and search other set of parameters that minimises the sum.
Appendix B: The effect of rules changes

When we propose the model’s rules, it was based on statistical observations made in the context of a data driven analysis. However, we had to made some assumption regarding the players motion and manoeuvres strategies. The idea in this section is to evaluate these assumptions by contrasting the model’s motion rules with alternative ones.

Firstly, let us evaluate the effect of change the manner upon the players choose a displacement $R$. The model propose the agents raw a value from an exponential distribution $P_a(r) = \frac{1}{a} e^{-r/a}$, where $a$ is the action radius (c.f. main text, section Methods, The model, rule 1). Here, we propose two alternative manners: (i) the agents raw a fixed (FX) displacement $R = a$, and (ii) they raw $R$ from an uniform distribution (UD) within the interval $[0, 3]$. The results are shown in Fig. 5. We can see that both random options, (UD) and (MO), agree with the observe in the DS, whereas the deterministic option (FX) seems to fail, particularly to capture the multimodal behaviour of $P(\Delta r, X)$. Therefore, we can conclude that an stochastic variable to describe the agents motions, is highly relevant into the foundations of the model.

Secondly, let us evaluate different strategies for agents movements. Focus on the agents without the ball, the model propose they move according to roles. The defender moves always towards the ball (TB), and the free player moves follow a random walk (RW). We propose to analyse two alternative strategies: the defender and the free player, both move (i) follow a RW, and (ii) always TB. The results are shown in Fig. 10. When they both follow a RW, because the absence of interaction, long possession times shows up in $P(T)$, as well as a high number of passes per BPI. On the other hand, when both move TB, the possession time becomes shorted as well as number of passes per BPI, moreover the system loose the multimodal behaviour of $P(\Delta r, X)$. Consequently, both strategies fail. Which seems to indicate that role– dependent strategies are needed into the model, and the agents movement rules can not be reduce to an unique rule for all of them in the system.

ACKNOWLEDGMENTS

This work was partially supported by grants from CONICET (PIP 112 20150 10028), FonCyT (PICT-2017-0973), SeCyT–UNC (Argentina) and MinCyT Córdoba (PID PGC 2018).

[1] Alexander M Petersen and Orion Penner. Renormalizing individual performance metrics for cultural heritage management of sports records. Chaos, Solitons & Fractals, 136:109821, 2020.
[2] Tal Neiman and Yonatan Loewenstein. Reinforcement learning in professional basketball players. Nature communications, 2(1):1–8, 2011.
[3] Satyam Mukherjee, Yun Huang, Julia Neidhardt, Brian Uzzi, and Noshir Contractor. Prior shared success predicts victory in team competitions. Nature human behaviour, 3(1):74–81, 2019.
[4] Sears Merritt and Aaron Clauset. Environmental structure and competitive scoring advantages in team competitions. Scientific reports, 3:3067, 2013.
[5] Radivoj Mandić, Sasa Jakovljević, Frane Erčulj, and Erik Štrumbelj. Trends in nba and euroleague basketball: Analysis and comparison of statistical data from 2000 to 2017. PloS one, 14(10), 2019.
[6] Juan I Perotti, H-H Jo, Ana L Schaigorodsky, and Orlando V Billoni. Innovation and nested preferential growth in chess playing behavior. EPL (Europhysics Letters), 104(4):48005, 2013.
[7] Ana L Schaigorodsky, Juan I Perotti, and Orlando V Billoni. Memory and long-range correlations in chess games. Physica A: Statistical Mechanics and its Applications, 394:304–311, 2014.
[8] Nahuel Almeira, Ana L Schaigorodsky, Juan I Perotti, and Orlando V Billoni. Structure constrained by metadata in networks of chess players. Scientific reports, 7(1):1–10, 2017.
[9] Felix Lebed and Michael Bar-Eli. Complexity and control in team sports: Dialectics in contesting human systems, volume 6. Routledge, 2013.
[10] Antônio M Lopes and JA Tenreiro Machado. Entropy analysis of soccer dynamics. Entropy, 21(2):187, 2019.
[11] Alessio Rossi, Luca Pappalardo, Paolo Cintia, F Marcello Iaia, Javier Fernández, and Daniel Medina. Effective injury forecasting in soccer with gps training data and machine learning. PloS one, 13(7), 2018.
[12] Lotte Bransen, Jan Van Haaren, and Michel van de Velden. Measuring soccer players’ contributions to chance creation by valuing their passes. Journal of Quantitative Analysis in Sports, 15(2):97–116, 2019.
[13] Javier M Buldu, J Busquets, Ignacio Echegoyen, et al. Defining a historic football team: Using network science to analyze guardiola’s fc barcelona. Scientific reports, 9(1):1–14, 2019.
[14] Claudio A Casal, Rubén Maneiro, Toni Ardá, Francisco J Marí, and José L Losada. Possession zone as a performance indicator in football. the game of the best teams. Frontiers in psychology, 8:1176, 2017.
[15] Ken Yamamoto and Takuma Narizuka. Examination of markov-chain approximation in football games based on time evolution of ball-passing networks. Physical Review E, 98(5):052314, 2018.
[16] Andrew H Hunter, Michael J Angilletta Jr, Theodore Pavlic, Glen Lichtwark, and Robbie S Wilson. Modeling the two-dimensional accuracy of soccer kicks. Journal of biomechanics, 72:159–166, 2018.
[17] Ali Cakmak, Ali Uzun, and Emrullah Delibas. Computational modeling of pass effectiveness in soccer. Advances in Complex Systems, 21(03n04):1850010, 2018.
[18] Luca Pappalardo, Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi, and Fosca Giannotti. A public data set of spatio-temporal match events in soccer competitions. *Scientific data*, 6(1):1–15, 2019.

[19] José Gama, Gonçalo Dias, Micael Couceiro, Tiago Sousa, and Vasco Vaz. Networks metrics and ball possession in professional football. *Complexity*, 21(S2):342–354, 2016.

[20] Joachim Gudmundsson and Michael Horton. Spatio-temporal analysis of team sports. *ACM Computing Surveys (CSUR)*, 50(2):1–34, 2017.

[21] Jonathan Bloomfield, Remco Polman, and Peter O’Donoghue. Physical demands of different positions in fa premier league soccer. *Journal of sports science & medicine*, 6(1):63, 2007.

[22] Andrew KC Wong and Manlai You. Entropy and distance of random graphs with application to structural pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(5):599–609, 1985.

[23] Wl Schöllhorn. Coordination dynamics and its consequences on sports. *International Journal of Computer Science in Sport*, 2(2):40–46, 2003.

[24] M Lames, J Ertmer, and F Walter. Oscillations in football—order and disorder in spatial interactions between the two teams. *International Journal of Sport Psychology*, 41(4):85, 2010.

[25] T Little and A C Williams. Specificity of acceleration, maximum speed, and agility in professional soccer players. *Journal of Strength and Conditioning Research*, 19:76–78, 2005.

[26] Irineu Loturco, Lucas A Pereira, Tomas T Freitas, Pedro E Alcaraz, Vinicius Zanetti, Chris Bishop, and Ian Jeffreys. Maximum acceleration performance of professional soccer players in linear sprints: Is there a direct connection with change-of-direction ability? *PloS one*, 14(5), 2019.

[27] D R Cox and H D Miller. The theory of stochastic processes, volume 134. Chapman & Hall/CRC press, 1980.
FIG. 1. Scheme summarizing the main parameters of the model (not to scale). Green circles represent the teammates, orange circle to the defender (A) We emphasize on the parameter: (i) \( d \), distance between the player with the ball and the defender, (ii) \( d_x \), distance between the player with the ball and the free player, and (iii) \( a \), action radius. (B) The circles placed at distance \( R_1 \) from the origin, represent the initial condition in the dynamics. Distance \( d_0 \), is the initial distance between the three agents. Radius \( R_2 \), delimits the border for the agents.

FIG. 2. Relevant statistical insight gathered from the dataset Events in ref. [18]. (A) Frequency by type of event. Blue bars, from the set of all the events. Red bars, only the events triggering a ball possession change (BPC). (B) The main plot shows the number of different players involve in a ball possession interval (BPI). The inset shows the number of different type of events in a BPI. (*) The acronym OOTB, stands for Others on the ball.
FIG. 3. Relevant statistical observables found in the dataset Events in ref. [18] (DS), compared with the model outcomes (MO). For the results shown in the four panels, we have set the parameters of the model with the values $a = 1$, $p = 0.3$, $R_1 = 2.25$ and $R_2 = 16$. (A) The main plot, shows the distribution of the total time of possession $T$ (red dot circles), whereas the inset shows the distribution of the time differences between two consecutive events, $P(\Delta t)$. (B) Distribution of the distance between two consecutive events, segmented in the groups (i) the whole set of events $P(\Delta r)$, (ii) the passes tagging as sub–type simple pass $P(\Delta r, X_1)$, (iii) the passes tagging with any other sub–type $P(\Delta r, X_2)$, and (iv) all the passes $P(\Delta r, X)$. Notice, the plot is in linear– log scale. (C) Distribution of the number of passes in the ball possession intervals, $P(N)$. (D) Mean value of the number of passes, as a function of the total time of possession. Solid blue line, indicates a linear fit performed on this region, $\langle N \rangle = \omega_p T$, with $\omega_p = 0.19 \pm 0.03$ (1/s).
FIG. 4. On the results of mapping the model in a Wiener process with drift and an absorbing barrier. (A) distribution of steps $\delta$, segmented in (i) all the data, $P(\delta)$ (dashed black line), (ii) those steps given in the context of a simple persecution, $P(\delta, S_1)$, and (iii) those steps in the context of a pass, $P(\delta, S_2)$. (B) non linear fit (red solid line) performed to distribution $P(T)$ (model), by using eq. 2.

FIG. 5. The effect of vary the parameter $p$, the probability of perform a pass, on distributions (A) $P(T)$, (B) $P(\Delta r, X)$, and (C) $P(N)$. Dots in the background, and solid black line ($p = 0.3$), help the reader to visualise the changes with respect to the DS analysis and the main results of the model (c.f. Fig. 3).
FIG. 6. The effect of vary the parameter $R_1$, the inner radius, on distributions (A) $P(T)$, (B) $P(\Delta r, X)$, and (C) $P(N)$. Dots in the background, and solid black line ($R_1 = 2.25$), help the reader to visualise the changes with respect to the DS analysis and the main results of the model (c.f. Fig. 3).

FIG. 7. The effect of vary the parameter $R_2$, the external radius, on distributions (A) $P(T)$, (B) $P(\Delta r, X)$, and (C) $P(N)$. Dots in the background, and solid black line ($R_2 = 16$), help the reader to visualise the changes with respect to the DS analysis and the main results of the model (c.f. Fig. 3).
FIG. 8. Jensen Shannon distance between the distributions measured from DS and MO. For the simulations, we set \( R_2 = 16 \) and varied the values of \( p \in [0.1, 0.9] \) and \( R_1 \in [1, 5] \). The maps indicate (A) \( D_{JS}^T \), (B) \( D_{JS}^S \), (C) \( D_{JS}^{\Delta r} \), and (D) \( D_{JS}^{\text{tot}} \). Colours in the map indicate the value of \( D_{JS} \), as is refereed in the colour bar at the right of D. On panels A, B, C, the vertical axis is labelled on panel B.

FIG. 9. Testing alternative manners of raw a new displacement \( \delta r \) (see Section IIB, rule 1): (i) from a fixed value (\( FX \)) given by the action radius \( a \), (ii) from an uniform distribution (\( UD \)) within the interval \((0, 3]\). The plots show the effect of the alternative strategies on distributions (A) \( P(T) \), (B) \( P(\Delta r, X) \), and (C) \( P(N) \). Dots in the background, and solid black line (MO), help the reader to visualise the changes with respect to the DS analysis and the main results of the model (c.f. Fig. 3).
FIG. 10. Assessing alternatives strategies for players motion: (i) the defender moves follow a random walk (RW), (ii) the free player moves toward the ball (TB). The plots show the effect of the alternative strategies on distributions (A) $P(T)$, (B) $P(\Delta r, X)$, and (C) $P(N)$. Dots in the background, and solid black line (MO), help the reader to visualise the changes with respect to the DS analysis and the main results of the model (c.f. Fig. 3).