

PAPER • OPEN ACCESS

Simple mechanism rules the dynamics of volleyball

To cite this article: Andrés Chacoma and Orlando V Billoni 2022 *J. Phys. Complex.* **3** 035006

View the [article online](#) for updates and enhancements.

You may also like

- [On the Practical Application of Computer Technology in the Field of Volleyball Training in China](#)
Liwei Rao
- [Vortex-induced dynamic loads on a non-spinning volleyball](#)
Wei Qing-ding, Lin Rong-sheng and Liu Zhi-jie
- [Physics of knuckleballs](#)
Baptiste Darbois Texier, Caroline Cohen, David Quéré et al.

OPEN ACCESS



PAPER

Simple mechanism rules the dynamics of volleyball

RECEIVED

15 March 2022

REVISED

17 August 2022

ACCEPTED FOR PUBLICATION

31 August 2022

PUBLISHED

15 September 2022

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Andrés Chacoma^{1,2,*}  and Orlando V Billoni^{1,2}¹ Instituto de Física Enrique Gaviola (IFEG-CONICET), Ciudad Universitaria, 5000 Córdoba, Argentina² Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Ciudad Universitaria, 5000 Córdoba, Argentina

* Author to whom any correspondence should be addressed.

E-mail: achacoma@famaf.unc.edu.ar**Keywords:** complex systems, self-organized systems, stochastic processesSupplementary material for this article is available [online](#)**Abstract**

In volleyball games, we define a rally as the succession of events observed since the ball is served until one of the two teams on the court scores the point. In this process, athletes evolve in response to physical and information constraints, spanning several spatiotemporal scales and interplaying co-adaptively with the environment. Aiming to study the emergence of complexity in this system, we carried out a study focused on three steps: data collection, data analysis, and modeling. First, we collected data from 20 high-level professional volleyball games. Then we conducted a data-driven analysis from where we identified fundamental insights that we used to define a parsimonious stochastic model for the dynamics of the game. On these bases, we show that it is possible to give a closed-form expression for the probability that the players perform n hits in a rally using only two stochastic variables. Our results fully agree with the empirical observations and represent a new advance in the comprehension of team-sports competition complexity and dynamics.

1. Introduction

The engaging nature of the complexity related to team sport competition has historically captured the attention of academics [1–12]. In recent years, the interest in studying this area has increased. Promoted by the demands of the thriving global sports market, the research community is now seeing the study of the phenomena related to these systems as challenges to face in the context of a hyper-competitive twenty-one-century industry [13]. However, most of the papers available in the literature use empirical data to perform statistical analysis aiming to generate key performance indicators [14–19]. This approach can be handy for coaches using data-driven based training systems, but it usually does not provide a deeper understanding of the phenomena. From an alternative perspective, our view focuses on use empirical data to uncovering the mechanisms that explain the emergence of global observables; based on a multidisciplinary framework that connects physics with complexity in the social sciences [20]. In previous works, we have successfully studied an modeled several aspects of the game of football [21, 22]. In this paper, we report a strikingly simple mechanism that governs the dynamics of volleyball.

To place the complexity of this sport in context, let us briefly discuss the basic concepts. In volleyball games, two teams of six players, separated by a high net, try to score points by grounding the ball on the other's team court. In possession of the ball, teams may hit the ball up to three times, but players must not hit the ball twice consecutively. During a rally, namely the time between the serve and the end of the play, the teams compete for the point using specially trained hits to control the ball, set, and attack. Blocks, likewise, are trained actions to try to stop or alter an opponent's attack, but unlike the others, the hit involved in blocks does not count as an 'official' hit. It means that a player who hits the ball by blocking can consecutively touch the ball again without infringing the rules. On the other hand, we want to point out that the emergence of complexity in the game is strictly related to players' ball control and accuracy. Note that if one of the players is not able to handle the ball,

their team misses the point; if one of the players cannot pass the ball accurately, it increases the probability that the teammates cannot control the ball, and if one of the players cannot attack accurately, it increases the probability that the opponents can handle the ball and eventually obtain the point.

In order to study this complex dynamic and unveil the mechanisms behind it, we proceed as follows. First we visualized and collected data from 20 high-level international games. Then we performed a data-driven analysis to get insight into the underlying mechanisms of the game. With this, we proposed a parsimonious model and developed an analytical approach to obtain a closed-form expression able to capture in remarkable good approximation one of the most relevant observables of the dynamics: the probability that the players perform n hits in a rally, P_n . A variable that is key to characterizing the system complexity, and it is also related to teams performances [23, 24], it is involved in self-regulated phenomena [25], may affect the motor activities during a top-level match [26], etc.

2. Data collection

To collect the data used in this work, we coded a visualization program that allows extracting information from video records of volleyball games. The program is designed for a trained operator to record the most relevant events observed during rallies. The information gathered at each event include,

- (a) An identification number for the player that hit the ball in the event.
- (b) The time of the event referred to the beginning of the game.
- (c) The position of the player that hit the ball. In two dimensions, referred to the court's floor.
- (d) The type of hit performed: pass, set, attack, block, etc.

We visualized 10 games of the 2018 FIVB Men's Volleyball Nations League and 10 games of the 2019 FIVB Men's Volleyball Challenger Cup, collecting the information of 3302 rallies in total. The visualization program and an anonymized version of the collected data used for this work can be found in [27]. For further information on the visualized games and details of the visualization process, please cf supplementary material section S1 (<https://stacks.iop.org/BB/3/035006/mmedia>).

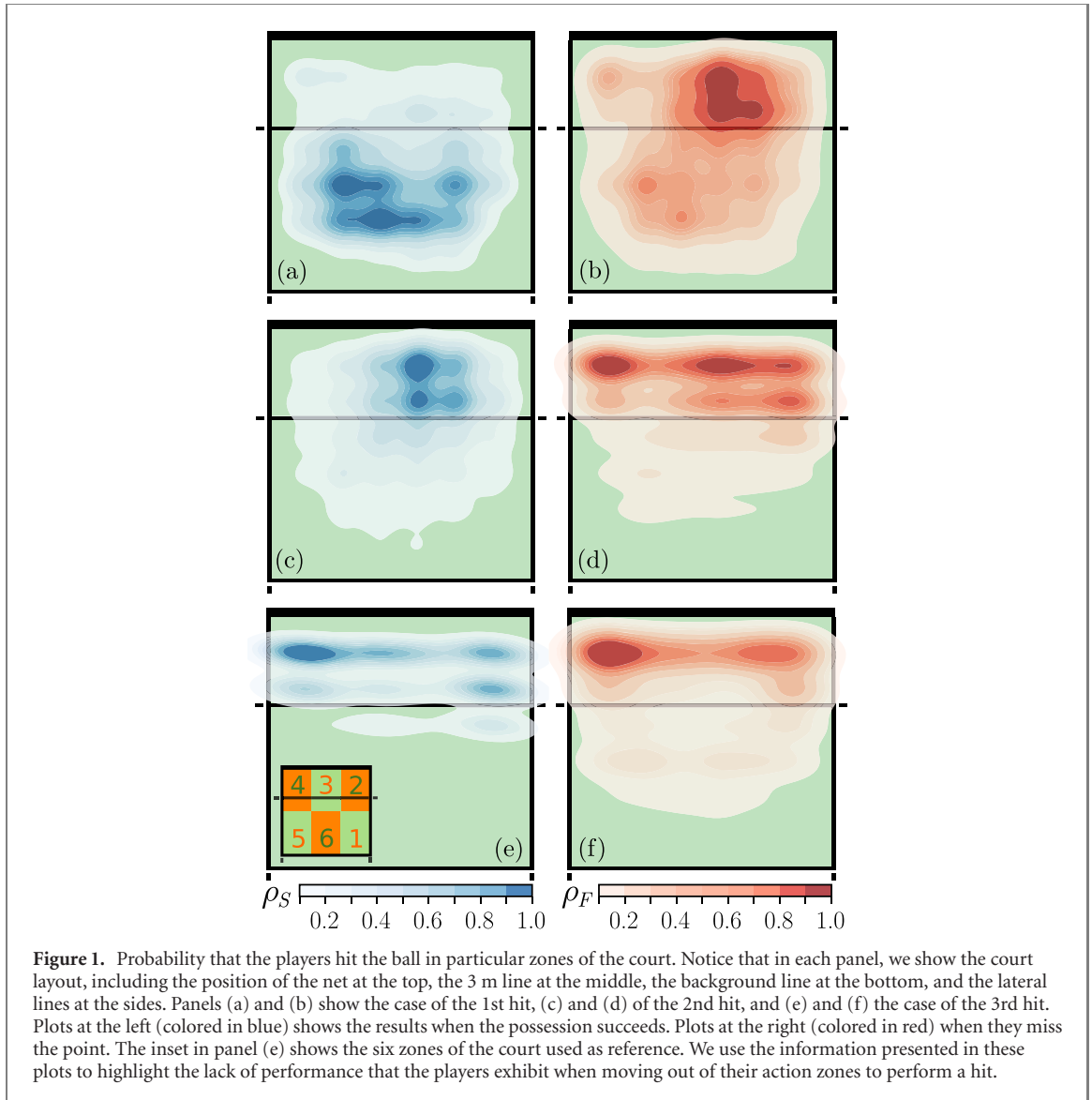
3. Data analysis

We analyzed the collected data to understand why the players succeed or fail when they intend to score a point. In figure 1 we show heatmaps indicating the probability that the players hit the ball in particular zones of the court. Panels (a) and (b) show the case of the 1st hit, (c) and (d) of the 2nd hit, and (e) and (f) the case of the 3rd hit. At the left (colored in blue), we show the results when the possession succeeds, namely when they scores the point. At the right (colored in red), when they miss the point. Parameters ρ_S and ρ_F give the probability normalized by the maximum value and are linked to the color intensity of the heatmaps. In the inset of panel (e), we show the court divided and numerated into zones. Notice that, in the following, we will use these references to discuss the results.

As a first observation, we can see that, in successful possessions, it is highly probable that the players hit the ball in their natural action zones: the first hit (reception) in zones 5-6-1, the second one (set up) around 2-3, and the attack in the zones 4-3-2 (at the front) and 6-1 (at the back). In unsuccessful possessions, we can see that the players have to move out of the actions zones to perform the hit. For instance, in panel (b), we can see that the probability of the players performing the 1st hit around zone 2 rises. Note, that it is not tactical convenient because this is the zone that the setter uses to perform the 2nd hit, and the tight presence of other players may hinder them and produces a fail or cause a reduction of accuracy. If the 1st hit is performed inaccurately, it produces what we see in panel (d), where the setter frequently has to go to the attackers' zones to handle the ball causing the same hinder effect as in the previous case. In this context, the attackers' options are limited. In extremes cases, they will not attack. They will pass an easy ball to the opponents just to keep playing. Therefore, in these cases, the players' performance diminishes. In conclusion, when the players have to move out of their action zones to handle the ball, it increases the probability of both missing the ball and hitting the ball imprecise. In the following, we will use these observations to define our model.

4. Model

We have observed that, during rallies, when the players have to move out of their action zone to perform a hit, the probability of missing the ball increases, and the precision decreases. In the light of these observations, to model the rallies' dynamic, we introduce two stochastic parameters, p and q , as follows,



- (a) *The probability of performing the hit, p .* If the players have to move out of their action zones to hit the ball, then, there is a probability p of performing the hit, and $1 - p$ of missing it.
- (b) *The probability of achieving precision, q .* If the players have to move out of their action zones to perform a hit, then there is a probability q of achieving precision in two situations:
1. Defending or setting, in the first and second hit passing the ball towards the teammate's action zone.
 2. Attacking or serving, sending the ball out of the action zone of the opponent in charge of taking the first hit.

With probability $1 - q$, likewise, the opposite occurs in both situations.

Following these rules, the players will move the ball around the court n times until one of them misses a hit, ending the rally.

5. Analytical approach to obtain P_n

We now focus on developing an analytical approach to obtain the probability distribution that the players perform n hits during rallies, P_n . First, we introduce two approximation: (i) we will suppose that the teams always try to use the three reglementary hits to score the point. In other words, they do not pass the ball to the other team at the first or second hit. Considering that these events are rare in the dataset ($<0.1\%$ of the observed cases), we understand that this is a reasonable approximation; and (ii) we will not consider the blocks as hits. By rule, when a team blocks an attack, they still have the three hits to control, set and attack. Therefore, in practice, the effects of blocks can be absorbed as inefficiencies in the attack.

Let us start calculating the cases P_1 , P_2 and P_3 . We obtain,

$$P_1 = q(1 - p), \quad (1a)$$

$$P_2 = qp(1 - q)(1 - p), \quad (1b)$$

$$P_3 = qp(1 - q)p(1 - q)(1 - p). \quad (1c)$$

In equation (1a), the probability q indicates that the service is performed outside the action zone of the player in charge of taking the first hit (a difficult service), and the probability $(1 - p)$ indicates that this player cannot perform the hit. Consequently, in the rally only one hit is performed: the service. In equation (1b), q indicates a difficult service, p indicates that the player can perform the second hit, $(1 - q)$ indicates the pass is performed outside the action zone of the player in charge of taking the second hit (setter), and $(1 - p)$ indicates that the later cannot perform the third hit. Consequently, in the rally two hits are performed: the service, and the first hit. With a similar analysis, we can obtain the probability of performing three hits, P_3 , that we show in equation (1c).

To obtain P_4 and further values of n is useful to define the transition probabilities linked to the teams' chances of performing the three hits that define a complete possession,

$$\begin{aligned} P_{dd} &= pq + p(1 - q)pq + p(1 - q)p(1 - q)pq, \\ P_{de} &= p(1 - q)p(1 - q)p(1 - q), \\ P_{ed} &= 1, \\ P_{ee} &= 0, \end{aligned} \quad (2)$$

where P_{dd} is the probability that a team receives a difficult ball and, after the three hits, delivers to the other team a difficult one, P_{de} is the probability that a team receives a difficult ball and delivers an easy one, P_{ed} the probability that a team receives an easy ball and delivers a difficult one, and P_{ee} the probability that a team receives and delivers an easy ball. In this frame, P_4 can be written as follows,

$$P_4 = qP_{dd}(1 - p) + (1 - q)P_{ed}(1 - p), \quad (3)$$

where the first term is the probability that a team receives a difficult ball, handles it, delivers a difficult ball, and the other team cannot achieve the first hit. The second term is the probability that a team receives an easy ball, delivers a difficult one, and the opponent team cannot achieve the first hit.

To calculate P_n for higher values of n , we use a binary decision tree. As an example, let us focus on calculating P_7 . In figure 2 we exhibit a three levels tree where the nodes indicate the number of hits performed and the edges the transition probabilities to the next level. Note, the level of the tree, l , is related to the number of performed hits by the expression $n = 3l - 2$. If we define the probability D_3 as,

$$D_3 = qP_{dd}P_{dd} + qP_{de}P_{ed} + (1 - q)P_{ed}P_{dd}, \quad (4)$$

obtained from the sum of all the paths leading to the third level leaves indicating that the team in possession is delivering a difficult ball (see circles in figure 2), then, to obtain P_7 we just multiply for $(1 - p)$, which indicates the team receiving the attack cannot perform the 8th hit. Therefore,

$$P_7 = D_3(1 - p). \quad (5)$$

To calculate the probability for values of n between two levels of the tree, for instance P_8 and P_9 , we use equation (4) as follows,

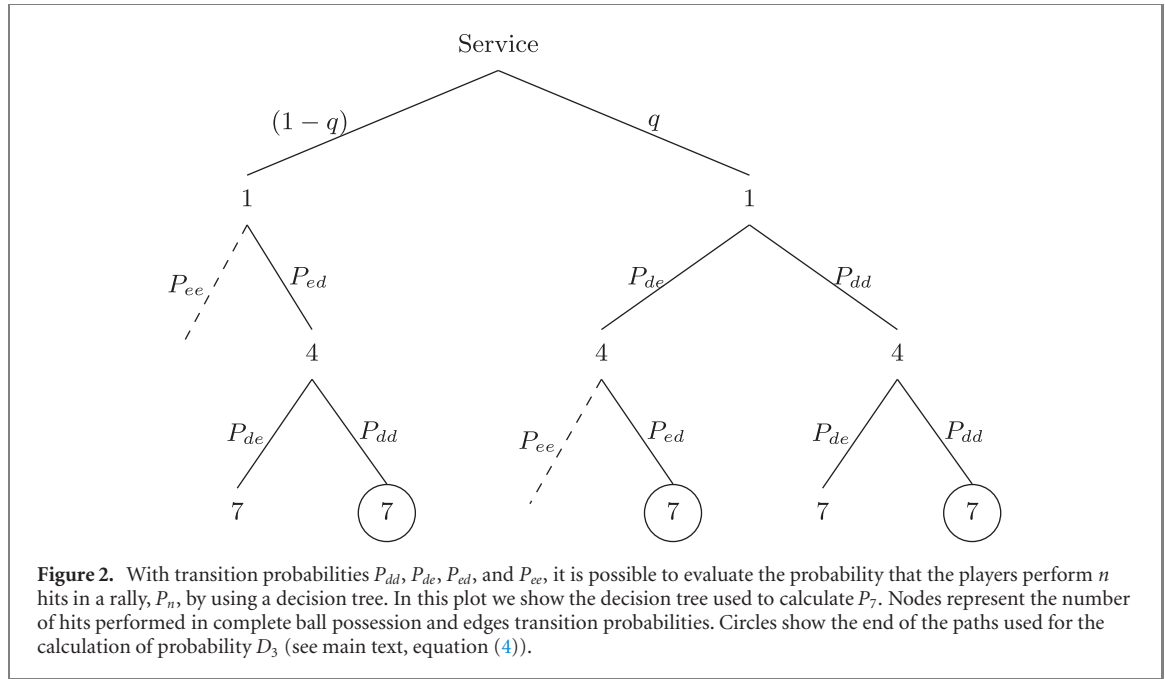
$$\begin{aligned} P_8 &= D_3p(1 - q)(1 - p) \\ P_9 &= D_3p(1 - q)p(1 - q)(1 - p). \end{aligned} \quad (6)$$

As the reader may be noted, the calculation of D_l is useful to obtain P_n . Therefore in the following, we focus on calculating $D_l \forall l \in \mathbb{N}$. We can write the probability of 'achieving' a level l of the tree, A_l , as,

$$A_l = E_l + D_l, \quad (7)$$

where probability E_l is the sum of all the paths leading to the l th level leaves which transition probabilities indicate that the team in possession is delivering an easy ball. Similarly, we can write D_{l+1} and A_{l+1} as,

$$\begin{aligned} D_{l+1} &= E_lP_{ed} + D_lP_{dd} \\ A_{l+1} &= E_lP_{ed} + D_l(P_{de} + P_{dd}). \end{aligned} \quad (8)$$



Combining equations (7) and (8), and using $P_{ed} = 1$ (see equation (2)), we obtain the following system of mutually recursive linear sequences,

$$\begin{pmatrix} D_{l+1} \\ A_{l+1} \end{pmatrix} = \begin{pmatrix} (P_{dd} - 1) & 1 \\ (P_{de} + P_{dd} - 1) & 1 \end{pmatrix} \begin{pmatrix} D_l \\ A_l \end{pmatrix}. \quad (9)$$

Then, the roots of the characteristic polynomial linked to the 2×2 matrix of system (9),

$$\lambda_{1,2} = \frac{P_{dd} \pm \sqrt{P_{dd}^2 + 4P_{de}}}{2}, \quad (10)$$

can be used to express the general solution of D_l as,

$$D_l = a\lambda_1^{l-1} + b\lambda_2^{l-1}, \quad (11)$$

where constants a and b can be found by using equations (10) and (11), and the expressions for $D_{l=1}$ and $D_{l=2}$,

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \lambda_1^0 & \lambda_2^0 \\ \lambda_1^1 & \lambda_2^1 \end{pmatrix}^{-1} \begin{pmatrix} q \\ qP_{dd} + (1-q)P_{ed} \end{pmatrix}. \quad (12)$$

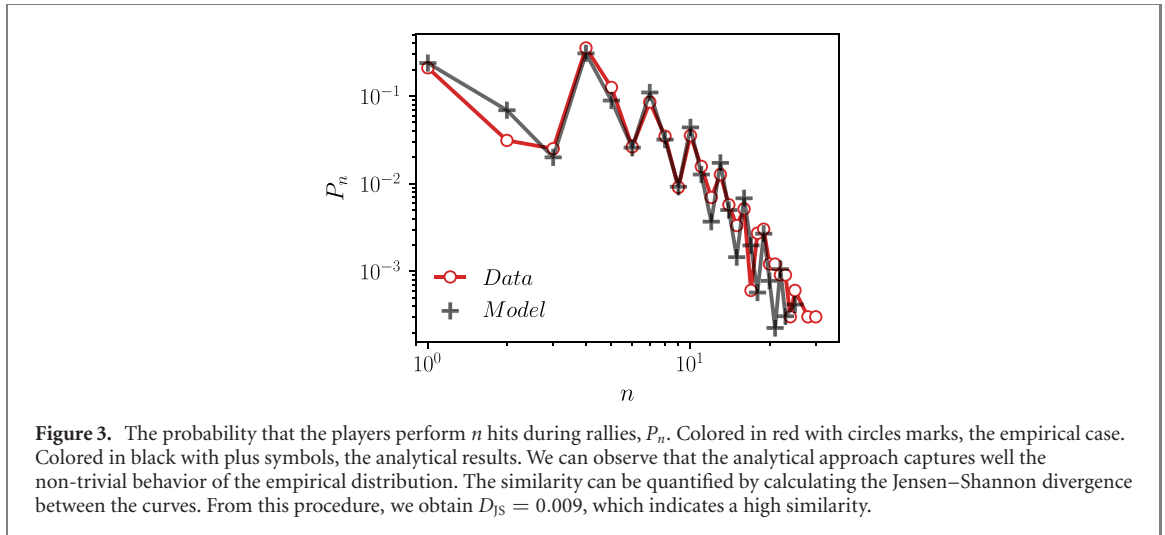
In the light of the described above, we have formally found the exact solution for P_n . In the following set of equations we summarize our result,

$$\begin{aligned} P_n &= D_{\frac{n+2}{3}}(1-p), \\ P_{n+1} &= D_{\frac{n+2}{3}}p(1-q)(1-p), \\ P_{n+2} &= D_{\frac{n+2}{3}}p(1-q)p(1-q)(1-p), \end{aligned} \quad (13)$$

with $n = 3l - 2$, and $l \in \mathbb{N}$.

In figure 3 we show the probability distribution P_n . The extracted from the data is colored in red and the theoretical calculated with equation (13) is colored in black. The values of parameters p and q were set by carrying out a minimization process using the Jensen–Shannon divergence (D_{JS}) between the curves as metrics. From this procedure, we obtained $p = 0.54$ and $q = 0.46$ with $D_{JS} = 0.009$, indicating a good similarity between the curves. Such correspondence can also be observed with the naked eye in the figure. On the other hand, the mean value and the standard deviation for the empirical curve are $\langle n \rangle = 4.58(70)$, $\sigma(n) = 3.40(45)$, and for the theoretical curve are $\langle n^{TH} \rangle = 4.74(83)$, $\sigma(n^{TH}) = 3.44(66)$. We can see that the values are smaller in the empirical case but still very similar to the calculations.

Regarding the obtained values of p and q , it is necessary to highlight that in a real games, each player on the court should have a different pair of values p and q . This is because distinct players should perform differently. Moreover, these pairs could depend on the set because players vary their performance during the



game. However, with these results, we show that the simplification proposed in our model allows us to represent each player as a single *average player*. In this context, since we construct the empirical probability P_n with the results of several matches, teams, and players, it is expected that throughout the minimization process, we obtain values close to 0.5.

The reader might note that the shape of the curves follows a zig-zag pattern with the peaks placed at the values $n = 1, 4, 7, 10, \dots$, which are the number of hits related to teams' complete ball possessions. We can explain this non-trivial behavior by analyzing the natural dynamics of the game. If a team can control an attack in the first hit, they will use the rest of the reglementary hits to set the ball and attack back. As we have previously mentioned, it is unlikely they attack at the first or second hit. Because of this, it is more probable to find in the dataset cases, for instance, with $n = 7$ than cases with $n = 6$.

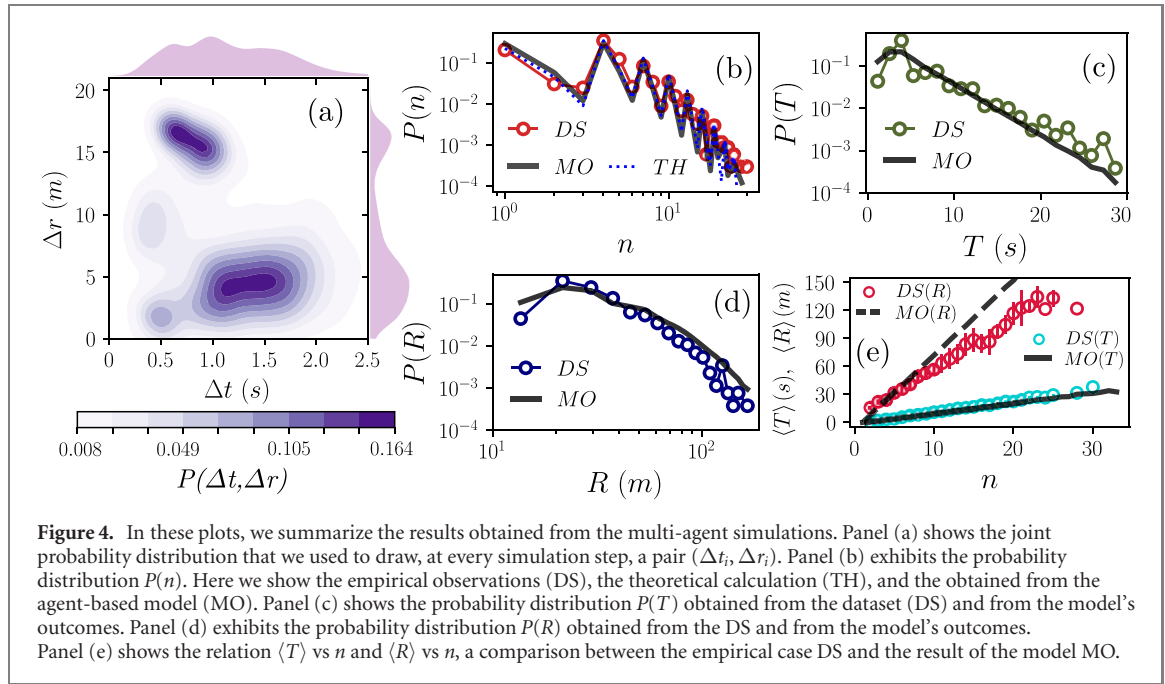
We want to highlight that despite the complex dynamics of the game, the analytical approach based on our parsimonious stochastic model succeeds in capturing the non-trivial behavior of this relevant observable.

6. Multi-agent simulations

We now apply the rules of our model to guide the dynamics of a minimalist 1D self-propelled agents system thought of to emulate volleyball rallies. We aim to confirm, with this agent-based model, the analytical results obtained for probability distribution P_n . Moreover, we show that it is possible to combine the analytical results with the data to give an approach to capture other empirical global observables related to spatiotemporal variables. The agent model is based on the following elements,

- Teams and players.* In this parsimonious system, there are two teams with two players (agents) by team.
- The court.* We represent the game's court as a 1D array with four sites. The 1st and the 2nd sites shape the first team side, and the 3rd and the 4th sites shape the second team side. The net is placed between sites 2 and 3, delimiting the teams' sides. Players are allowed to occupy any site of their zones. They can also overlap, but they cannot invade the rival's side.
- The teammates' roles.* To emulate the player's tactical dependency in volleyball teams, we propose that one of the two players manages the 1st and 3rd hit (defender/attacker), and the other manages the 2nd hit (setter).
- Initial conditions.* The player's sites at the beginning of the rally are randomly set. The player at the 1st site serves the ball.
- Dynamics.* If the agents have to change their current place to hit the ball, then the parameters p and q give the probability of performing the hit and achieving precision, respectively (see section model). Achieving precision in this context means, in the 1st hit, sending the ball to the partner place, and in the 3rd hit, sending the ball to the site where the rival in charge of taking the 1st hit is not occupying. The rally ends when one of the players misses a hit.

We define T as the rallies' total time and R as the length of the projection of the ball trajectory on the court's floor. These two variables can be empirically measured from the collected data by simply adding the succession of temporal and spatial intervals, Δt and Δr , observed between rallies' events. To obtain T and R from simulations we propose the following,



- (f) *Global observables T and R .* At every step of the dynamics, the players hitting the ball draws from the empirical joint distribution $P(\Delta t, \Delta r)$, shown in figure 4(a), a pair of values $(\Delta t_i, \Delta r_i)$ that we will link to the i th event. Then, at the end of the rallies, we compute $T = \sum_i^n \Delta t_i$ and $R = \sum_i^n \Delta r_i$, where n is the total number of events, to obtain the simulated values of these observables.

Notice we sample Δt and Δr from the joint distribution because these variables are correlated. This is evidenced in figure 4(a), where we can see that the multimodal behavior of $P(\Delta r)$ and $P(\Delta t)$ results in a non-trivial joint distribution with several local maxima.

With the model defined, we performed 10^5 rally simulations setting the parameters p and q with the values calculated in section 5. In figure 4 panels (b) to (e), we compare the outcomes with the empirical data. Additionally, in panel (b), we plot the theoretical curve given by the equation (13). In the following, we discuss these results. In panel (b), we show the probability distribution $P(n)$. We can see that it agrees with the theoretical curve, confirming the calculations. Panel (c) show the probability distribution $P(T)$. In this case, we can see that the model captures very well the empirical data. The plot is in linear-log scale to emphasize the probability decay as an exponential curve. The calculation for the mean value and the standard deviation for the empirical data gives $\langle T^{DS} \rangle = 5.47$ (s) and $\sigma(T^{DS}) = 4.00$ (s), and for the simulations $\langle T^{MO} \rangle = 3.52$ (s) and $\sigma(T^{MO}) = 3.86$ (s). The similarity between the values of the first and second moment supports the existence of an exponential behavior. In panel (d), the probability distribution $P(R)$ is shown. As well as in previous cases, we see that the model satisfactorily captures the global aspects of the empirical data. The mean and standard deviation for both cases are $\langle R^{DS} \rangle = 34.42$ (m), $\sigma(R^{DS}) = 18.23$ (m), and $\langle R^{MO} \rangle = 39.17$ (m), $\sigma(R^{MO}) = 25.52$ (m). Lastly, in panel (e), we show the evolution of $\langle T \rangle$ and $\langle R \rangle$ with the number of hits n . We can see a linear behavior in both cases. For $\langle T \rangle$ the empirical observations agree well with the model's outcomes. For $\langle R \rangle$, on the contrary, we observe a deviation at higher values of n . This deviation may be related to the emergence of some type of complexity in the actual dynamics that our simple model cannot capture. For instance, sometimes during the game the play becomes unstable, and the players hit the ball several times in reduced space until one team is able to control the ball. Notice we cannot capture this kind of 'burstiness effect' through our simple model. According to the stated rules, at every step, we have to randomly draw a pair $(\Delta t, \Delta r)$ which implies following a memoryless process. In this frame, since we have a joint probability distribution that is bimodal in the variable Δr , the probability of obtaining a long sequence of pairs with small values of Δr is low. Consequently, for a given value of n , the calculated value of $\langle R \rangle$ in the model could tend to be slightly higher than in the empirical case, as we observe in figure 4 panel (e).

7. Conclusions

Our investigation focused on studying the dynamics of volleyball's rallies. We proposed a framework based on games visualization, collection of relevant information, and through data-driven analysis aiming to obtain insights to define the rules of a mathematical model able to capture the underlying dynamics in volleyball

games. We found that the players are more likely to fail and become imprecise if they have to move out of their actions zones to perform a hit. With this in mind, we proposed a model based on two stochastic parameters: p and q , where the first is the probability of performing the hit, and the latter is the probability of achieving precision. Then we calculated a closed-form expression for the probability that players perform n hits in the rally, P_n , that agree remarkably well with the empirical observations. Therefore, we understand that we have uncovered two stochastic variables able to partially generate the level of complexity observed in the volleyball dynamics. In this regard, we consider that this work represents a new step towards a broad understanding of volleyball games as complex adaptive systems. Moreover, the collected data that we make publicly available with this work is, as far as we know, the largest open collection of volleyball-logs ever released, an invaluable resource for the research community that opens the door for further research in this area. Finally, we want to point out that our findings provide new knowledge that should be actively taken into account for sports scientists and coaches. Since it has been shown that the length of rallies may affects the team behavior through many indirect variables [23–26], our results can be handy for designing new efficient data-driven training systems aiming to enhance the performance in competitive scenarios.

Acknowledgment

We acknowledge enlightening discussions with Lucía Pedraza. This work was partially supported by CONICET under Grant No. PIP 112 20200 101100; FonCyT under Grant Nos. PICT-2017-0973; and SeCyT-UNC (Argentina).

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/chacoma/volleyball>.

ORCID iDs

Andrés Chacoma  <https://orcid.org/0000-0002-9879-494X>

References

- [1] Blasius B and Tönjes R 2009 Zipf's law in the popularity distribution of chess openings *Phys. Rev. Lett.* **103** 218701
- [2] Clauset A, Kogan M and Redner S 2015 Safe leads and lead changes in competitive team sports *Phys. Rev. E* **91** 062815
- [3] Hollecsek T and Tröster G 2012 Particle-based model for skiing traffic *Phys. Rev. E* **85** 056101
- [4] Baek S K, Son S-W and Jeong H-C 2015 Nash equilibrium and evolutionary dynamics in semifinalists' dilemma *Phys. Rev. E* **91** 042144
- [5] Ben-Naim E and Hengartner N W 2007 Efficiency of competitions *Phys. Rev. E* **76** 026106
- [6] Yamamoto K and Narizuka T 2021 Preferential model for the evolution of pass networks in ball sports *Phys. Rev. E* **103** 032302
- [7] Neiman T and Loewenstein Y 2011 Reinforcement learning in professional basketball players *Nat. Commun.* **2** 569
- [8] Mukherjee S, Huang Y, Neidhardt J, Uzzi B and Contractor N 2019 Prior shared success predicts victory in team competitions *Nat. Hum. Behav.* **3** 74–81
- [9] Mandić R, Jakovljević S, Erčulj F and Štrumbelj E 2019 Trends in NBA and Euroleague basketball: analysis and comparison of statistical data from 2000 to 2017 *PLoS One* **14** e0223524
- [10] Merritt S and Clauset A 2013 Environmental structure and competitive scoring advantages in team competitions *Sci. Rep.* **3** 3067
- [11] Gudmundsson J and Horton M 2017 Spatio-temporal analysis of team sports *ACM Comput. Surv.* **50** 1–34
- [12] Fister I Jr, Ljubić K, Suganthan P N, Perc M and Fister I 2015 Computational intelligence in sports: challenges and opportunities within a new research domain *Appl. Math. Comput.* **262** 178–86
- [13] Patel D, Shah D and Shah M 2020 The intertwine of brain and body: a quantitative analysis on how big data influences the system of sports *Ann. Data. Sci.* **7** 1–16
- [14] Drikos S, Kountouris P, Laios A and Laios Y 2009 Correlates of team performance in volleyball *Int. J. Perform. Anal. Sport* **9** 149–56
- [15] Buldu J M *et al* 2019 Defining a historic football team: using network science to analyze Guardiola's FC Barcelona *Sci. Rep.* **9** 13602
- [16] Bransen L, Van Haaren J and van de Velden M 2019 Measuring soccer players' contributions to chance creation by valuing their passes *J. Quant. Anal. Sports* **15** 97–116
- [17] Ali C, Ali U and Delibas E 2018 Computational modeling of pass effectiveness in soccer *Adv. Complex Syst.* **21** 1850010
- [18] Gama J, Dias G, Couceiro M, Sousa T and Vaz V 2016 Networks metrics and ball possession in professional football *Complexity* **21** 342–54
- [19] Garrido D, Antequera D R, Busquets J, López Del Campo R, Serra R R, Vielcazat S J and Buldú J M 2020 Consistency and identifiability of football teams: a network science perspective *Sci. Rep.* **10** 1–10
- [20] Jusup M *et al* 2022 Social physics *Phys. Rep.* **948** 1–148
- [21] Chacoma A, Almeida N, Perotti J I and Billoni O V 2020 Modeling ball possession dynamics in the game of football *Phys. Rev. E* **102** 042120
- [22] Chacoma A, Almeida N, Perotti J I and Billoni O V 2021 Stochastic model for football's collective dynamics *Phys. Rev. E* **104** 024110

- [23] Sánchez-Moreno J, Marcelino R, Mesquita I and Ureña A 2015 Analysis of the rally length as a critical incident of the game in elite male volleyball *Int. J. Perform. Anal. Sport* **15** 620–31
- [24] Link D and Wenninger S 2019 Performance streaks in elite beach volleyball-does failure in one sideout affect attacking in the next? *Front. Psychol.* **10** 919
- [25] Sánchez-Moreno J, Afonso J, Mesquita I and Ureña A 2016 Dynamics between playing activities and rest time in high-level men's volleyball *Int. J. Perform. Anal. Sport* **16** 317–31
- [26] Mroczek D, Januszkiewicz A, Kawczyński A S, Borysiuk Z and Chmura J 2014 Analysis of male volleyball players' motor activities during a top level match *J. Strength Condit Res.* **28** 2297–305
- [27] Chacoma A 2022 *Github Repository* <https://github.com/chacoma/volleyball>