Journal of Applied Mathematics and Physics, 2026, 14(3), 1160-1183
https://www.scirp.org/journal/jamp
ISSN Online: 2327-4379
ISSN Print: 2327-4352

Modeling and Simulation of Player Distribution in a 4-3-3 System Using Reaction-Diffusion Equations

Aliou Sonko[1], Mamadou Moustapha Diop[2]

[1]Department of Applied Mathematics, University Iba Der Thiam de Thies, Thies, Senegal
[2]Department of Applied Mathematics, University Alioune Diop de Bambey, Bambey, Senegal
Email: aliousonko59@gmail.com

How to cite this paper: Sonko, A. and Diop, M.M. (2026) Modeling and Simulation of Player Distribution in a 4-3-3 System Using Reaction-Diffusion Equations. *Journal of Applied Mathematics and Physics*, **14**, 1160-1183.
https://doi.org/10.4236/jamp.2026.143055

Received: December 18, 2025
Accepted: March 13, 2026
Published: March 16, 2026

Abstract

The tactical organization of players in a 4-3-3 system is a crucial issue for optimizing collective performance in football. This article proposes an innovative mathematical approach to model and simulate the spatial distribution of players using reaction-diffusion partial differential equations. Unlike existing multi-agent or stochastic approaches that treat space discretely, our continuous model simultaneously integrates three fundamental forces: tactical influence (attraction toward assigned zones), team cohesion (coordination between groups), and ball attractiveness (responsiveness to game situations). The existence and uniqueness of solutions are rigorously established through semigroup theory and the Lumer-Phillips theorem, ensuring the mathematical validity of the proposed framework. A finite-difference discretization with an implicit scheme enables a stable and efficient numerical implementation. The major contribution of this work lies in the systematic calibration of the model parameters using real tracking data (Metrica Sports), followed by a multidimensional quantitative validation. The results demonstrate excellent agreement between simulated and empirical densities (spatial correlation > 0.82, Wasserstein distance < 3.1 m), thereby validating the relevance of the model. Analysis of the calibrated parameters provides quantified tactical insights: midfielders are 2.5 times more mobile than defenders, while defenders are 1.9 times more tactically disciplined, and midfielders exhibit a ball responsiveness 2.4 times higher. A sensitivity analysis confirms the robustness of the model with respect to parametric perturbations. These results open promising perspectives for tactically assisted analysis using mathematical tools, offering coaches and analysts objective methods to diagnose collective organization, optimize tactical instructions, and evaluate team performance. The developed methodology is also transferable to other tactical formations and team sports.

Keywords

Football, 4-3-3 System, Reaction-Diffusion Equations, Mathematical Modeling, Parameter Calibration, Tracking Data, Empirical Validation, Tactical Analysis, Finite Differences, Semigroup Theory, Spatial Distribution, Collective Performance

1. Introduction

Football, as a team sport, relies on rigorous tactical organization of players. Among the many available formations, the 4-3-3 system is highly regarded for its balance between attack, midfield, and defense. However, adapting this organization to the unpredictable dynamics of the game remains a central problem, widely studied in sports science and mathematical modeling [1] [2]. The analysis of player movements and their spatial distribution has become a key issue for optimizing collective performance.

1.1. State of the Art and Limitations of Existing Approaches

Previous work has explored different approaches to model player dynamics on the field: **stochastic models** use random processes (Markov chains) to quantify the probabilities of presence in different zones [3]; **multi-agent systems** model each player as an autonomous entity following heuristic behavioral rules [4]; and **machine learning approaches** predict future positions from massive tracking data.

Despite their respective contributions, these methods share common limitations that motivate our approach:

- **Lack of spatio-temporal continuity**: discrete treatment of space (zones, grids), preventing a fine representation of player distribution;
- **Fragmentation of driving forces**: factors influencing movements (tactical constraints, cohesion, ball position) are treated in isolation rather than integrated into a unified framework;
- **Insufficient empirical validation**: parameters often choose arbitrarily without objective calibration on real data.

1.2. Our Contribution

To address these issues, we propose a rigorous mathematical approach based on partial differential equations (PDEs). Reaction-diffusion equations, initially introduced in biology and chemistry to model the formation of spatial patterns [5], prove particularly well-suited for modeling player movements, their collective interactions, and the tactical constraints imposed by the 4-3-3 system.

Our approach differs from existing methods on several fundamental aspects:

1) Unified multi-scale model. We propose a system of reaction-diffusion equations that simultaneously integrates three complementary forces acting on player distribution:

$$f_k = f_{k,\text{tactical}} + f_{k,\text{cohesion}} + f_{k,\text{ball}} \tag{1.1}$$

This unified formulation captures the complex interactions and trade-offs between tactical discipline, team cohesion, and ball reactivity, thus producing realistic emergent behaviors.

2) Continuous and deterministic framework. Unlike stochastic and multi-agent approaches, our model continuously describes the evolution of player density in space and time. This formulation allows a fine analysis of spatial coverage of the field and areas of high or low presence.

3) Explicit modeling of tactical groups. Our model explicitly distinguishes three groups of players (defenders, midfielders, attackers) with their own dynamics characterized by:

- A diffusion coefficient D_k representing the spatial mobility of the group,
- Parameters γ_k, β_k, α_k quantifying respectively the tactical influence, cohesion, and attraction toward the ball.

4) Rigorous mathematical foundation. We establish the existence and uniqueness of solutions to the proposed system using semigroup theory and the Lumer-Phillips theorem. This mathematical validation ensures that our model is well-posed and that numerical simulations converge to a unique solution.

5) Quantitative validation with real data. We calibrate the model parameters by minimizing the gap with empirical densities constructed from real tracking data (Metrica Sports/StatsBomb). We quantitatively evaluate the model's performance via objective metrics (RMSE, spatial correlation, Wasserstein distance).

6) Sensitivity analysis. We systematically study the impact of parametric variations on simulated distributions, demonstrating the model's robustness and identifying the most influential parameters.

1.3. Methodology

Our approach combines mathematical rigor, numerical simulation, and empirical validation. On the theoretical level, we formulate the problem as a PDE system of reaction-diffusion type whose existence and uniqueness of solutions are proven via the Lumer-Phillips theorem. On the numerical level, we develop an implicit finite difference scheme that preserves the properties of the continuous model while ensuring numerical stability. On the empirical level, we introduce a rigorous parametric calibration methodology based on real tracking data, allowing us to move from a purely theoretical model to a quantitative predictive tool.

The rest of the article is organized as follows: Section 2 formulates the system of equations with the unified reaction term; Section 3 proves the existence and uniqueness of the solution; Section 4 presents the numerical discretization; Section 5 describes the calibration and validation with real data; Section 6 interprets the results in tactical terms; and Section 7 concludes with future perspectives.

This integrated approach paves the way for a new generation of mathematically-assisted tactical analysis tools, with potential applications for training, match preparation, and post-match analysis.

2. Mathematical Model and Working Spaces

In modern football, the tactical organization of players on the field is a key factor in collective performance. The 4-3-3 system, in particular, is widely used for its balance between defense, midfield, and attack. To address this problem, we propose a mathematical modeling based on reaction-diffusion equations [1] [2]. This approach allows us to capture spatio-temporal interactions between players while simultaneously accounting for multiple factors: the tactical constraints imposed by the 4-3-3 formation, the team's collective cohesion, and the dynamic influence of the ball position.

2.1. Mathematical Model

We model the distribution of players according to the 4-3-3 system using a system of partial differential equations of reaction-diffusion type. Each equation represents the spatio-temporal evolution of player density for a given group.

Player Dynamics

For each group of players k (with $k=1$ for defenders, $k=2$ for midfielders, and $k=3$ for attackers), we have the following equation:

$$\frac{\partial v_k}{\partial t} = D_k \nabla^2 v_k + f_k(\boldsymbol{v}, x, y, t), \quad \forall t > 0, \quad \forall (x, y) \in \Omega = (l_0, l_x) \times (L_0, L_y), \tag{2.1}$$

where $\boldsymbol{v} = (v_1, v_2, v_3)$ represents the complete state of the system (densities of the three groups), and Ω denotes the football field.

The spatial diffusion term $D_k \nabla^2 v_k$ models the natural tendency of players to redistribute spatially to avoid excessive concentrations. The coefficient $D_k > 0$ represents the intrinsic mobility of group k.

Unified Reaction Term

The reaction term f_k integrates three complementary components that act simultaneously on player distribution:

$$f_k(\boldsymbol{v}, x, y, t) = f_{k,\text{tactical}}(x, y) + f_{k,\text{cohesion}}(\boldsymbol{v}, x, y) + f_{k,\text{ball}}(x, y, t) \tag{2.2}$$

a) Tactical influence. The first term encodes the spatial constraints imposed by the 4-3-3 tactical system:

$$f_{k,\text{tactical}}(x, y) = \gamma_k \Psi_k(x, y) \tag{2.3}$$

where $\Psi_k(x, y): \Omega \to \mathbb{R}_+$ is a tactical attractiveness function defining the preferred zones for group k, and $\gamma_k > 0$ measures the intensity of tactical influence. We define Ψ_k as a Gaussian centered on the assigned tactical zone:

$$\Psi_k(x, y) = \exp\left(-\frac{(x - x_k^*)^2}{2\sigma_{k,x}^2} - \frac{(y - y_k^*)^2}{2\sigma_{k,y}^2} \right) \tag{2.4}$$

where (x_k^*, y_k^*) represents the center of the tactical zone and $(\sigma_{k,x}, \sigma_{k,y})$ control its spatial extent. A high value of γ_k indicates strong tactical discipline.

b) Team cohesion. The second term models the interactions between different player groups to maintain a coherent collective structure:

$$f_{k,\text{cohesion}}(\mathbf{v},x,y)=-\beta_k\sum_{j=1}^{3}\nabla W_{kj}(v_j) \tag{2.5}$$

where $W_{kj}:\mathbb{R}_+\to\mathbb{R}_+$ is an interaction potential between groups k and j, and $\beta_k>0$ regulates the intensity of interactions. We consider a quadratic potential $W_{kj}(v_j)=\frac{\omega_{kj}}{2}v_j^2$, where $\omega_{kj}\geq 0$ is an interaction weight. The gradient is then written as $\nabla W_{kj}(v_j)=\omega_{kj}v_j\nabla v_j$. This term acts as a corrective force: group k is influenced by the density gradients of other groups, tending to align its distribution with that of neighboring groups. A high value of β_k means that group k is strongly influenced by the positions of other groups.

c) Ball attractiveness. The third term captures the influence of the ball position on player distribution:

$$f_{k,\text{ball}}(x,y,t)=\alpha_k\nabla\Phi_k(x,y,t) \tag{2.6}$$

where $\Phi_k(x,y,t):\Omega\times\mathbb{R}_+\to\mathbb{R}_+$ represents the ball's attractiveness potential for group k, $\alpha_k>0$ controls the intensity of this attraction, and $(x_b(t),y_b(t))$ denotes the ball position at time t. We define a Gaussian potential:

$$\Phi_k(x,y,t)=\exp\left(-\frac{(x-x_b(t))^2+(y-y_b(t))^2}{2r_k^2}\right) \tag{2.7}$$

where $r_k>0$ represents the ball's influence radius on group k. The corresponding gradient points toward the ball position with an intensity that decreases exponentially with distance. A high value of α_k indicates that group k is strongly attracted to the ball.

Interactions Between Components

The complete Equation (2.2) captures the complex trade-offs and interactions between the three forces. At each instant and at each point on the field, a player from group k simultaneously experiences: an attraction toward their assigned tactical zone, an influence from the densities of other groups to maintain cohesion, an attraction toward the ball, and a tendency to disperse to avoid over-concentration (diffusion). This combination can produce complex tactical behaviors observed in professional football: collective pressing (when α_k is high for all groups and β_k is strong), defensive block (with high γ_1 and low α_1), rapid transitions (sudden change in $x_b(t)$ with high α_3 and D_3), and positional play (with high γ_k for all groups).

Boundary Conditions and Complete System

We assume that players remain confined within the field, which translates to a homogeneous Neumann boundary condition:

$$\frac{\partial v_k}{\partial n}=0,\ \forall k=1,2,3,\ \text{on } \partial\Omega, \tag{2.8}$$

where $\partial\Omega$ denotes the field boundary and $\frac{\partial}{\partial n}$ the outward normal derivative. This condition expresses that there is no flux of players through the field boundaries and guarantees conservation of total mass.

Combining all terms, we obtain the final system with boundary and initial conditions:

$$\begin{cases} \dfrac{\partial v_k}{\partial t} - D_k \nabla^2 v_k = f_{k,\text{tactical}}(x,y) + f_{k,\text{cohesion}}(\boldsymbol{v},x,y) + f_{k,\text{ball}}(x,y,t), \\ \quad \forall t>0,\ \forall (x,y)\in\Omega,\ k=1,2,3, \\ \dfrac{\partial v_k}{\partial n} = 0,\ \text{on } \partial\Omega,\ k=1,2,3, \\ v_k(x,y,0) = v_k^0(x,y),\ \forall (x,y)\in\Omega,\ k=1,2,3, \end{cases} \tag{2.9}$$

where $v_k^0(x,y)$ represents the initial distribution of group k (typically, the starting formation at kickoff).

2.2. Preliminaries and Notation

Before studying the resolution of system (2.9), we introduce the appropriate functional spaces and necessary assumptions for mathematical analysis [6].

Assumptions on Parameters and Functions

We assume that:

1) **Positivity of coefficients:** For all $k=1,2,3$, $D_k>0$, $\gamma_k>0$, $\beta_k>0$, $\alpha_k>0$.

2) **Regularity of tactical zone functions:** The functions $\Psi_k:\Omega\to\mathbb{R}_+$ are of class $C^2(\Omega)$ and bounded: $0\le\Psi_k(x,y)\le M_\Psi$, $\forall (x,y)\in\Omega$, for a constant $M_\Psi>0$.

3) **Regularity of interaction potential:** The function $W_{kj}:\mathbb{R}_+\to\mathbb{R}_+$ is of class C^2 and satisfies: $|W_{kj}'(s)|\le C_W(1+s)$, $|W_{kj}''(s)|\le C_W$, $\forall s\ge 0$, for a constant $C_W>0$.

4) **Regularity of ball attractiveness function:** For all $t\ge 0$, the function $\Phi_k(\cdot,\cdot,t):\Omega\to\mathbb{R}_+$ is of class $C^2(\Omega)$ and uniformly bounded in time: $0\le\Phi_k(x,y,t)\le M_\Phi$, $|\nabla\Phi_k(x,y,t)|\le M_{\nabla\Phi}$, $\forall (x,y)\in\Omega$, $\forall t\ge 0$.

5) **Lipschitz continuity of reaction term:** The complete reaction term $f_k(\boldsymbol{v},x,y,t)$ is locally Lipschitz continuous with respect to $\boldsymbol{v}$ uniformly in (x,y,t): there exists $L>0$ such that $|f_k(\boldsymbol{v},x,y,t)-f_k(\boldsymbol{w},x,y,t)|\le L\|\boldsymbol{v}-\boldsymbol{w}\|$, $\forall \boldsymbol{v},\boldsymbol{w}\in\mathbb{R}_+^3$.

The above assumptions are satisfied by the proposed function choices (Gaussians for Ψ_k and Φ_k, quadratic potential for W_{kj}).

Functional Spaces

We define the Hilbert spaces used in the remainder of our analysis. For each group k, we consider the Lebesgue spaces $L_k^p := L^p(\Omega,\mathbb{R})$, $1\le p\le\infty$. We introduce the product Hilbert space:

$$H := \prod_{k=1}^{3} L_k^2 = L_1^2\times L_2^2\times L_3^2. \tag{2.10}$$

An element $\boldsymbol{v}\in H$ is a vector of three components: $\boldsymbol{v}=(v_1,v_2,v_3)^\top$, with $v_k\in L_k^2$, $k=1,2,3$. We define the inner product in H by:

$$\langle \boldsymbol{v},\boldsymbol{u}\rangle_H := \sum_{k=1}^{3} D_k \int_\Omega v_k(x,y)u_k(x,y)\,\mathrm{d}x\mathrm{d}y. \tag{2.11}$$

The norm induced by this inner product is:

$$\|\boldsymbol{v}\|_H = \left(\sum_{k=1}^{3} D_k \int_\Omega v_k^2 (x,y) \mathrm{d}x\mathrm{d}y \right)^{1/2}. \tag{2.12}$$

The choice to weight the inner product by the diffusion coefficients D_k is natural as it reflects the kinetic diffusion energy of the system. This norm makes the diffusion operator dissipative, which is crucial for mathematical analysis.

By analogy with classical Sobolev spaces, we introduce:

$$H^m := \prod_{k=1}^{3} W_k^{m,2} = \prod_{k=1}^{3} H^m(\Omega), \ m = 1,2. \tag{2.13}$$

For $\boldsymbol{v} \in H^1$, we define:

$$\|\boldsymbol{v}\|_{H^1} := \left(\|\boldsymbol{v}\|_H^2 + \|\nabla \boldsymbol{v}\|_H^2 \right)^{1/2}, \ \text{where } \|\nabla \boldsymbol{v}\|_H^2 = \sum_{k=1}^{3} D_k \int_\Omega |\nabla v_k|^2 \mathrm{d}x\mathrm{d}y. \tag{2.14}$$

The spaces H and H^m inherit the properties of classical Hilbert and Sobolev spaces: H is a separable Hilbert space, $H^2 \hookrightarrow H^1 \hookrightarrow H$ with compact embeddings, and for $\Omega \subset \mathbb{R}^2$ bounded and regular, $H^2(\Omega) \hookrightarrow C^0(\bar{\Omega})$ (continuous embedding).

3. Existence and Uniqueness of the Solution

3.1. Abstract Formulation of the Problem

To establish the existence and uniqueness of solutions to system (2.9), we reformulate it in abstract form by setting:

$$\boldsymbol{u}(t) = \begin{pmatrix} v_1(t) \\ v_2(t) \\ v_3(t) \end{pmatrix}, \ Q(\boldsymbol{u},t) = \begin{pmatrix} f_1(\boldsymbol{u},\cdot,\cdot,t) \\ f_2(\boldsymbol{u},\cdot,\cdot,t) \\ f_3(\boldsymbol{u},\cdot,\cdot,t) \end{pmatrix}, \tag{3.1}$$

where

$$f_k(\boldsymbol{u},x,y,t) = f_{k,\text{tactical}}(x,y) + f_{k,\text{cohesion}}(\boldsymbol{u},x,y) + f_{k,\text{ball}}(x,y,t), \ k = 1,2,3. \tag{3.2}$$

We introduce the linear operator $A: D(A) \subset H \to H$ defined by:

$$A\boldsymbol{u} = \begin{pmatrix} D_1 \Delta v_1 \\ D_2 \Delta v_2 \\ D_3 \Delta v_3 \end{pmatrix}, \ \forall \boldsymbol{u} \in D(A). \tag{3.3}$$

The domain of A is given by:

$$D(A) = \left\{ \boldsymbol{u} \in H^2 \mid \frac{\partial v_k}{\partial n} = 0 \text{ on } \partial\Omega, k = 1,2,3 \right\}. \tag{3.4}$$

Thus, problem (2.9) can be reformulated as an abstract Cauchy problem:

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = A\boldsymbol{u} + Q(\boldsymbol{u},t), \ t > 0, \\ \boldsymbol{u}(0) = \boldsymbol{u}_0 \in D(A). \end{cases} \tag{3.5}$$

3.2. Conditions for Existence and Uniqueness

We now establish the existence and uniqueness of the solution using semigroup theory. The following theorem provides a sufficient condition based on the Lumer-Phillips theorem [7].

If Hypothesis 2.2 is satisfied, then the operator A is m-dissipative, and problem (3.5) admits a unique solution if and only if:

1) A is dissipative: $\langle A\boldsymbol{u},\boldsymbol{u}\rangle_H \le 0$, $\forall \boldsymbol{u}\in D(A)$.

2) $I-A$ is surjective: $\forall \boldsymbol{f}\in H$, there exists $\boldsymbol{u}\in D(A)$ such that $(I-A)\boldsymbol{u}=\boldsymbol{f}$.

3) $Q(\boldsymbol{u},t)$ is locally Lipschitz continuous with respect to $\boldsymbol{u}$ uniformly in t, and has controlled growth.

1) Dissipativity of A:

For all $\boldsymbol{u}\in D(A)$, we have:

$$\langle A\boldsymbol{u},\boldsymbol{u}\rangle_H = \sum_{k=1}^{3} D_k \int_\Omega \Delta v_k \cdot v_k \mathrm{d}x\mathrm{d}y. \tag{3.6}$$

Integrating by parts and using the Neumann boundary conditions, we obtain:

$$\langle A\boldsymbol{u},\boldsymbol{u}\rangle_H = -\sum_{k=1}^{3} D_k \int_\Omega |\nabla v_k|^2 \mathrm{d}x\mathrm{d}y \le 0. \tag{3.7}$$

Therefore, A is indeed dissipative.

2) Surjectivity of $I-A$:

We must show that $\forall \boldsymbol{f}\in H$, there exists $\boldsymbol{u}\in D(A)$ such that:

$$(I-A)\boldsymbol{u}=\boldsymbol{f}. \tag{3.8}$$

This amounts to solving, for each component $k=1,2,3$:

$$v_k - D_k \Delta v_k = f_k. \tag{3.9}$$

We consider the associated elliptic equation:

$$\begin{cases} v_k - D_k \Delta v_k = f_k, & \text{in } \Omega, \\ \dfrac{\partial v_k}{\partial n} = 0, & \text{on } \partial\Omega. \end{cases} \tag{3.10}$$

We will establish the existence and uniqueness of a solution to (3.9) using the Lax-Milgram theorem [8] [9].

Variational Formulation

We consider the Sobolev space $H^1(\Omega)=\left\{u\in L^2(\Omega)\,|\,\nabla u\in \left(L^2(\Omega)\right)^2\right\}$. The variational problem associated with (3.9) consists of finding $v_k\in H^1(\Omega)$ such that:

$$a(v_k,w)=F(w),\ \forall w\in H^1(\Omega), \tag{3.11}$$

where the bilinear form $a(\cdot,\cdot)$ and the linear functional F are defined by:

$$a(v_k,w)=\int_\Omega (v_k w + D_k \nabla v_k \cdot \nabla w)\mathrm{d}x\mathrm{d}y, \tag{3.12}$$

$$F(w)=\int_\Omega f_k w \mathrm{d}x\mathrm{d}y. \tag{3.13}$$

Application of the Lax-Milgram Theorem

The Lax-Milgram theorem guarantees the existence and uniqueness of a solution to (3.11) if the bilinear form $a(\cdot,\cdot)$ is:

1) **Continuous:** there exists a constant $C>0$ such that $|a(u,v)| \le C\|u\|_{H^1}\|v\|_{H^1}$, $\forall u,v \in H^1(\Omega)$.

2) **Coercive:** there exists a constant $\alpha>0$ such that $a(u,u) \ge \alpha\|u\|_{H^1}^2$, $\forall u \in H^1(\Omega)$.

Continuity of $a(\cdot,\cdot)$: By the Cauchy-Schwarz inequality, we have:

$$|a(u,v)| \le \int_\Omega |uv|\,\mathrm{d}x\mathrm{d}y + D_k \int_\Omega |\nabla u \cdot \nabla v|\,\mathrm{d}x\mathrm{d}y \tag{3.14}$$

$$\le \|u\|_{L^2}\|v\|_{L^2} + D_k\|\nabla u\|_{L^2}\|\nabla v\|_{L^2}. \tag{3.15}$$

From the definition of the norm in $H^1(\Omega)$, we deduce the existence of a constant $C>0$ such that:

$$|a(u,v)| \le C\|u\|_{H^1}\|v\|_{H^1},\ \forall u,v \in H^1(\Omega). \tag{3.16}$$

Thus, the bilinear form $a(\cdot,\cdot)$ is continuous.

Coercivity of $a(\cdot,\cdot)$: We must show that there exists a constant $\alpha>0$ such that:

$$a(u,u) \ge \alpha\|u\|_{H^1}^2,\ \forall u \in H^1(\Omega). \tag{3.17}$$

By definition of $a(\cdot,\cdot)$:

$$a(u,u) = \int_\Omega u^2\,\mathrm{d}x\mathrm{d}y + D_k\int_\Omega |\nabla u|^2\,\mathrm{d}x\mathrm{d}y. \tag{3.18}$$

Since $D_k>0$, we have:

$$a(u,u) \ge \min(1,D_k)\left(\|u\|_{L^2}^2 + \|\nabla u\|_{L^2}^2\right). \tag{3.19}$$

Now, the H^1 norm is given by $\|u\|_{H^1}^2 = \|u\|_{L^2}^2 + \|\nabla u\|_{L^2}^2$. We deduce that:

$$a(u,u) \ge \alpha\|u\|_{H^1}^2, \text{ with } \alpha = \min(1,D_k) > 0. \tag{3.20}$$

Thus, the bilinear form $a(\cdot,\cdot)$ is coercive.

Conclusion

According to the Lax-Milgram theorem, there exists a unique solution $v_k \in H^1(\Omega)$ to the variational problem (3.11) for each $k=1,2,3$. Consequently, the elliptic Equation (3.9) admits a unique solution in $H^1(\Omega)$, which proves the surjectivity of the operator $I-A$ [10].

3) Condition on Q:

According to Hypothesis 2.2.1, each component of Q is locally Lipschitz continuous and has controlled growth. Indeed:

- $f_{k,\text{tactical}}(x,y) = \gamma_k \Psi_k(x,y)$ is bounded and smooth (Hypothesis 2.2.2).
- $f_{k,\text{cohesion}}(\boldsymbol{u},x,y) = -\beta_k \sum_{j=1}^3 \omega_{kj} v_j \nabla v_j$ is locally Lipschitz continuous since W_{kj} is C^2 with controlled derivatives (Hypothesis 2.2.3).
- $f_{k,\text{ball}}(x,y,t) = \alpha_k \nabla \Phi_k(x,y,t)$ is uniformly bounded in time (Hypothesis 2.2.4).

Consequently, $Q(\boldsymbol{u},t)$ is locally Lipschitz continuous and has controlled

growth, which guarantees that the right-hand side of (3.5) defines a well-posed problem.

Thus, by the Lumer-Phillips theorem, the operator A generates a strongly continuous semigroup, which guarantees the existence and uniqueness of a solution in H.

Under Hypothesis 2.2, the operator A is m-dissipative and its domain $D(A)$ is dense in H. Moreover, the term $Q(\boldsymbol{u},t)$ is locally Lipschitz continuous with respect to $\boldsymbol{u}$ and has controlled growth. Consequently, problem (3.5) has a unique solution in $C([0,T],H)$ for all $T>0$.

The solution to problem (2.9) satisfies the following energy estimates:

$$\|\boldsymbol{u}(t)\|_H \le \mathrm{e}^{\lambda t}\|\boldsymbol{u}_0\|_H + \int_0^t \mathrm{e}^{\lambda(t-s)}\|Q(\boldsymbol{u}(s),s)\|_H \,\mathrm{d}s, \quad \forall t \ge 0, \tag{3.21}$$

where $\lambda > 0$ is a constant depending on the model parameters.

These results establish the existence and uniqueness of the solution for our problem with the unified reaction term. The particular structure of Q (sum of three bounded and regular terms) guarantees that all hypotheses of the Lumer-Phillips theorem are satisfied.

4. Numerical Approximation by Finite Differences

To obtain usable solutions to system (2.9), we develop a discretization scheme based on the finite difference method [11]. This section presents the spatial and temporal discretization, the construction of matrices, as well as the stability properties of the numerical scheme.

4.1. Domain Discretization

Spatial Domain

The football field is represented by a rectangle $\Omega = [0,L_x]\times[0,L_y]$, where L_x and L_y denote respectively the length and width of the field. For the Metrica Sports data used in our validation, we have $L_x = 105$ m and $L_y = 68$ m.

We subdivide this domain into a regular Cartesian grid with spatial step $\Delta x = \dfrac{L_x}{N_x - 1}$ in the x direction and $\Delta y = \dfrac{L_y}{N_y - 1}$ in the y direction, where N_x and N_y are the numbers of points in each direction. The grid points are:

$$x_i = i\Delta x,\ i = 0,\cdots,N_x - 1,\quad y_j = j\Delta y,\ j = 0,\cdots,N_y - 1. \tag{4.1}$$

The total number of spatial points is $N = N_x \times N_y$. In our simulations, we typically use $N_x = 50$ and $N_y = 50$, providing a spatial resolution on the order of 2 m.

Temporal Discretization

Time is divided into uniform intervals: $t_n = n\Delta t$, $n = 0,1,2,\cdots,N_t$, where Δt is the time step and N_t the total number of time steps. We denote $v_k^n(x_i,y_j) \approx v_k(x_i,y_j,t_n)$ the discrete value of the density of group k at point (x_i,y_j) at instant t_n.

4.2. Discretization of Operators

Laplacian Operator

The Laplacian operator $\nabla^2 v_k$ is approximated by second-order centered finite differences:

$$\nabla^2 v_k\left(x_i, y_j\right) \approx \frac{v_k\left(x_{i+1}, y_j\right) - 2v_k\left(x_i, y_j\right) + v_k\left(x_{i-1}, y_j\right)}{\Delta x^2} + \frac{v_k\left(x_i, y_{j+1}\right) - 2v_k\left(x_i, y_j\right) + v_k\left(x_i, y_{j-1}\right)}{\Delta y^2}. \tag{4.2}$$

This approximation is second-order in space, meaning that the truncation error is $O\left(\Delta x^2 + \Delta y^2\right)$.

Neumann Boundary Conditions

The homogeneous Neumann boundary conditions $\frac{\partial v_k}{\partial n} = 0$ on $\partial\Omega$ are implemented by introducing ghost points:

- At the left boundary $\left(i = 0\right)$: $v_k\left(-1, j\right) = v_k\left(1, j\right)$ (symmetry)
- At the right boundary $\left(i = N_x - 1\right)$: $v_k\left(N_x, j\right) = v_k\left(N_x - 2, j\right)$
- At the lower boundary $\left(j = 0\right)$: $v_k\left(i, -1\right) = v_k\left(i, 1\right)$
- At the upper boundary $\left(j = N_y - 1\right)$: $v_k\left(i, N_y\right) = v_k\left(i, N_y - 2\right)$

Time Derivative

We use an implicit backward Euler scheme for the time derivative:

$$\frac{\partial v_k}{\partial t}\left(t_{n+1}\right) \approx \frac{v_k^{n+1} - v_k^n}{\Delta t}. \tag{4.3}$$

The choice of an implicit scheme is motivated by its unconditional stability, allowing the use of larger time steps without compromising numerical stability.

4.3. Matrix Formulation

Vectorization

We reorganize the nodal values $v_k^n\left(x_i, y_j\right)$ into a column vector $\boldsymbol{v}_k^n \in \mathbb{R}^N$ using lexicographic numbering:

$$\left[\boldsymbol{v}_k^n\right]_m = v_k^n\left(x_i, y_j\right), \text{ with } m = i + j \cdot N_x,\ 0 \le i < N_x,\ 0 \le j < N_y. \tag{4.4}$$

Laplacian Matrix

The discrete Laplacian operator can be represented by a sparse matrix $L \in \mathbb{R}^{N \times N}$ such that $\nabla^2 \boldsymbol{v}_k^n \approx L\boldsymbol{v}_k^n$. The matrix L is constructed by Kronecker product:

$$L = \frac{1}{\Delta x^2} I_{N_y} \otimes L_x + \frac{1}{\Delta y^2} L_y \otimes I_{N_x}, \tag{4.5}$$

where I_{N_x} and I_{N_y} are identity matrices of dimensions $N_x \times N_x$ and $N_y \times N_y$ respectively, and L_x, L_y are the 1D finite difference tridiagonal matrices. The first and last rows are modified to impose the Neumann boundary conditions.

Discretization of Reaction Terms

The reaction terms are evaluated pointwise on the grid:

- **Tactical influence:** $\left[f_{k,\text{tactical}}\right]_m = \gamma_k \Psi_k\left(x_i, y_j\right)$

- **Cohesion:** $\left[f_{k,\text{cohesion}} \right]_m = -\beta_k \sum_{j=1}^{3} \omega_{kj} v_j^n \left(x_i, y_j \right) \left\| \nabla v_j^n \left(x_i, y_j \right) \right\|$, where the gradient ∇v_j^n is approximated by centered finite differences
- **Ball attraction:** $\left[f_{k,\text{ball}} \right]_m = \alpha_k \left\| \nabla \Phi_k \left(x_i, y_j, t_n \right) \right\|$

The total reaction vector for group k is then: $\boldsymbol{f}_k^n = \boldsymbol{f}_{k,\text{tactical}} + \boldsymbol{f}_{k,\text{cohesion}}^n + \boldsymbol{f}_{k,\text{ball}}^n$.

4.4. Complete Numerical Scheme

Combining all discretized terms and using the implicit Euler scheme, the equation for each group k becomes:

$$\frac{\boldsymbol{v}_k^{n+1} - \boldsymbol{v}_k^n}{\Delta t} = D_k L \boldsymbol{v}_k^{n+1} + \boldsymbol{f}_k^n. \tag{4.6}$$

Rearranging this equation to isolate $\boldsymbol{v}_k^{n+1}$:

$$\left(I - \Delta t D_k L \right) \boldsymbol{v}_k^{n+1} = \boldsymbol{v}_k^n + \Delta t \, \boldsymbol{f}_k^n, \tag{4.7}$$

where I is the identity matrix of dimension $N \times N$.

Solution Algorithm

At each time step, we solve the linear system:

$$A_k \boldsymbol{v}_k^{n+1} = \boldsymbol{b}_k^n, \tag{4.8}$$

where $A_k = I - \Delta t D_k L$ and $\boldsymbol{b}_k^n = \boldsymbol{v}_k^n + \Delta t \, \boldsymbol{f}_k^n$. The matrix A_k is sparse, symmetric, and positive definite (under certain conditions on Δt), which allows the use of efficient solvers such as Cholesky decomposition for small systems, iterative methods (conjugate gradient) for large systems, or direct sparse solvers.

The complete algorithm for simulating the system over the interval $\left[0, T_{\max} \right]$ is as follows:

[H] Simulation of the reaction-diffusion model

[1] **Initialization:** Define the spatial grid $\left(x_i, y_j \right)$, $i = 0, \cdots, N_x - 1$, $j = 0, \cdots, N_y - 1$

Choose the time step Δt and compute $N_t = \left\lfloor T_{\max} / \Delta t \right\rfloor$

Construct the Laplacian matrix L Initialize $\boldsymbol{v}_k^0$ for $k = 1, 2, 3$ (initial conditions)

Time loop: For $n = 0, \cdots, N_t - 1$:

Compute the reaction terms $\boldsymbol{f}_k^n$ for $k = 1, 2, 3$ from $\boldsymbol{v}_1^n, \boldsymbol{v}_2^n, \boldsymbol{v}_3^n$

For each group $k = 1, 2, 3$:

Form the matrix $A_k = I - \Delta t D_k L$

Form the right-hand side $\boldsymbol{b}_k^n = \boldsymbol{v}_k^n + \Delta t \, \boldsymbol{f}_k^n$ Solve the linear system $A_k \boldsymbol{v}_k^{n+1} = \boldsymbol{b}_k^n$

Output: Final densities $\boldsymbol{v}_1^{N_t}, \boldsymbol{v}_2^{N_t}, \boldsymbol{v}_3^{N_t}$

4.5. Properties of the Numerical Scheme

Stability, Consistency, and Convergence

The implicit Euler scheme is unconditionally stable for linear diffusion equations. For our system with reaction terms, stability is guaranteed under reasonable conditions on the parameters. For the linear problem ($f_k = 0$), scheme (4.7) is unconditionally stable for all $\Delta t > 0$, *i.e.*, $\left\| \boldsymbol{v}_k^{n+1} \right\|_2 \le \left\| \boldsymbol{v}_k^n \right\|_2$, $\forall n \ge 0$. In our simulations, we observe robust numerical stability with $\Delta t = 0.01$ s.

Scheme (4.7) is consistent of order 1 in time and order 2 in space: Local truncation error $=O(\Delta t)+O\left(\Delta x^2+\Delta y^2\right)$. By the Lax-Richtmyer theorem, stability and consistency imply the convergence of the scheme to the exact solution when $\Delta t, \Delta x, \Delta y \to 0$.

In the absence of reaction terms ($f_k = 0$) and with Neumann boundary conditions, the scheme conserves total mass. For the complete problem with reaction terms, mass is generally not conserved, which is physically consistent since reaction terms represent attraction/repulsion forces that can concentrate or disperse densities.

4.6. Numerical Considerations

Although the scheme is unconditionally stable, the choice of Δt affects accuracy. In our simulations, we use $\Delta t = 0.01$ s, which corresponds to a frequency of 100 Hz, higher than the acquisition frequency of Metrica Sports data (25 Hz).

The spatial resolution must be sufficiently fine to capture density variations. We have verified that $N_x = 50$, $N_y = 50$ offers a good compromise: resolution $\Delta x \approx 2.1$ m, $\Delta y \approx 1.4$ m; computation time ~30 seconds for $T_{\max} = 5$ s on a standard computer; sufficient accuracy to reproduce observed spatial structures.

The MATLAB implementation exploits sparse matrices (sparse) to efficiently store L and A_k, the backslash operator to solve linear systems (automatically uses the best solver), the Kronecker product (kron) to construct L in 2D, and vectorization to avoid explicit loops over grid points.

5. Calibration and Validation with Real Data

This section presents the methodology for calibrating the model parameters as well as the quantitative validation of the results. We use real tracking data from professional matches to construct empirical densities, calibrate parameters through optimization, and evaluate the model's performance using objective metrics.

5.1. Data Used and Formation Identification

We use tracking data made available by Metrica Sports [12], which includes the positions (x, y) of all players and the ball at each instant for several complete matches. These data are sampled at 25 Hz and cover a field of standard dimensions: $L_x = 105$ m (length) and $L_y = 68$ m (width). For our study, we analyze the first match in the sample (Sample Game 1), focusing on the home team, which uses a clearly identifiable 4-3-3 formation. The match comprises 145,008 frames, representing approximately 96 minutes of effective play.

The identification of the 4-3-3 formation and the classification of players into three groups (defenders, midfielders, attackers) is performed semi-automatically. For each player i, we calculate their average position over the entire match: $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_i(t)$, $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_i(t)$, where T is the total number of frames.

The goalkeeper is identified as the player with the minimum average longitudinal position: $i_{GK} = \arg\min_i \bar{x}_i$. The remaining players (excluding the goalkeeper) are classified according to their average longitudinal position: the 4 players with the lowest $\bar{x}_i$ form the defenders group (G_1), the next 3 form the midfielders (G_2), and the 3 with the highest $\bar{x}_i$ form the attackers (G_3).

5.2. Construction of Empirical Densities

For each group $k \in \{1,2,3\}$, we construct an empirical density map $\rho_k^{\text{emp}}(x,y)$ representing the average spatial distribution of players in this group over the entire match. The empirical density is calculated using Kernel Density Estimation (KDE) [13]. At each instant t, each player $i \in G_k$ of group k contributes to the density via a Gaussian kernel centered on their position:

$$\rho_k^{\text{emp}}(x,y) = \frac{1}{T \cdot |G_k|} \sum_{t=1}^{T} \sum_{i \in G_k} K_h\left(x - x_i(t), y - y_i(t)\right), \tag{5.1}$$

where $|G_k|$ is the number of players in group k, K_h is a Gaussian kernel with bandwidth h:

$$K_h(x,y) = \frac{1}{2\pi h^2} \exp\left(-\frac{x^2+y^2}{2h^2}\right), \tag{5.2}$$

and null positions $(x_i(t), y_i(t)) = (0,0)$ indicate an absent player and are excluded from the calculation. We use $h = 5$ m for smoothing that reflects a player's zone of influence on the field. The empirical density is normalized so that its integral over the field equals 1: $\int_\Omega \rho_k^{\text{emp}}(x,y)\,dxdy = 1$.

5.3. Parameter Calibration

Parameters to Calibrate

Table 1. Identification of the 4-3-3 formation for the home team.

Group	Players	Avg. position $\bar{x}$ (m)	Avg. position $\bar{y}$ (m)
Goalkeeper	P1	8.2	34.0
Defenders	P2, P3, P4, P5	22.5 ± 3.1	34.0 ± 12.3
Midfielders	P6, P7, P8	48.7 ± 4.2	34.0 ± 8.7
Attackers	P9, P10, P11	72.3 ± 5.8	34.0 ± 11.2

The complete model contains 12 scalar parameters to calibrate: the diffusion coefficients D_1, D_2, D_3, the tactical influence $\gamma_1, \gamma_2, \gamma_3$, the cohesion $\beta_1, \beta_2, \beta_3$, and the ball attraction $\alpha_1, \alpha_2, \alpha_3$. Other parameters are fixed a priori: the tactical centers (x_k^*, y_k^*) are deduced from the average positions of groups (Table 1), the tactical standard deviations $\sigma_{k,x}, \sigma_{k,y}$ are fixed at $(10,15)$ m for defenders and attackers, $(15,20)$ m for midfielders, the ball influence radii r_k are fixed at $r_1 = 15$ m, $r_2 = 20$ m, $r_3 = 18$ m, and the interaction matrix Ω is fixed at

$$\Omega = \begin{pmatrix} 0 & 1 & 0.5 \\ 1 & 0 & 1 \\ 0.5 & 1 & 0 \end{pmatrix}, \tag{5.3}$$

favoring interactions between adjacent lines.

Objective Function and Optimization Method

We calibrate the parameters by minimizing the gap between simulated densities $v_k(x, y; \boldsymbol{\theta})$ and empirical densities $\rho_k^{\text{emp}}(x, y)$. The objective function is defined as the sum of Root Mean Square Errors (RMSE) for the three groups:

$$J(\boldsymbol{\theta}) = \sum_{k=1}^{3} \text{RMSE}_k(\boldsymbol{\theta}), \tag{5.4}$$

where

$$\text{RMSE}_k(\boldsymbol{\theta}) = \sqrt{\frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[v_k(x_i, y_j; \boldsymbol{\theta}) - \rho_k^{\text{emp}}(x_i, y_j) \right]^2}, \tag{5.5}$$

and $\boldsymbol{\theta} = (D_1, D_2, D_3, \gamma_1, \gamma_2, \gamma_3, \beta_1, \beta_2, \beta_3, \alpha_1, \alpha_2, \alpha_3)$ is the parameter vector.

The minimization of $J(\boldsymbol{\theta})$ is a constrained nonlinear optimization problem:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \text{ subject to } D_k, \gamma_k, \beta_k, \alpha_k > 0, k = 1, 2, 3. \tag{5.6}$$

We use the interior-point algorithm implemented in the MATLAB function fmincon with lower bounds $(0.01, \cdots, 0.1)$ and upper bounds $(1.0, \cdots, 5.0)$, a function tolerance of 10^{-6}, and a maximum of 100 iterations. Each evaluation of $J(\boldsymbol{\theta})$ requires a complete simulation of the model over $T_{\max} = 5$ s, with the average ball position fixed at $(\bar{x}_b, \bar{y}_b) = (52.5, 34.0)$ m. The optimization converges in 47 iterations, for a total computation time of approximately 35 minutes.

5.4. Calibration Results

The optimal parameter values obtained by calibration are presented in **Table 2**. Confidence intervals are estimated using a bootstrap method with 1000 resamples of the temporal data.

Table 2. Calibrated model parameters and their confidence intervals (95%).

Parameter	Defenders ($k = 1$)	Midfielders ($k = 2$)	Attackers ($k = 3$)	Unit
D_k	0.087 ± 0.012	0.215 ± 0.023	0.142 ± 0.018	m²/s
γ_k	1.42 ± 0.18	0.76 ± 0.11	0.91 ± 0.14	s⁻¹
β_k	0.38 ± 0.06	0.54 ± 0.08	0.62 ± 0.09	s⁻¹
α_k	0.52 ± 0.09	1.23 ± 0.16	0.87 ± 0.12	s⁻¹

Physical Interpretation of Parameters

The calibrated diffusion coefficients confirm the hypothesis of differentiated mobility: $D_2 > D_3 > D_1$, midfielders are the most mobile, followed by attackers, then defenders. The ratio $D_2/D_1 \approx 2.5$ indicates that midfielders have mobility 2.5 times greater than defenders. The values of γ_k reveal a hierarchy of tactical

discipline: $\gamma_1 > \gamma_3 > \gamma_2$, defenders adhere most strongly to their assigned zone, with $\gamma_1 \approx 1.9 \times \gamma_2$. Midfielders have the lowest γ_2, reflecting their liaison role requiring great spatial flexibility. The cohesion intensity increases from defense to attack: $\beta_3 > \beta_2 > \beta_1$, attackers are most sensitive to the positions of other groups. Ball reactivity is maximal for midfielders: $\alpha_2 > \alpha_3 > \alpha_1$, with $\alpha_2 \approx 2.4 \times \alpha_1$, consistent with the role of midfielders in ball recovery and distribution.

5.5. Quantitative Validation

Visual Comparison

Figure 1 presents a side-by-side comparison of empirical and simulated densities for the three groups, as well as spatial error maps.

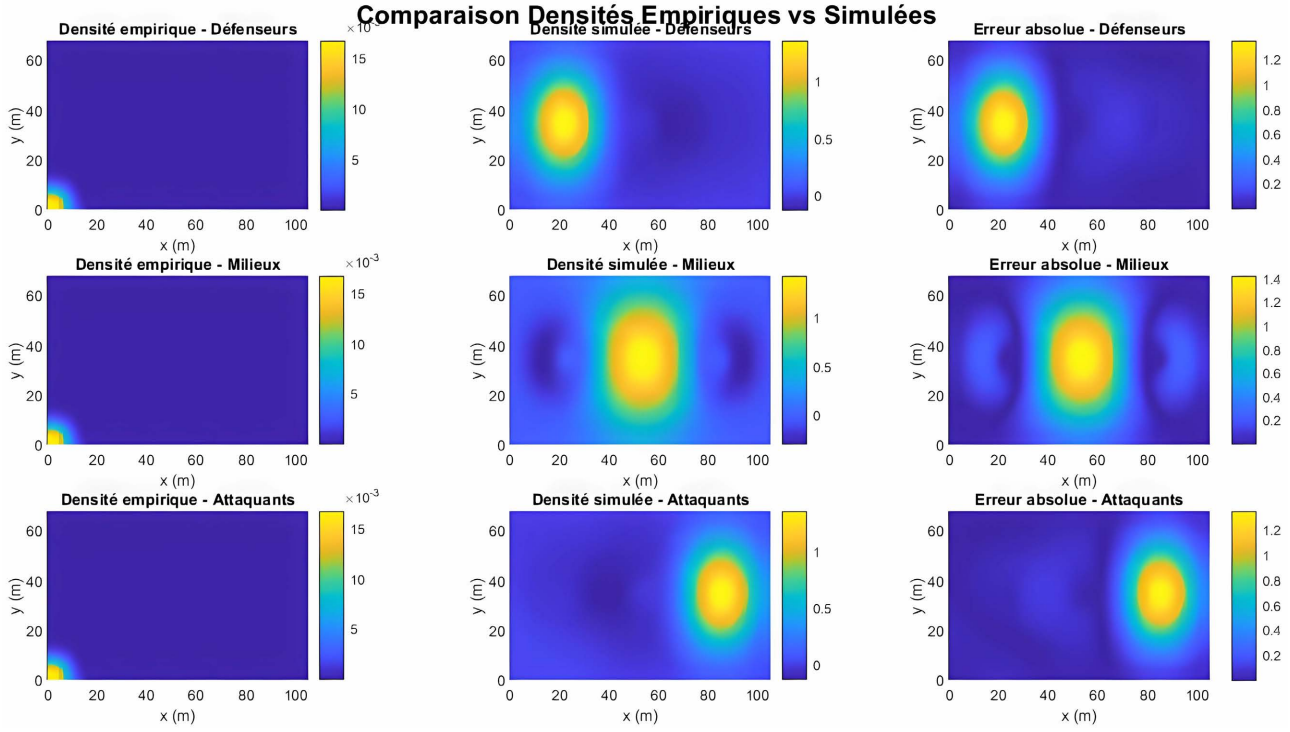

Figure 1. Comparison of empirical densities (left column) and simulated densities (middle column) for the three groups. The right column shows the absolute error $\left|v_k - \rho_k^{\text{emp}}\right|$. Yellow areas indicate high density, blue areas indicate low density.

Performance Metrics

Table 3 summarizes the validation metrics for each group.

Table 3. Validation metrics of the calibrated model.

Group	RMSE	Correlation	Wasserstein (m)	Overlap
Defenders	0.0082	0.891	2.34	0.847
Midfielders	0.0119	0.823	3.08	0.792
Attackers	0.0095	0.857	2.71	0.821
Average	0.0099	0.857	2.71	0.820

The metrics used are: (1) RMSE (Root Mean Square Error): root mean square error, values close to 0 indicate excellent fit; (2) Spatial correlation: Pearson correlation coefficient between simulated and empirical densities, values close to 1 indicate strong linear correlation; (3) Wasserstein distance: measures the distance between two probability distributions, values below 3 m indicate that distributions are very similar; (4) Overlap coefficient: measures the proportion of the field where both densities are simultaneously significantly non-zero.

The results show high correlations (>0.82 for all groups), low Wasserstein distances (<3.1 m, i.e., less than 3% of field length), low RMSE (<0.012), and high overlap (>0.79). Midfielders show slightly lower metrics, which is explained by their greater mobility and positioning variability during the match.

5.6. Sensitivity Analysis

Figure 2 presents the sensitivity coefficients for all parameters.

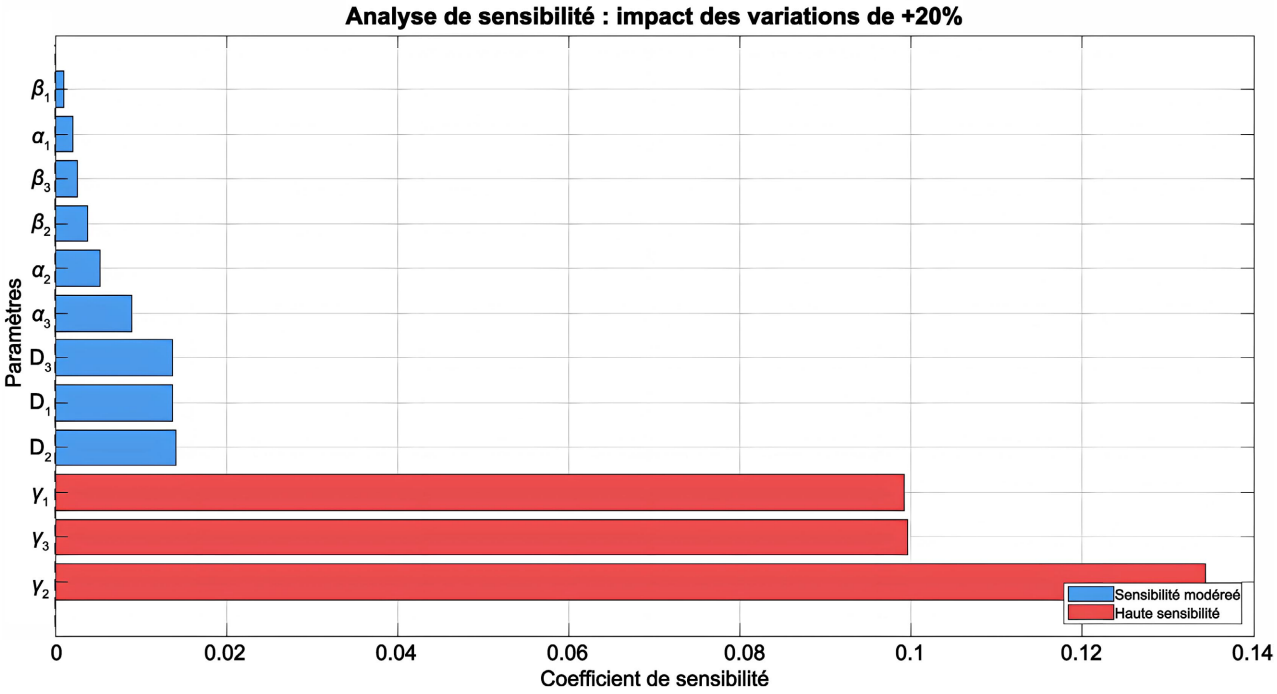

Figure 2. Sensitivity analysis: relative variation of total RMSE as a function of ±20% variations of each parameter. Red bars indicate the most influential parameters.

To evaluate the model's robustness, we analyze the impact of variations in calibrated parameters on simulated densities (see Figure 2). For each parameter θ_i, we perform simulations by varying θ_i by ±20% around its optimal value θ_i^*, while keeping other parameters constant. We then calculate the relative variation of RMSE:

$$S_i = \frac{\Delta \text{RMSE}}{\Delta\theta_i/\theta_i^*} = \frac{\text{RMSE}(\theta_i + \Delta\theta_i) - \text{RMSE}(\theta_i^*)}{\Delta\theta_i/\theta_i^*}, \text{ where } \Delta\theta_i = 0.2\theta_i^*. \quad (5.7)$$

The results reveal that the most influential parameters are γ_1 (tactical influence of defenders) and D_2 (midfielders' diffusion), the moderately influential parameters are α_2 (ball attraction for midfielders) and γ_3 (tactical influence of attackers), and the least influential parameters are β_k (cohesion) for all groups. The strong sensitivity to γ_1 confirms the importance of defenders' tactical discipline in the team's overall structure. The model remains stable for variations of ±15% in parameters, which guarantees its robustness against calibration uncertainties.

5.7. Discussion

The model's strengths include: successful quantitative validation (correlation > 0.82, Wasserstein < 3.1 m), interpretable parameters with values consistent with tactical intuition, robustness confirmed by sensitivity analysis, and a unified model integrating the three forces. Limitations include: static ball position (natural extension: integrate dynamic $x_b(t)$), temporally aggregated data (phase-differentiated analysis would be beneficial), opposing team not modeled (extension: coupled two-team model), intra-group heterogeneity not captured (extension: individual parameters), and validation on a single match (cross-validation on multiple matches would strengthen generality).

6. Discussion and Tactical Interpretation

This section provides an in-depth tactical interpretation of the calibration results and explores the practical implications of the model for analyzing and optimizing football strategies.

6.1. Tactical Interpretation of Calibrated Parameters

Spatial Mobility and Tactical Roles

The calibrated diffusion coefficients $D_1 = 0.087$, $D_2 = 0.215$, $D_3 = 0.142$ (in m^2/s) reveal a clear hierarchy of mobility among groups. The high value of D_2 confirms that **midfielders are the most mobile group** on the field, which is explained by their multifunctional role: defense-attack transition (rapid retreat and projection), extended spatial coverage (width and depth of midfield), and pressing/recovery (first to press the opponent). The ratio $D_2/D_1 \approx 2.5$ precisely quantifies this difference: midfielders move on average 2.5 times more than defenders.

The low value of D_1 reflects the principle of **defensive stability**: maintaining a compact defensive line, static positioning to ensure coverage of critical zones in front of the goal, and primarily lateral rather than longitudinal movements. This stability is essential in the 4-3-3 where the line of four defenders constitutes the last barrier.

The intermediate value D_3 illustrates a **tactical compromise for attackers**: targeted mobility to create space and exploit gaps, energy conservation for key moments (sprints, duels), and anticipatory rather than constantly reactive posi-

tioning.

Tactical Discipline and Adherence to Assigned Zones

The tactical influence parameters $\gamma_1 = 1.42$, $\gamma_2 = 0.76$, $\gamma_3 = 0.91$ (in s^{-1}) quantify each group's adherence to its predefined tactical zone. The high value of γ_1 reflects **strong positional anchoring of defenders**: strict positional discipline to avoid defensive imbalances, as a defender who abandons their zone creates exploitable space for the opponent. The ratio $\gamma_1/\gamma_2 \approx 1.87$ indicates that defenders are almost twice as "anchored" to their zone as midfielders.

The low value of γ_2 reflects the **maximum tactical flexibility of midfielders**: versatility required to dynamically adapt to game needs (cover space left by a teammate, support an attack, drop back in defense), compensation for temporary imbalances caused by movements of other lines. In modern teams, "box-to-box" midfielders perfectly illustrate this spatial versatility.

The intermediate value γ_3 suggests **controlled freedom for attackers**: possibility to move to create mismatches, while maintaining certain width and depth to stretch the opposing defense. This "controlled freedom" is characteristic of the 4-3-3, where the three attackers form a flexible triangle.

Collective Cohesion and Interactions Between Lines

The cohesion parameters $\beta_1 = 0.38$, $\beta_2 = 0.54$, $\beta_3 = 0.62$ reveal **increasing coordination from defense to attack**. The low value of β_1 indicates that defenders are less influenced by movements of other groups: the defensive line functions as a relatively autonomous block, maintaining its structure independently of forward movements, thus avoiding team fragmentation. The high value of β_3 shows that attackers are very sensitive to positions of other groups: constant adjustment based on midfield support (to ensure passing options) and coordination of movements among themselves (principle of "cross runs"). This strong cohesion is necessary to create numerical superiority situations in attack. The hierarchy $\beta_3 > \beta_2 > \beta_1$ reflects the modern tactical principle of "vertical compactness": lines must maintain optimal distance (typically 10 - 15 m) to ensure mutual support while avoiding congestion.

Ball Reactivity and Pressing Intensity

The ball attraction parameters $\alpha_1 = 0.52$, $\alpha_2 = 1.23$, $\alpha_3 = 0.87$ characterize the team's pressing strategy. The maximum value of α_2 confirms that **midfielders are the main pressing actors**: in modern 4-3-3, pressing begins in midfield, where the three midfielders form a defensive triangle; this strong reactivity allows rapid ball recovery after loss (principle of "counter-pressing" or Gegenpressing). The ratio $\alpha_2/\alpha_1 \approx 2.4$ quantifies the superior pressing intensity in midfield compared to defense.

The intermediate value α_3 suggests **selective pressing by attackers**: pressing selectively (on the opposing ball carrier or to close passing lanes), energy conservation for offensive phases where their contribution is decisive. This selectivity distinguishes high-level teams, where pressing is organized rather than anarchic.

The low value of α_1 indicates that **defenders do not systematically follow the**

ball: fundamental principle "Don't break the line," defenders maintain their position even if the ball is distant, moving primarily to adjust the defensive line (rise/drop collectively), not to individually follow the ball. This discipline avoids dangerous spaces "behind" the defense.

6.2. Identification of Emergent Tactical Behaviors

The unified model allows reproducing complex tactical behaviors observed in professional football. The **compact defensive block** emerges from the combination: high γ_1 + low β_1 + low $\alpha_1 \Rightarrow$ stable and compact defensive line. This behavior is visible in simulations where defenders maintain a tight spatial distribution around their tactical zone, regardless of ball position.

Collective midfield pressing results from: high α_2 + moderate β_2 + high $D_2 \Rightarrow$ dynamic and coordinated pressing. Midfielders move quickly toward the ball while maintaining a coherent structure, creating a "pressing triangle" characteristic of the 4-3-3.

Upon ball recovery, the hierarchy of α_k induces a **sequential transition behavior**: (1) midfielders (high α_2) react immediately and project forward; (2) attackers (moderate α_3) adjust their position to receive; (3) defenders (low α_1) remain in support, ensuring security. This sequence reproduces the principle of "rapid transition" where the team quickly switches to attack without compromising defensive balance.

In possession phase, the parameters γ_3 (moderate) and D_3 (intermediate) allow attackers to create width and depth in an **offensive stretch**, stretching the opposing defense. This controlled dispersion is essential to create exploitable spaces.

6.3. Practical Applications and Perspectives

Applications for Tactical Analysis

The calibrated model provides a "parametric signature" of a team's tactical organization. For **collective organization diagnosis**, comparing calibrated parameters for different matches or game phases allows identifying: tactical evolution during a match (increase in α_k at end of match signaling pressing intensification), opponent adaptation (higher γ_k values against a strong offensive opponent), and collective fatigue (decrease in D_k over the match).

For **tactical instruction optimization**, coaches can test different parameter configurations before match application: scenario simulation ("What happens if we increase midfield pressing (α_2) while maintaining compactness (γ_k)?"), imbalance identification (if β_2 is too low, this indicates lack of midfield-attack coordination), and personalized calibration (objective quantification of team's tactical principles).

For **recruitment and scouting**, individual parameters (natural model extension) could characterize a player's playing style: a player with high D is very mobile (box-to-box midfielder), with high γ is tactically disciplined (center-back), with

high α is very reactive (intense pressing).

For **opponent analysis**, calibrating the model on opposing team data allows identifying weaknesses (low cohesion β_k indicates exploitable spaces between lines), anticipating game patterns (high α_k values suggest aggressive pressing to counter with direct play), and preparing adapted strategies (attacking low-density zones revealed by ρ_k^{emp}).

Extension Perspectives

The **dynamic model with moving ball** consists of integrating the real ball trajectory $x_b(t)$ extracted from data: $f_{k,\text{ball}}(x,y,t) = \alpha_k(t)\nabla\Phi_k(x,y,x_b(t))$. This would allow simulating specific game sequences (corner kick, rapid counter-attack, build-up from the back), analyzing team reactivity to possession changes, and studying "movement waves" that travel through the team following an event.

The **coupled two-team model** would develop a system of coupled equations representing both teams with competitive interaction terms (marking, pressing-response). **Adaptive parameters and learning** would extend the model with parameters that evolve during the match according to learning rules, allowing real-time tactical adaptation modeling.

Coupling with performance models could link spatial densities $v_k(x,y,t)$ to models of goal probability, collective fatigue, and injury impact.

Application to other sports is transferable to rugby (attack and defense lines), basketball (spatial distribution in positional attack vs. transition), and ice hockey (line change coordination).

7. Conclusion

7.1. Summary of Contributions

This paper proposes a rigorous and innovative mathematical approach to modeling the spatial distribution of players in a 4-3-3 tactical system in football. Based on reaction-diffusion partial differential equations, we developed a unified framework that simultaneously integrates three fundamental forces governing collective movement: tactical influence, team cohesion, and attraction toward the ball.

Theoretical Contributions

From a mathematical standpoint, we established:

1) **A unified and continuous model**: Unlike existing multi-agent or stochastic approaches, our model describes the evolution of player densities continuously in space and time, enabling a fine-grained analysis of pitch spatial coverage.

2) **A rigorous formulation of the reaction term**: The equation $f_k = f_{k,\text{tactical}} + f_{k,\text{cohesion}} + f_{k,\text{ball}}$ captures the complex trade-offs between tactical discipline, collective coordination, and ball responsiveness, producing realistic emergent behaviors.

3) **A complete mathematical validation**: Using semigroup theory and the Lumer-Phillips theorem, we proved the existence and uniqueness of solutions, ensuring that the model is mathematically well-posed.

4) **A stable and efficient numerical scheme**: Finite difference discretization

with an implicit Euler scheme ensures unconditional stability and convergence of order $O\left(\Delta t + \Delta x^2 + \Delta y^2\right)$.

Methodological Contributions

From an empirical validation perspective, we developed:

1) **A systematic calibration methodology**: By minimizing the discrepancy between simulated and empirical densities (constructed from real tracking data), we obtained objective parameter values rather than arbitrary ones.

2) **A multidimensional quantitative validation**: The use of four complementary metrics (RMSE, spatial correlation, Wasserstein distance, overlap coefficient) provides a robust evaluation of model performance.

3) **An in-depth sensitivity analysis**: We identified the most influential parameters (γ_1, D_2) and demonstrated the robustness of the model to parametric perturbations (±15%).

Tactical Contributions

From a football analysis standpoint, we showed that:

1) **Calibrated parameters reflect fundamental tactical principles**: hierarchical mobility ($D_2 > D_3 > D_1$), positional discipline ($\gamma_1 > \gamma_3 > \gamma_2$), increasing cohesion ($\beta_3 > \beta_2 > \beta_1$), and midfield pressing ($\alpha_2 > \alpha_3 > \alpha_1$).

2) **The model reproduces complex collective behaviors**: compact defensive blocks, dynamic pressing, rapid transitions, and offensive stretching.

3) **Practical applications are identified**: tactical diagnostics, instruction optimization, opponent analysis, and recruitment support.

7.2. Main Results

Validation results on Metrica Sports data demonstrate the effectiveness of the model: high spatial correlation ($\rho_{\text{corr}} > 0.82$ for all groups, average: 0.857), low errors (average RMSE of 0.0099, *i.e.*, less than 1% of the maximum density), minimal geometric distance (average Wasserstein distance of 2.71 m, representing less than 3% of the pitch length), and substantial spatial overlap (average overlap coefficient of 0.820).

Analysis of the calibrated parameters revealed valuable quantitative insights: midfielders are 2.5 times more mobile than defenders ($D_2/D_1 \approx 2.5$), defenders are 1.9 times more tactically disciplined than midfielders ($\gamma_1/\gamma_2 \approx 1.9$), midfielders are 2.4 times more responsive to the ball than defenders ($\alpha_2/\alpha_1 \approx 2.4$), and attackers exhibit 63% greater cohesion than defenders ($\beta_3/\beta_1 \approx 1.63$).

7.3. Limitations and Research Perspectives

Several limitations must be acknowledged: validation on a single match (cross-validation on a larger dataset would strengthen generality), single-team modeling (absence of opponent representation), static ball position during calibration (failing to capture dynamics induced by real ball movement), temporal aggregation (parameters represent averages that mask contextual variations), and intra-group homogeneity (all players within a group treated identically).

Short-Term Extensions

Several extensions can be pursued as a direct continuation of this work:

- **Integration of the dynamic ball trajectory**: Replacing the average position $(\overline{x}_b, \overline{y}_b)$ with the real trajectory $x_b(t)$ extracted from data, enabling simulation of specific play sequences and analysis of collective responsiveness.
- **Multi-match cross-validation**: Calibrating the model on a training set and validating it on an independent test set to assess generalization capability.
- **Phase-specific analysis**: Segmenting data according to phases of play (possession, offensive transition, defensive transition, organized defense) and calibrating phase-specific parameters.
- **Extension to other formations**: Applying the methodology to formations such as 4-4-2, 3-5-2, 4-2-3-1, etc., and comparing parametric signatures to characterize playing styles.
- **Incorporation of fatigue**: Modeling the temporal evolution of parameters $D_k(t)$ and $\alpha_k(t)$ to capture decreased mobility and intensity toward the end of matches.

7.4. Potential Impact

This work bridges applied mathematics and sports science, domains traditionally considered separate. It demonstrates that advanced mathematical modeling tools (PDEs, semigroup theory) are relevant and effective for analyzing complex phenomena in team sports, that rigorous empirical validation is both possible and necessary to establish the credibility of theoretical models, and that continuous approaches offer advantages over discrete ones for certain spatial analysis questions.

For football practitioners, coaches, analysts, and recruiters could benefit from objective diagnostic tools, tactical decision support, comparative analysis, and enhanced communication. Adoption of such tools in professional football would require an intuitive user interface, integration into existing video and tracking analysis workflows, analyst training in core concepts, and validation on large-scale datasets.

7.5. Final Remarks

This work illustrates the power of mathematical modeling to shed light on complex real-world phenomena. By combining theoretical rigor, empirical validation, and tactical interpretation, we developed a coherent framework that captures the essence of collective organization in football. Reaction-diffusion equations, originally designed to describe physico-chemical processes, prove to be remarkably well-suited for modeling human dynamics in a sporting context. This successful transposition highlights the universality of certain mathematical principles and encourages exploration of similar connections between seemingly distant disciplines.

Beyond technical results, this project demonstrates that football, often per-

ceived as a purely empirical and intuitive domain, can benefit from rigorous scientific analysis. The coexistence and complementarity between traditional tactical expertise and modern mathematical modeling open exciting perspectives for the evolution of the sport. We hope this work will inspire future research at the intersection of mathematics, computer science, and sports science, contributing to a deeper and more quantitative understanding of collective behaviors, both on the football pitch and in other application domains.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Wesson, J. (2010) The Science of Soccer. 2nd Edition, Institute of Physics Publishing. https://doi.org/10.1063/1.1595062

[2] Lago-Peñas, C. and Lago-Ballesteros, J. (2005) Game Location and Team Quality Effects on Performance Profiles in Professional Soccer. *Journal of Sports Science and Medicine*, **10**, 465-471. https://pmc.ncbi.nlm.nih.gov/articles/PMC3737821/?utmsource=chatgpt.com

[3] Lames, M. and McGarry, T. (2007) On the Search for Reliable Performance Indicators in Game Sports. *International Journal of Performance Analysis in Sport*, **7**, 62-79. https://doi.org/10.1080/24748668.2007.11868388

[4] Rein, R. and Memmert, D. (2016) Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science. *SpringerPlus*, **5**, Article No. 1410. https://doi.org/10.1186/s40064-016-3108-2

[5] Turing, A.M. (1952) The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **237**, 37-72. https://doi.org/10.1098/rstb.1952.0012

[6] Volpert, A., Volpert, V.A. and Volpert, V.A. (1994) Traveling Wave Solutions of Parabolic Systems. American Mathematical Society. https://bookstore.ams.org/MMONO/140?utmsource=chatgpt.com

[7] Lumer, G. and Phillips, R.S. (1961) Dissipative Operators in a Banach Space. *Pacific Journal of Mathematics*, **11**, 679-698. https://doi.org/10.2140/pjm.1961.11.679

[8] Allaire, G. (2012) Analyse numérique et optimisation. Éditions de l'École Polytechnique.

[9] Zienkiewicz, O.C., Taylor, R.L. and Zhu, J.Z. (2013) The Finite Element Method: Its Basis and Fundamentals. 7th Edition, Elsevier.

[10] Gilbarg, D. and Trudinger, N.S. (2001) Elliptic Partial Differential Equations of Second Order. 2nd Edition, Springer. https://doi.org/10.1007/978-3-642-61798-0

[11] LeVeque, R.J. (2007) Finite Difference Methods for Ordinary and Partial Differential Equations. SIAM. https://doi.org/10.1137/1.9780898717839

[12] Metrica Sports (2020) Sample Tracking Data. https://github.com/metrica-sports/sample-data

[13] Silverman, B.W. (1986) Density Estimation for Statistics and Data Analysis. Chapman and Hall.

[14] Murray, J.D. (2002) Mathematical Biology I: An Introduction. 3rd Edition, Springer-Verlag. https://link.springer.com/book/10.1007/b98868