

Article

Not peer-reviewed version

A New Linear Two-State Dynamical Model for Athletic Performance Prediction in Elite-Level Soccer Players

[Nicolò Colistra](#)*, [Vincenzo Manzi](#)*, Samir Maikano, [Francesco Laterza](#), Rosario D'Onofrio, [Cristiano Maria Verrelli](#)

Posted Date: 9 September 2025

doi: 10.20944/preprints202509.0709.v1

Keywords: Mathematical modeling; nonlinear dynamics; performance prediction; training load; GPS technology; metabolic power; soccer



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A New Linear Two-State Dynamical Model for Athletic Performance Prediction in Elite-Level Soccer Players

Nicolò Colistra ^{1,2,*}, Vincenzo Manzi ^{3,*}, Samir Maikano ⁴, Francesco Laterza ⁵,
Rosario D'Onofrio ⁶ and Cristiano Maria Verrelli ¹

¹ Department of Electronic Engineering, University of Rome Tor Vergata, 00133 Rome, Italy

² Division of Biotechnologies, ENEA, 00123 Rome, Italy

³ Department of Wellbeing, Nutrition and Sport, Pegaso Open University, 80143 Naples, Italy

⁴ Open Fiber S.p.A., PMO – Operational Reporting Cluster C&D, 00142 Rome, Italy

⁵ Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, 37131 Verona, Italy

⁶ Department of Wellbeing, Nutrition and Sport, Pegaso Open University, 80143 Naples, Italy

* Correspondence: colistra@ing.uniroma2.it; vincenzo.manzi@unipegaso.it

[†] These authors contributed equally to this work.

Abstract

Recent advancements in wearable technology have allowed researchers to collect high-resolution data on athletes' workloads and performance, paving the way for more accurate mathematical models in sports science. In this paper, inspired by the modeling of heart rate during exercise, we introduce a novel linear, time-varying, two-state discrete-time dynamical model for predicting athletic performance in elite-level soccer players. Model parameters are estimated *via* the Differential Evolution optimization algorithm, and GPS-derived metrics such as metabolic power and equivalent distance index are incorporated. The model originally accounts for complex interactions between a performance-related state variable and a second lumped variable, whose dynamics are intertwined. This model was compared to the most effective one in the literature, namely the (nonlinear) Busso model. Results, concerning two professional soccer players over a half-season period, show that the proposed model outperforms the traditional approach in predictive accuracy, with significantly higher correlation coefficients and lower prediction errors across all players. These findings suggest that integrating two-state dynamics and fine-grained GPS metrics provides a more biologically realistic framework for load monitoring in team sports. The proposed model thus represents a powerful tool for training optimization and athlete readiness assessment, with potential applications in real-time decision support systems for coaching staff. By predicting the effects of training load on future performance, it might also contribute to injury risk reduction and the prevention of maladaptive responses to excessive workload.

Keywords: mathematical modeling; nonlinear dynamics; performance prediction; training load; GPS technology; metabolic power; soccer

1. Introduction

In recent decades, mathematical modeling has become a fundamental tool in sports science, allowing sports and data scientists to better understand the close relationship between training load and athletes' performance capacity [1]. The data obtained from such modeling have provided coaches with valuable support for the prescription of increasingly effective conditioning programs aimed at maximizing athletes' performance capacity at key moments, while avoiding illness, injury or overtraining.

The origins of these mathematical models date back to the 1970s through the studies of Banister, whose fitness-fatigue model represented the first attempt to formalize the relationship between training stimulus and performance [1]. The model proposed by Banister is based on the assumption that athletes'

performance capacity is the result of two competing factors: a first factor, called fitness, represents the positive effects of training on performance, while a second factor, called fatigue, describes the negative effects that reduce performance capacity. From a practical point of view, the model predicts that after a workload there is a brief decrease in performance due to accumulated fatigue, followed after adequate recovery by an improvement in physical capacity thanks to the adaptations induced by training [1].

Subsequent to the publication of this model, numerous studies have confirmed its effectiveness in predicting the performance capacity of athletes in several individual sports. In fact, this model has been applied in disciplines such as running [2], cycling [3], swimming [1,4], demonstrating in many of these studies a notable accuracy in predicting the actual performance of athletes [1,3–5].

More recently, research conducted by Busso has allowed to further develop and refine Banister's model, successfully introducing new parameters specifically designed to capture the effects of training monotony—defined as low variability in daily workloads—and the cumulative fatigue that arises when intense sessions are performed consecutively without adequate recovery. These additions enhance the model ability to simulate real-world training responses [6]. Although these models have demonstrated their validity in many individual sports, they have proven to be less accurate when applied to athletes in team sports such as soccer, where performance is influenced by multiple physiological factors [7,8]. Indeed, a soccer match requires players to undergo numerous phases of intermittent exercise characterized by periods of high intensity alternating with phases of lower intensity or recovery [9]. This highly variable metabolic demand, to which players are subjected, represents a complex challenge for traditional models, which are often based on the assumption of a linear relationship between training load and performance.

Furthermore, soccer players are often engaged in multiple daily training sessions, each characterized by different intensities and physiological objectives (e.g. strength and power training in the morning and technical/tactical training in the afternoon). The variability of the physiological objectives of the training sessions can produce cumulative effects on the performance capacity of athletes, which are difficult to be predicted [2]. Recent studies have increasingly focused on developing predictive models of athletic performance in soccer, adopting approaches such as time-series analysis, machine learning techniques, and piecewise-linear formulations to improve the accuracy of performance forecasts [10–12]. Nevertheless, these approaches present relevant limitations such as reliance on subjective self-reports instead of GPS-derived external load [10], a focus on short-window extreme events with small sample sizes and acknowledged reliability constraints [11], or consideration of only the last training sessions without an explicit dynamical representation of training–performance interactions [12].

Consequently, building effective mathematical models in a sport such as soccer requires the inclusion of multiple and complex variables that reflect the athlete's performance ability in a dynamic context typical of game actions. Recent advances in wearable technology have provided sports scientists with tools able to acquire more detailed data on athletes' performance both during training sessions and during competitions [13,14]. Devices such as GPS systems and accelerometers enable real-time monitoring of critical performance parameters, including speed, acceleration, metabolic power and workload [15].

These technologies have generated enormous amounts of data, providing new opportunities to refine existing models or create more sophisticated ones. For example, some studies have shown that soccer performance is more accurately predicted when these parameters are combined together to calculate training loads, which take into account the intermittent nature of the sport [15,16].

Based on these considerations, the original contribution of this paper is to demonstrate that a new linear, time-varying, two-state model—linking the dynamics of a performance-related state variable with those of a lumped state variable, and whose parameters are estimated *via* the Differential Evolution optimization method—achieves more accurate performance predictions than the (nonlinear) Busso model. This improvement is made possible by integrating a wider range of performance metrics derived from GPS technology. The proposed model provides a more comprehensive evaluation of the

interplay between training and recovery, offering a more reliable tool for predicting performance in team sports such as soccer.

Importantly, by modeling the effects of training load on future performance, the proposed modeling approach can also support the early identification of excessive or poorly distributed training sessions, helping coaches optimize training loads and prevent maladaptive responses and reduce the risk of injury. This makes it not only a predictive model, but also a preventive tool for performance sustainability and long-term athlete's health.

2. Materials and Methods

2.1. Subjects

Two male professional soccer players were recruited from the Italian Serie A League in order to monitor their readiness during a competitive season. Players were active members of a Serie A team and had at least 5 years of competitive experience in the top division.

2.2. Video Match Analysis

A Match analysis was conducted using the STATS SportVU® system, a validated multicamera video analysis tool from STATS LLC in Chicago, United States. The STATS SportVU® system is a sophisticated and widely used technology in sports analytics, especially soccer. It utilizes multiple high-definition cameras installed at various angles around the stadium to track players and the ball at rates up to 25 Hz, with a measurement accuracy determined by the Technical University of Munich (TUM) to have a typical error of 2.7% for total distance [17]. K-Sport provided raw data in Cartesian coordinates, and the primary data were smoothed at 5 Hz [18]. STATS SportVU® uses high-definition cameras, advanced software, and statistical algorithms to extract and process player (X, Y) and ball (X, Y, Z) coordinates [19]. Roof-level cameras captured player movements during matches, and the data were analyzed using STATS Viewer and STATS Dynamix, both provided by STATS Perform, to create a dataset of each player's physical and technical performance. A specific software (<https://www.gpexe.com>) was used to download and analyze data extracted from STATS SportVU, including metabolic power ($\text{W} \cdot \text{kg}^{-1}$), equivalent distance index (%), and other metabolic-energetic metrics. Data were filtered according to the theoretical model based on an energetic approach where the energy cost of accelerations and decelerations plays a central role, as suggested by di Prampero et al. [20].

2.3. GPS Technology

The external load during each training session was evaluated using high-frequency GPS Pro2 devices operating at 18 Hz (GPEXE®, Exelio srl, Udine, Italy). The reliability, accuracy, and validity of the GPEXE® system have been reported in previous studies [21,22]. This level of resolution allows for a detailed assessment of players' running patterns, directional changes, and speed profiles during both training sessions and official matches. The technology provides coaches with real-time performance metrics, supporting data-driven decision-making. The system is capable of calculating over 300 metrics in real-time and ensures rapid and accurate data transfer. Additionally, it supports three-dimensional tracking of individual players or teams over time, enabling assessments of positional workloads, training intensities, and physiological demand changes. All GPS devices were activated 15 minutes prior to data collection to ensure optimal signal quality and satellite lock. Signal quality was evaluated based on the number of connected satellites (11 ± 1.4) and the horizontal dilution of precision (0.9 ± 0.1), according to the criteria established by Witte and Wilson [23].

2.4. Parameter Estimation via Differential Evolution (DE)

To identify the optimal parameter values for both the performance prediction models (namely, the proposed two-state model and the comparative Busso model), we employed the Differential Evolution (DE) optimization algorithm. DE is particularly well-suited for solving constrained nonlinear

optimization problems and has been implemented in Python to automate the parameter estimation process.

The DE algorithm operates by maintaining a population of candidate solutions and iteratively combining them through mutation, crossover, and selection mechanisms to explore the search space. This evolutionary approach ensures a global search capability while progressively minimizing the objective function, defined here as the Sum of Squared Errors (SSE) between the model-predicted performance and the actual observed performance data.

In our implementation, the SSE cost function was evaluated over the entire time series, and the algorithm was initialized with random values within the physiologically meaningful bounds subsequently reported in Table 1. The optimization was repeated multiple times (with different seeds) to avoid convergence to local minima and ensure robustness of the estimated parameters.

Table 1. Model parameters, their description, and the suggested bounds used during model estimation.

Parameter	Description	Bound
τ_1	Time constant related to the decay of the lumped variable over time in the performance dynamics	[5, 30]
τ_2	Time constant representing the decay of the training stimulus effectiveness in the performance dynamics	[30, 100]
τ_3	Time constant related to the decay of the performance over time in the lumped variable dynamics	[5, 60]
a_1	Coefficient that regulates how the performance evolves over time	[0.6, 1.1]
a_2	Coefficient weighting the lumped variable influence on the performance	[-8, 8]
a_3	Coefficient that regulates how the lumped variable evolves over time	[-1.2, 1.2]
a_4	Scaling parameter that adjusts the influence of the performance on the lumped variable dynamics	[-20, 20]
b_1	Gain parameter that controls the effect of the current training load $u(k)$ on the performance	[-0.2, 0.8]
b_2	Gain parameter that controls the contribution of training load to the lumped variable	[-6, 2.5]

The DE-based estimation procedure was applied individually for each subject and for both models, allowing direct comparison of model fidelity across different formulations.

This parameter identification process ensures that each model configuration is optimally fitted to the athlete’s data, enabling reliable assessment of predictive accuracy and physiological interpretability.

2.5. Study Design

This study aimed at evaluating a novel mathematical model for predicting performance in elite-standard soccer players. Specifically, the study focuses on players participating in the Italian Premier Division (Serie A) during the 2018–2019 season. They were monitored over half a season (6 months), including the pre-season period, the first part of the regular season, and all official matches (Serie A and National League matches), which took place from July to December during the 2018–2019 season. During the pre-season, players underwent training sessions seven times a week, with friendly matches typically scheduled on Thursdays or over the weekend. These training sessions mainly focused on developing technical-tactical skills and physical fitness. During the league season, players trained six times a week, with an official match played on the weekend. Friendly and National Cup matches were usually played on Thursdays and Wednesdays, respectively.

2.6. Quantification of Training

Training load quantification was performed using the equation proposed by [16], which combines metabolic power (MP) and equivalent distance index (EDI) to assess external training load (ETL). Notably, this method demonstrated a significant correlation with internal training load (ITL), as measured using Edwards’ heart rate-based TL method. This finding supports the validity of the MP-based training load calculation in effectively reflecting internal physiological demands during soccer activities. Briefly, metabolic power metrics were developed to accurately describe the physical demands placed on soccer players during matches or training sessions [15]. Integrating metabolic power with kinematic parameters provides an integrated measurement of acceleration and velocity, making it a more precise indicator of the physiological effort exerted by soccer players during both training and match events [15]. In particular, to calculate metabolic power (MP), researchers applied the following formula [20]:

$$MP = EC \cdot V \quad (1)$$

where:

- EC: energetic cost of running uphill at an inclination derived from the running acceleration vector;
- V: instantaneous velocity.

To better assess the metabolic intensity of the training or game, some authors propose incorporating both the equivalent distance (ED) and the equivalent distance index (EDI) [16,24,25]. ED refers to the distance an athlete would have covered at a constant pace on grass using the total energy expended during the match, while EDI is the ratio between the ED and the total distance covered in the game. It's important to note that the EDI provides an indirect estimate of the metabolic intensity of the match, with a high index indicating a greater contribution from anaerobic energy stores due to high-intensity acceleration efforts. To calculate ED and EDI, the following formulas were applied [15]:

$$ED = \frac{W}{EC_c \cdot KT}, \quad EDI = \frac{ED}{TD} \quad (2)$$

where:

- W: total energy expenditure measured in joules per kilogram;
- EC_c : energetic cost of running at a constant pace on flat, compact terrain (assumed to be $3.6 \text{ J} \cdot \text{kg}^{-1} \cdot \text{m}^{-1}$);
- KT: grassy terrain constant;
- ED: equivalent distance measured in meters;
- TD: total distance covered in meters.

Based on existing data, the average ratio between ED and total distance covered during soccer matches was approximately 1.20 [15], though this value varied among players (range: 1.15 to 1.35) depending on positional role [26]. In our study, we calculated the external training load (ETL) on the basis of the product of MP and EDI [16]. It defines the intensity of the external load (I) as follows:

$$ETL = \int V(t) dt \times \int I(t) dt, \quad I = MP \times EDI \quad (3)$$

2.7. Performance Test

All performance data were obtained from official matches and collected through video match analysis. As a performance criterion, we used the ratio between the total distance covered by the athlete and the distance covered at a metabolic power output exceeding $25 \text{ W} \cdot \text{kg}^{-1}$ (indicative of high-intensity activity). This threshold effectively identifies efforts with elevated metabolic demands and is considered a reliable indicator for assessing high-intensity performance in soccer players [27]. For the analysis, only players who completed the entire match or were substituted no earlier than five minutes before the final whistle (i.e., ≥ 85 minutes of play) were included.

3. Modeling

The modeling approach employed in this study aims at originally describing and predicting athletes' performance capacity in response to ETL derived from GPS-based metrics. We developed a novel linear, time-varying, two-state, discrete-time dynamical model that draws inspiration from the model of heart rate dynamics during variable-intensity exercises [3,4,20].

The proposed model includes two state variables: $x_1(k)$, representing a performance-related variable at time step k , and $x_2(k)$, a lumped variable at time step k whose variations directly affect the temporal dynamics of $x_1(k)$. The model is governed by the following recursive equations:

$$\begin{cases} x_1(k+1) = a_1 x_1(k) - a_2 x_2(k) e^{-k/\tau_1} + b_1 e^{-k/\tau_2} u(k) \\ x_2(k+1) = a_3 x_2(k) + b_2 \left(u(k) - a_4 e^{-k/\tau_3} x_1(k) \right) \end{cases} \quad (4)$$

where $u(k)$ is the (%-normalized) training load input, and $k \in \mathbb{N}$. The model incorporates multiple time constants τ_i to account for decaying memory effects and parameters a_1, a_2, a_3, a_4 to fit the $(x_1(k), x_2(k))$ profiles satisfactorily. To ensure physiological plausibility and numerical stability, we introduced weak bounding constraints on all model parameters. These bounds were based on prior modeling studies [1,2,5]. Table 1 summarizes the description of the model parameters and the bounds applied during model optimization. Note that when b_1 is negative, the load $u(k)$ directly contributes negatively to $x_1(k+1)$, whereas it may still contribute positively to $x_1(k+2)$, in accordance with the following auto-regressive moving-average expression¹:

$$x_1(k+2) = \left(a_1 + e^{-1/\tau_1}a_3\right)x_1(k+1) - \Sigma_1x_1(k) + b_1e^{-(k+1)/\tau_2}u(k+1) - \Sigma_2u(k) \quad (5)$$

with

$$\begin{aligned} \Sigma_1 &= a_3a_1e^{-1/\tau_1} - a_2a_4b_2e^{-(k+1)/\tau_1}e^{-k/\tau_3} \\ \Sigma_2 &= a_3b_1e^{-1/\tau_1}e^{-k/\tau_2} + a_2b_2e^{-(k+1)/\tau_1} \end{aligned} \quad (6)$$

This formulation allows for a more biologically plausible representation of the complex relationship between training and performance in soccer players.

On the other hand, the input term $b_1e^{-k/\tau_2}u(k)$ introduces a form of temporal “memory” in the system: past training loads are exponentially attenuated over time, meaning that, a training session at time k will progressively lose its influence on performance estimation at time $k+1$ as k increases. This mechanism reflects the fading physiological impact of past stimuli and allows the model to mimic the adaptive dynamics of athletic response.

From a mathematical standpoint, the proposed formulation also marks a structural departure from prior models. Whereas performance was previously estimated *via* single or double summations of exponentially decaying terms related to training and fatigue, here it emerges from a linear, time-varying, two-state system, with a lumped variable. This architecture enhances both the flexibility and interpretability of the model under dynamic load conditions.

Moreover, the model generates an asynchronous output – related to the performance achieved in match events – so that the visible effects of training inputs are not reflected in the instantaneous performance. This behavior captures the delayed and distributed nature of physiological adaptations and fatigue-recovery cycles, thus aligning the model behavior more closely with real-world athletes’ response.

More specifically, the model exhibits asynchronous behavior by distributing the influence of each training input over multiple future time steps. This is governed by a set of estimated time constants that rule how long the effects of a training session persist in the athlete’s physiological state. As a result, a single high-load session may influence performance across several subsequent days—initially increasing fatigue, and only later contributing to fitness improvements. This mechanism allows the model to capture both the delayed onset of adaptation and the extended recovery period, offering a more biologically realistic depiction of training responses.

To evaluate its effectiveness and highlight the value of its novel features, the proposed two-state model was systematically compared with a well-established reference, namely the three-component (Busso) model [6], which introduces a dynamic fatigue term accounting for the monotony and cumulative load of prior sessions.

Researchers started from the one-component model, which represents the most basic approach, assuming that performance is solely driven by the accumulation of fitness as a function of prior training loads [1]. This model neglects fatigue effects and uses a single exponential decay term to reflect the residual impact of each training session. Despite its simplicity, it provides a useful baseline for evaluating the predictive power of more complex models.

¹ Start from the first equation in (4) specialized to $x_1(k+2)$ and then use the second equation in (4). Finally eliminate $x_2(k)$ by again using the first equation in (4).

In general, a fitness-fatigue model conceptualizes the performance as the sum of a positive training effect (fitness) and a negative one (fatigue). It estimates the performance through the convolution of exponentially decaying responses to training impulses, weighted by individual-specific coefficients and time constants [1,2].

The mathematical formulation of the comparative model used in this study is reported below for the sake of exhaustiveness:

- **Three-component (Busso) model [28]:**

$$p(k) = p(0) + k_1 \sum_{i=1}^{k-1} w(i) e^{-\frac{(k-i)}{\tau_1}} + \sum_{i=1}^{k-1} k_2 w(i) e^{-\frac{(k-i)}{\tau_2}} + \sum_{i=1}^{k-1} k_2(i) w(i) e^{-\frac{(k-i)}{\tau_2}} \quad (7)$$

$$k_2(i) = k_3 \sum_{j=1}^i w(j) e^{-\frac{(i-j)}{\tau_3}} \quad (8)$$

where $k_2(i)$ is a dynamic fatigue term representing the cumulative training monotony, and k_3 and τ_3 model its influence.

4. Results

4.1. Athletic Performance Prediction

As previously mentioned, the predictive capacity of the implemented models was evaluated by analyzing data from two professional soccer players.

Athletic performance was assessed, for both Player 1 and Player 2, as the ratio between the total distance covered and the distance covered at a metabolic power greater than $25 \text{ W} \cdot \text{kg}^{-1}$, which was considered indicative of high-intensity activity. Data were obtained *via* video analysis of official matches and high-frequency GPS devices (GPEXE©).

The predictions were made for each player at 14 official competitive matches, held on the following dates: 2, 15, 19, and 22 September; 1, 22, and 28 October; 4 and 25 November; 1, 8, 16, 22, and 26 December 2018. Furthermore, for Player 2, performance prediction was also assessed during the pre-season phase, considering 8 friendly matches held on 14, 21, 25, and 28 July; 1, 4, 12, and 26 August 2018.

4.1.1. Model Accuracy Comparison Across Seasonal Phases

The results reported in this section show that the proposed model significantly outperforms the Busso model in predictive accuracy. Specifically, for Player 1, the proposed model achieved a Pearson correlation coefficient of $r = 0.89$ and a Sum of Squared Errors ($\text{SSE} = 5.84$). In contrast, the Busso model yielded lower correlations and higher errors ($r = 0.75$, $\text{SSE} = 11.95$).

The improved predictive performance of the proposed two-state dynamical model compared to the Busso model is consistent across the two athletes analyzed (Table 2), confirming that the inclusion of a lumped variable allows for a more accurate modeling of the complex relationship between training load and performance in professional soccer. Estimated model parameters for each player are reported in Table 3.

Figures 1a and 1b show the performance predictions for Player 1 during the competitive season, obtained using the newly proposed two-state model and the Busso model, respectively. The corresponding predictions for Player 2 are presented in Figures 2a and 2b. Furthermore, as shown in Figures 3a-b and in Table 4, the proposed model also achieves superior predictive accuracy during the pre-season (Player 2). The corresponding parameter estimates are provided in Table 5.

These findings confirm the generalizability and robustness of the proposed model across different athletes and seasonal contexts, including both the competitive and pre-season phases.

The following initial conditions at time step $k = 0$ were adopted for the two-state variables $x_1(k)$ and $x_2(k)$ in the newly proposed model: $x_1(0) = 31.28$, $x_2(0) = 2.78$ for Player 1 (competitive season),

$x_1(0) = 27.98, x_2(0) = 3.18$ for Player 2 (competitive season), and $x_1(0) = 14.44, x_2(0) = 6.75$ for Player 2 (pre-season). In all cases, $x_1(0)$ was set equal to the measured performance at $k = 0$, while $x_2(0)$ was involved in the DE optimization procedure until convergence to the global minimum.

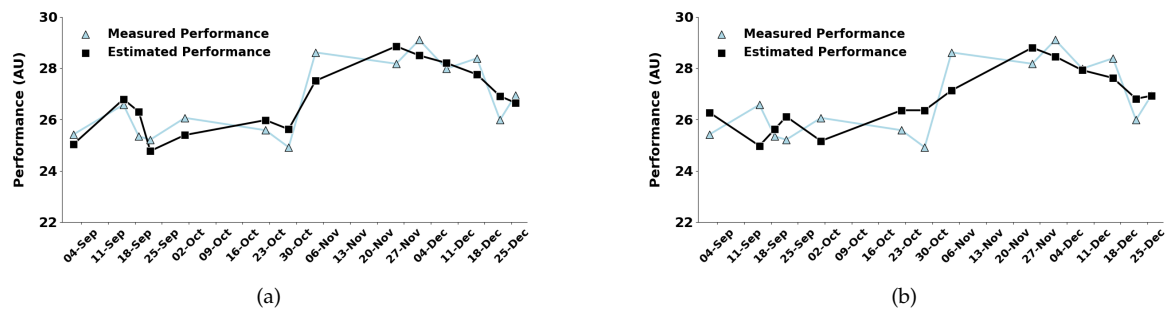


Figure 1. Athletic performance estimation for Player 1 obtained using (a) the proposed two-state model and (b) the Busso model during the competitive season.

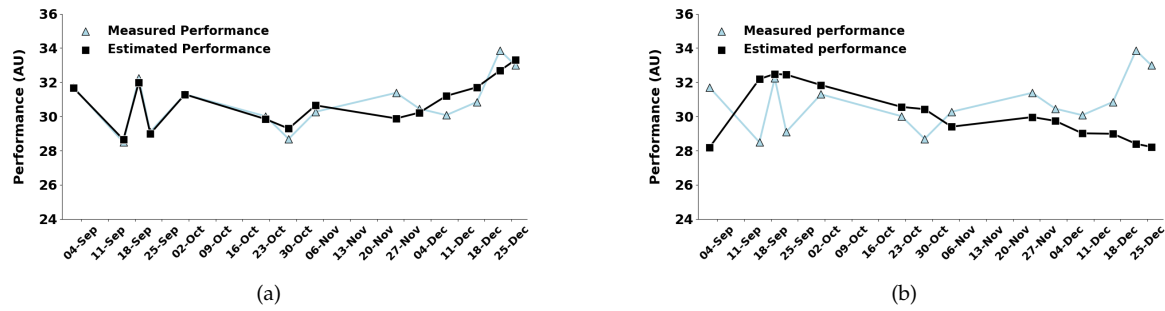


Figure 2. Athletic performance estimation for Player 2 obtained using (a) the proposed two-state model and (b) the Busso model during the competitive season.

Table 2. Correlation coefficient (r) and Sum of Squared Errors (SSE) for each model and player during the competitive season.

Player	Busso Model		Proposed Model	
	r	SSE	r	SSE
1	0.75	11.95	0.89	5.84
2	0.58	14.60	0.90	6.41

Table 3. Identified model parameters for each player during the competitive season.

Model	Parameters
Player 1	
Proposed Model	$\tau_1 = 19.69, \tau_2 = 90.58, \tau_3 = 28.74, a_1 = 0.97, a_2 = 1.97, a_3 = 0.59, a_4 = 1.45, b_1 = 0.07, b_2 = 0.05$
Busso Model	$\tau_1 = 55.66, \tau_2 = 59.99, \tau_3 = 7.20, k_1 = 0.28775, k_2 = -0.34632, k_3 = 0.00025$
Player 2	
Proposed Model	$\tau_1 = 6.86, \tau_2 = 85.31, \tau_3 = 21.43, a_1 = 1.01, a_2 = -1.27, a_3 = -0.89, a_4 = 0.96, b_1 = -0.02, b_2 = 0.39$
Busso Model	$\tau_1 = 30, \tau_2 = 15, \tau_3 = 2, k_1 = -0.0119915, k_2 = 0.0187640, k_3 = 0.0000730$

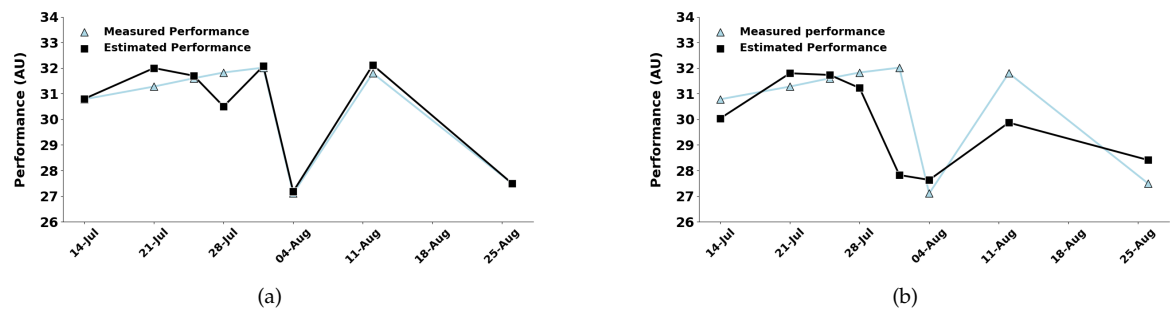


Figure 3. Athletic performance estimation for Player 2 obtained using (a) the proposed two-state model and (b) the Busso model during the pre-season.

Table 4. Correlation coefficient (*r*) and Sum of Squared Errors (SSE) for player 2 and each model during the pre-season.

Player	Busso Model		Proposed Model	
	<i>r</i>	SSE	<i>r</i>	SSE
2	0.50	12.31	0.87	5.45

Table 5. Identified model parameters for Player 2 during the pre-season.

Model	Parameters
Player 2	
Proposed Model	$\tau_1 = 23, \tau_2 = 47, \tau_3 = 8, a_1 = 0.86, a_2 = 8, a_3 = -0.94, a_4 = 17, b_1 = 0.24, b_2 = -0.051$
Busso Model	$\tau_1 = 60, \tau_2 = 5, \tau_3 = 5, k_1 = 0.0018, k_2 = 0.029, k_3 = -0.00038$

5. Discussion

This study evaluated the performance of a novel linear, time-varying, two-state discrete-time mathematical model designed to predict athletic performance in professional soccer players. The proposed approach outperformed the (nonlinear) Busso model, showing higher Pearson correlation coefficients and lower SSE values across different athletes and seasonal phases, both during the competitive and pre-season periods (Tables 2–4; Figures 1a-b, 2a-b, 3a-b).

Building on this structural insight, the model also introduces a predictive-preventive perspective. In fact, beyond simply describing performance retrospectively, the model enables early identification of suboptimal load trends. This capability supports timely intervention and may help mitigate injury risks. Such anticipatory modeling is of particular interest to training staff who aim to monitor and optimize load progression proactively.

To ensure consistency with biological plausibility, particular care was taken during the model-fitting phase. Specifically, parameter estimation was constrained within physiologically meaningful ranges. This approach enhances both model robustness and transferability, especially when applied to athletes with limited data availability.

On the other hand, the integration of GPS-derived external load metrics—specifically, MP and EDI—further increased the model responsiveness. These metrics allowed a more sensitive tracking of high-intensity activities and training density, which are often underrepresented by traditional indicators. Their inclusion highlights the importance of selecting input variables that reflect the complexity of the training stimulus.

The obtained results suggest that the proposed model offers a biologically informed, data-driven framework for monitoring and optimizing performance in elite soccer. Its structure allows for adaptive

modeling under dynamic training regimes, and its parameters provide interpretable insights that align with established physiological mechanisms.

Additionally, implementing real-time model updating—*via* sliding-window mechanisms—may support dynamic performance tracking throughout entire competitive calendars.

6. Limitations

This paper presents a proof of concept, based on a relatively small sample of two professional players. As such, the observed differences in predictive accuracy between models should be interpreted with caution. Future studies will involve larger and more diverse athlete populations.

7. Conclusions

This study introduced a novel linear, time-varying, two-state discrete-time mathematical model that outperforms a classical model in predicting athletic performance among elite soccer players.

A key innovation is the inclusion of GPS-derived metrics such as metabolic power and equivalent distance index, which enhanced the model sensitivity to the physiological intensity of training sessions.

The results obtained in this paper suggest that the proposed model is a powerful tool for performance prediction while suggesting that it is also a practical framework for guiding training load optimization in high-intensity team sports. It offers a biologically grounded and data-driven approach to enhance athlete readiness and reduce the risk of overtraining. Moreover, by predicting how training load impacts future performance, the model provides valuable support in injury prevention strategies, helping to avoid maladaptive responses due to training errors or excessive load accumulation.

Future extensions of this work may include real-time applications, integration with sliding window updates, and broader validation across diverse sports and athletic populations.

Author Contributions: Conceptualization, C.M.V. and V.M.; methodology, C.M.V. and V.M.; software, S.M. and N.C.; validation, V.M., S.M., N.C., F.L., and R.D.; formal analysis, C.M.V. and V.M.; investigation, S.M., N.C., F.L. and R.D.; resources, V.M.; data curation, S.M. and N.C.; writing—original draft preparation, N.C.; writing—review and editing, N.C. and C.M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: The players gave informed consent regarding the purposes of the research.

Data Availability Statement: The data will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Banister, E.W.; Calvert, T.W.; Savage, M.V.; Bach, T.M. A systems model of training for athletic performance. *Aust J Sports Med* **1975**, *7*, 57–61. <https://doi.org/10.1109/TSMC.1976.5409179>.
2. Morton, R.H.; Fitz-Clarke, J.R.; Banister, E.W. Modeling human performance in running. *J Appl Physiol* **1990**, *69*, 1171–1177. <https://doi.org/10.1152/jappl.1990.69.3.1171>.
3. Busso, T.; Carasso, C.; Lacour, J.R. Adequacy of a systems structure in the modeling of training effects on performance. *J Appl Physiol* **1991**, *71*, 2044–2049. <https://doi.org/10.1152/jappl.1991.71.5.2044>.
4. Mujika, I.; Busso, T.; Lacoste, L.; Barale, F.; Geyssant, A.; Chatard, J.C. Modeled responses to training and taper in competitive swimmers. *Med Sci Sports Exerc* **1996**, *28*, 251–258. <https://doi.org/10.1097/00005768-199602000-00015>.
5. Busso, T.; Hakkinen, K.; Pakarinen, A.; Carasso, C.; Lacour, J.R.; Komi, P.V.; Kauhanen, H. A systems model of training responses and its relationship to hormonal responses in elite weight-lifters. *Eur J Appl Physiol Occup Physiol* **1990**, *61*, 48–54. <https://doi.org/10.1249/01.MSS.0000074465.13621.37>.
6. Busso, T. Variable dose-response relationship between exercise training and performance. *Med Sci Sports Exerc* **2003**, *35*, 1188–1195. <https://doi.org/10.1007/BF00236693>.
7. Bangsbo, J. *The physiology of soccer—with special reference to intense intermittent exercise*; Vol. 619, Acta Physiol Scand Suppl, 1994; pp. 1–155.

8. Rampinini, E.; Coutts, A.J.; Castagna, C.; Sassi, R.; Impellizzeri, F.M. Variation in top level soccer match performance. *Int J Sports Med* **2007**, *28*, 1018–1024. <https://doi.org/10.1055/s-2007-965158>.
9. Stølen, T.; Chamari, K.; Castagna, C.; Wisløff, U. Physiology of soccer: An update. *Sports Med* **2005**, *35*, 501–536. <https://doi.org/10.2165/00007256-200535060-00004>.
10. Kulakou, S.; Ragab, N.; Midoglu, C.; Boeker, M.; Johansen, D.; Riegler, M.A.; Halvorsen, P. Exploration of Different Time Series Models for Soccer Athlete Performance Prediction. *Eng Proc* **2022**, *18*, 37. <https://doi.org/10.3390/engproc2022018037>.
11. Nowak, M.; Bok, B.; Wilczek, A.; Oleksy, L.; Kamola, M. Forecasting extremes of football players' performance in matches. *Sci Rep* **2024**, *14*, 27319. <https://doi.org/10.1038/s41598-024-78708-5>.
12. Javadpour, L.; Khazaeli, M.; Molenaar, R. From Practice To Performance: Predicting Soccer Match Outcomes from Training Data. *SN Comput Sci* **2025**, *6*, 324. <https://doi.org/10.1007/s42979-025-03870-0>.
13. Aughey, R.J. Applications of GPS Technologies to Field Sports. *Int. J. Sports Physiol. Perform.* **2011**, *6*, 295–310. <https://doi.org/10.1123/ijsp.6.3.295>.
14. Cummins, C.; Orr, R.; O'Connor, H.; West, C. Global Positioning Systems (GPS) and Microtechnology Sensors in Team Sports: A Systematic Review. *Sports Med.* **2013**, *43*, 1025–1042. <https://doi.org/10.1007/s40279-013-0069-2>.
15. Osgnach, C.; Poser, S.; Bernardini, R.; Rinaldo, R.; di Prampero, P.E. Energy cost and metabolic power in elite soccer: a new match analysis approach. *Med Sci Sports Exerc* **2010**, *42*, 170–178. <https://doi.org/10.1249/MSS.0b013e3181ae5cfd>.
16. Manzi, V.; Savoia, C.; Padua, E.; Edriss, S.; Iellamo, F.; Caminiti, G.; Annino, G. Exploring the interplay between metabolic power and equivalent distance in training games and official matches in soccer: a machine learning approach. *Front Physiol* **2023**, *14*, Oct 24. <https://doi.org/10.3389/fphys.2023.1230912>.
17. Linke, D.; Link, D.; Lames, M. Validation of electronic performance and tracking systems EPTS under field conditions. *PLoS One* **2018**, *13*, e0199519. <https://doi.org/10.1371/journal.pone.0199519>.
18. Gaudino, P.; Iaia, F.M.; Alberti, G.; Strudwick, A.J.; Atkinson, G.; Gregson, W. Monitoring training in elite soccer players: systematic bias between running speed and metabolic power data. *Int J Sports Med* **2013**, *34*, 963–968. <https://doi.org/10.1055/s-0033-1337943>.
19. Linke, D.; Link, D.; Lames, M. Football-specific validity of TRACAB's optical video tracking systems. *PLoS One* **2020**, *15*, e0230179. <https://doi.org/10.1371/journal.pone.0230179>.
20. di Prampero, P.E.; Fusi, S.; Sepulcri, L.; Morin, J.B.; Belli, A.; Antonutto, G. Sprint running: a new energetic approach. *J Exp Biol* **2005**, *208*, 2809–2816. <https://doi.org/10.1242/jeb.01700>.
21. Nagahara, R.; Botter, A.; Rejc, E.; Koido, M.; Shimizu, T.; Samozino, P.; et al. Concurrent validity of GPS for deriving mechanical properties of sprint acceleration. *Int J Sports Physiol Perform* **2017**, *12*, 129–132. <https://doi.org/10.1123/ijsp.2015-0566>.
22. Hoppe, M.W.; Baumgart, C.; Polglaze, T.; Freiwald, J. Validity and reliability of GPS and LPS for measuring distances covered and sprint mechanical properties in team sports. *PLoS One* **2018**, *13*, e0192708. <https://doi.org/10.1371/journal.pone.0192708>.
23. Witte, T.H.; Wilson, A.M. Accuracy of non-differential GPS for the determination of speed over ground. *J Biomech* **2004**, *37*, 1891–1898. <https://doi.org/10.1016/j.jbiomech.2004.02.031>.
24. Guerrero-Calderón, B.; Klemp, M.; Castillo-Rodríguez, A.; Morcillo, J.A.; Memmert, D. A new approach for training-load quantification in elite-level soccer: contextual factors. *Int J Sports Med* **2021**, *42*, 716–723. <https://doi.org/10.1055/a-1289-9059>.
25. Guerrero-Calderón, B.; Morcillo, J.A.; Chena, M.; Castillo-Rodríguez, A. Comparison of training and match load between metabolic and running speed metrics of professional Spanish soccer players by playing position. *Biol Sport* **2022**, *39*, 933–941. <https://doi.org/10.5114/biolSport.2022.110884>.
26. di Prampero, P.E.; Botter, A.; Osgnach, C. The energy cost of sprint running and the role of metabolic power in setting top performances. *Eur J Appl Physiol* **2015**, *115*, 451–469. <https://doi.org/10.1007/s00421-014-3086-4>.
27. Martín-García, A.; Castellano, J.; Méndez Villanueva, A.; Gómez-Díaz, A.; Cos, F.; Casamichana, D. Physical Demands of Ball Possession Games in Relation to the Most Demanding Passages of a Competitive Match. *J Sports Sci Med* **2020**, *19*, 1–9.
28. Kolossa, D.; Bin Azhar, M.A.; Rasche, C.; Endler, S.; Hanakam, F.; Ferrauti, A.; Pfeiffer, M. Performance Estimation using the Fitness-Fatigue Model with Kalman Filter Feedback. *International Journal of Computer Science in Sport* **2017**, *16*, 105–120. <https://doi.org/10.1515/ijcss-2017-0010>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.