# Validation and Optimisation of Player Motion Models in Football

Moritz Renkin[1] [0000-0002-1404-857X], Jonas Bischofberger[2] [0000-0002-0878-0641], Erich Schikuta[1] [0000-0002-4126-4243], and Arnold Baca[2] [0000-0002-1704-0290]

[1] Research Group Workflow Systems and Technology, Faculty of Computer Science, University of Vienna, Austria
a11807211@unet.univie.ac.at

[2] Department of Biomechanics, Kinesiology and Computer Science in Sport, Institute of Sports Science, University of Vienna, Austria
jonas.bischofberger@univie.ac.at

**Abstract.** Modelling the trajectorial motion of humans along the ground is a foundational task in the quantitative analysis of sports like association football. Most existing models of football player motion have not been validated yet with respect to actual data. One of the reasons for this lack is that such a validation is not straightforward, because the validation typically needs to be performed with respect to noisy extreme values rather than expected values.

This paper proposes a validation routine for trajectorial motion models that measures and optimises the ability of a motion model to accurately predict all possibly reachable positions by favoring the smallest predicted area that encompasses all observed reached positions up to a manually defined threshold. We demonstrate validation and optimisation on four different motion models, assuming (a) motion with constant speed, (b) motion with constant acceleration, (c) motion with constant acceleration with a speed limit, and (d) motion along two segments with constant speed. Our results show that assuming motion with constant speed or constant acceleration without a limit on the achievable speed is particularly inappropriate for an accurate distinction between reachable and unreachable locations. Motion along two segments of constant speed provides by far the highest accuracy among the tested models and serves as an efficient and accurate approximation of real-world player motion.

**Keywords:** Football · Positional data · Motion models · Performance analysis · Model validation · Complex systems

## 1 Introduction

Recently, professional association football has seen a surge in the availability of positional data of the players and the ball, typically collected by GPS, radar or camera systems. The growing availability of such data has opened up an exciting new avenue for performance analysis. High-quality measures of performance that

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_4

include positional information are invaluable for effective training, opposition scouting, and player recruitment.

The modelling of human motion is a foundational component of many performance metrics based on positional data. For example, algorithms that compute space control [7] or simulate passes [6] implicitly or explicitly make assumptions about human kinematics. These kinematic assumptions have never been verified so far, which calls the validity of these assumptions and the resulting models into question.

Rather than predicting actual human motion, many applications merely require the prediction of possibly reachable positions. This requirement essentially shifts the purpose of a motion model from predicting expected positions towards estimating the most remote reachable positions. Estimating such extreme values from real-world data can be difficult, because extreme values are typically rare and particularly likely to include a component of measurement error with an often unknown distribution that cannot easily be accounted for.

The contributions of this paper are twofold: First, we formally propose a validation routine for the quality of player motion models. Second, we use this routine to evaluate and optimise the parameters of four models of motion.

The rest of this paper is structured as follows: Section 2 provides some background on motion models in football and their validation. Section 3 formally presents our validation routine. Section 4 describes our exemplary model validation and optimisation based on a real data set and discusses its results. Section 5 summarises the contributions of this paper and points out possible directions of further research.

## 2 Motion models in football: state of the art

Assumptions about the trajectorial motion of players are inherent to many performance indicators within the analysis of sports games. One example is the commonly used concept of space control which assigns control or influence over different areas on the pitch to players. It is used, for example, as a context variable to rate football actions [5] and for time series analyses [3]. Controlled space is often defined as the area that a player is able to reach before any other player, given a specific model of motion for each player. Commonly used for this purpose are motion models assuming constant and equal speed, which results in a Voronoi partition of the pitch, or accelerated movement with limited speed [7]. Spearman et al. [6] assume accelerated player motion with a limit on acceleration and velocity in the context of modeling ground passes.

Motion models have also been estimated directly from positional data [1, 2]. However, such empirical models can be computationally expensive, prone to outliers and their current versions lend themselves less naturally to extreme value estimation than theoretically derived models. Attempts to validate trajectorial player motion models are rare. Notably, Caetano et al. [2] performed a validation of their space control model, and thus indirectly also the underlying

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_4

motion model, by checking how many future positions of players fall within their associated controlled area for a number of time horizons.

# 3 Player motion model and validation procedure

We propose a validation procedure rating a player motion model on how well it fits some real positional data. In order to abstract our validation procedure from the underlying positional data, we introduce the concept of a *trail*. A trail represents a slice of a player's trajectory over some duration $\Delta t$. Formally, a trail is defined as the quadruple: $(\vec{x}_0, \vec{v}_0, \vec{x}_t, \Delta t)$

- $\vec{x}_0$: (2D) position of a given player at some arbitrary time $t_0$
- $\vec{v}_0$: (2D) velocity of the player at time $t_0$
- $\vec{x}_t$: (2D) position of the player at time $t = t_0 + \Delta t$
- $\Delta t$: time horizon (predefined)

Since every reached position is trivially contained in a large enough area, the validation function should take not only correctness but also precision of the model into account. The correctness of a motion model measures its ability to make true predictions, i.e. to predict reachable areas that contain the true target position $\vec{x}_t$. Precision refers to how well narrowed-down the predicted areas of a model are. There is a trade-off relationship between correctness and precision.

## 3.1 Measuring Correctness.

Considering only a single trail, a motion model $m$ makes a prediction for the reachable area using $\vec{x}_0$, $\vec{v}_0$ and $\Delta t$. If $\vec{x}_t$ is contained in the predicted reachable area, the model has made a correct prediction. Following this logic, a motion model achieves the highest possible correctness if and only if for every trail, the model predicts a reachable area in which $\vec{x}_t$ is contained. The ratio between the number of correct predictions $n_{correct}$ and the number of total predictions $n_{total}$ of a model $m$ for a sample of trails $T$ will be called $hit\_ratio$.

$$hit\_ratio(m, T) = \frac{n_{correct}}{n_{total}} \tag{1}$$

We can use the $hit\_ratio$ of a model as an indicator for its correctness. A high $hit\_ratio$ corresponds to a high correctness and vice-versa.

## 3.2 Measuring Precision.

In the context of this paper, the precision of a motion model represents how much it narrows down the reachable area of a player. Smaller reachable areas imply a higher precision of the model and are generally preferable, given an equal $hit\_ratio$.

To determine the precision of a model across multiple evaluated trails, we use the inverse of the mean surface area of all correctly predicted reachable

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_4

areas. Incorrect predictions, where the target position $\vec{x_t}$ is not contained in the predicted reachable area are excluded from this average, since the precision of a model would otherwise increase inappropriately for very narrow, incorrect predictions. The precision of model $m$ across a sample of trails $T$ is given by:

$$precision(m, T) = \frac{1}{\frac{1}{n_{correct}} \cdot \sum areas_{correct}} = \frac{n_{correct}}{\sum areas_{correct}} \tag{2}$$

where $\sum areas_{correct}$ is the sum of all correctly predicted reachable areas.

### 3.3 Defining an overall Validation Score.

Since we aim for a single numerical value as a score for player motion models, correctness and precision have to be balanced in some way. Due to the fact that some measurement-related extreme outliers can usually be expected in positional data from football games, a model with a $hit\_ratio$ of 100% might not necessarily be desirable. Therefore, we introduce a minimum level of correctness $hit\_ratio_{min}$, which represents a minimal required ratio between correct and total predictions of a model. We propose that if a motion model $m$ satisfies the condition $hit\_ratio(m, T) \geq hit\_ratio_{min}$ for a trail sample $T$, the exact $hit\_ratio(m, T)$ should be indifferent for the overall validation score of $m$. This way, extreme outliers in the positional data caused by measurement-related errors have no influence on the validation score, as long as $hit\_ratio_{min}$ is chosen adequately.

Consequently, for a motion model $m$ that exceeds $hit\_ratio_{min}$, the validation score is only determined by the precision of the model (2). We define the *score* of a motion model $m$ with the sample of trails $T$ as:

$$score(m, T) = \begin{cases} 0 & \text{if } hit\_ratio(m, T) < hit\_ratio_{min} \\ precision(m, T) & \text{else} \end{cases} \tag{3}$$

The *score* measures how well a motion models fits a sample of positional data. $hit\_ratio_{min}$ can be considered a free parameter of this validation procedure. It should be chosen to accommodate for the error distribution of the positional data.

## 4 Experiment & Evaluation of results

### 4.1 Data set

For the evaluation, we use the public sample data set provided by Metrica Sports which consists of three anonymised games of football [4]. The positional data has been collected using a video-based system and is provided at a frequency of 25 Hz. For this experiment, we use a constant time horizon of $\Delta t = 1s$

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_4

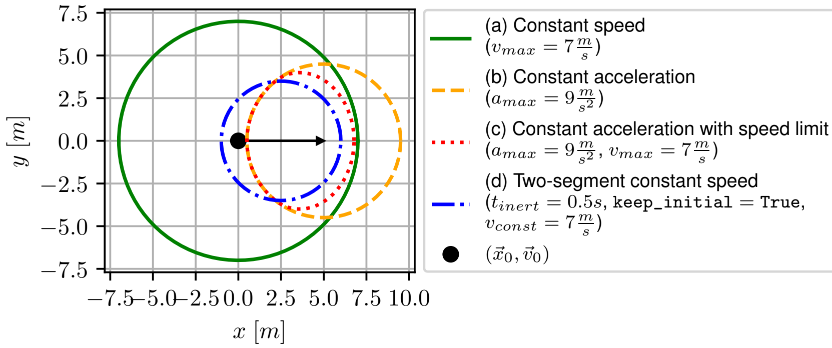


**Fig. 1.** Exemplary boundaries of the reachable area defined by different motion models when the player starts at $\vec{x}_0 = \vec{0}$ with velocity $\vec{v}_0 = \begin{bmatrix} 5\frac{m}{s} \\ 0 \end{bmatrix}$. The time horizon is $\Delta t = 1s$.

After visual inspection of the data, the minimal required hit ratio is set to $hit\_ratio_{min} = 99.975\%$. We evaluate the models on a random sample of $5 \cdot 10^5$ trails across all three games and all participating players.

### 4.2 Preparation of motion models

We optimize and evaluate the following models of motion where each one defines a reachable area, depending on specific parameter values. These areas are exemplarily visualized in Figure 1.

(a) Constant speed: Motion with constant speed $v_{max}$ in any direction
(b) Constant acceleration: Motion with constant acceleration $a_{max}$ in any direction
(c) Constant acceleration with speed limit: Motion with constant acceleration $a_{max}$ in any direction until a maximal speed $v_{max}$ is reached
(d) Two-segment constant speed: Motion along two segments of constant speed. During the first segment, the player is simulated to run in the direction of $\vec{v}_0$ with constant speed $v_{seg1}$ which is set to either $|\vec{v}_0|$ or to some fixed value $v_{const}$, depending on the boolean parameter `keep_initial`. During the second segment, the speed of the player is set to either $v_{const}$ or $\min(v_{seg1} + a_{max}t_{inert}, v_{max})$, depending on whether the parameters $a_{max}$ and $v_{max}$ are set.

Using the evaluation routine outlined in section 3, we find the optimal parameter configuration for each model via Bayesian optimization. Discrete parameters like `keep_initial` are handled by performing one round of Bayesian optimisation for each combination of discrete parameter values and using the best score across those results.

### 4.3 Evaluation of results

The performance of the optimised models (a) - (d) and their parameter values are shown in Figure 2.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_4

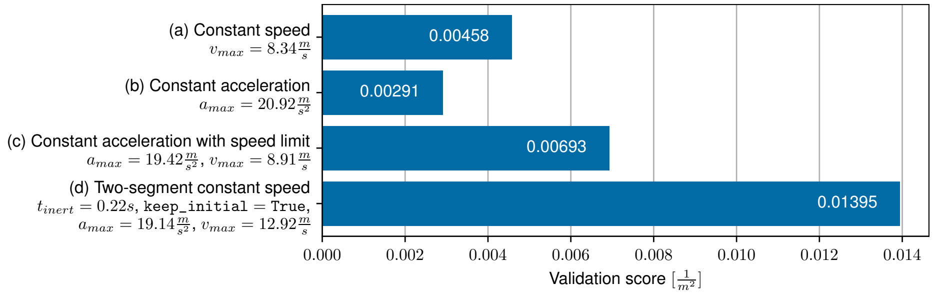


**Fig. 2.** Comparison of the performance of the four models (a) - (d) with their optimised parameter values.

The constant-speed model (a) unsurprisingly shows a weaker performance ($score^{-1} = 218m^2$) than the more sophisticated models (c) and (d), since it does not factor in the initial kinematic state of the player.

The naive constant acceleration model (b) ($score^{-1} = 344m^2$) performs even worse than model (a), likely because it makes the unrealistic assumption that the possible magnitude of acceleration is independent of the magnitude and direction of a player's current velocity. This implies in particular that for high speeds, the amount of reachable space in the direction that a player is moving towards will be heavily overestimated since the model assumes that the player's speed can increase unboundedly.

The model assuming constant acceleration with a speed limit (c) ($score^{-1} = 144m^2$) outperforms models (a) and (b). However, the optimised value of the maximally possible acceleration of a player $a_{max}$ is physically unrealistic. A value of $a_{max} = 19.42\frac{m}{s^2}$ assumes that a player can accelerate from zero to the top speed $v_{max} = 8.91\frac{m}{s}(= 32.08\frac{km}{h})$ within about half a second, which is implausibly fast. Therefore, the model still overestimates the reachable area.

The two-segment constant speed model (d) ($score^{-1} = 71.7m^2$) is able to account for all reachable positions by predicting only about half the area of model (c). It successfully narrows down the area that a player can reach within one second to a circle with an average radius of 4.8 meters which is highly accurate. Model (d) not only achieves the best score in our evaluation, but is also mathematically simpler than model (c). For that reason, it is also computationally more efficient across the various tasks that motion models are used for, like the computation of reachable areas or the shortest time to arrive at a specific location.

## 5 Conclusion

We presented a novel approach to the validation and optimisation of models of trajectorial player motion in football and similar sports. We also presented an empirical comparison of the accuracy of various such models. While more sophisticated kinematic assumptions tend to be reflected in better predictive performance, the best-performing model is our proposed approximate model

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_4

which assumes motion along two segments with constant speed. Using this model allows researchers to compute complex performance indicators more efficiently and accurately over large data sets.

The validation and optimisation approach described in this paper can be applied to data with arbitrary distributions of measurement error. However, this is also a disadvantage, since the threshold for the amount of outliers that are attributed to measurement error has to be determined manually. This threshold also has to be set for each distinguished population, depending on the frequency of extrema and the distribution of measurement error in the population. For example, if motion models are individualized, it would be misleading to use the same threshold for goalkeepers and outfield players, because goalkeepers produce far less positional extrema and thus outliers. As a solution, one could contextualize validation and optimization with various thresholds or derive an optimal threshold from a known error distribution.

In the future, we plan to search for motion models that further exceed the presented ones in accuracy and computational efficiency. A key towards this goal is to estimate motion models from positional data. Many problems addressed in this paper are mirrored in empirical model fitting, for example the need to exclude outliers and the lack of generalisability across populations [1]. In the context of validation, empirical models can serve as a highly informative benchmark to reveal how well theoretical models are able to approximate actual human motion.

## References

1. Brefeld, U., Lasek, J., Mair, S.: Probabilistic movement models and zones of control. Machine Learning **108**(1), 127–147 (2019). https://doi.org/10.1007/s10994-018-5725-1
2. Caetano, F.G., Barbon Junior, S., Torres, R., Cunha, S., Ruffino, P., Martins, L., Moura, F.: Football player dominant region determined by a novel model based on instantaneous kinematics variables. Scientific Reports **11**(1), 1–10 (2021). https://doi.org/10.1038/s41598-021-97537-4
3. Fonseca, S., Milho, J., Travassos, B., Araújo, D.: Spatial dynamics of team sports exposed by voronoi diagrams. Human movement science **31**(6), 1652–1659 (2012). https://doi.org/10.1016/j.humov.2012.04.006
4. Metrica sports sample data. https://github.com/metrica-sports/sample-data/commit/e706dd506b360d69d9d123d5b8026e7294b13996. GitHub repository (2021)
5. Rein, R., Raabe, D., Memmert, D.: "which pass is better?" novel approaches to assess passing effectiveness in elite soccer. Human movement science **55**, 172–181 (2017). https://doi.org/10.1016/j.humov.2017.07.010
6. Spearman, W., Basye, A., Dick, G., Hotovy, R., Pop, P.: Physics-based modeling of pass probabilities in soccer. In: 11th MIT Sloan Sports Analytics Conference (2017)
7. Taki, T., Hasegawa, J.: Visualization of dominant region in team games and its application to teamwork analysis. In: Proceedings computer graphics international 2000. pp. 227–235. IEEE (2000)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_4