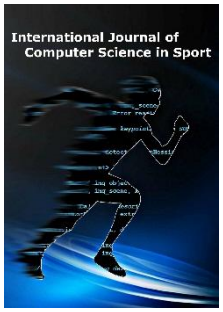
International Journal of Computer Science in Sport

Volume 25, Issue 1, 2026

Journal homepage: http://iacss.org/index.php?id=30

DOI: 10.2478/ijcss-2026-0004



Short-term performance prediction in team sports using tracking data

Arnaud Odet[1,2], Thomas Bechard[2], Sebastien Déjean[1] and Cristian Pasquaretta[2]

[1] Institut de Mathématiques de Toulouse, UMR 5219 Université de Toulouse et CNRS, Toulouse, France

[2] Centre de Recherches sur la Cognition Animale (CRCA), Centre de Biologie Integrative (CBI), CNRS, Toulouse, France

Abstract

This research addresses short-term performance prediction in Rugby Union using player tracking data. Rugby's unique characteristics as a conquest sport with strict offside rules make the ball position a crucial performance indicator. We used event data to identify ball carriers in offensive phase and considered their locations as proxy for the ball. We then aim to predict this performance indicator. We formulated this challenge as a time series prediction problem where player positions (multivariate) predict performance (univariate). Our results suggest using player tracking data improves the prediction ability.

KEYWORDS: DEEP LEARNING, TRACKING DATA, TIME SERIES, RUGBY UNION.

Introduction

The prediction of performance in sport science has become increasingly important with the advancement of player tracking technologies and analytics (Kovalchik, 2022). Performance in sport can be defined and analyzed at different scales. Studies about long-term performance (e.g. over a complete season) focus on problems such as roster constitution (Pérez-Toledano et al., 2019) or fatigue management (Vanrenterghem et al., 2017). For medium-term performance (e.g. over a game), most studies focus on feature importance (Miljkovic et al., 2010, Odet et al., 2025). Finally, for short-term performance (e.g. over an action), one main aim is to quantify "goodness" of the action (Cervone et al., 2014). This study focuses on this third prediction task in the context of team sports, with a specific focus on Rugby Union, where the dynamic nature of play presents unique analytical opportunities and challenges. Our aim is to incorporate collective player behavior to predict a performance feature. We use in this paper the ball position over the x-axis as proxy for performance.

Rugby Union, as a conquest sport, is characterized by strict rules governing player positioning, particularly the offside rule (in short, a player is offside when positioned in front of the teammate who last played the ball and cannot dispute it). These rules make ball position a crucial indicator of territorial advantage and play development. The relevance of ball position as a performance indicator in Rugby Union is strongly justified due to its direct influence on match outcomes and tactical strategies (Bennett et al., 2019; Colomer et al., 2020). In addition, line breaks are identified as a crucial offensive indicator, as they directly contribute to advancing the ball's position into advantageous opposition territory (Schoeman & Schall, 2019). Despite their importance, ball tracking data is often unavailable or incomplete in professional sports settings, creating a significant challenge for performance analysis. To address this limitation, we combined manually annotated event data with player GPS tracking data and provided a method to estimate ball position in Rugby Union.

We formulate the performance prediction objective as a time series prediction task, in which multivariate tracking data predicts univariate performance outcome. This formulation is particularly challenging in the context of team sports because the tracking data encode not only the temporal evolution of player movements but also the spatial interactions and collective coordination among them. Players' movements in rugby are interconnected. Each player's behavior is shaped by their role and by how they adapt to the movements of others, creating complex patterns that evolve over time. To properly capture these dependencies, models must go beyond basic statistical methods and represent both temporal dynamics and interactions between players. Recent studies have shown that deep learning models perform well in modeling time-dependent data, and that multivariate approaches can directly learn relationships between multiple movement series and time-varying interactions, outperforming traditional forecasting methods (Cao et al., 2020). These advances motivate our use of multivariate time series models to capture the effect of collective player movement on performance prediction in Rugby Union.

By leveraging tracking data in a predictive framework, it becomes possible to quantify aspects of team coordination and tactical decision-making, allowing for "what if" simulation analyses and continuous time play valuation. Such an approach has potential applications in both real-time game analysis and post-match performance evaluation, providing valuable tools for coaches, analysts, and sports scientists.

Contributions: we identified three main contributions of this work:

(i) Short-term Performance Prediction: We formulate short-term performance prediction in Rugby Union as a multivariate to univariate time series forecasting task and demonstrate the relevance of using complex high-frequency tracking data.

(ii) Publicly Available Dataset: We provide our dataset obtained by a preprocessing pipeline that includes a ball position estimation algorithm combining tracking and event data, addressing the challenge of incomplete ball tracking data in team sports.

(iii) Benchmarking of several state-of-the-art time series models: We benchmark several state-of-the-art models, including an exploration of the hyperparameters space, versus popular vanilla models and naive baselines.

Related Work

Tracking Data for Sports Performance Prediction

Initial applications of tracking data in team sports primarily focused on quantifying physical performance and conducting time-motion analysis (Ferraz et al., 2023). This field has since undergone a paradigmatic shift, moving towards using high-resolution spatiotemporal data for more sophisticated tactical analysis and understanding collective movement patterns among players (Low et al., 2020; Perše et al., 2009). A key area of research involves deriving performance indicators from spatial and temporal aggregations of tracking data, utilizing metrics such as team centroid, space control, and other domain-specific features (Clemente et al., 2013; Low et al., 2020). Building upon this, a significant development is the concept of continuous-time within-play valuation. This includes the development of frameworks for estimating Expected Possession Value (EPV), which quantifies the expected number of points an offense will score on a possession conditional on its evolution, using a multiresolution stochastic process model for basketball (Cervone et al., 2014). The EPV concept has been adapted to soccer with a decoupled deep learning approach that provides a frame-by-frame estimation of the expected outcome of a possession (Fernández, 2020). Similarly, models for estimating Expected Points (EP) and Win Probability (WP) continuously throughout a play have been developed for American football, using Long Short-Term Memory (LSTM) networks to model outcomes based on player locations and trajectories (Yurko et al., 2019). Machine learning techniques, including Graph Neural Networks (GNNs) and Recurrent Neural Networks (RNNs), allows to automate the identification of tactical patterns and summarize complex player interactions, reducing the reliance on manual annotation (Anzer et al., 2022; Tian et al., 2019). These advanced data-driven models enable the rating of specific actions not just by their outcome probability, but more comprehensively by quantifying the value they add or subtract from the expected outcome of a possession or play (Dick et al., 2021).

While machine learning has been increasingly applied to performance prediction in sports, the specific task of trajectory prediction, i.e. estimating the future position of a ball or player, remains relatively underexplored in the literature. This task is particularly important in team sports like rugby, where anticipating the movement of the ball and players is critical for strategic planning, performance analysis, and real-time decision making. Trajectory prediction has been successfully applied in other domains such as robotics, autonomous vehicles, and aerospace, where forecasting object motion is essential for navigation and control (references). Inspired by these approaches, we developed a machine learning model that predicts the trajectory of the rugby ball using both player and ball coordinates, enabling the estimation of ball position several seconds into the future. This approach captures the complex interactions on the field and provides a data-driven tool to support tactical analysis and game understanding. The present work aims at leveraging the sequential nature of tracking data for modeling short-term performance in Rugby Union and brings novelty by benchmarking state-of-the-art time series models on this task.

Time series Forecasting

Time series analysis involves studying collections of observations ordered by time, with applications across various fields including energy, economics, and meteorology. The main challenge is to understand the temporal dynamics in time series, which includes decomposition, seasonality, long-term dependencies, etc. It has been a very dynamic, long-standing field of study (Klein, 1997), with traditional statistical methods (ARMA and descendants, see Box et al., 1978) and advances in Deep Learning with RNN (Rumelhart et al., 1986), LSTM (Hochreiter & Schmidhuber, 1997), Transformer (Vaswani et al., 2017) and more recently Mamba (Gu & Dao, 2023). Transformer models have been adapted for Time series analysis for their ability to capture long-range dependencies, with a wide focus on the attention mechanism (see review by Wang et al. 2024).

PatchTST (Nie et al., 2023) advances long-term forecasting by introducing a patching strategy that segments each univariate time series into fixed-length subseries, which serve as tokens for a channel-independent Transformer. This design preserves local semantic structure within each patch, reduces attention complexity, and enables much longer look-back windows. By sharing embeddings and Transformer weights across all channels, PatchTST effectively scales to high-dimensional multivariate data.

Non-stationary Transformers (Y. Liu et al., 2022) tackle the inherent distribution shifts in real-world series—issues poorly addressed by classical stationarization techniques. Their framework integrates two components: Series Stationarization, which unifies and later restores statistics to improve predictability, and De-stationary Attention, which reintroduces non-stationary patterns into the attention mechanism to prevent over-stationarization. This leads to more expressive temporal dependencies.

Pyraformer (S. Liu et al., 2022) proposes a pyramidal, multi-resolution attention mechanism that models temporal dependencies at different scales. Its structured attention graph dramatically shortens maximum path length, enabling efficient long-range modeling without the prohibitive memory cost of standard Transformers. Through coarse-to-fine representations and carefully designed cross-scale connectivity, Pyraformer captures periodic patterns and long-term structure effectively.

Crossformer (Zhang & Yan, 2023) directly targets the overlooked *cross-dimension dependency* in multivariate time series. It employs a **Two-Stage Attention (TSA)** mechanism: a cross-time attention applied independently within each variate, followed by a cross-dimension attention—with an efficient router mechanism—to capture inter-variable correlations without prohibitive cost. Built atop these components, a **Hierarchical Encoder–Decoder (HED)** aggregates multi-scale temporal patterns, from coarse trends to fine details.

In contrast to these temporal-token-based approaches, iTransformer (Y. Liu et al., 2023) inverts the conventional time-centric input representation by treating variates as tokens. This architecture applies self-attention across variate tokens to model cross-variable interactions, while the feed-forward network captures temporal evolution.

Zeng et al. (2022) questioned the necessity of using Transformers for long-term forecasting, suggesting that simpler linear models can also achieve competitive results. These models are compared on reference datasets (energy, foreign exchange, weather, etc. see Wang et al., 2024) The present study applies these models to sequences expected to present distinct temporal dynamics than those of the reference datasets.

Ball position estimation

While tacking ball position from videos is a flourishing topic in computer vision (Ohno et al., 2000, Ren et al., 2009, Silva et al., 2012, Chakraborty et Meher, 2012), effort to estimate the ball trajectory without optical or tracking data remain limited, as most studies using tracking data dispose of the ball tracking (e.g. Tian et al., 2019 and Goes et al., 2020). Our work differs from this previous effort as we do not rely on computer vision, but only on player tracking data obtained by GPS, allowing us to be less dependent on video quality and computing power. In a similar setting, Amirli and Alemdar (2022) used neural network on Soccer players' tracking data and reached a Mean Absolute Error (MAE) greater than 5 meters (from tracking data obtained by video). Kim et al. (2023) combined set transformers and hierarchical architecture to achieve a 3.6 meters MAE over GPS data. Our work differs as we used heuristics leveraging event data.

Tracking and events data were combined in Hughes et al. (2025) in soccer and event data were found to provide contextual understanding of the game and considerably improve the task of reconstructing noisy trajectories. Biermann et al. (2023) highlighted "only a few applications perform a joint analysis of both data sources despite the various involved advantages emerging from such an approach". In a review of the use of tracking data in Team Sport, Gudmundsson and Horton (2016) also found only a few papers combined tracking and event data. Most of studies obtain tracking data through computer vision and focus on sports such as Soccer (e.g. Anzer et Bauer, 2021) or Basketball (e.g. Ghasemzadeh et al., 2021), where occlusions are less frequent and concern fewer player than in Rugby Union, and it is worth noting that a flourishing body of literature intends to build event datasets from tracking data (e.g. Khaustov et Mozgovoy, 2020 and Hauri et Vucetic,2022).

Methodology

Problem Formulation

Let $X_{(j,t)}$ be the feature vector of player $j \in 1,\dots,15$ at time t. $X_{(j,t)}$ is 4- dimensional, as it contains the position ($(x_j)_t$ and $(y_j)_t$ coordinates), the velocity ($(v_j)_t = \sqrt{\Delta^2_{(x_j)_t} + \Delta^2_{(y_j)_t}}$ where $\Delta_{(x_j)_t} = (x_j)_{t+1} - (x_j)_t$ and $\Delta_{(y_j)_t} = (y_j)_{t+1} - (y_j)_t$) and the goal angle $\phi_j = \arctan\left(\Delta_{(y_j)}/\Delta_{(x_j)}\right)$ of player j (the goal angle is the angle between the actual trajectory and a trajectory consisting in running straight to the opponent try line, i.e. with $\Delta_{(x_j)} > 0$ and $\Delta_{(y_j)} = 0$). We note s the sequence length, p the prediction length (both in tenth of second), and P_t the relative value of our performance feature at time t. It is obtained by subtracting the value of the performance feature at the end of the known sequence (time t_0+s for any sequence starting at time t_0) to the value of the performance feature at any time $t \in [t_0, t_0+s+p]$, materializing the progress of the action. P_t is 1-dimensional. Here, we chose the ball coordinate over the x-axis as performance feature, as it reflects the territorial advantage. Let $S_t = P_{t+s+p}$ be the performance variable associated with the sequence $X_t,\dots,X_{t+s}$ where $X_t = \text{concat}(X_{1,t},\dots,X_{15,t},P_t)$ is the feature vector at time$t$($X_t$has 61 features). *concat* represents the concatenation operator.

Our objective can be formulated as predicting S_t given $X_t,\dots,X_{t+s}$, in other words, predicting a single performance metric from an observed feature sequence.

Equivalently, we can consider (i) predicting the whole sequence of performance metrics from the same observed feature sequence and (ii) extracting a single point of interest in the predicted sequence (here, the last point). The objective can be reformulated as: Predicting $P_{t+s+1},\dots,P_{t+s+p}$ given $X_t,\dots,X_{t+s}$ (multivariate to univariate time series prediction task), where methods are

compared using the Final Displacement Error (FDE) metric, as the FDE only considers the last point of a trajectory. The FDE is the Mean Squared Error between the last points of the trajectories.

Formally, the FDE is defined as $\text{FDE}=\frac{1}{n}\sqrt{\sum\left(\hat{P}_{t_i+\text{s}+\text{p}}-P_{t_i+\text{s}+\text{p}}\right)^2}$ where $\hat{P}$ is the estimated performance and t_i indicates the start of sequence $i \in 1,\ldots,\text{n}$.

Dataset construction

We used a dataset containing 47 French Pro D2 games of a given team. A distribution of these games by location and outcome can be found in Table 1.

Table 1: Games locations and outcomes

	Win	Draw	Loss	**Total**
Away	4	-	13	**17**
Home	24	2	4	**30**
Total	**28**	**2**	**17**	**47**

For these games, we dispose of (i) event data and (ii) tracking data. Event data are manually labeled by the team's video analysts. Each entry corresponds to an event, and comprises its type (pass, tackle, kick, substitution, yellow or red card, etc.), timestamp, and the identity of the player performing it. Tracking data consists of the GPS position of all 23 players (including 8 substitutes) during the game, with a frequency of 10 Hz. This data is only available for the team which provided it. We have no data about the opposing team.

From event data, we identified the specific time intervals at which an action starts and finishes. We gathered 22 hours of playtime (4542 actions), from which we only considered offensive situations (11 hours - 2248 actions) as our ball position estimation method can only be applied when the analyzed team is in possession of the ball. We also used event data to remove from our analysis sequences in which the team is in numerical inferiority (due to yellow or red card) for the sake of simplicity, we set aside sequences with missing or aberrant values and actions either shorter than 3.2 seconds (our shortest time horizon, see below) or without identified ball handler (required to quantify performance, see below). After this cleaning step, 8 hours (1332 actions) of offensive play remain available for analysis.

We then split our data into sequences of different time horizons, to verify the robustness of findings. A time horizon consists in sequence length (the number of historical time steps used as input for prediction) and prediction length (the number of following time steps to be forecasted). Time horizons are expressed in tenth of seconds – tenth of seconds, as the frequency is 10Hz. For instance, Time Horizon 32-32 therefore refers to the task of observing 3.2 seconds and predicting the next 3.2 seconds. Sequences are then chronologically split into train (70%), validation (15%) and test (15%) sets. Table 2 displays the number of actions and sequences for each time horizon.

We transformed tracking Latitude and Longitude to x and y coordinates where the (0, 0) coordinate is the lower-left corner from top view of the pitch. We rotate second halves, so the analyzed team always attacks from left to right (increasing x for successful attacks), and we computed the velocity and goal angle of all 15 players playing the sequence (identified based on the starting lineup and the changes registered in event data).

Table 2: Number of actions and sequences per time horizon

Time Horizon	Actions	Sequences
16-16	1,332	112,325
32-32	1,093	90,033
64-64	795	57,629
128-128	303	23,771

Ball position extraction method

We used event data to estimate the position of the ball. For each possible type of event (approx. 30 are used by the technical staff), we identified if the event is a ball handler event (e.g. pass, breakthrough, etc.) or not (e.g. support, tackle, etc.). For ball handler events, we added a distinction between "give" events, where the ball handler loses control of the ball soon after the event is recorded (e.g. passes, kicks, etc.) and "keep" events, where the ball handler maintains control of the ball (e.g. breakthrough, contacts, etc.). This event classification was obtained from the technical staff.

We then used for each ball handler event the coordinates of the ball handler as coordinates of the ball for the timestamp of the corresponding event. For all timestamps between two ball handler events (meaning the vast majority of timestamps, as ball handler events do not occur at the same frequency the tracking data is captured), we used the coordinates of the ball handler of the first (respectively second) event if the event is considered a "keep" (respectively "give") event, and applied Gaussian filter to smooth ball trajectories. We then performed visual inspection of randomly selected reconstructed sequences versus the corresponding videos of the game.

To the best of our knowledge, this dataset is the only publicly available dataset containing Rugby Union Tracking data. We believe its size (47 games) and frequency (10Hz) make it valuable for the broader scope of team sports performance analysis.

Experiments

Experimental Setup

To assess the relevance of our suggested approach, we compare state-of-the-art time series models with naive baselines and vanilla models. As baselines, we used (i) a Linear Regression model and (ii) two naive approaches: Stand Still: $P_t = 0 \forall t > t_0 + s$ where $t_0$is the start of the sequence (considering the action makes no progress during the prediction period) and Constant Velocity: $P_t = i * v_p \forall t > t_0 + s$ where $v_p = \frac{P_{t_0+s} - P_{t_0}}{s}$,$i \in 1,\ldots,p$ (considering the action progresses during the prediction period at the same average speed than during the observed sequence). An illustration of target series and naive baselines is displayed in Figure 1. We selected as vanilla models LSTM (Hochreiter and Schmidhuber, 1997) and Transformer (Vaswani et al., 2017). Finally, we selected several state-of-the-art time series models, highlighted in Wang et al. (2024): iTransformer (Y Liu et al., 2023), PatchTST (Nie et al., 2023), Non-Stationary Transformer (Y Liu et al., 2022), Crossformer (Zhang & Yan, 2023), Pyraformer (S Liu et al., 2022) and DLinear (Zeng et al., 2022).

For the time series and vanilla models, we ran the two experiments: the first where only our prediction variable is used to predict itself (univariate to univariate prediction task) and the second where the whole feature vector is used to predict the target variable (multivariate to

univariate prediction task). The comparison between both experiments aims at quantifying the prediction ability gained by using tracking data.

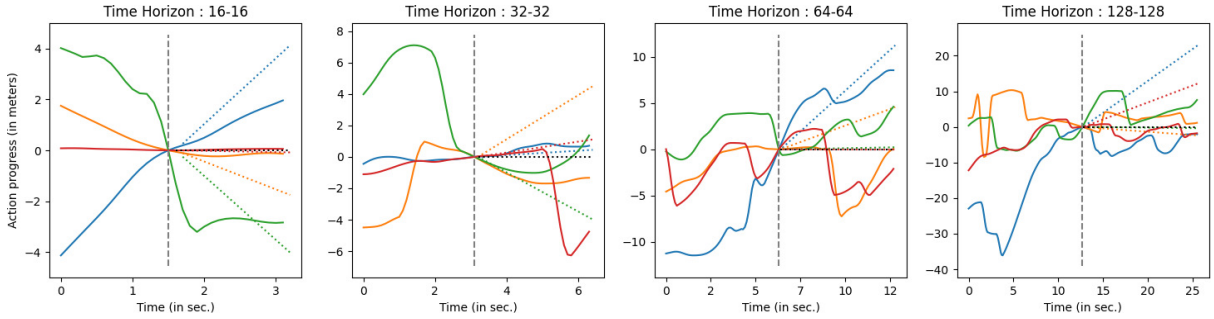


Fig. 1: Illustration of 4 randomly selected target series P over different time horizons. The vertical dashed gray line represents the separation between known sequence on the left and prediction sequence on the right, the horizontal dotted black line represents the Stand Still baseline and the colored dotted lines the Constant Velocity baseline associated with the series displayed with the same color.

Training procedure

Due to the lack of prior knowledge about which model would best fit our data, the considerable size of the hyperparameters space, and for a reasonable use of computing resources, we designed a halving / pruning protocol (Jamieson & Talwalkar, 2015). For each model, we conducted various experiments based on common hyperparameters ranges, taking into consideration authors' suggestions. We limited the training budget to 100 epochs (an epoch is an iteration of gradient descent in which all observations of the training set are used once) with an early stopping criterion set at 8, and every 4 epochs, we compared the validation loss of all "alive" models and pruned half of them. We used *adamw* optimizer (Loshchilov & Hutter, 2017), MSE loss, and a customized LR scheduler with warm-up and restarts (applying a decay of 0.5 at each restart) aligned with our pruning cycle.

Testing procedure

At the end of training, we evaluated over the test set the models of which training was terminated either by early stopping or by reaching the training budget (either case meaning the experiment completed the training). We found that in rare cases, during training, a given model M having reached a validation loss z_0 earlier during its training arrives at pruning step with a validation loss $z_1 > z_p > z_0$, where z_p is the pruning threshold. M is therefore pruned, and budget is allocated to other models, of which some will never reach z_0 and yet be tested. In such cases, we reintegrated M for testing.

We finally compared results from univariate and multivariate input models using the Mann-Whitney U test.

Results and Discussions

Training and pruning

Figure 2 (and Appendix 1) displays, for each time horizon, the best epoch and the MSE on the validation set for the 5 best experiments by model, for univariate and multivariate inputs. For most models, the best epoch is earlier for multivariate inputs than for univariate inputs. This observation is not surprising as considering more features to predict the same outputs can be expected to lead more easily to overfitting. This finding gets more visible with longer time horizons, which can be explained by the lower number of sequences (see Table 2). The consequence of the best epoch being reached earlier is fewer multivariate inputs models will be pruned, and therefore more will reach testing. During training, in most cases, univariate inputs models reached lower MSE on the validation set than their multivariate counterparts, which could indicate a better ability to fit the data. However, for shorter time horizons (left part of Figure 2), multivariate models reached lower MSE, with Crossformer (Zhang & Yan, 2023) for both univariate and multivariate reaching the lowest MSE. For longer time horizons (right part of Figure 2), Pyraformer (S Liu et al., 2022) for univariate and Crossformer (Zhang & Yan, 2023) and Patch TST (Nie et al., 2023) for multivariate performed best.

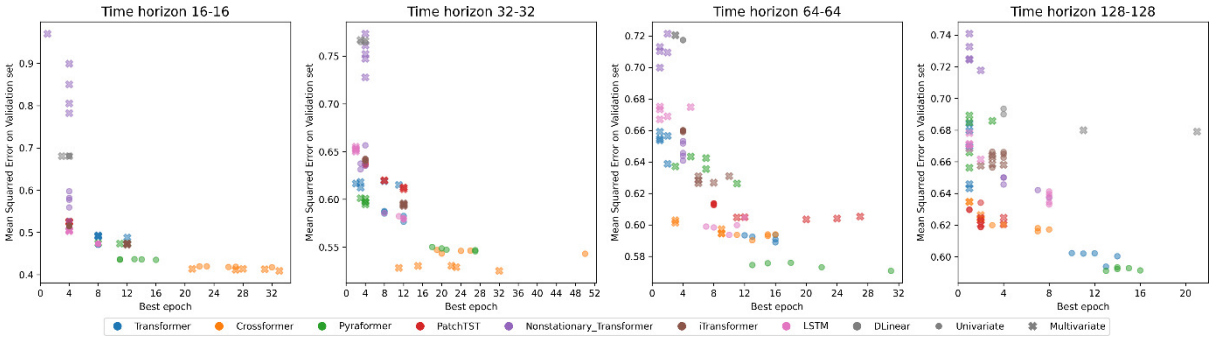


Fig. 2: MSE and best epoch of the 5 best experiments for each model over the validation set.

Testing

Figure 3 (and Tables 1 to 4 of Appendix for comprehensive figures) displays, for each time horizon, the distribution of relative improvements in FDE over the test set against the best performing baseline for models that reached the test step. Those models are those that either completed training or were reintegrated.

Our results show robustness across time horizons: in all four-time horizons tested, Linear Regression is the best performing baseline, multivariate input models outperforms both baseline and univariate input models, and in three-time horizons out of four, univariate input models outperforms baseline. In 16-16 (respectively 32-32) time horizon, Crossformer (Zhang & Yan, 2023) ranked best in multivariate and univariate reaching minimum FDE of 3.90 meters and 3.95 meters (respectively 6.60 meters and 6.86 meters) versus baseline reaching 4.50 meters (respectively 7.82 meters). In 64-64 time horizon, Crossformer (Zhang & Yan, 2023) with multivariate input still reached the lowest FDE with 9.83 meters while Pyraformer (S Liu et al., 2022) was the best univariate input model with FDE of 10.43 meters (vs 11.14 meters for baseline). In 128-128 time horizon, no univariate input model reached the baseline FDE of 13.46 meters (lowest FDE among univariate is 13.59 meters for Pyraformer) but multivariate input models still perform best reaching a 12.29 meters FDE for Patch TST (Nie et al., 2023).

We also reported the number of models tested. Surprisingly, while during the training step

univariate input models reached lower MSE than multivariate inputs models for longer time horizons, during the testing phase, for all time horizons, the lowest FDE model is a multivariate model, and visual inspection of the boxplots indicates multivariate models outperform univariate models. These observations suggest univariate models over longer time horizons are likely to be overfitting. Due to the lower number of actions and sequences used in longer time horizons (see Table 2), we are probably missing data to validate findings over longer periods.

To assess the statistical significance of our findings, we performed a Mann-Whitney U test. Its results appear in Figure 3. In line with the visual inspection of boxplots, the test reports significant difference in performance between multivariate and univariate inputs models for all time horizons except 16-16, with p-values below 10^{-5}. It confirms superiority of multivariate input models in these time horizons.

In line with what we observed during the training step, Crossformer (Zhang & Yan, 2023) models achieve the best performance in shorter time horizon in both univariate and multivariate inputs, while in longer time horizons, no model reached the same dominance. The good performance of Crossformer (Zhang & Yan, 2023) indicates the relevance of using a cross-dimension attention mechanism: the consideration of interactions between the players increases the ability to model and predict performance.

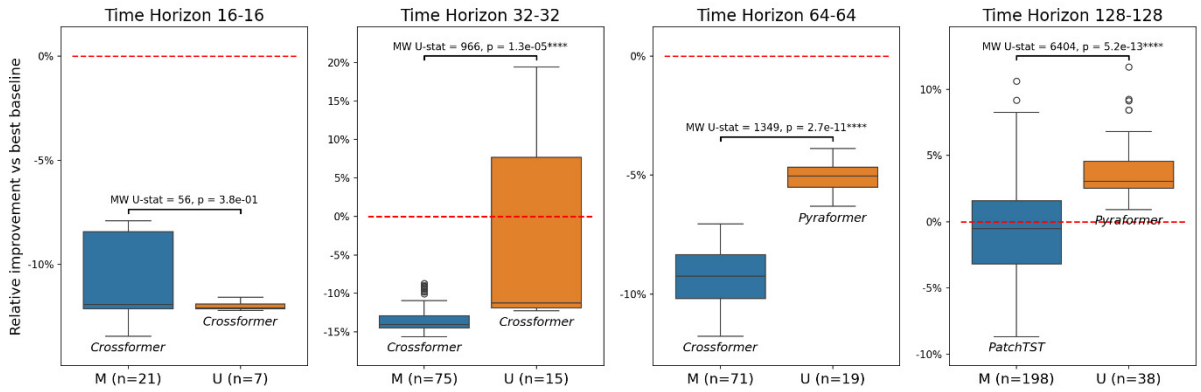


Fig. 3: Distribution of FDE relative improvement (in %) vs best baseline (Linear Regression for all cases) for Multivariate (M) and Univariate (U) inputs models for each time horizon. Each graph represents the results of the test procedure over the corresponding time horizon. The number in parenthesis indicates the number of models tested (see 4.1) and the model names located below the boxplots indicate the models reaching the lowest minimum FDE over the test set. The dashed red line at y = 0 materializes the score of the best performing baseline. Results from Mann-Whitney U tests are reported above the boxplots. Significance code: * p<0.05, ** p< 0.01, *** p<0.001, **** p<0.0001.

Feature importance

To assess the relative importance of the input features, we conducted a permutation analysis on the best performing models over the train set. The result of this analysis for time horizon 128-128 is displayed in Figure 4. This permutation analysis aims at quantifying the importance of feature $i \in \{1,\ldots,61\}$ by computing the mean difference between $FDE(f(X_{test}), Y_{test})$ and $FDE\left(f\left(X^i_{test}\right), Y_{test}\right)$ over k iterations where f is the fitted model, X_{test} represents the test set, Y_{test} the corresponding ground truth labels and X^i_{test} the test set with values of feature i randomly permuted and all features $j \neq i$ maintained in the correct order. k is a hyperparameter defined by the user and was set to 3 in our experiment. In other words, we consider the importance of a given feature for the model to be proportional to the degradation of the predictive

power of the model when the given feature is randomly permuted.

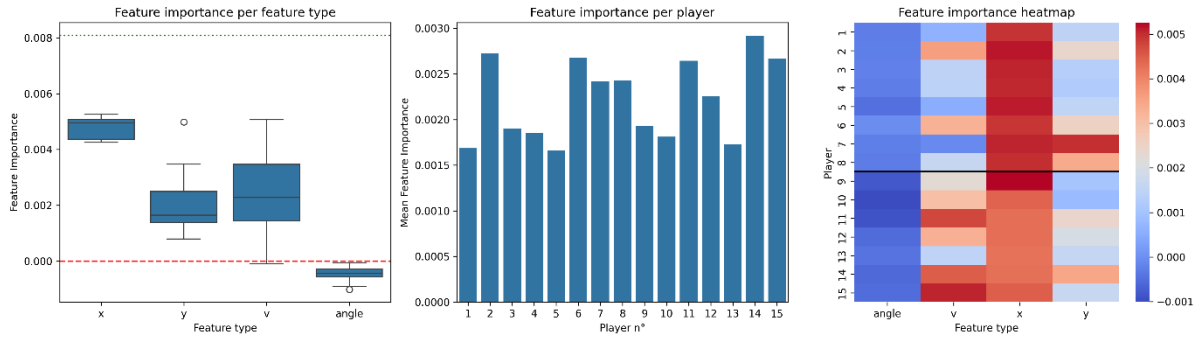


Fig. 4: Analysis of Permutation importance for time horizon 128-128. The left panel displays the importance per feature type (position over the x-axis and y-axis, velocity and goal angle). The dashed red line materializes the 0-impact level while the dotted green line the importance of the performance feature. The middle panel indicates the importance of features per player. The heatmap of the right panel indicates feature importance over the two combined dimensions. The black line separates the forwards (player 1 to 8) and backs (players 9 to 15).

The left panel indicates position on the x-axis and in a lesser extent velocity features generate higher degradation of prediction when permuted, indicating higher importance for the model. All values of the boxplots fell below the dotted green line indicating the feature importance of the performance feature, indicating the later seems to be the feature bringing most explicative power to the model. This finding is intuitive as difference between univariate input models (where only past values of the performance feature are provided) perform rather close to multivariate (even if multivariate inputs models perfom best). When analyzed by player (center panel), player 6 to 8 (third line) and 11 and 14 (wings) appear to be more important when computing feature importance. Visual inspection of the heatmap (right panel) reveals a distinction between forwards (player 1 to 8) for whom x position is the most impacting feature and backs (players 9 to 15) for whom x position appears of lesser importance than for forwards but for whom velocity appears to be relevant. Angle and y features seems to be of considerably lower importance.

Conclusion

This study presents several novel contributions to the field of sports analytics. First, we leveraged existing work in time series analysis to model short-term performance in Rugby Union. Second, we constructed a dataset using a preprocessing pipeline that includes a method to estimate ball position, addressing a common limitation in sports analysis. By making our dataset publicly available, we contribute to the advancement of sports analytics research. Furthermore, we propose a comprehensive benchmark of state-of-the-art time series models on this prediction task, highlighting which models are best fitted to predict short-term performance in Rugby Union from tracking data .

On our experiments, we found that multivariate to univariate prediction task outperforms univariate to univariate and baselines, indicating the relevance of using players tracking data to predict short term performance. We used here the position of the ball over the x axis of the field as performance indicator, due to its relevance (see introduction section). This approach could be extended to alternative indicators that capture team dynamics more broadly such as the team's barycenter, or dynamics of specific subgroups, like the barycenter of the forward pack for Rugby

Union (players 1 to 8).

Our work also emphasizes models designed to deal with long sequences and decomposition challenges (e.g. electric consumption and weather forecast) can work on shorter time-periods, with data displaying different characteristics.

Our work presents room for improvement. The validation of our ball position estimation pipelines only relies on visual inspection of randomly selected sequences, and not on a robust metric. The robustness of disclosed results might be strengthened, as we only gathered data from a single team and miss data for longer time horizons. Moreover, considering only the offensive team is a clear limitation for two reasons. First, the attacking and defensive teams constantly adapt to one another, we are therefore missing precious information about play development. Second, a team in offense in Rugby Union can either try to advance with (by carrying and / or passing) or without (by kicking) the ball. Only considering actions where the analyzed team is in possession of the ball leads us to fail to capture the success of a team trying to advance without the ball, as sometimes, losing the ball is a desired outcome (e.g. when kicking). Finally, we would benefit from using ball tracking data instead of our ball extraction pipeline, as it would (i) be more precise as previous research highlighted the asynchrony challenge of combining automatically captured tracking and manually annotated event data (Biermann et al. 2023 and Kim et al. 2025) and (ii) allow us to consider defensive actions.

This study opens up several promising research directions. Performance estimation could support the analysis of interpersonal **coordination**, including fluctuations in running correlations between players, which have been shown to be an important parameter for success in Rugby Union (Rodrigues et al., 2013). Furthermore, predicted offense outcomes could be used in **game simulation and invented scenario to** test alternative offensive plays. Our research also relates with multiagent trajectory prediction, which could make us benefit from deep learning architectures specifically designed for spatio-temporal data (Yu et al., 2020, Wu et al., 2023). These models leverage Graph Neural Networks to handle the spatial component of the data and are notably used to model pedestrian trajectories. Additionally, applying explainability techniques,which aim to identify the key factors driving model's prediction, or generating counterfactuals could provide valuable insights for coaching staffs (Odet et al., 2025). While it has not yet been systematically demonstrated, it is plausible that players contribute unequally to the short-term prediction of ball trajectory. Certain specific players, such as the ball carrier and the supporting runners, are likely to exert a stronger influence on where the ball will go next. Focusing on these players could improve prediction accuracy and model interpretability. For example by modifying the feature vector of a given subset of players (for instance, the forward players during a scrum or a lineout and analyzing the resulting changes in performance prediction, one could gain actionable insights into the quality of the initial movements of these key players. This could be helpful in play development simulations to enhance tactical understanding of players and staffs. Best performing models can also be used as reward models in Reinforcement Learning tasks to identify which collective choreography yields the best outcome.

Acknowledgments

We thank the US Colomiers Rugby Union club for providing us with the data used in this study. This work was granted access to the HPC resources of CALMIP supercomputing center under the allocation P25003.

References

Amirli, A, & Alemdar, H. (2022). Prediction of the ball location on the 2d plane in football using optical tracking data. *Academic Platform Journal of Engineering and Smart Systems, 10* 1-8. https://doi.org/10.21541/apjess.1060725

Anzer, G., & Bauer, P. (2021). A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Frontiers in Sports and Active Living*, *3*. https://doi.org/10.3389/fspor.2021.624475

Anzer, G, Bauer, P, Brefeld, U, & Fassmeyer, D. (2022). Detection of tactical patterns using semi-supervised graph neural networks. *MIT Sloan Sports Analytics Conference*.

Bennett, M., Bezodis, N., Shearer, D. A., Locke, D., & Kilduff, L. P. (2019). Descriptive conversion of performance indicators in rugby union. *Journal of Science and Medicine in Sport*, *22*(3), 330–334. https://doi.org/10.1016/j.jsams.2018.08.008

Biermann, H., Komitova, R., Raabe, D., Müller-Budack, E., Ewerth, R., & Memmert, D. (2023). Synchronization of passes in event and spatiotemporal soccer data. *Scientific Reports*, *13*(1), 15878. https://doi.org/10.1038/s41598-023-39616-2

Box, GEP, Jenkins, GM, Reinsel, GC, & Ljung, GM. (1978). Time series analysis: Forecasting and control. *The Statistician, 27*, 265–265. https://doi.org/10.1111/jtsa.12194

Cao, D., Wang, Y., Duan, J., Zhang, C., Zhu, X., Huang, C., Tong, Y., Xu, B., Bai, J., Tong, J., & Zhang, Q. (2021). Spectral Temporal Graph Neural Network for Multivariate Time-series Forecasting. *ArXiv*. https://doi.org/10.48550/arXiv.2103.07719

Cervone, D, D'Amour, A, Bornn, L, & Goldsberry, K. (2014). A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association, 111*, 585–599. https://doi.org/10.1080/01621459.2016.1141685

Chakraborty, B., & Meher, S. (2012). Real-time position estimation and tracking of a basketball. *2012 IEEE International Conference on Signal Processing, Computing and Control*, 1–6. https://doi.org/10.1109/ISPCC.2012.6224370

Clemente, M. F., Couceiro, S. M., Martins, F. M. L., Mendes, R., & Figueiredo, A. J. (2013). Measuring Collective Behaviour in Football Teams: Inspecting the impact of each half of the match on ball possession. *International Journal of Performance Analysis in Sport*, *13*(3), 678–689. https://doi.org/10.1080/24748668.2013.11868680

Colomer, C. M. E., Pyne, D. B., Mooney, M., McKune, A., & Serpell, B. G. (2020). Performance Analysis in Rugby Union: a Critical Systematic Review. Sports Medicine - Open, 6(1), 4. https://doi.org/10.1186/s40798-019-0232-x

Dick, U., Tavakol, M., & Brefeld, U. (2021). Rating Player Actions in Soccer. *Frontiers in Sports and Active Living*, *3*, 682986. https://doi.org/10.3389/fspor.2021.682986

Fernández, J. (2020). Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. *MIT Sloan Sports Analytics Conference*.

Ferraz, A, Duarte-Mendes, P, Sarmento, H, Valente-dos-Santos, J, & Travassos,B. (2023). Tracking devices and physical performance analysis in team sports: A comprehensive framework for research—trends and future directions. *Frontiers in Sports and Active Living, 5,* 1284086. https://doi.org/10.3389/fspor.2023.1284086

Ghasemzadeh, S.A., Zandycke, G.V., Istasse, M., Sayez, N., Moshtaghpour, A., & Vleeschouwer, C.D. (2021). DeepSportLab: a Unified Framework for Ball Detection, Player Instance Segmentation and Pose Estimation in Team Sports Scenes. *British Machine Vision Conference*. https://doi.org/10.48550/arXiv.2112.00627

Goes, F.R., Brink, M.S., Elferink-Gemser, M.T., Kempe, M., & Lemmink, K.A. (2020). The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data. *Journal of Sports Sciences*, 39, 523 - 532.https://doi.org/10.1080/02640414.2020.1834689

Gu, A, & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. *ArXiv, abs/2312.00752*.https://doi.org/10.48550/arXiv.2312.00752

Gudmundsson, J., & Horton, M. (2016). Spatio-Temporal Analysis of Team Sports. *ACM Computing Surveys (CSUR)*, 50, 1 – 34.https://doi.org/10.1145/3054132

Hauri, S., & Vucetic, S. (2022). Group Activity Recognition in Basketball Tracking Data - Neural Embeddings in Team Sports (NETS). *European Conference on Artificial Intelligence*.https://arxiv.org/pdf/2209.00451

Hochreiter, S, & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*, 1735–1780.https://doi.org/10.1162/neco.1997.9.8.1735

Hughes, A, Barnes, A, Churchill, SM, & Stone, JA. (2017). Performance indicators that discriminate winning and losing in elite men's and women's rugby union. *International Journal of Performance Analysis in Sport, 17*, 534–544.https://doi.org/10.1080/24748668.2017.1366759

Hughes, H., Horton, M., Wei, X., Gammulle, H., Fookes, C., Sridharan, S., & Lucey, P. (2025). Event2Tracking: Reconstructing Multi-Agent Soccer Trajectories Using Long-Term Multimodal Context. *AAAI Conference on Artificial Intelligence*.https://doi.org/10.1609/aaai.v39i11.33289

Jamieson, KG, & Talwalkar, A. (2015). Non-stochastic best arm identification and hyperparameter optimization. *International Conference on Artificial Intelligence and Statistics*.https://doi.org/10.48550/arXiv.1502.07943

Khaustov, V., & Mozgovoy, M. (2020). Recognizing Events in Spatiotemporal Soccer Data. *Applied Sciences*.https://doi.org/10.3390/app10228046

Kim, H, Choi, HJ, Kim, CJ, Yoon, J, & Ko, SK. (2023). Ball trajectory inference from multi-agent sports contexts using set transformer and hierarchical bi-lstm. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.https://doi.org/10.1145/3580305.3599779

Kim, H., Choi, H., Seo, S., Boomstra, T., Yoon, J., & Park, C. (2025). ELASTIC: Event-Tracking Data Synchronization in Soccer Without Annotated Event Locations. *ArXiv, abs/2508.09238.*https://doi.org/10.48550/arXiv.2508.09238

Klein, JL. (1997). Statistical visions in time: A history of time series analysis, 1662-1938.https://doi.org/10.2307/2669730

Kovalchik, SA. (2022). Player tracking data in sports. *Annual Review of Statistics and Its Application, 10,* 677-697.https://doi.org/10.1146/annurev-statistics-033021-110117

Liu, S, Yu, H, Liao, C, Li, J, Lin, W, Liu, AX, & Dustdar, S. (2022). Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. *International Conference on Learning Representations*.

Liu, Y, Hu, T, Zhang, H, Wu, H, Wang, S, Ma, L, & Long, M. (2023). Itransformer: Inverted transformers are effective for time series forecasting. *International Conference on Learning Representations*.https://doi.org/10.48550/arXiv.2310.06625

Liu, Y, Wu, H, Wang, J, & Long, M. (2022). Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting. Neural Information Processing Systems.https://doi.org/10.48550/arXiv.2205.14415

Loshchilov, I, & Hutter, F. (2017). Decoupled weight decay regularization. *International Conference on Learning Representations*.https://doi.org/10.48550/arXiv.1711.05101

Low, B, Coutinho, D, Gonçalves, B, Rein, R, Memmert, D, & Sampaio, J. (2020). A Systematic Review of Collective Tactical Behaviours in Football Using Positional Data. *Sports Medicine, 50*(2), 343–385.https://doi.org/10.1007/s40279-019-01194-7

Miljkovic, D, Gajic, L, Kovacevic, A, & Konjovic, Z. (2010). The use of data mining for basketball matches outcomes prediction. *IEEE 8th International Symposium on*

Intelligent Systems and Informatics, 309–312. https://doi.org/10.1109/SISY.2010.5647440

Nie, Y, Nguyen, NH, Sinthong, P, & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. *International Conference on Learning Representations.* https://doi.org/10.48550/arXiv.2211.14730

Odet, A., Bechard, T., Moretto, P., Dejean, S., & Pasquaretta, C. (2025). A machine learning and explainability-driven methodology for identifying winning strategies in Rugby Union. *Decision Analytics Journal, 15*. https://doi.org/10.1016/j.dajour.2025.100568

Ohno, Y., Miura, J., & Shirai, Y. (2000). Tracking players and estimation of the 3D position of a ball in soccer games. *Proceedings 15th International Conference on Pattern Recognition*. ICPR-2000, 1, 145-148 vol.1.https://doi.org/10.1109/ICPR.2000.905293

Pérez-Toledano, MÁ, Rodriguez, FJ, García-Rubio, J, & Ibáñez, SJ. (2019). Players' selection for basketball teams, through performance index rating, using multiobjective evolutionary algorithms. *PLoS ONE, 14*(9). https://doi.org/10.1371/journal.pone.0221258

Perše, M, Kristan, M, Kovačič, S, Vučkovič, G, & Perš, J. (2009). A trajectory based analysis of coordinated team activity in a basketball game. *Computer Vision and Image Understanding, 113*(5), 612–621.https://doi.org/10.1016/j.cviu.2008.03.001

Ren, J, Orwell, Jones, J, Xu, G, Orwell, J, Jones, GA, & Xu, M. (2009). Tracking the soccer ball using multiple fixed cameras. Strathprints Institutional Repository, *Computer Vision and Image Understanding, 113* (5). 633-642. https://doi.org/10.1016/j.cviu.2008.01.007

Rodrigues, M.M., & Passos, P. (2013). Patterns of Interpersonal Coordination in Rugby Union: Analysis of Collective Behaviours in a Match Situation. Advances in Physical Education, 03, 209-214.http://dx.doi.org/10.4236/ape.2013.34034

Rumelhart, DE, Hinton, GE, & Williams, RJ. (1986). Learning representations by back-propagating errors. *Nature, 323,* 533–536.https://doi.org/10.1038/323533a0

Schoeman, R, & Schall, R. (2019). Team performance indicators as predictors of final log position and team success in aviva premiership, guinness pro 14, french top 14 and super rugby. *International Journal of Performance Analysis in Sport, 19*, 763–777. https://doi.org/10.1080/24748668.2019.1655337

Silva, H., Dias, A., Almeida, J.M., Martins, A., & Silva, E.P. (2012). Real-Time 3D Ball Trajectory Estimation for RoboCup Middle Size League Using a Single Camera. *Robot Soccer World Cup*. https://doi.org/10.1007/978-3-642-32060-6_50

Tian, C, De Silva, V, Caine, M, & Swanson, S. (2019). Use of Machine Learning to Automate the Identification of Basketball Strategies Using Whole Team Player Tracking Data. *Applied Sciences, 10*(1), 24. https://doi.org/10.3390/app10010024

Vanrenterghem, J, Nedergaard, NJ, Robinson, MA, & Drust, B. (2017). Training load monitoring in team sports: A novel framework separating physiological and biomechanical load-adaptation pathways. *Sports Medicine, 47*, 2135–2142. https://doi.org/10.1007/s40279-017-0714-2

Vaswani, A, Shazeer, NM, Parmar, N, Uszkoreit, J, Jones, L, Gomez, AN, Kaiser, L, & Polosukhin, I. (2017). Attention is all you need. *Neural Information Processing Systems*. https://doi.org/10.48550/arXiv.1706.03762

Wang, Y, Wu, H, Dong, J, Liu, Y, Long, M, & Wang, J. (2024). Deep Time Series Models: A Comprehensive Survey and Benchmark. *ArXiv, abs/2407.13278*. https://doi.org/10.48550/arXiv.2407.13278

Wu, Y., Wang, L., Zhou, S., Duan, J., Hua, G., & Tang, W. (2023). Multi-Stream Representation Learning for Pedestrian Trajectory Prediction. AAAI Conference on Artificial Intelligence. https://doi.org/10.1609/aaai.v37i3.25389

Yu, C., Ma, X., Ren, J., Zhao, H., & Yi, S. (2020). Spatio-Temporal Graph Transformer Networks for Pedestrian Trajectory Prediction. *European Conference on Computer Vision* https://doi.org/10.48550/arXiv.2005.08514

Yurko, R, Matano, F, Richardson, LF, Granered, N, Pospisil, T, Pelechrinis, K, & Ventura, SL. (2019). Going deep: Models for continuous-time within-play valuation of game outcomes in american football with tracking data. *Journal of Quantitative Analysis in Sports, 16*, 163–182. https://doi.org/10.48550/arXiv.1906.01760

Zeng, A, Chen, M, Zhang, L, & Xu, Q. (2022). Are Transformers Effective for Time Series Forecasting? *AAAI Conference on Artificial Intelligence.* https://doi.org/10.48550/arXiv.2205.13504

Zhang, Y, & Yan, J. (2023). Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. *International Conference on Learning Representations.*

Appendix

Tables 1 to 4 summarize the results of the testing phase.
The metric displayed is the FDE (on non-scaled values) reached over the test set.
For each model, n indicates the number of model tested, Min. FDE the lowest FDE reached by the corresponding model, μ the average of all model tested, and the σ the standard deviation when more than one model (i.e. with different hyperparameters) were tested. For Min. FDE and μ, the best value is indicating in bold and the second and third best are underlined.

Table 1: Testing summary - Time horizon 16-16

Model	Inputs	n	Min. FDE	μ	σ
Crossformer	Multivariate	12	**3.90**	**3.94**	±0.03
Crossformer	Univariate	7	<u>3.95</u>	<u>3.96</u>	±0.01
Pyraformer	Multivariate	9	<u>4.11</u>	<u>4.13</u>	±0.01
StandStill	Baseline	1	5.23	5.23	-
ConstantVelocity	Baseline	1	6.13	6.13	-
Linear Regression	Baseline	1	4.50	4.50	-

Table 2: Testing summary - Time horizon 32-32

Model	Inputs	n	Min. FDE	μ	σ
Crossformer	Multivariate	57	**6.60**	**6.71**	±0.05
Crossformer	Univariate	6	<u>6.86</u>	<u>6.88</u>	±0.02
Non Stationary Transformer	Univariate	6	7.46	8.49	±0.61
Pyraformer	Multivariate	13	7.04	7.10	±0.03
Pyraformer	Univariate	3	<u>6.91</u>	<u>6.93</u>	±0.02
iTransformer	Multivariate	5	6.93	6.96	±0.02
StandStill	Baseline	1	8.44	8.44	-
ConstantVelocity	Baseline	1	11.41	11.41	-
Linear Regression	Baseline	1	7.82	7.82	-

Table 3: Testing summary - Time horizon 64-64

Model	Inputs	n	Min. FDE	μ	σ
Crossformer	Multivariate	40	**9.83**	<u>10.13</u>	±0.13
Crossformer	Univariate	1	10.48	10.48	-
PatchTST	Multivariate	15	<u>10.01</u>	<u>10.13</u>	±0.07
Pyraformer	Multivariate	2	10.20	10.22	±0.03
Pyraformer	Univariate	18	10.43	10.57	±0.07
iTransformer	Multivariate	14	<u>9.85</u>	**9.95**	±0.06
StandStill	Baseline	1	11.79	11.79	-
ConstantVelocity	Baseline	1	17.56	17.56	-
Linear Regression	Baseline	1	11.14	11.14	-

Table 4: Testing summary - Time horizon 128-128

Model	Inputs	n	Min. FDE	μ	σ
Crossformer	Multivariate	18	13.48	13.68	±0.11
Crossformer	Univariate	3	13.68	13.78	±0.09
DLinear	Multivariate	2	13.65	13.72	±0.10
LSTM	Multivariate	7	13.93	14.35	±0.28
PatchTST	Multivariate	88	**12.29**	**13.07**	±0.44
Pyraformer	Multivariate	7	13.51	13.99	±0.31
Pyraformer	Univariate	30	13.59	13.94	±0.25
Transformer	Multivariate	5	13.70	13.71	±0.01
Transformer	Univariate	4	14.38	14.68	±0.27
iTransformer	Multivariate	71	<u>12.74</u>	<u>13.50</u>	±0.35
StandStill	Baseline	1	15.09	15.09	-
ConstantVelocity	Baseline	1	25.17	25.17	-
Linear Regression	Baseline	1	<u>13.46</u>	<u>13.46</u>	-