Prediction of Physical Performance

Proof of Concept

Code: <https://github.com/Lentz92/physical-performance-metric>

How to read the code: There are 4 important directories: data, notebooks, src, visualisations.

* Data: all the data raw and processed
* Notebooks: The main work. They are in numerical order of execution.
* Src: the source files used in the notebooks
* Visualisations: All the figures created.

## Introduction

The game speed in football has increased over the last 50 years [1]. There has been an increase of approximately 50% in the quantity of short bouts of high-intensity running and a 40% increase in the number of ball passes, while the total distance travelled has decreased by about 2% [2]. Given that football matches often are decided by just one goal, it is particularly important to identify players who experience a drop in physical performance.

One of the most powerful tools a coach can use to influence a match is substitutions. Therefore, it is crucial for the coach to have a deep understanding of which players are experiencing a decline in physical performance. However, IT and data solutions often struggle to provide helpful and concrete feedback to the coach due to the complexity of the data, resulting in inconclusive communication [3].

Therefore, my proposal is to engineer a comprehensive Physical Performance Metric (PPM) that incorporates both external data (distance covered, high-intensity running, etc.) and internal data (player load, accelerations, etc.). While this metric will not capture all the complexity, it can be a powerful tool by providing clear and conclusive communication to the coach. By basing the PPM on the historical data of the player and their position, it can aid in making quick decisions during a match.

To build on this, it is important to predict when a player will experience a decline in physical performance ahead of time. Therefore, I also propose a forecasting solution that can predict the expected performance of a player during a match.

## Method

For a more in-depth documentation of the analysis, please refer to the code provided. It includes additional visualizations and intermediate work that contributed to this paper.

### Data cleaning

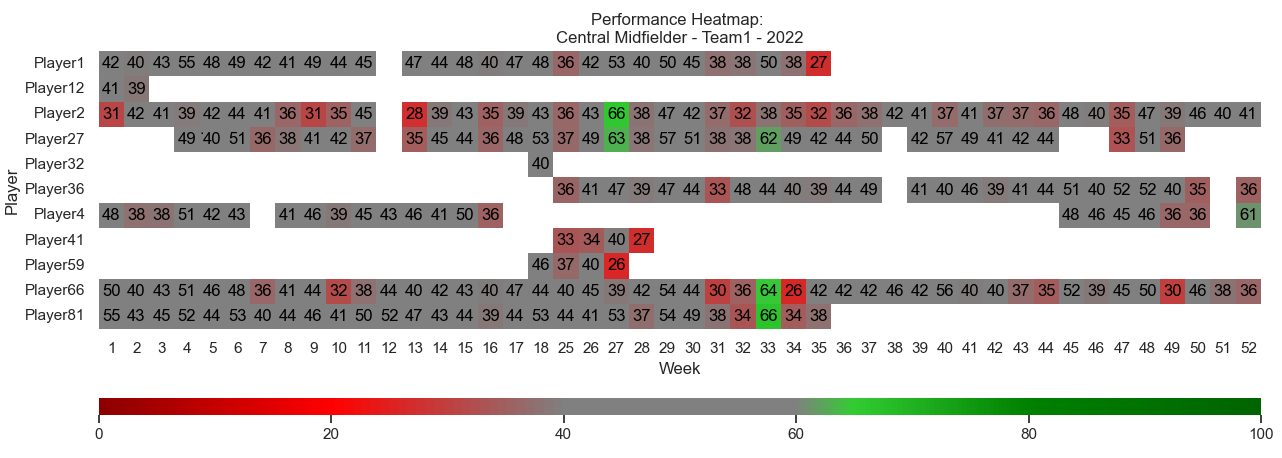
All data from training sessions shorter than 30 minutes were excluded, as they were considered too brief to be meaningful. Additionally, players with fewer than 100 data points were excluded to avoid significant gaps in their monitoring data.

### Physical Performance Metric

The Physical Performance Metric was created using a straightforward approach: applying weights to different columns based on my personal reasoning. I assigned weightings to the variables according to their perceived importance, with total player load and high-intensity running distance being the most significant variables. Additionally, "ima\_cod" is a summation of both left and right changes of direction (Table 1).

*Table 1: Variable weightings for the physical performance metric*

|  |  |
| --- | --- |
| Variable | Weight |
| total\_distance\_m | 0.05 |
| total\_player\_load | 0.25 |
| acc\_2m\_s\_s\_total\_efforts | 0.1 |
| acc\_3\_m\_s\_s\_total\_efforts | 0.1 |
| dec\_2\_m\_s\_s\_total\_efforts | 0.1 |
| dec\_3m\_s\_s\_total\_efforts | 0.1 |
| high\_intensity\_distance\_m\_v5\_v6\_m | 0.25 |
| sprint\_distance\_m\_m | 0.05 |
| maximum\_velocity\_km\_h | 0.05 |
| ima\_cod | 0.1 |

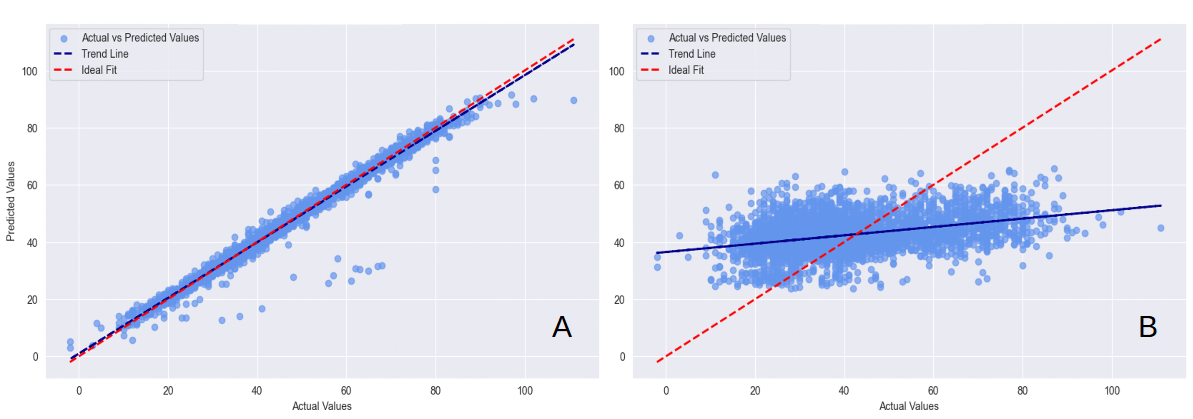
The PPM was then normalized to a score between 0 and 100, with 0 representing the lowest possible performance, 50 representing the mean performance, and 100 representing the absolute best performance. The normalization was performed with respect to the playing position to ensure fair comparisons. For example, normalizing across the entire team would make it seem like goalkeepers never performed very well. This normalization makes it easy to visualize with a heatmap (figure 1).

*Figure 1: Heatmap of the physical performance metric for the central midfielder on Team 1 in 2022. The y-axis represents the players, while the x-axis represents the weeks. The color gradient ranges from shades of red (poor performance) to grey (normal performance) to shades of green (good performance).*

## Modeling

The modeling approach used a simple random forest regression algorithm to predict the PPM for football players on the next monitored training day. Before applying the algorithm, some feature engineering was conducted. This involved using all the variables listed in Table 1, as well as creating lagged versions of these variables, representing their values from one day up to seven days prior. These engineered features were then used in the model to improve its predictive accuracy, allowing coaches to forecast the PPM for players for the following day.

In an ideal scenario, the predictions would resemble Figure 2A, where the dots align almost perfectly with the ideal line. However, the initial model yielded poor performance, as shown in Figure 2B. Here, the dots and the blue line diverge significantly from the red ideal line.

*Figure 2: Figure A represents the ideal fit, while Figure B shows the actual performance of the model. The x-axis displays the actual values, and the y-axis shows the predicted values. Each dot represents a Physical Performance Metric (PPM) value. The blue dashed line is the best-fit line between the points, and the red dashed line represents an ideal fit.*

## Evaluation

Even though the modeling results were poor, they establish a foundation for predicting physical performance. Imagine a match where the coach must decide whether to substitute Player A or Player B for Player C. Having a tool that can reliably forecast a player's physical performance 10 minutes in advance could provide crucial insights for making informed decisions. Dijkhuis et al. (2021) accomplished a similar goal by predicting whether a player would perform or underperform [4].

The tool will function as a decision support system, helping the coach by providing a simple, concise metric. The coach can then combine this information with tactical and technical insights from the game to make informed decisions.

To enhance the capabilities of this system, more raw data, especially from the IMU, will be required. As stated in the introduction, the game of football increasingly revolves around fast, dynamic movements rather than linear running. The IMU excels at capturing this information, which will be crucial for developing an accurate Physical Performance Metric (PPM) and making reliable predictions.

## References

[1] Wallace, J.L.; Norton, K.I. Evolution of World Cup soccer final games 1966-2010: Game structure, speed and play patterns. *J. Sci.*

*Med. Sport* **2014**, *17*, 223–228

[2] Barnes C, Archer D, Hogg B, Bush M, Bradley P (2014) The evolution of physical and technical performance parameters in English Premier League. Int J Sports Med 35:1095–1100.

[3] Nosek, P.; Brownlee, T.E.; Drust, B.; Andrew, M. Feedback of GPS training data within professional English soccer: A comparison of decision making and perceptions between coaches, players and performance staff. Sci. Med. Footb. 2021, 5.1, 35–47.

[4] Dijkhuis, Talko B., Matthias Kempe, and Koen APM Lemmink. "Early Prediction of Physical Performance in Elite Soccer Matches—A Machine Learning Approach to Support Substitutions." Entropy 23.8 (2021): 952.