UiT The Arctic University of Norway

School of Sports Sciences, Faculty of Health Sciences

Developing a Linear Mixed Model to Predict RPE and sRPE in Female Elite Football Players Using External Load Measures

Extended introduction, previous literature, and methods

Jakob Kjos Vollstad

Master's thesis in Sports Sciences…3901 May 2023

Preface

As I complete my master's thesis, I would like to express my gratitude to those who have supported me throughout this journey. Balancing my role as a full-time football coach with my studies has been a challenging yet rewarding experience. I am grateful to my colleagues for their understanding and assistance when needed.

I would like to extend my thanks to UiT - The Arctic University of Norway for their guidance and support, as well as the Female Football Research Centre (FFRC) for providing the data essential for my study.

A special thank you goes to my supervisor, Andreas Winther, whose help has been instrumental in the completion of this thesis. I would also like to acknowledge my fellow students for their engaging and inspiring discussions that have enriched this experience.

Jakob Kjos Vollstad

Table of content

List of Figures

List of Tables

1 Introduction

Professional football has experienced significant advancements in sports science and technology in recent years, leading to enhanced player performance strategies (Brocherie & Beard, 2020). Despite these advancements, existing research predominantly focuses on male athletes, creating a substantial knowledge gap regarding female players; as stated by Okholm Kryger et al. (2022), *"the numbers are far from comparable to current research output levels in men's football"*.

Kirkendall and Krustrup (2022) report that a mere 15% of football research includes women, with previous studies primarily addressing physical demands, talent identification, body composition, injury risk mitigation, health, and nutrition. However, research on training load monitoring and management in female players remains scarce (Luteberget et al., 2021; Okholm Kryger et al., 2022). Consequently, traditional training and match principles have been designed primarily with male athletes in mind, potentially neglecting female athletes' specific needs and characteristics (Harkness-Armstrong et al., 2022; Kirkendall, 2020; Luteberget et al., 2021).

As technology advances and competition demands increase (Bullock et al., 2022; FIFA, 2022; Randell et al., 2021), coaching practices must evolve accordingly (Brocherie & Beard, 2020). With numerous factors influencing the coaches' daily planning process, the need for methods to help elicit the necessary training load is ever-present. Monitoring training load enables practitioners to tailor training programs to athletes' needs, balancing workload and recovery to reduce injury risk and enhance performance (Impellizzeri et al., 2019).

To aid such challenge, research has suggested using predictions to evaluate medical, training and performance data in football (Bullock et al., 2022; Bullock et al., 2023; McCall et al., 2017; Rico-González et al., 2023; Seshadri et al., 2021). Prediction models can support practitioners in making daily decisions related to performance and health by using data from multiple key predictors measured at a specific time. These models help estimate an individual's probability of experiencing a health- or performance-related outcome, either at the time of measurement (diagnosis) or in the future (prognosis) (Bullock et al., 2022; Collins et al., 2015; Impellizzeri et al., 2021; McCall et al., 2017).

Indeed, integrating sophisticated technology and analytics offer significant potential to help coaches predict players' training load, allowing for more individualised and effective training prescription (Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022; Seshadri et al., 2021). Combining this data-informed approach with practitioners' expertise and intuition can optimise player performance, well-being and help reduce injury risk for both male and female footballers (Bullock et al., 2022; Bullock et al., 2023; Impellizzeri, McCall, et al., 2020; Impellizzeri, Ward, et al., 2020; Luteberget et al., 2021; McCall et al., 2017; Rico-González et al., 2023; West et al., 2021).

1.1 Issue

The purpose of this study is to develop a prediction model for estimating rating of perceived exertion (RPE) and session rating of perceived exertion (sRPE) in elite female football players using GPS-derived metrics and to identify key predictors of RPE and sRPE. We employ a linear mixed model (LMM) to analyse missing data points and model nonlinear individual characteristics (Krueger & Tian, 2004). This statistical approach accounts for the variability in individual player responses to external load exposures while examining the relationships between GPS metrics and RPE and sRPE (Iannaccone et al., 2021). LMMs incorporate fixed and random effects, making them suitable for analysing repeated measurements on the same statistical units, such as longitudinal data from individual players over time (Iannaccone et al., 2021; Krueger & Tian, 2004).

Research question

Given the study's focus, we propose the following research question: "Can a linear mixed model using GPS-derived metrics accurately predict the RPE and sRPE in elite female football players?"

Aim

The primary aims of the study are to:

1) Investigate whether external load variables, such as total distance (TD), high-speed running distance (HSRD), sprint distance (SpD), and peak speed ($Peak_{speed}$), can accurately predict RPE and sRPE for female football players using LMM.

2) Identify which external load variables contributes the most to RPE and sRPE.

3) Provide practitioners with a simple yet effective method for predicting RPE and sRPE based on these relationships.

The prediction model can serve as a valuable tool for detecting anomalies in the data, allowing practitioners to quickly identify unusual patterns that may indicate increased injury risk, fatigue, illness, or underperformance (Agrawal & Agrawal, 2015; Chandola et al., 2009). This information can help practitioners make more informed decisions regarding training management, ultimately contributing to a more efficient and effective training program for elite female football players.

1.2 Outline

Chapter 2, Theory: The second chapter reviews the existing literature to establish a solid foundation for understanding the significance of load monitoring and the potential of predicting RPE and sRPE in elite female football players. Furthermore, the chapter will summarise the reviewed literature, highlighting the current state of knowledge in the field and identifying potential limitations and gaps in existing research.

Chapter 3, Methods: The aim of Chapter 3 is to provide the reader with a comprehensive overview of the methodologies employed to address the research question posed by the study.

Chapter 4, Strengths and Limitations: The final chapter discusses the strengths and limitations of the study's methods, and the implementation of prediction models to practice, particularly from a practitioner's perspective.

2 Theory

2.1 Load Monitoring

Workload monitoring has become prevalent in professional football, allowing practitioners to assess and adjust athletes' training load to optimise performance and reduce injury risk (Akenhead & Nassis, 2016; Bourdon et al., 2017; S. L. Halson, 2014). By measuring key training load variables, practitioners can evaluate if players' training loads are progressing according to plan and adopt corrective actions when these measures deviate from a predefined threshold (Impellizzeri, Menaspà, et al., 2020).

Deciding on how to progress an athlete's training load is primarily a subjective decision, reliant on the practitioner's knowledge about the athlete, sports-specific movement patterns, training principles, and prior experience (Bourdon et al., 2017). However, technological advancements and new analytical approaches have significantly impacted load monitoring practice, providing valuable tools and strategies for practitioners (Bourdon et al., 2017; Shona L. Halson, 2014; Seshadri et al., 2021). Therefore, incorporating data-informed approaches with expert knowledge could help refine decision-making and ensure that athletes are exposed to appropriate training load, minimising the risk of injury and facilitating performance improvements (Luteberget et al., 2021; Malone et al., 2020; West et al., 2021).

2.2 Training load

Impellizzeri et al. (2005) define *training load* as an input variable manipulated to elicit a desired training response. Training load is commonly represented as internal and external components to identify the quality and quantity of workloads (Impellizzeri et al., 2019). External training load (ETL) represents the physical exposure of training, such as the distance covered and intensity of the workout, while internal training load (ITL) reflects the psycho-physiological response of the individual to that workload (Impellizzeri et al., 2019). The response experienced by players during the training process is the stimulus for the biological and psychological adaptations, also known as training outcome (Impellizzeri, Menaspà, et al., 2020).

Monitoring ITL and ETL can provide insight into whether athletes adapt to a training program, understand athletes' training responses and assess fatigue and the associated need for recovery (Bourdon et al., 2017; Impellizzeri et al., 2019). Monitoring only one of these components may give limited information, as both ITL and ETL are important for understanding the cumulative stress placed on athletes (Bourdon et al., 2017; Impellizzeri et al., 2019; Weaving et al., 2017). Only monitoring external load does not capture the player's psychological responses to a specific load, as players with different individual characteristics may experience the same external load differently (Hoff et al., 2002; Impellizzeri et al., 2019; Impellizzeri et al., 2004; Schwellnus et al., 2016; Viru & Viru, 2000). Conversely, only monitoring the internal load may not provide information about the nature and volume of the external load, which helps in understanding the type and magnitude of the physiological stress imposed on the athlete (Impellizzeri et al., 2019).

Assessing these components allows practitioners to understand whether the external load has induced the planned psycho-physiological response (internal load) and whether that load has induced the expected adaptations (training outcome). To help achieve the desired training load and training adaptation, the use of technology and methods such as global position systems (GPS) and RPE to monitor workloads is shared within the football environment (Akenhead & Nassis, 2016; Almulla et al., 2020; Hennessy & Jeffreys, 2018; Impellizzeri et al., 2019; Impellizzeri, Menaspà, et al., 2020; Impellizzeri et al., 2004).

2.2.1 Rating of perceived exertion

The rating of perceived exertion is considered a helpful tool in sports (Foster et al., 2021). Originally developed by Borg (1982), the RPE scale is a method for individuals to rate the intensity of their physical exertion subjectively. Recognising the need for a more intuitive scale, a modification to Borg's RPE scale was proposed, resulting in the development of the Category-Ratio 10 (CR10) RPE scale Borg (1982). This modified scale simplified the range from Borg's 6-20 to a more straightforward 0-10, with 0 representing 'no exertion at all' and 10 signifying 'extremely strong' (Borg, 1970; Borg, 1982).

Building on these foundations, Foster et al. (2001) adapted the RPE scale further to create the session RPE (sRPE) method. The sRPE is calculated by multiplying the player's rating with the duration of the exercise. This results in an arbitrary unit of the average RPE acquired across an entire training session incorporating both intensity and volume of the training (Foster et al., 2021; Foster et al., 2001). In some contexts, sRPE may be referred to as 'RPE load' or 'sRPE-load', but for the purposes of this thesis, the term 'sRPE' will be consistently used.

RPE is considered a valid indicator of ITL as it accounts for physiological and psychological factors, such as physical work rate, injury, illness, and daily fluctuations in players' psycho-physiological status (Borresen & Lambert, 2009; Foster et al., 2021; Impellizzeri et al., 2004). sRPE provides a good estimation of training intensity and has been found to correlate with objective internal and external load measures, including heart rate, blood lactate, and GPS-derived metrics (Askow et al., 2021; Coutts et al., 2009; Foster et al., 2021; Impellizzeri et al., 2004).

The simplicity of RPE and sRPE allows practitioners to understand better the training intensity and load experienced by athletes (Foster et al., 2021). However, while RPE and sRPE are practical and valid tools for assessing ITL in football, research investigating their relationship with GPS measures in elite female players remains limited (Costa et al., 2022; Torres-Ronda et al., 2022).

2.2.2 GPS as a monitoring tool

Following the 2015 FIFA rule change allowing GPS use in official competitions, practitioners have increasingly employed GPS technology to align ETL with desired training outcomes (Akenhead & Nassis, 2016; Buchheit & Simpson, 2017; Hennessy & Jeffreys, 2018; Pons et al., 2019). GPS enables accurate live and post monitoring, offering a wide range of variables for analysis, which supports detailed ETL planning and provides insights into the physical demands of training and match play (Akenhead & Nassis, 2016; Buchheit & Simpson, 2017; Hennessy & Jeffreys, 2018; Scott et al., 2016).

Additionally, GPS gives detailed information on position-specific demands, which has significant application for individualised training and return-to-play interventions (Akenhead & Nassis, 2016; Hennessy & Jeffreys, 2018). By considering factors such as match turnover, position and game minutes, practitioners can prescribe ETLs that challenge current fitness levels or taper load during pre-game training to optimise individual and team performance (Akenhead & Nassis, 2016; Buchheit et al., 2021; Buchheit et al., 2023b; Hennessy & Jeffreys, 2018).

2.3 GPS metrics in football

Research on GPS demonstrates inconsistency in data collection and interpretation methods (Costa et al., 2022; Harkness-Armstrong et al., 2022; Hennessy & Jeffreys, 2018; Torres-Ronda et al., 2022). Impellizzeri, McCall, et al. (2020) highlighted that the lack of conceptual frameworks allows researchers excessive freedom in selecting metrics and measurement methods. This leads to varying results across studies, as GPS manufacturers, speed and velocity thresholds, and modelling approaches all shape research findings, ultimately limiting generalisability.

Harkness-Armstrong et al. (2022) have called for a standardised approach for determining velocity thresholds in women's football to address this issue. Other researchers have suggested using lower thresholds for the female population to reflect better their physical performance capacities (Buchheit et al., 2010; J. J. Malone et al., 2017; Mujika et al., 2009). However, the literature lacks sufficient justification and definitions for the selected velocity thresholds (Costa et al., 2022; Hennessy & Jeffreys, 2018; Sweeting et al., 2017).

2.3.1 Absolute and individual thresholds

Velocity thresholds refer to predefined speed zones used to categorise and quantify players' movements during training or matches (Buchheit & Simpson, 2017). These thresholds help differentiate between intensity zones, enabling practitioners to analyse athletes' physical outputs based on the distance covered and speed (Buchheit & Simpson, 2017).

Absolute thresholds are fixed speed zones to quantify physical outputs for various velocity and acceleration metrics (Gualtieri et al., 2023). Absolute velocity thresholds are suitable for making between-player comparisons but fail to account for individual physical capacity differences (Abt & Lovell, 2009). J. J. Malone et al. (2017) argue that the sport-specific nature, varying demands, and contextual factors affecting external loading patterns may render the standardisation (absolute) of thresholds "*academic*" and of limited practical relevance.

As an alternative, researchers have proposed using individual thresholds, which allows for a more accurate and personalised assessment of players' workload (Abt & Lovell, 2009; Beato et al., 2021; Jastrzębski & Radzimiński, 2015). For example, Gualtieri et al. (2020) found a significant difference between starters and non-starters for sprint distance only when individual thresholds (i.e., 80% of the maximum peak velocity) were used.

However, the current evidence does not allow for definitive conclusions regarding using individual velocity thresholds (Gualtieri et al., 2023). While using individual thresholds seems to offer the advantage of a more precise quantification of individual external load, it may hinder comparisons between players, training sessions, and matches, or even over time when the same players have changed their velocity thresholds (Gualtieri et al., 2023).

2.3.2 Acceleration and deceleration metrics

A survey of practitioners from high-level football clubs revealed that acceleration variables are among the most used metrics when monitoring training load (Akenhead & Nassis, 2016). This is likely due to the widespread use of SSGs, which involve high acceleration and deceleration actions, resulting in high-intensity efforts that impose significant physiological and mechanical loading demands (Bloomfield et al., 2007; Dalen et al., 2021; Dello Iacono et al., 2023; Douchet et al., 2021; Osgnach et al., 2010; Verheul et al., 2021). However, there is limited evidence regarding the accuracy of the higher-intensity acceleration and deceleration frequencies (Ellens et al., 2022).

The minimal effort duration (MED) defines the minimum time a player must maintain acceleration or deceleration above a predefined threshold to identify as an effort (Harper et al., 2019). Low and high MED can result in over- and under-estimates, raising doubt about the accuracy of reported higher-intensity acceleration and deceleration frequencies (Harper et al., 2019). Small changes like 0.1 s in MED can result in substantial differences in the frequency of high-intensity efforts. A lower MED can detect shorter and higher rates of acceleration and deceleration whilst also being more susceptible to measurement error, potentially attributing multiple accelerations or decelerations given to a single effort (Buchheit & Simpson, 2017; Varley et al., 2017).

An alternative is to use an average acceleration-deceleration metric (Ave Acc/Dec), calculated by taking the absolute value of all raw acceleration and deceleration values and then averaging them for a selected period. This approach has been found to have better reliability and sensitivity across a range of GPS devices than threshold-based approaches (Delaney et al., 2018; Thornton et al., 2019). In addition, it represents a total multi-directional load, which could be helpful (Taylor et al., 2017). However, while indicating the absolute acceleration and deceleration demands, the approach does not differentiate between different magnitudes of acceleration or deceleration. Furthermore, it does not enable the identification of acceleration and deceleration density, and when acceleration and deceleration values are combined, it fails to differentiate the unique physiological and mechanical loading demands of these activities (Harper et al., 2019).

Given these limitations, several authors propose that future research should aim to quantify acceleration and decelerations into zones or carefully consider the criterion used to describe the starting and finishing velocities of acceleration and deceleration, i.e. when acceleration falls below a certain threshold (Buchheit & Simpson, 2017; Harper et al., 2019; Mara et al., 2017; Varley et al., 2017). Harper et al. (2019) propose establishing distinct high-intensity acceleration and deceleration thresholds that consider relative variations in maximal acceleration and deceleration abilities rather than relying on absolute thresholds.

2.3.3 Total distance

Total distance (TD) refers to a player's cumulative distance during a training session or match (Hennessy & Jeffreys, 2018). This metric indicates a player's overall physical activity and can provide insights into their work rate and physical capacity (Harkness-Armstrong et al., 2022). A strength of TD is its simplicity and ease of interpretation, making it a helpful starting point for assessing a player's overall workload (Akenhead & Nassis, 2016). However, TD does not differentiate between different intensity zones, limiting its ability to provide detailed insights into the specific physiological demands of a session or match (Hennessy & Jeffreys, 2018).

2.3.4 High-speed and sprint distance

High-speed running distance (HSRD) and sprinting activities are strongly associated with goal situations, match-winning outcomes and the most demanding phases of play (Carling et al., 2012; Chmura et al., 2018; Dello Iacono et al., 2023; Faude et al., 2012; Gualtieri et al., 2023). GPS has proven to be a valid tool for measuring HSRD and peak speed in sports (M. Beato et al., 2018) and demonstrates excellent inter-unit reliability for linear sprint distance (Beato & de Keijzer, 2019) and sports-specific movement (Marco Beato et al., 2018).

HSRD entry velocities are typically set between 12.2 $km \cdot h^{-1}$ and 15.6 $km \cdot h^{-1}$ for females, with 12.5 $km \cdot h^{-1}$ being the most common (Gualtieri et al., 2023). Sprint distance entry velocity is often set between 17.8 km·h^{-1} and 22.5 km·h^{-1} (22.5 km·h^{-1} being the most frequent) for females (Gualtieri et al., 2023). FIFA set women's HSRD and sprinting at 19 km·h^{-1} and 23 km·h^{-1}, respectively, showing variability in velocity thresholds for the same metrics used among researchers and practitioners (Bradley & Scott, 2020; Gualtieri et al., 2023; Sweeting et al., 2017).

Monitoring HSRD and SpD is commonly used by practitioners to inform training activities and prepare players for match demands (Akenhead & Nassis, 2016; Buchheit et al., 2021). Research has found relationships between hamstring strain injuries and near-maximal sprinting in team sports (Buchheit et al., 2023a; S. Malone et al., 2017). This emphasises the importance of both exposing players to high-speed and sprinting activities and closely monitoring these exposures (Buchheit et al., 2023a). By tracking HSRD and SpD, practitioners can validate the training process and optimise physical development, preparing players for demanding game phases requiring near-maximum speed (Gualtieri et al., 2023).

2.3.5 Peak speed

Peak speed represents the maximum speed achieved by a player during a training session or match (Harkness-Armstrong et al., 2022). This metric is especially relevant for assessing a player's sprinting ability and capacity for high-intensity actions, which are crucial components of physical and match performance (Gualtieri et al., 2023; Harkness-Armstrong et al., 2022; Hennessy & Jeffreys, 2018). Practitioners can use peak speed to monitor sprinting progress, inform strategies designed to enhance sprint performance and guide return-to-play interventions (e.g., percentages of peak speed).

However, this metric can be influenced by factors such as fatigue, pitch dimensions, and match context (Harkness-Armstrong et al., 2022; Riboli et al., 2020; Trewin et al., 2018; Varley et al., 2012; Varley et al., 2017). Additionally, peak speed does not provide information about the frequency or duration of high-intensity actions. However, coupled with Acc, Dec, HSRD, and sprint distance gives a comprehensive picture of high-intensity actions completed (Hennessy & Jeffreys, 2018).

2.3.6 The relationship between GPS metrics and RPE and RPE

In the existing literature, the relationship between RPE and sRPE and GPS-derived metrics in female footballers remains little investigated (Costa et al., 2022; Torres-Ronda et al., 2022). However, available evidence highlights a strong correlation between RPE/sRPE and training volume and a moderate correlation with session intensity (Askow et al., 2021; Douchet et al., 2021; McLaren et al., 2018).

Askow et al. (2021) investigated the association between sRPE and GPS-derived external load measures in 21 NCAA Division I women's football players. Their results demonstrated a strong association between sRPE and TD ($p < 0.001$), the distance between 18-25 $km \cdot h^{-1}$ ($p < 0.001$), and accelerations (-2-1 $m \cdot s^{-2}$ and 1-2 $m \cdot s^{-2}$) ($p < 0.001$).

Douchet et al. (2021) examined the association between ETL, RPE, and sRPE in 12 elite female football players. Their study found a significant association between changes in ETL and corresponding RPE and sRPE reporting ($p < 0.001$).

Conversely, Scott and Lovell (2018) found that relative thresholds failed to enhance the dose-response relationship in a study of international female players during a 21-day training camp. Using peak sprinting speed to quantify external load yielded weaker associations with RPE, indicating a diminished capacity to determine training dose response. However, they did find a large correlation between HSRD and RPE (r = 0.53-0.67).

Table 1: *Overview of correlation studies on female football players. Abbreviations: ACC = acceleration, DEC = deceleration, Z4 = velocity zone 4, HSR = High-speed running, VHSR = very high-speed running.*

Study	*N*, age, competition level	Condition	Training/match duration	Internal measures and instruments	External measures and instruments	Coefficient
(Douchet et al., 2021)	12 (24.2 ± 2.3) Professional Females	Training	2 in-season weeks (6 training days)	RPE (CR-10) sRPE (CR-10)	Polar Team Pro sensor (Electro, Kempele, Finland): TD m·min−1 ACC >2 $m \cdot s^{-2}$ (n) DEC <2 $m \cdot s^{-2}$ (n)	($p<0.001$)
(Askow et al., 2021)	21 (20.3 ± 1.5) Amateur Females	Training and match	16 in-season weeks	sRPE CR-10)	OptimEye X4 (Catapult Innovations, Melbourne, Australia): TD Z4 18-25 $km \cdot h^{-1}$ (m) >25 $km \cdot h^{-1}$ (m) ACC Z3 (-2-1 m·s−2) ACC Z6 (1-2 m·s−2)	($p<0.001$)
(Scott & Lovell, 2018)	22 International female players	Training	21-day training camp	RPE (CR-10)	OptimEye S5 (Catapult Sports, Melbourne, Australia): HSR 12.67 $km \cdot h^{-1}$ VHSR 17.82 $km \cdot h^{-1}$	($r = 0.53$-0.67)

2.4 Prediction model

In recent years, load monitoring research has increasingly adopted advanced methods such as multileveled models and machine learning (ML) techniques to identify key predictors of training load (Bartlett et al., 2017; Carey et al., 2016; Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022; Newans et al., 2022). These methods offer the advantage of individual player differences when analysing internal and external load relationships in team sports (Seshadri et al., 2021). Notably, these methods differ from correlation analysis, as they aim to predict outcomes based on numerous variables rather than merely identifying relationships (Asuero et al., 2006; Newans et al., 2022; Pepe et al., 2004). Thus, predictive models have the potential to act as automated data analysts, providing valuable insights into the athletes' condition (Seshadri et al., 2021).

Shmueli (2011) defines predictive modelling *as "the process of applying a statistical model or data mining algorithm to data for the purpose of predicting new or future observations."* Predictive modelling is a form of data mining that analyses historical data to identify trends or patterns to predict future outcomes (Hernán et al., 2019). Constructing a prediction model starts with identifying historical data representative of the desired predictive outcome; hence the model cannot predict what has never been observed (Shmueli, 2011). Therefore, the sample size used to train the model is important in ensuring accurate predictions. Additionally, proper data management and organisation are necessary to avoid overfitting, which occurs when a model memorises critical data points rather than generalising them (Shmueli, 2011). Overfitting is implied if the model performs significantly better on the training set than on the holdout set (Shmueli, 2011).

Shmueli (2011) explains that predictive modelling does not need to explore each variable's exact role in an underlying causal structure because of the focus on association rather than causation. Instead, prediction criteria are the quality of association between the predictors and the outcome in question, data quality, and availability of the predictors at the time of prediction (Hernán et al., 2019; Shmueli, 2011).

Bias in a prediction model refers to the model's tendency to overestimate or underestimate outcomes. This bias can be introduced through the historical data used to train the model, as past outcomes often reflect existing biases (Shmueli, 2011). For instance, if the model is trained with a smaller sample size, it may not accurately represent the full scope of possible outcomes, leading to higher bias (Shmueli, 2011).

In order to mitigate this, a data partitioning procedure can be employed. This involves dividing the data into separate subsets for training and testing the model. This procedure helps balance bias and variance — two crucial elements influencing a model's accuracy. While it might slightly increase the bias (since less data is used to train the model), it significantly reduces sampling variance, which is the model's sensitivity to fluctuations in the training data (Shmueli, 2011).

2.4.1 Root mean squared error and Mean absolute error

Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are commonly used methods to evaluate the performance of prediction models (Chai & Draxler, 2014). RMSE is the square root of the mean squared differences between the predicted and the actual values, while MAE is the mean of the absolute differences between the predicted and the actual values (Hodson, 2022). RMSE and MAE can be used in prediction to evaluate the model's accuracy by comparing the predicted values to the observed values (Hodson, 2022; Iannaccone et al., 2021). The smaller the values of RMSE and MAE, the better the model's performance, indicating a closer fit between the predicted and actual values (Hodson, 2022).

The strength of RMSE lies in its sensitivity to large errors, which can be useful in situations where detecting and penalising large deviations is essential (Chai & Draxler, 2014; Hodson, 2022). However, this sensitivity can also be a limitation, as it may exaggerate the impact of outliers, leading to overestimating the model's error (Chai & Draxler, 2014; Hodson, 2022). In contrast, the strength of MAE is its robustness to outliers, providing a more consistent measure of model accuracy across different datasets (Chai & Draxler, 2014; Hodson, 2022). The primary limitation of MAE is its insensitivity to large errors, which may lead to an underestimation of the model's error in cases where large deviations are of critical importance (Chai & Draxler, 2014; Hodson, 2022).

RMSE and MAE can provide valuable insights into model performance when evaluating the model's accuracy. Given the nature of RPE, which contains a degree of subjectivity and potential for large deviations, the combination of RMSE and MAE can help identify the model's overall accuracy and its sensitivity to extreme errors (Hodson, 2022; Iannaccone et al., 2021).

2.4.2 Coefficient of determination

The coefficient of determination, commonly known as R^2, is a statistical measure that quantifies the proportion of the variance in the dependent variable that the independent variables can explain in a regression model (Nakagawa & Schielzeth, 2013). In the context of our study on predicting RPE and sRPE, R^2 helps us assess how well the external load factors can explain the variation in the RPE and sRPE models. We only report on the conditional R^2, which is concerned with variance explained by fixed and random factors (Nakagawa & Schielzeth, 2013). A thorough description of the R^2 calculation can be found in Nakagawa and Schielzeth (2013), which will, for practical purposes, be reiterated here.

R^2 ranges from 0 to 1, where higher values indicate that the independent variables account for a greater proportion of the variability in the dependent variable (Nakagawa & Schielzeth, 2013). An R^2 value of 1 signifies a perfect fit of the model, explaining 100% of the variability, while an R^2 value of 0 suggests that the model does not explain any of the variability (Nakagawa & Schielzeth, 2013).

However, a high R^2 value can sometimes be misleading, as it may not necessarily indicate that the model accurately represents the underlying data. Additionally, R^2 is sensitive to the inclusion of additional variables, which may inflate its value without truly contributing to the model's accuracy (Nakagawa & Schielzeth, 2013).

2.4.3 Intraclass correlation coefficient

The intraclass correlation coefficient (ICC) is a statistical measure that assesses the degree of similarity or reliability between observations within the same group or class relative to the total variation across all observations (Liljequist et al., 2019). In our study, ICC helps us evaluate the consistency of individual players' RPE and sRPE responses to external load exposure, accounting for their differences.

ICC values range from 0 to 1, where higher values indicate greater consistency or reliability within a group, and lower values suggest less consistency and more variability between groups (Koo & Li, 2016). An ICC value of 1 signifies perfect consistency within a group, with no variability between groups, while an ICC value of 0 suggests that there is no consistency within a group and that all variability is due to differences between groups (Koo & Li, 2016).

However, interpreting ICC values should be done with caution, as they can be influenced by the variability of the data and the number of observations per group. A high ICC value may not necessarily indicate that the model accurately represents the underlying data, and it is also essential to consider other model fit indices (Koo & Li, 2016; Liljequist et al., 2019). Additionally, ICC is sensitive to the sample size and the number of groups, which may impact its value and the interpretation of the results (Koo & Li, 2016; Liljequist et al., 2019).

2.4.3.1 The difference between ICC and R^2

ICC and R^2 serve different purposes and provide different information about the relationship between the variables and the consistency of the data.

For instance, a high ICC value for the sRPE model indicates that a substantial proportion of the total variability in sRPE responses can be attributed to player differences. On the other hand, R^2 helps assess how well the external load variables can explain the variation in RPE and sRPE models. A high R^2 value indicates that the external load variables account for a significant proportion of the variability in RPE and sRPE, suggesting that the model accurately represents the underlying data.

In summary, while ICC focuses on the consistency of RPE and sRPE responses within individual players, R^2 evaluates the ability of external load variables to explain the variability in RPE and sRPE. Both measures are important in understanding the relationship between external load variables and perceived exertion, but they provide different insights into the data's structure and the model's performance.

2.4.4 Prediction in a practical context

By providing accurate predictions of athletes' RPE and sRPE in response to ETL, prediction models can facilitate individualised and optimised training programs that account for varying intensity, volume, and frequency (Geurkink et al., 2019; Marynowicz et al., 2022).

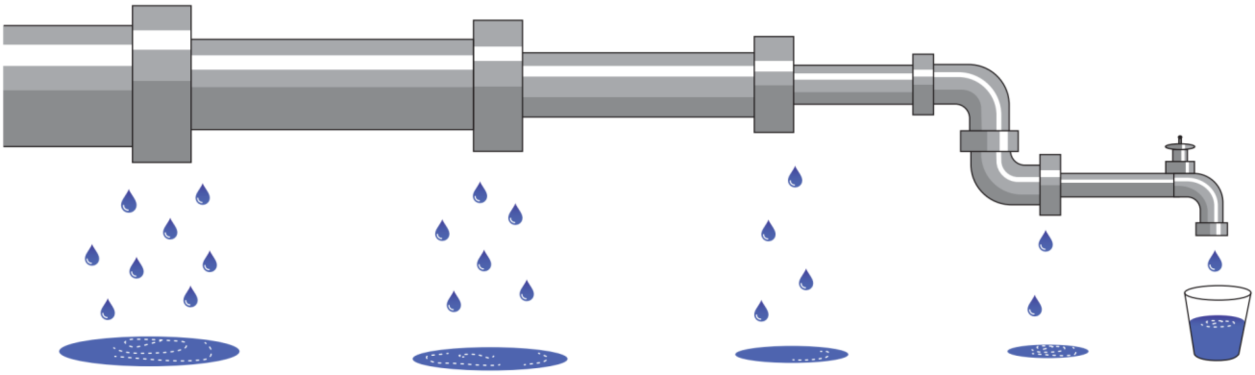
Predicting sRPE can be particularly useful when planning training sessions and determining the appropriate training load. For example, a coach can plan a training session and adjust the training drills (type or volume) and associated distance and speed covered to achieve the desired predicted sRPE. This approach enables coaches to carefully manage workload distribution across a training week, ensuring that players receive sufficient stimulus for adaptation without excessive fatigue (Geurkink et al., 2019; Marynowicz et al., 2022). This method, however, requires prior knowledge of the distances and speeds players achieve in specific training drills. Such information can be cumbersome to retrieve but is available in most GPS software.

However, while there could be benefits to using prediction models, such as explained above, implementing them in clinical practice can be problematic (van Royen et al., 2022). As exemplified with their pipeline figure (see Figure 1), van Royen et al. (2022) illustrate why prediction model implementation is challenging. Primarily, the authors highlight four reasons why prediction models fail to be adopted:

1. Not fit for purpose
2. No validation
3. No implementation
4. Not adopted

Essentially, a successful prediction model must be practical, cost-effective, and specific to the target population. It requires accurate predictors, reliable outcomes, and thorough validation—especially on other individuals than from which it was developed. Importantly, it should be transparent, trustworthy, actionable, and significantly influence decision-making processes (van Royen et al., 2022).

All the above-mentioned factors impact the implementation and usefulness of prediction models. The complexity of developing and evaluating a prediction model may not only limit practitioners' ability to adapt it to their specific needs but also affect their willingness to adopt it due to the perceived complexity and the required time and resources. An extended discussion on implementing prediction models from a practitioner’s perspective will be discussed in the strength and limitation chapter under Perspectives.



Not fit for purpose	No validation	No implementation	Not adopted
Developed on wrong patient population	Lack of data or incentive to pursue validation studies	No impact on decision making or patient (health) outcomes	Prediction (perceived as) not useful
Expensive or non-available predictors	Incompletely reported prediction model	No software developed to implement and use the model	Predictions not trusted
Time intensive to use model	Poorly developed or overfitted model	Requirements for adherence to (medical device) regulations	Model not transparent enough, or no tools available to enhance its use in practice
Outcome measured unreliably	Proprietary model code	Cost(-effectiveness) of using proprietary model	Model (perceived as) outdated

Figure 1: *“Leaky prognostic model adoption pipeline. Examples of reasons for failed prediction model adoption in clinical practice.” From van Royen et al. (2022).*

2.4.4.1 Anomaly detection

Anomaly detection refers to the process of identifying data points that do not conform to the expected patterns or behaviour, indicating potential issues or outliers in the data (Agrawal & Agrawal, 2015; Chandola et al., 2009). Prediction models can be used in anomaly detection by identifying unusual RPE and sRPE patterns or deviations from expected values in athletes' responses to training load.

Prediction models are particularly suitable for analysing RPE and sRPE data, as they can account for both fixed and random effects, allowing for the modelling of individual variability in response to training load while also considering group-level trends (Iannaccone et al., 2021). This flexibility allows for developing more accurate and individualised estimates of the expected RPE and sRPE, catering to each athlete's unique characteristics and needs (Geurkink et al., 2019).

The prediction capabilities of prediction models can be leveraged to detect (time-series) anomalies in RPE and sRPE by comparing the predicted values to the observed values. Significant deviations between the predicted and observed RPE and sRPE values may indicate unusual responses to training load or potential anomalies in the data. Such deviations could indicate an underlying issue, such as an emerging injury, illness, or psychological factor affecting the athlete's performance (Iannaccone et al., 2021). As discussed previously, these deviations can be quantified using metrics such as RMSE and MAE to determine the magnitude of the discrepancies between predicted and observed values. Early identification of these anomalies enables practitioners to intervene promptly, addressing potential problems before they escalate and negatively impact the athlete's performance or well-being. Notably, anomaly detection is performed post-training to identify unusual patterns.

2.4.5 Previous research

Jaspers et al. (2018) utilised machine learning models to investigate the relationships between external load indicators and RPE in 38 professional male football players. They collected GPS and RPE data, applying various machine learning algorithms, such as Artificial Neural Networks (ANN) and Least Absolute Shrinkage and selection operator (LASSO), to predict RPE from a set of 67 external load indicators. The average RPE for all 5917 analysed training sessions was 3.59 (1.46) AU. The ANN model performed an MAE of 1.09 (1.07-1.11), and the LASSO had an MAE of 0.80 (0.78-0.82).

Geurkink et al. (2019) used Generalized Additive models, multivariate Adaptive regression Splines, Decision Tree, Random Forest, Linear Regression and Support vector Regression to predict RPE. *Gradient Boosting Machines* were used to identify the RPE's leading predictive indicators and predict RPE. In total, 70 external load indicators, internal load indicators, individual characteristics and supplementary variables were used to build prediction models. The average training RPE was 4.34 ± 1.06 AU, with a median of 18 records per player. The predictive model had an MAE of 0.67 (0.09) AU and an RMSE of 0.93 (0.16) AU. In addition, TD was the strongest predictor.

Marynowicz et al. (2022) used a 'decision tree model' to predict RPE from external measures, TD (m/min), HSRD (m/min), PlayerLoad (n/min), impacts (n/min), Accdist (m/min), Decdist (m/min), Acc counts (n/min), and Dec counts (n/min) at individual and group levels in elite youth male football players. The results showed that HSRD per minute was the strongest predictor of RPE. The individualised models' prediction error (RMSE 0.755 ± 0.014) was lower than the population model (RMSE 1.621 ± 0.001).

All studies exhibit strengths in their methodological approaches. Jaspers et al. (2018) leveraged multiple ML techniques to examine the associations between internal load and external load indicators, providing valuable insights into the predictive power of these relationships. Interestingly, the analysis found that group models (ANN, LASSO) demonstrated equivalent or superior accuracy compared with individual models when predicting RPE, which is in contrast to Geurkink et al. (2019) and Marynowicz et al. (2022).

Geurkink et al. (2019) took a comprehensive approach to identify the most relevant factors in predicting RPE by accounting for external load indicators, internal load indicators, individual characteristics, and supplementary variables, which enhanced the accuracy of their models. In addition, individual deviations in several external load indicators compared to the group mean were derived based on historical data. Accounting for individual differences in ETL compared to group means. Importantly, missing data handling was presented, explaining that missing values were replaced with the group mean.

However, these studies also present limitations. The machine learning algorithms employed in their study may be too complex for practical use by practitioners, who typically require simpler, more interpretable methods (Malone et al., 2020; West et al., 2021). It is essential to consider the practicality and interpretability of these models, ensuring that practitioners can effectively use these tools alongside their expert knowledge to make informed decisions (Luteberget et al., 2021; Malone et al., 2020; West et al., 2021).

In addition, the small sample sizes in the studies can affect the accuracy of prediction models (Shmueli, 2011). Limited data may not capture the full range of population variability, potentially leading to biased estimates and reduced generalisability (Shmueli, 2011). For instance, Jaspers et al. (2018) observed differences in model performance when constructing group models with over 2,000 data points compared to individual models with fewer than 100 data points.

Furthermore, Geurkink et al. (2019) could not predict RPE values of 8-10 because of the limited number of high RPE values observed. Therefore, they adapted the model to predict values between 1 and 7. This may affect the model's accuracy for predicting RPE values within this range.

Table 2: *Overview of prediction studies. Abbreviations: Ann = Artificial Neural Networks, LASSO = Least Absolute Shrinkage and selection operator, RMSE = root mean squared error, MAE = mean absolute error.*

Study	*N*, age, competition level	Condition	Training/match duration	Internal measures and instruments	External measures and instruments	Coefficient
(Jaspers et al., 2018)	38 (22.7) professional male players	Training	2014-2015 and 2015-2016 Pre-season and in-season training session. In total, 5917 training sessions analysed.	RPE (CR-10, AU)	OptimEye S5 (Catapult Sports, Melbourne, Australia) 67 GPS metrics categorised into: Duration (1) Distance (17) Speed (8) Acc and Dec (18) PlayerLoad (10) Repeated high-intensity efforts (13)	ANN, MAE (90% CI) = 1.09 (1.07-1.11) LASSO, MAE (90% CI) = 0.80 (0.78-0.82)
(Marynowicz et al., 2022)	18 (17.81 ± 0.96) Amateur Male	Training	1 full in-season training session only containing one game (5-6 sessions per week) Training duration: 68 ± 15 minutes	RPE (CR-10, AU)	GPS PLAYTEK (Catapult Innovations, Melbourne, Australia) Total distance (m) HSR >19.8 $km \cdot h^{-1}$ (m) PlayerLoad (AU) Impacts >3 g (n) Acc > 2 $m \cdot s^{-2}$ (m and n) Dec > 2 $m \cdot s^{-2}$ (m and n)	RMSE: Individual model 0.755 ± 0.014 Group model 1.621 ± 0.001
(Geurkink et al., 2019)	46 (25.6 ± 4.2) professional male players	Training	61 training sessions, a total of 913 individual observations.	RPE (CR-10, AU)	Polar Team Pro (Kempele, Finland) Total distance (m) Distance 3-6.99, 7-10.99, 11-14.99, 15-18.99, >19.00 km/h (m). Acc 0.50-0.99, 1-1.99, 2-2.99, 3-50.00 m/s^2 (n). Dec 0.50-0.99, 1-1.99, 2-2.99, 3-50.00 m/s^2 (n). Sprints >25 km/h (n)	MAE 0.67 (0.09) RMSE 0.93 (0.16)

2.5 Summary

Load monitoring can be a useful tool in optimising training outcomes by ensuring the appropriate prescription of internal and external training load. Using GPS and RPE/sRPE could help practitioners when periodising short- and long-term physical goals, especially when planning periods of recovery, overload and tapering to improve physiological capacity and maximise readiness for competition. However, choosing which GPS metrics and thresholds to use is challenging, considering the inconsistency in the literature, collection methods and the lack of conceptual frameworks.

Although the relationship between RPE and sRPE and GPS metrics in elite female footballers remains underexplored, existing evidence indicates a strong correlation with training volume and a moderate correlation with session intensity (Askow et al., 2021; Douchet et al., 2021; McLaren et al., 2018).

The existing evidence in predicting RPE and sRPE is scarce, and to the author's knowledge, there is currently no attempt to predict RPE and sRPE in female football players. Internal load prediction could help practitioners to enhance individualised training prescriptions and better understand players' adaptation, fatigue, and recovery needs. Ultimately contributing to optimising performance and reducing injury risk. In this context, prediction models are practical as they can detect anomalies of RPE and sRPE by identifying deviations from the expected values, which could help practitioners detect potential issues impacting performance and well-being.

The use of RMSE and MAE helps to quantify the accuracy of the predictive models, while R^2 is used to determine the proportion of variance explained in the models. ICC is used to assess the reliability of the measurements, which is important for ensuring the validity of the results.

However, maximising the benefits of load monitoring and prediction requires balancing technology and reliance on practitioner experience, knowledge and intuition. Such an approach benefits from the "best of both worlds", ultimately helping to optimise player performance, well-being and injury risk reduction (Nassis et al., 2023).

3 Methods

3.1 Female football research centre (FFRC)

The present study is conducted in collaboration with the Female Football Research Centre (FFRC), a research facility dedicated to advancing knowledge in women's football. The primary goal of the FFRC is to gain new fundamental insights into the factors that affect the performance and overall health of elite female football players (UiT, 2022). The overarching objective is to develop novel methodologies for epidemiological research that can inform both sports and medical fields, as well as to create non-invasive, privacy-preserving, and practical technology that quantifies and monitors athlete behaviour from various perspectives, including biomechanics, sports science, medicine, coaches, and athletes (UiT, 2022).

3.2 Ethical considerations

All participants have been informed and have given written consent, and thus can withdraw from the study at any time. Personal identifiers were removed from the data files when used for statistical analyses. All FFRC protocols have been submitted for approval to the Norwegian Social Science Data Services and Regional Committees for Medical and Health Research Ethics (REK), and NSD- Norwegian Centre for research data (number 296155).

3.3 Data collection

58 female football players (22 ± 4 years of age) from two top-level Norwegian clubs were included in the study. A thorough description of the data collection can be found in Winther et al. (2022) and Baptista et al. (2022) and will, for practical purposes, be reiterated here.

Locomotor data from the two clubs' training sessions in the 2020 and 2021 seasons (Figure 4) were collected using GPS APEX (STATSports), with a sampling frequency of 10 Hz. The validity and levels of accuracy (bias <5%) of this tracking system has been previously presented (Marco Beato et al., 2018). During training and matches, each player wore a tight vest with the GPS unit on the back of their upper body between the scapula as described by the manufacturer. To minimise inter-device error, each player used the same GPS unit during the entire collection period (Marco Beato et al., 2018).Doppler-derived speed data was exported from manufacturer software (STATSport Sonra 2.1.4) into Python 3.7.6 for processing (linearly interpolating any missing raw data) and to derive metrics. Raw acceleration was then calculated for 0.6 s. After deriving all the metrics, the data were transferred to R (R.4.0.5, R Core Team, 2021) for statistical analysis.

Approximately 30 minutes post-training session, players reported RPE values, following Foster et al. (2001) recommendations to minimise the effect of immediate fatigue after training and reduce the influence of peer pressure (Malone et al., 2015). The scores were recorded individually using the RPE CR-10 scale (Foster et al., 2001) via the PM Reporter Pro smartphone application. Each player was fully familiarised with the scale before reporting. After submitting their responses, data were automatically uploaded to cloud-based software.

3.3.1 STATSports

Originating from Ireland, STATSports is a customer-driven platform providing sports-specific software for football, basketball, American football, rugby, and athletics (Statsports, 2022). The platform offers products suited for elite, professional, and grassroots teams and allows for valid and reliable live data collection and tracking using GPS technology (Marco Beato et al., 2018; Beato & de Keijzer, 2019; Crang et al., 2022). The built-in software includes advanced algorithms that can identify poor-quality data and automatically filter, smooth, or extract data, making it a practical tool for fast evaluation of performance, workloads, and exercise prescription. Additionally, the platform provides players and staff with visualisation tools through its web platform and application services and allows for raw data export for custom analysis (Statsports, 2022).

3.4 Physical performance variables

The physical parameters analysed included total distance (TD), high-speed running distance (HSRD) (19-23 km/h), sprint distance (SpD) (>23 km/h), and peak speed ($Peak_{speed}$). The speed thresholds were chosen according to previous research (Park et al., 2019) and are the same thresholds used by FIFA (Bradley & Scott, 2020). Acceleration and deceleration metrics were not included in our analysis. Although acceleration and decelerations are probably predictive of both SRPE and RPE, this decision was made because of the wide range of methods available for calculating these metrics (Buchheit et al., 2014; Ellens et al., 2022). Thus, to keep the model as parsimonious and user-friendly as possible, these metrics were left out.

Table 3: *Variable description.*

	Variable	Type	Units
Dependent	Total distance (TD)	Continuous	Complete load, Meters (m)
	High-speed running distance (HSRD)	Continuous	19-23 km/h
	Sprint distance (SpD)	Continuous	>23 km/h
	Peak speed ($Peak_{speed}$)	Continuous	Highest speed achieved during training
Fixed	Position	Nominal	Central defender (CD), Wide defender (WD), Central midfielder (CM), Wide midfielder (WM), Striker (S)
	In match squad	Nominal	Playing status (starter, sub, unused sub)
	Match	Nominal	1 = Match, 0 =Training
	Minutes played Last match	Continuous	Number of minutes played last match
	Session duration	Continuous	The length of the training session (min)
Random	Player ID	Nominal	Unique ID

3.4.1 PMsys

PMSys, developed by Forzasys through a collaboration between Simula Research Laboratory and the University of Tromsø, is an athlete monitoring system designed to collect and analyse training load, wellness, illness, and injury data (Forzasys, 2022; Thambawita et al., 2020). The system uses a smartphone application to facilitate data collection through a subjective questionnaire, with push notifications reminding players to complete the questionnaire. This application allows athletes to track their progress, while coaches can monitor team and individual player workload and performance over time. In addition, PMSys features a web-based interface for coaches to visualise team and individual player statistics, providing basic pattern analysis based on the collected data.

3.5 Rating of perceived exertion

The rating of perceived exertion is considered a helpful tool in sports (Foster et al., 2021). Originally developed by Borg (1982), the RPE scale is a method for individuals to rate the intensity of their physical exertion subjectively. Recognising the need for a more intuitive scale, a modification to Borg's RPE scale was proposed, resulting in the development of the Category-Ratio 10 (CR10) RPE scale Borg (1982). This modified scale simplified the range from Borg's 6-20 to a more straightforward 0-10, with 0 representing 'no exertion at all' and 10 signifying 'extremely strong' (Borg, 1970; Borg, 1982).

3.5.1 Validity of RPE

RPE has been demonstrated to be a valid indicator of ITL, as it correlates with objective internal measures such as VO_2, heart rate and blood lactate (Borg & Kaijser, 2006; Coutts et al., 2009; Eston, 2012; Impellizzeri et al., 2004). Moreover, sRPE is a simple and effective method for quantifying training load in high-intensity, intermittent team sports such as football (Borresen & Lambert, 2008; Foster et al., 2021; Gaudino et al., 2015; Haddad et al., 2013; Impellizzeri et al., 2004; Little & Williams, 2007). Further association and correlation between sRPE and GPS-derived metrics have been presented previously (chapter 2.3.6).

3.5.2 Reliability of RPE

The reliability of RPE and sRPE can be influenced by various factors, such as individual characteristics, training content, and players’ recovery before training sessions (Beato et al., 2023; Foster et al., 2021; Impellizzeri et al., 2019; Wiig et al., 2020). Recently, Wiig et al. (2020) demonstrated that several external load variables had substantial within-player effects on sRPE, substantial between-player effects on sRPE, and a large between-session variability in sRPE for external load variables. The substantial between-session variability in sRPE indicates that single external load variables do not fully explain sRPE in training sessions.

3.6 Data preprocessing

3.6.1 Data cleaning

For commercial GPS systems used in individual- and team sports to accurately determine the position of the GPS receiver and send precise time information on the duration of the signal transit, a minimum of four satellites must be connected to the device (Larsson, 2003; J. J. Malone et al., 2017). However, J. J. Malone et al. (2017) have argued that the connection and data quality may be compromised when fewer than six satellites are connected to the GPS device.

Furthermore, the signal quality during data collection affects the accuracy of the recorded data and may be influenced by location (indoor/outdoor), obstructions (stadiums, high buildings) and the orientation of satellites in the atmosphere (Buchheit & Simpson, 2017; Williams & Morgan, 2009; Witte & Wilson, 2004). The horizontal dilution of precision (HDOP) is also a measure of the accuracy of the GPS horizontal positional signal and is determined by the geometrical organisation of the satellites (J. J. Malone et al., 2017). High HDOP values indicate poor precision, while low values indicate good precision. It is considered ideal when HDOP values are less than one (J. J. Malone et al., 2017).

Data collected with fewer than 12 satellites, HDOP higher than 5, session duration over 210 minutes (3 hours) and less than 1 minutes were treated as missing. These thresholds were chosen based on visually inspecting the histograms of each variable for inaccurate and unlikely data. In addition, RPE values equal to 0 were treated as missing since PmSys currently does not have a dedicated missing value. Positions were grouped into CD, WD, CM, WM, and S. Goalkeepers were excluded from the analysis.

***Table 4**: Distribution of data after the cleaning process*

Position	**Match/Training**	**Total number of observations**	**Number of players**	**Mean number of observations per player**	**Minimum**	**Maximum**
CD	Training	547	15	36.5	1	139
CD	Match	161	13	12.4	1	39
CM	Training	2231	49	45.5	2	141
CM	Match	526	46	11.4	1	38
ST	Training	517	25	20.7	1	66
ST	Match	124	22	5.6	1	18
WD	Training	841	28	30	1	144
WD	Match	175	23	7.6	1	33
WM	Training	175	15	11.7	2	28
WM	Match	34	13	2.6	1	9

3.6.2 Missing data

The initial sample contained 12900 observations, with 58% (7510 observations) of the data missing. The initial sample assumes that there is activity for all players in the squad on all days from the start till the end of the season.

Figure 2: *Proportion of missing data from the full dataset with percentage of variables missingness. Abbreviations: meanNumSatellites = the mean number of satellites, meanHdop = the mean horizontal dilution of precision, z5 = SpD (>23 km/h), z4 = HSRD (19-23 km/h).*

The missing data are assumed to be both missing at random (MAR) and missing completely at random (MCAR), potentially due to factors such as a lack of GPS devices, forgetting to turn on pods, injuries, illnesses, or days off. A Complete Case Analysis (CCA) was performed to remove the missing data. Results are presented in Table 4.

3.6.2.1 Complete case analysis

Missing data is a common issue in longitudinal studies that must be addressed to produce accurate and reliable results (Azur et al., 2011; Van Buuren, 2018). One traditional approach to handling missing data is to remove incomplete cases, also known as *complete case analysis*.

CCA is a method for handling missing data by only including observations with complete information for all variables included in the analysis (Dong & Peng, 2013). In our study, CCA was employed by removing instances with missing data for predictor variables, outcome variables, or other relevant covariates and then performing the LMM analysis on the remaining complete cases.

The simplicity of this approach is its primary strength, as it does not require complex imputation methods or statistical assumptions to address missing data (Graham, 2009). CCA is easily implemented in most statistical software packages and can provide unbiased estimates of model parameters if the data are MCAR (Dong & Peng, 2013; Schafer & Graham, 2002).

However, CCA has significant limitations. First, it can lead to a substantial loss of statistical power and reduced generalisability if a large proportion of cases are removed due to missing data (Graham, 2009). This can result in less precise estimates and decreased model prediction reliability. Second, if data are not MCAR but instead are MAR, CCA may introduce bias in the parameter estimates, as the observed data no longer accurately represent the underlying population (Graham, 2009; Schafer & Graham, 2002). In the context of internal load prediction, this could lead to inaccurate predictions and potentially misinformed decisions about training load management (Impellizzeri, Menaspà, et al., 2020).

Our study assumes that the missing data is MAR and MCAR. Under the MCAR assumption, the probability of missing data is unrelated to the observed or unobserved data, and the missingness is considered non-informative (Rubin, 1976). When data is MAR, the probability of missing observation is related to observed variables (Bhaskaran & Smeeth, 2014). The CCA can handle MCAR data without introducing bias in the parameter estimates (Dong & Peng, 2013). However, a distinct pattern of missingness is observed, particularly on the day following matches (MD+1). This occurrence is potentially attributed to players being granted a day off following x amount of playing minutes the day before (i.e., match minutes explain missing data on the training day preceding the match). This leads to bias in the estimates (Bhaskaran & Smeeth, 2014).

3.7 Linear mixed model

In sports science, relationships between variables are commonly examined through correlation analysis and analysis of variance (ANOVA) with repeated measures when analysing longitudinal datasets (Iannaccone et al., 2021). However, datasets are often complicated and characterised by multiple dependent observations across training and matches and imbalanced data due to injuries, illness and team selection (Newans et al., 2022). In the specific context of football, it is important to consider not only the team-level variability but also the inter-individual variability when analysing workload data, as exercise responses may vary not only between players but also within the same players, creating multilevel hierarchical data (Gelman & Hill, 2006; Iannaccone et al., 2021; Impellizzeri et al., 2004; Nakagawa & Schielzeth, 2013; Newans et al., 2022).

Using a LMM allows for linear regression with both fixed and random effects, thereby accounting for intrasubject variability and making it possible to handle missing data without having to remove subjects from the analysis (Atkinson et al., 2011; Iannaccone et al., 2021; Lininger et al., 2015; Little et al., 2014).

The random effects are variables that are not measured directly. Instead, estimated on their impact on models based on how much they cause the data to vary. This is summarised using estimated variance (how far the data spread out from their average) and covariances (how much two variables change together) (Iannaccone et al., 2021). Including random effects in a LMM is beneficial because it helps prevent "inflation" and "type 1 error" (Iannaccone et al., 2021). Inflation, in this context, refers to overestimating the precision of our estimates. Type 1 error is when we incorrectly conclude that there is a significant effect or relationship when, in fact, there is not. By summarising random effects according to their estimated variances and covariances, an LMM helps keep these errors in check, leading to more accurate and reliable results (Iannaccone et al., 2021).

The structure of LMMs also provides flexibility, allowing for the construction of models with random intercepts that calculate intraindividual variability and random slopes that model distinct slopes for different types of training sessions, which can be correlated, independent or independent with equal variances (Iannaccone et al., 2021; Laird & Ware, 1982; Newans et al., 2022). In addition, LMM can separately estimate the predictive effects of an individual predictor and its group-level mean (Gelman & Hill, 2006).

3.8 Statistical analysis

In our study, we've tried to adhere to the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines, a key framework for enhanced transparency and reliability in prediction model studies (Collins et al., 2015).

An LMM analysis was used to investigate associations between RPE and sRPE and selected independent variables. After the data cleaning process was completed, the dataset was divided into a training set (70% of sessions) and a holdout set (30% of sessions). The full model consisted of session duration, TD, HSRD, SpD, $Peak_{speed}$, squad status (starter, substitute, or unused in current or last match), playing position (in current or last match), and minutes played in current or last match. Backward elimination with a significance level of 0.05 was used to remove non-significant predictors. The final model consisted of session duration, TD, HSRD, SpD, $Peak_{speed}$, starters, substitutes and unused substitutes.

As RPE and GPS data were nested within players (multiple training sessions for each player), and players were nested within teams, the data are hierarchical. This creates dependent RPE and sRPE scores within players. The chosen method can handle such multilevel data. RMSE and MAE were computed for the observed RPE and sRPE values and values predicted by the final model to evaluate predictive accuracy. An "explained" variance measure, a conditional R^2, was computed for the same observed and predicted values (Nakagawa & Schielzeth, 2013). Moreover, the intraclass correlation (ICC) was used to quantify the proportion of variance in RPE and sRPE data that can be attributed to differences between individual players (Nakagawa et al., 2017). Additionally, RMSE, MAE, ICC and R^2 were computed on the holdout set to evaluate how the model performed on unseen data.

4 Strength and limitations

The primary aims of the study were to **(1)** investigate whether external load variables can accurately predict RPE and sRPE for female football players using LMM; **(2)** identify which external load variables contribute the most to RPE and sRPE; and **(3)** provide practitioners with a simple yet effective method for predicting RPE and sRPE based on these relationships.

The results revealed that the final LMM demonstrated accurate values of R^2 (0.65 and 0.66), RMSE (1.18 and 142.2) and MAE (0.93 and 104.9) for RPE and sRPE, respectively. The ICC revealed that 84.79% and 73.44% of the total variance in sRPE and RPE responses could be attributed to between-player differences, while the remaining was due to within-player differences or variability between training sessions. This suggests that a few external load variables can accurately predict RPE and sRPE for female players.

Strength

This study has several strengths. Primarily, the study contributes new insight into a population that has been little researched. Contributing to research on female football players is important to enhance the development and knowledge, especially as women's football becomes increasingly popular and professionalised worldwide. Moreover, the study aligns with the FFRC goal of gaining new fundamental insights into the factors that affect the performance and overall health of elite female football players and developing novel methodologies that quantify and monitors athlete behaviour from various perspectives (UiT, 2022).

Furthermore, the study offers a robust approach to investigating the relationships between RPE, sRPE, and various physical performance variables in female football players. We utilised a reliable GPS and consistent procedures for data collection, which enhanced the study's replicability and the findings' accuracy. The use of LMM is particularly effective when handling hierarchical or nested data structures, as seen in our study. Regarding model evaluation, we incorporated RMSE, MAE, R^2 and ICC. We comprehensively evaluated the model's predictive accuracy and explanatory power using these measures.

Limitations

Despite these strengths, our methodology also has some limitations. A key concern is the significant amount of missing data in our study. The significant amount of missing data likely leads to overfitting in the prediction model, as indicated by its diminished performance on the holdout set compared to the training set. Consequently, this model is expected to have limited accuracy when applied to individuals not included in the dataset used for its development (Newans et al., 2022). In addition, the uneven distribution of player positions in the dataset, particularly the larger number of midfielders (Table 4), may introduce bias in the predictive accuracy of the models, as they may be more accurate in estimating RPE and sRPE for midfielders than other positions.

Furthermore, using CAA to handle missing data assumes the data are MCAR, which may not be accurate in our study considering the missingness pattern following matches. This could potentially introduce bias in our results. Further research could improve missing data treatment by using imputation techniques, which allows for unbiased and statistical valid missing data handling (Bhaskaran & Smeeth, 2014; Greenland & Finkle, 1995; Sterne et al., 2009)

Incorporating acceleration and deceleration metrics could have improved the model's accuracy, offering a more comprehensive understanding of the relationship between external and internal training loads. While this decision was made to ensure consistency across GPS systems, including these variables might have offered additional insights into players' training load response.

Lastly, the scope of our study is focused on elite female football players. Consequently, our findings may not directly apply to other populations or different sports. This limitation should be acknowledged when interpreting our results.

Perspective

The complexity of managing player performance can present unique challenges to practitioners. In this context, prediction models, such as those in the present study, can offer valuable insights (Bullock et al., 2023). However, their use and application should be understood from a user perspective.

Prediction models are not simply plug-and-play tools. They require significant customisation and fine-tuning to offer accurate predictions. This is contingent on extensive data collection over time, specific to the team and individual players. However, the high turnover rate of players and coaching staff in football complicates this process. The ever-changing composition of teams can disrupt data continuity, potentially compromising the reliability of these models.

In this regard, prediction models may find greater application in more stable environments such as academies, where turnover rates for players and coaching staff are lower. This stability encourages comprehensive data collection on individuals, enhancing the specificity and accuracy of prediction models.

Furthermore, the competitive nature of football presents another challenge. Football clubs are often protective of their data, seeing it as a strategic asset that provides them with a competitive edge. This reluctance to share data hampers the validation and further improvement of prediction models, slowing down their evolution and refinement. A more open data-sharing culture could accelerate advancements in this area, but this runs counter to the deeply ingrained competitive instincts of the sport (Bullock et al., 2023).

Despite the potential of prediction models, the implementation can be resource-intensive, requiring substantial time and expertise. This may be a luxury that is unaffordable in certain contexts, such as women's football. Current prediction models may need to be more straightforward, cost-effective, and less time-consuming to be ready for immediate adoption by practitioners.

Moreover, there is a risk of over-reliance on these models at the expense of practitioner expertise and judgement. These models should be seen as tools to support, not replace, the decision-making process. Practitioners must maintain a critical perspective on the assumptions and limitations inherent in each method, ensuring that their conclusions and recommendations are grounded in sound scientific principles and a comprehensive understanding of the multifaceted nature of athletic performance.

5 References

Abt, G., & Lovell, R. (2009). The use of individualized speed and intensity thresholds for determining the distance run at high-intensity in professional soccer. *J Sports Sci*, *27*(9), 893-898. https://doi.org/10.1080/02640410902998239

Agrawal, S., & Agrawal, J. (2015). Survey on Anomaly Detection using Data Mining Techniques. *Procedia Computer Science*, *60*, 708-713. https://doi.org/https://doi.org/10.1016/j.procs.2015.08.220

Akenhead, R., & Nassis, G. P. (2016). Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *International Journal of Sports Physiology and Performance*, *11*(5), 587-593. https://doi.org/10.1123/ijspp.2015-0331

Almulla, J., Takiddin, A., & Househ, M. (2020). The use of technology in tracking soccer players' health performance: a scoping review. *BMC Med Inform Decis Mak*, *20*(1), 184. https://doi.org/10.1186/s12911-020-01156-4

Askow, A. T., Lobato, A. L., Arndts, D. J., Jennings, W., Kreutzer, A., Erickson, J. L., Esposito, P. E., Oliver, J. M., Foster, C., & Jagim, A. R. (2021). Session Rating of Perceived Exertion (sRPE) Load and Training Impulse Are Strongly Correlated to GPS-Derived Measures of External Load in NCAA Division I Women's Soccer Athletes. *J Funct Morphol Kinesiol*, *6*(4). https://doi.org/10.3390/jfmk6040090

Asuero, A. G., Sayago, A., & González, A. G. (2006). The Correlation Coefficient: An Overview. *Critical Reviews in Analytical Chemistry*, *36*(1), 41-59. https://doi.org/10.1080/10408340500526766

Atkinson, G., Batterham, A. M., Jones, H., Taylor, C. E., Willie, C. K., & Tzeng, Y. C. (2011). Appropriate within-subjects statistical models for the analysis of baroreflex sensitivity. *Clin Physiol Funct Imaging*, *31*(1), 80-82. https://doi.org/10.1111/j.1475-097x.2010.00974.x

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*, *20*(1), 40-49. https://doi.org/10.1002/mpr.329

Baptista, I., Winther, A. K., Johansen, D., Randers, M. B., Pedersen, S., & Pettersen, S. A. (2022). The variability of physical match demands in elite women's football. *Science and Medicine in Football*, 1-7. https://doi.org/10.1080/24733938.2022.2027999

Bartlett, J. D., O'Connor, F., Pitchford, N., Torres-Ronda, L., & Robertson, S. J. (2017). Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an Individualized Approach. *International Journal of Sports Physiology and Performance*, *12*(2), 230-234. https://doi.org/10.1123/ijspp.2015-0791

Beato, M., Coratella, G., Stiff, A., & Iacono, A. D. (2018). The Validity and Between-Unit Variability of GNSS Units (STATSports Apex 10 and 18 Hz) for Measuring Distance and Peak Speed in Team Sports [Original Research]. *Frontiers in Physiology*, *9*(1288). https://doi.org/10.3389/fphys.2018.01288

Beato, M., & de Keijzer, K. L. (2019). The inter-unit and inter-model reliability of GNSS STATSports Apex and Viper units in measuring peak speed over 5, 10, 15, 20 and 30 meters. *Biology of sport*, *36*(4), 317-321. https://doi.org/10.5114/biolsport.2019.88754

Beato, M., Devereux, G., & Stiff, A. (2018). Validity and Reliability of Global Positioning System Units (STATSports Viper) for Measuring Distance and Peak Speed in Sports. *J Strength Cond Res*, *32*(10), 2831-2837. https://doi.org/10.1519/jsc.0000000000002778

Beato, M., Drust, B., & Iacono, A. D. (2021). Implementing High-speed Running and Sprinting Training in Professional Soccer. *Int J Sports Med*, *42*(4), 295-299. https://doi.org/10.1055/a-1302-7968

Beato, M., Vicens-Bordas, J., Peña, J., & Costin, Andrew J. (2023). Training load comparison between small, medium, and large-sided games in professional football [Brief Research Report]. *Frontiers in Sports and Active Living*, *5*. https://doi.org/10.3389/fspor.2023.1165242

Bhaskaran, K., & Smeeth, L. (2014). What is the difference between missing completely at random and missing at random? *Int J Epidemiol*, *43*(4), 1336-1339. https://doi.org/10.1093/ije/dyu080

Bloomfield, J., Polman, R., & O'Donoghue, P. (2007). Physical demands of different positions in FA Premier League soccer. *Journal of sports science & medicine*, *6*(1), 63.

Borg, E., & Kaijser, L. (2006). A comparison between three rating scales for perceived exertion and two different work tests. *Scandinavian Journal of Medicine & Science in Sports*, *16*(1), 57-69. https://doi.org/https://doi.org/10.1111/j.1600-0838.2005.00448.x

Borg, G. (1970). Perceived exertion as an indicator of somatic stress. *Scand J Rehabil Med*, *2*(2), 92-98.

Borg, G. A. V. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, *14*(5), 377-381. https://journals.lww.com/acsm-msse/Fulltext/1982/05000/Psychophysical_bases_of_perceived_exertion.12.aspx

Borresen, J., & Lambert, M. I. (2008). Quantifying training load: a comparison of subjective and objective methods. *Int J Sports Physiol Perform*, *3*(1), 16-30. https://doi.org/10.1123/ijspp.3.1.16

Borresen, J., & Lambert, M. I. (2009). The quantification of training load, the training response and the effect on performance. *Sports Med*, *39*(9), 779-795. https://doi.org/10.2165/11317780-000000000-00000

Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M. C., Gabbett, T. J., Coutts, A. J., Burgess, D. J., Gregson, W., & Cable, N. T. (2017). Monitoring Athlete Training Loads: Consensus Statement. *International Journal of Sports Physiology and Performance*, *12*(s2), S2-161-S162-170. https://doi.org/10.1123/ijspp.2017-0208

Bradley, P., & Scott, D. (2020, 6. july 2020). *Physical abalysis of France 2019 shows increase in speed and intensity* Fédération Internationale de Football Association (FIFA) https://www.fifa.com/tournaments/womens/womensworldcup/france2019/news/physical-analysis-of-france-2019-shows-increase-in-speed-and-intensity

Brocherie, F., & Beard, A. (2020). All Alone We Go Faster, Together We Go Further: The Necessary Evolution of Professional and Elite Sporting Environment to Bridge the Gap Between Research and Practice. *Front Sports Act Living*, *2*, 631147. https://doi.org/10.3389/fspor.2020.631147

Buchheit, M., Haddad, H. A., Simpson, B. M., Palazzi, D., Bourdon, P. C., Salvo, V. D., & Mendez-Villanueva, A. (2014). Monitoring Accelerations With GPS in Football: Time to Slow Down? *International Journal of Sports Physiology and Performance*, *9*(3), 442-445. https://doi.org/10.1123/ijspp.2013-0187

Buchheit, M., Mendez-Villanueva, A., Simpson, B. M., & Bourdon, P. C. (2010). Match running performance and fitness in youth soccer. *Int J Sports Med*, *31*(11), 818-825. https://doi.org/10.1055/s-0030-1262838

Buchheit, M., Sandua, M., Berndsen, J., Shelton, A., Smith, S., Norman, D., McHugh, D., & Hader, K. (2021). Loading patterns and programming practises in elite football: insights from 100 elite practitioners. *Sports Performance & Science Reports*, *153*, v1.

Buchheit, M., Settembre, M., Hader, K., & McHugh, D. (2023a). Exposures to near-to-maximal speed running bouts during different turnarounds in elite football: association with match hamstring injuries [journal article]. *Biology of sport*, 1057-1067. https://doi.org/10.5114/biolsport.2023.125595

Buchheit, M., Settembre, M., Hader, K., & McHugh, D. (2023b). Planning the Microcycle in Elite Football: To Rest or Not to Rest? *International Journal of Sports Physiology and Performance*, 1-7. https://doi.org/10.1123/ijspp.2022-0146

Buchheit, M., & Simpson, B. M. (2017). Player-Tracking Technology: Half-Full or Half-Empty Glass? *International Journal of Sports Physiology and Performance*, *12*(s2), S2-35-S32-41. https://doi.org/10.1123/ijspp.2016-0499

Bullock, G. S., Hughes, T., Arundale, A. H., Ward, P., Collins, G. S., & Kluzek, S. (2022). Black Box Prediction Methods in Sports Medicine Deserve a Red Card for Reckless Practice: A Change of Tactics is Needed to Advance Athlete Care. *Sports Med*, *52*(8), 1729-1735. https://doi.org/10.1007/s40279-022-01655-6

Bullock, G. S., Ward, P., Impellizzeri, F. M., Kluzek, S., Hughes, T., Dhiman, P., Riley, R. D., & Collins, G. S. (2023). The Trade Secret Taboo: Open Science Methods are Required to Improve Prediction Models in Sports Medicine and Performance. *Sports Medicine*. https://doi.org/10.1007/s40279-023-01849-6

Carey, D. L., Ong, K., Morris, M. E., Crow, J., & Crossley, K. M. (2016). Predicting ratings of perceived exertion in Australian football players: methods for live estimation. *International Journal of Computer Science in Sport*, *15*(2), 64-77. https://doi.org/doi:10.1515/ijcss-2016-0005

Carling, C., Le Gall, F., & Dupont, G. (2012). Analysis of repeated high-intensity running performance in professional soccer. *Journal of Sports Sciences*, *30*(4), 325-336. https://doi.org/10.1080/02640414.2011.652655

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, *7*(3), 1247-1250. https://doi.org/10.5194/gmd-7-1247-2014

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Comput. Surv.*, *41*, 15:11-15:58.

Chmura, P., Konefał, M., Chmura, J., Kowalczuk, E., Zając, T., Rokita, A., & Andrzejewski, M. (2018). Match outcome and running performance in different intensity ranges among elite soccer players. *Biology of sport*, *35*(2), 197-203. https://doi.org/10.5114/biolsport.2018.74196

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine*, *13*(1), 1. https://doi.org/10.1186/s12916-014-0241-z

Costa, J. A., Rago, V., Brito, P., Figueiredo, P., Sousa, A., Abade, E., & Brito, J. (2022). Training in women soccer players: A systematic review on training load monitoring. *Front Psychol*, *13*, 943857. https://doi.org/10.3389/fpsyg.2022.943857

Coutts, A. J., Rampinini, E., Marcora, S. M., Castagna, C., & Impellizzeri, F. M. (2009). Heart rate and blood lactate correlates of perceived exertion during small-sided soccer games. *J Sci Med Sport*, *12*(1), 79-84. https://doi.org/10.1016/j.jsams.2007.08.005

Crang, Z. L., Duthie, G., Cole, M. H., Weakley, J., Hewitt, A., & Johnston, R. D. (2022). The inter-device reliability of global navigation satellite systems during team sport movement across multiple days. *Journal of Science and Medicine in Sport*, *25*(4), 340-344. https://doi.org/https://doi.org/10.1016/j.jsams.2021.11.044

Dalen, T., Sandmæl, S., Stevens, T. G. A., Hjelde, G. H., Kjøsnes, T. N., & Wisløff, U. (2021). Differences in Acceleration and High-Intensity Activities Between Small-Sided Games and Peak Periods of Official Matches in Elite Soccer Players. *The Journal of Strength & Conditioning Research*, *35*(7), 2018-2024. https://doi.org/10.1519/jsc.0000000000003081

Delaney, J. A., Thornton, H. R., Rowell, A. E., Dascombe, B. J., Aughey, R. J., & Duthie, G. M. (2018). Modelling the decrement in running intensity within professional soccer players. *Science and Medicine in Football*, *2*(2), 86-92. https://doi.org/10.1080/24733938.2017.1383623

Dello Iacono, A., McLaren, S. J., Macpherson, T. W., Beato, M., Weston, M., Unnithan, V. B., & Shushan, T. (2023). Quantifying Exposure and Intra-Individual Reliability of High-Speed and Sprint Running During Sided-Games Training in Soccer Players: A Systematic Review and Meta-analysis. *Sports Medicine*, *53*(2), 371-413. https://doi.org/10.1007/s40279-022-01773-1

Dong, Y., & Peng, C. Y. (2013). Principled missing data methods for researchers. *Springerplus*, *2*(1), 222. https://doi.org/10.1186/2193-1801-2-222

Douchet, T., Humbertclaude, A., Cometti, C., Paizis, C., & Babault, N. (2021). Quantifying Accelerations and Decelerations in Elite Women Soccer Players during Regular In-Season Training as an Index of Training Load. *Sports*, *9*(8), 109. https://www.mdpi.com/2075-4663/9/8/109

Ellens, S., Middleton, K., Gastin, P. B., & Varley, M. C. (2022). Techniques to derive and clean acceleration and deceleration data of athlete tracking technologies in team sports: A scoping review. *J Sports Sci*, *40*(16), 1772-1800. https://doi.org/10.1080/02640414.2022.2054535

Eston, R. (2012). Use of Ratings of Perceived Exertion in Sports. *International Journal of Sports Physiology and Performance*, *7*(2), 175-182. https://doi.org/10.1123/ijspp.7.2.175

Faude, O., Koch, T., & Meyer, T. (2012). Straight sprinting is the most frequent action in goal situations in professional football. *J Sports Sci*, *30*(7), 625-631. https://doi.org/10.1080/02640414.2012.665940

FIFA. (2022, 16. december). *Women's Football: Accelerate the growth of women's football* https://www.fifa.com/womens-football

Forzasys. (2022). *pmSys*. https://forzasys.com/pmSys.html

Foster, C., Boullosa, D., McGuigan, M., Fusco, A., Cortis, C., Arney, B. E., Orton, B., Dodge, C., Jaime, S., Radtke, K., van Erp, T., de Koning, J. J., Bok, D., Rodriguez-Marroyo, J. A., & Porcari, J. P. (2021). 25 Years of Session Rating of Perceived Exertion: Historical Perspective and Development. *International Journal of Sports Physiology and Performance*, *16*(5), 612-621. https://doi.org/10.1123/ijspp.2020-0599

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A new approach to monitoring exercise training. *J Strength Cond Res*, *15*(1), 109-115.

Gaudino, P., Iaia, F. M., Strudwick, A. J., Hawkins, R. D., Alberti, G., Atkinson, G., & Gregson, W. (2015). Factors influencing perception of effort (session rating of

perceived exertion) during elite soccer training. *Int J Sports Physiol Perform*, *10*(7), 860-864. https://doi.org/10.1123/ijspp.2014-0518
Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
Geurkink, Y., Vandewiele, G., Lievens, M., de Turck, F., Ongenae, F., Matthys, S. P. J., Boone, J., & Bourgois, J. G. (2019). Modeling the Prediction of the Session Rating of Perceived Exertion in Soccer: Unraveling the Puzzle of Predictive Indicators. *International Journal of Sports Physiology and Performance*, *14*(6), 841-846. https://doi.org/10.1123/ijspp.2018-0698
Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annu Rev Psychol*, *60*, 549-576. https://doi.org/10.1146/annurev.psych.58.110405.085530
Greenland, S., & Finkle, W. D. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol*, *142*(12), 1255-1264. https://doi.org/10.1093/oxfordjournals.aje.a117592
Gualtieri, A., Rampinini, E., Dello Iacono, A., & Beato, M. (2023). High-speed running and sprinting in professional adult soccer: Current thresholds definition, match demands and training strategies. A systematic review [Review]. *Frontiers in Sports and Active Living*, *5*. https://doi.org/10.3389/fspor.2023.1116293
Gualtieri, A., Rampinini, E., Sassi, R., & Beato, M. (2020). Workload Monitoring in Top-level Soccer Players During Congested Fixture Periods. *Int J Sports Med*, *41*(10), 677-681. https://doi.org/10.1055/a-1171-1865
Haddad, M., Chaouachi, A., Wong del, P., Castagna, C., Hambli, M., Hue, O., & Chamari, K. (2013). Influence of fatigue, stress, muscle soreness and sleep on perceived exertion during submaximal effort. *Physiol Behav*, *119*, 185-189. https://doi.org/10.1016/j.physbeh.2013.06.016
Halson, S. L. (2014). Monitoring training load to understand fatigue in athletes. *Sports Med*, *44 Suppl 2*(Suppl 2), S139-147. https://doi.org/10.1007/s40279-014-0253-z
Halson, S. L. (2014). Monitoring Training Load to Understand Fatigue in Athletes. *Sports Medicine*, *44*(2), 139-147. https://doi.org/10.1007/s40279-014-0253-z
Harkness-Armstrong, A., Till, K., Datson, N., Myhill, N., & Emmonds, S. (2022). A systematic review of match-play characteristics in women's soccer. *PloS one*, *17*(6), e0268334. https://doi.org/10.1371/journal.pone.0268334
Harper, D. J., Carling, C., & Kiely, J. (2019). High-Intensity Acceleration and Deceleration Demands in Elite Team Sports Competitive Match Play: A Systematic Review and Meta-Analysis of Observational Studies. *Sports Medicine*, *49*(12), 1923-1947. https://doi.org/10.1007/s40279-019-01170-1
Hennessy, L., & Jeffreys, I. (2018). The Current Use of GPS, Its Potential, and Limitations in Soccer. *Strength & Conditioning Journal*, *40*(3), 83-94. https://doi.org/10.1519/ssc.0000000000000386
Hernán, M. A., Hsu, J., & Healy, B. (2019). A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *CHANCE*, *32*(1), 42-49. https://doi.org/10.1080/09332480.2019.1579578
Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geosci. Model Dev.*, *15*(14), 5481-5487. https://doi.org/10.5194/gmd-15-5481-2022
Hoff, J., Wisloff, U., Engen, L. C., Kemi, O. J., & Helgerud, J. (2002). Soccer specific aerobic endurance training. *Br J Sports Med*, *36*(3), 218-221. https://doi.org/10.1136/bjsm.36.3.218

Iannaccone, A., Conte, D., Cortis, C., & Fusco, A. (2021). Usefulness of Linear Mixed-Effects Models to Assess the Relationship between Objective and Subjective Internal Load in Team Sports. *Int J Environ Res Public Health*, *18*(2). https://doi.org/10.3390/ijerph18020392

Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2019). Internal and External Training Load: 15 Years On. *Int J Sports Physiol Perform*, *14*(2), 270-273. https://doi.org/10.1123/ijspp.2018-0935

Impellizzeri, F. M., McCall, A., Ward, P., Bornn, L., & Coutts, A. J. (2020). Training Load and Its Role in Injury Prevention, Part 2: Conceptual and Methodologic Pitfalls. *J Athl Train*, *55*(9), 893-901. https://doi.org/10.4085/1062-6050-501-19

Impellizzeri, F. M., Menaspà, P., Coutts, A. J., Kalkhoven, J., & Menaspà, M. J. (2020). Training Load and Its Role in Injury Prevention, Part I: Back to the Future. *J Athl Train*, *55*(9), 885-892. https://doi.org/10.4085/1062-6050-500-19

Impellizzeri, F. M., Rampinini, E., Coutts, A. J., Sassi, A., & Marcora, S. M. (2004). Use of RPE-based training load in soccer. *Med Sci Sports Exerc*, *36*(6), 1042-1047. https://doi.org/10.1249/01.mss.0000128199.23901.2f

Impellizzeri, F. M., Rampinini, E., & Marcora, S. M. (2005). Physiological assessment of aerobic training in soccer. *J Sports Sci*, *23*(6), 583-592. https://doi.org/10.1080/02640410400021278

Impellizzeri, F. M., Ward, P., Coutts, A. J., Bornn, L., & McCall, A. (2020). Training Load and Injury Part 2: Questionable Research Practices Hijack the Truth and Mislead Well-Intentioned Clinicians. *J Orthop Sports Phys Ther*, *50*(10), 577-584. https://doi.org/10.2519/jospt.2020.9211

Impellizzeri, F. M., Woodcock, S., Coutts, A. J., Fanchini, M., McCall, A., & Vigotsky, A. D. (2021). What Role Do Chronic Workloads Play in the Acute to Chronic Workload Ratio? Time to Dismiss ACWR and Its Underlying Theory. *Sports Medicine*, *51*(3), 581-592. https://doi.org/10.1007/s40279-020-01378-6

Jaspers, A., De Beéck, T. O., Brink, M. S., Frencken, W. G. P., Staes, F., Davis, J. J., & Helsen, W. F. (2018). Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn From Machine Learning? *International Journal of Sports Physiology and Performance*, *13*(5), 625-630. https://doi.org/10.1123/ijspp.2017-0299

Jastrzębski, Z., & Radzimiński, Ł. (2015). Individual vs General Time-Motion Analysis and Physiological Response in 4 vs 4 and 5 vs 5 Small-Sided Soccer Games. *International Journal of Performance Analysis in Sport*, *15*(1), 397-410. https://doi.org/10.1080/24748668.2015.11868801

Kirkendall, D. T. (2020). Evolution of soccer as a research topic. *Prog Cardiovasc Dis*, *63*(6), 723-729. https://doi.org/10.1016/j.pcad.2020.06.011

Kirkendall, D. T., & Krustrup, P. (2022). Studying professional and recreational female footballers: A bibliometric exercise. *Scand J Med Sci Sports*, *32 Suppl 1*, 12-26. https://doi.org/10.1111/sms.14019

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Krueger, C., & Tian, L. (2004). A Comparison of the General Linear Mixed Model and Repeated Measures ANOVA Using a Dataset with Multiple Missing Data Points. *Biological Research For Nursing*, *6*(2), 151-157. https://doi.org/10.1177/1099800404267682

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963-974.
Larsson, P. (2003). Global positioning system and sport-specific testing. *Sports Med*, *33*(15), 1093-1101. https://doi.org/10.2165/00007256-200333150-00002
Liljequist, D., Elfving, B., & Skavberg Roaldsen, K. (2019). Intraclass correlation - A discussion and demonstration of basic features. *PloS one*, *14*(7), e0219854. https://doi.org/10.1371/journal.pone.0219854
Lininger, M., Spybrook, J., & Cheatham, C. C. (2015). Hierarchical linear model: thinking outside the traditional repeated-measures analysis-of-variance box. *J Athl Train*, *50*(4), 438-441. https://doi.org/10.4085/1062-6050-49.5.09
Little, T., & Williams, A. G. (2007). Measures of exercise intensity during soccer training drills with professional soccer players. *J Strength Cond Res*, *21*(2), 367-371. https://doi.org/10.1519/r-19445.1
Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. (2014). On the joys of missing data. *J Pediatr Psychol*, *39*(2), 151-162. https://doi.org/10.1093/jpepsy/jst048
Luteberget, L. S., Houtmeyers, K. C., Vanrenterghem, J., Jaspers, A., Brink, M. S., & Helsen, W. F. (2021). Load Monitoring Practice in Elite Women Association Football. *Front Sports Act Living*, *3*, 715122. https://doi.org/10.3389/fspor.2021.715122
Malone, J. J., Barrett, S., Barnes, C., Twist, C., & Drust, B. (2020). To infinity and beyond: the use of GPS devices within the football codes. *Science and Medicine in Football, 4*(1), 82-84. https://doi.org/10.1080/24733938.2019.1679871
Malone, J. J., Di Michele, R., Morgans, R., Burgess, D., Morton, J. P., & Drust, B. (2015). Seasonal training-load quantification in elite English premier league soccer players. *Int J Sports Physiol Perform*, *10*(4), 489-497. https://doi.org/10.1123/ijspp.2014-0352
Malone, J. J., Lovell, R., Varley, M. C., & Coutts, A. J. (2017). Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *Int J Sports Physiol Perform*, *12*(Suppl 2), S218-s226. https://doi.org/10.1123/ijspp.2016-0236
Malone, S., Roe, M., Doran, D. A., Gabbett, T. J., & Collins, K. (2017). High chronic training loads and exposure to bouts of maximal velocity running reduce injury risk in elite Gaelic football. *J Sci Med Sport*, *20*(3), 250-254. https://doi.org/10.1016/j.jsams.2016.08.005
Mara, J. K., Thompson, K. G., Pumpa, K. L., & Morgan, S. (2017). The acceleration and deceleration profiles of elite female soccer players during competitive matches. *Journal of Science and Medicine in Sport*, *20*(9), 867-872.
Marynowicz, J., Lango, M., Horna, D., Kikut, K., & Andrzejewski, M. (2022). Predicting ratings of perceived exertion in youth soccer using decision tree models [journal article]. *Biology of sport*, *39*(2), 245-252. https://doi.org/10.5114/biolsport.2022.103723
McCall, A., Fanchini, M., & Coutts, A. J. (2017). Prediction: The Modern-Day Sport-Science and Sports-Medicine "Quest for the Holy Grail". *Int J Sports Physiol Perform*, *12*(5), 704-706. https://doi.org/10.1123/ijspp.2017-0137
McLaren, S. J., Macpherson, T. W., Coutts, A. J., Hurst, C., Spears, I. R., & Weston, M. (2018). The Relationships Between Internal and External Measures of Training Load and Intensity in Team Sports: A Meta-Analysis. *Sports Medicine, 48*(3), 641-658. https://doi.org/10.1007/s40279-017-0830-z
Mujika, I., Santisteban, J., Impellizzeri, F. M., & Castagna, C. (2009). Fitness determinants of success in men's and women's football. *J Sports Sci*, *27*(2), 107-114. https://doi.org/10.1080/02640410802428071

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R(2) and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J R Soc Interface*, *14*(134). https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142. https://doi.org/https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nassis, G. P., Verhagen, E., Brito, J., Figueiredo, P., & Krustrup, P. (2023). A review of machine learning applications in soccer with an emphasis on injury risk. *Biology of sport*, *40*(1), 233-239. https://doi.org/10.5114/biolsport.2023.114283

Newans, T., Bellinger, P., Drovandi, C., Buxton, S., & Minahan, C. (2022). The Utility of Mixed Models in Sport Science: A Call for Further Adoption in Longitudinal Data Sets. *International Journal of Sports Physiology and Performance*, 1-7. https://doi.org/10.1123/ijspp.2021-0496

Okholm Kryger, K., Wang, A., Mehta, R., Impellizzeri, F. M., Massey, A., & McCall, A. (2022). Research on women's football: a scoping review. *Science and Medicine in Football*, *6*(5), 549-558. https://doi.org/10.1080/24733938.2020.1868560

Osgnach, C., Poser, S., Bernardini, R., Rinaldo, R., & di Prampero, P. E. (2010). Energy cost and metabolic power in elite soccer: a new match analysis approach. *Med Sci Sports Exerc*, *42*(1), 170-178. https://doi.org/10.1249/MSS.0b013e3181ae5cfd

Park, L. A., Scott, D., & Lovell, R. (2019). Velocity zone classification in elite women's football: where do we draw the lines? *Science and Medicine in Football*, *3*(1), 21-28.

Pepe, M. S., Janes, H., Longton, G., Leisenring, W., & Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*, *159*(9), 882-890. https://doi.org/10.1093/aje/kwh101

Pons, E., García-Calvo, T., Resta, R., Blanco, H., López del Campo, R., Díaz García, J., & Pulido, J. J. (2019). A comparison of a GPS device and a multi-camera video technology during official soccer matches: Agreement between systems. *PloS one*, *14*(8), e0220729.

Randell, R. K., Clifford, T., Drust, B., Moss, S. L., Unnithan, V. B., De Ste Croix, M. B. A., Datson, N., Martin, D., Mayho, H., Carter, J. M., & Rollo, I. (2021). Physiological Characteristics of Female Soccer Players and Health and Performance Considerations: A Narrative Review. *Sports Med*, *51*(7), 1377-1399. https://doi.org/10.1007/s40279-021-01458-1

Riboli, A., Coratella, G., Rampichini, S., Cé, E., & Esposito, F. (2020). Area per player in small-sided games to replicate the external load and estimated physiological match demands in elite soccer players. *PloS one*, *15*(9), e0229194. https://doi.org/10.1371/journal.pone.0229194

Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F. M., & Baca, A. (2023). Machine learning application in soccer: a systematic review. *Biology of sport*, *40*(1), 249-263. https://doi.org/10.5114/biolsport.2023.112970

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592. https://doi.org/10.1093/biomet/63.3.581

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychol Methods*, *7*(2), 147-177.

Schwellnus, M., Soligard, T., Alonso, J. M., Bahr, R., Clarsen, B., Dijkstra, H. P., Gabbett, T. J., Gleeson, M., Hägglund, M., Hutchinson, M. R., Janse Van Rensburg, C., Meeusen, R., Orchard, J. W., Pluim, B. M., Raftery, M., Budgett, R., & Engebretsen, L. (2016).

How much is too much? (Part 2) International Olympic Committee consensus statement on load in sport and risk of illness. *Br J Sports Med*, *50*(17), 1043-1052. https://doi.org/10.1136/bjsports-2016-096572

Scott, D., & Lovell, R. (2018). Individualisation of speed thresholds does not enhance the dose-response determination in football training. *J Sports Sci*, *36*(13), 1523-1532. https://doi.org/10.1080/02640414.2017.1398894

Scott, M. T. U., Scott, T. J., & Kelly, V. G. (2016). The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review. *The Journal of Strength & Conditioning Research*, *30*(5), 1470-1490. https://doi.org/10.1519/jsc.0000000000001221

Seshadri, D. R., Thom, M. L., Harlow, E. R., Gabbett, T. J., Geletka, B. J., Hsu, J. J., Drummond, C. K., Phelan, D. M., & Voos, J. E. (2021). Wearable Technology and Analytics as a Complementary Toolkit to Optimize Workload and to Reduce Injury Burden [Review]. *Frontiers in Sports and Active Living*, *2*. https://doi.org/10.3389/fspor.2020.630576

Shmueli, G. (2011). To Explain or to Predict? *Statistical Science*, *25*. https://doi.org/10.1214/10-STS330

Statsports. (2022). *About Statsports*. https://statsports.com/careers

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, *338*, b2393. https://doi.org/10.1136/bmj.b2393

Sweeting, A. J., Cormack, S. J., Morgan, S., & Aughey, R. J. (2017). When Is a Sprint a Sprint? A Review of the Analysis of Team-Sport Athlete Activity Profile. *Front Physiol*, *8*, 432. https://doi.org/10.3389/fphys.2017.00432

Taylor, J. B., Wright, A. A., Dischiavi, S. L., Townsend, M. A., & Marmon, A. R. (2017). Activity Demands During Multi-Directional Team Sports: A Systematic Review. *Sports Med*, *47*(12), 2533-2551. https://doi.org/10.1007/s40279-017-0772-5

Thambawita, V., Hicks, S. A., Borgli, H., Stensland, H. K., Jha, D., Svensen, M. K., Pettersen, S.-A., Johansen, D., Johansen, H. D., Pettersen, S. D., Nordvang, S., Pedersen, S., Gjerdrum, A., Grønli, T.-M., Fredriksen, P. M., Eg, R., Hansen, K., Fagernes, S., Claudi, C., . . . Halvorsen, P. (2020). *PMData: a sports logging dataset* Proceedings of the 11th ACM Multimedia Systems Conference, Istanbul, Turkey. https://doi.org/10.1145/3339825.3394926

Thornton, H. R., Nelson, A. R., Delaney, J. A., Serpiello, F. R., & Duthie, G. M. (2019). Interunit reliability and effect of data-processing methods of global positioning systems. *International Journal of Sports Physiology and Performance*, *14*(4), 432-438.

Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A., & Clubb, J. (2022). Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis. *Sports Med Open*, *8*(1), 15. https://doi.org/10.1186/s40798-022-00408-z

Trewin, J., Meylan, C., Varley, M. C., & Cronin, J. (2018). The match-to-match variation of match-running in elite female soccer. *J Sci Med Sport*, *21*(2), 196-201. https://doi.org/10.1016/j.jsams.2017.05.009

UiT. (2022). *FFRC*

Female Football Research Centre https://uit.no/research/ffrc

Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC press.
van Royen, F. S., Moons, K. G. M., Geersing, G. J., & van Smeden, M. (2022). Developing, validating, updating and judging the impact of prognostic models for respiratory diseases. *Eur Respir J*, *60*(3). https://doi.org/10.1183/13993003.00250-2022
Varley, M. C., Fairweather, I. H., & Aughey, R. J. (2012). Validity and reliability of GPS for measuring instantaneous velocity during acceleration, deceleration, and constant motion. *Journal of Sports Sciences*, *30*(2), 121-127. https://doi.org/10.1080/02640414.2011.627941
Varley, M. C., Jaspers, A., Helsen, W. F., & Malone, J. J. (2017). Methodological considerations when quantifying high-intensity efforts in team sport using global positioning system technology. *International Journal of Sports Physiology and Performance*, *12*(8), 1059-1068.
Verheul, J., Nedergaard, N. J., Pogson, M., Lisboa, P., Gregson, W., Vanrenterghem, J., & Robinson, M. A. (2021). Biomechanical loading during running: can a two mass-spring-damper model be used to evaluate ground reaction forces for high-intensity tasks? *Sports biomechanics*, *20*(5), 571-582.
Viru, A., & Viru, M. (2000). Nature of training effects. *Exercise and sport science*, *6795*.
Weaving, D., Jones, B., Till, K., Abt, G., & Beggs, C. (2017). The Case for Adopting a Multivariate Approach to Optimize Training Load Quantification in Team Sports. *Front Physiol*, *8*, 1024. https://doi.org/10.3389/fphys.2017.01024
West, S. W., Clubb, J., Torres-Ronda, L., Howells, D., Leng, E., Vescovi, J. D., Carmody, S., Posthumus, M., Dalen-Lorentsen, T., & Windt, J. (2021). More than a Metric: How Training Load is Used in Elite Sport for Athlete Management. *Int J Sports Med*, *42*(4), 300-306. https://doi.org/10.1055/a-1268-8791
Wiig, H., Andersen, T. E., Luteberget, L. S., & Spencer, M. (2020). Individual Response to External Training Load in Elite Football Players. *Int J Sports Physiol Perform*, *15*(5), 696-704. https://doi.org/10.1123/ijspp.2019-0453
Williams, M., & Morgan, S. (2009). Horizontal positioning error derived from stationary GPS units: A function of time and proximity to building infrastructure. *International Journal of Performance Analysis in Sport*, *9*(2), 275-280. https://doi.org/10.1080/24748668.2009.11868483
Winther, A. K., Baptista, I., Pedersen, S., Randers, M. B., Johansen, D., Krustrup, P., & Pettersen, S. A. (2022). Position specific physical performance and running intensity fluctuations in elite women's football. *Scand J Med Sci Sports*, *32 Suppl 1*, 105-114. https://doi.org/10.1111/sms.14105
Witte, T. H., & Wilson, A. M. (2004). Accuracy of non-differential GPS for the determination of speed over ground. *J Biomech*, *37*(12), 1891-1898. https://doi.org/10.1016/j.jbiomech.2004.02.031

UiT The Arctic University of Norway

School of Sports Sciences, Faculty of Health Sciences

Developing a Linear Mixed Model to Predict RPE and sRPE in Female Elite Football Players Using External Load Measures

Jakob Kjos Vollstad

Master's article in Sports Sciences…3901 May 2023

Abstract

Introduction: Despite the broad consensus on the importance of training load monitoring and management for footballers' performance, existing research on female football players remains scarce. The present study aimed to develop and evaluate a linear mixed model (LMM) to predict the rating of perceived exertion (RPE) and session RPE (sRPE) in female elite football players using GPS-derived external load measures.

Methods: A sample of 58 players from two clubs in Norway were monitored across the 2020 and 2021 seasons (n = 4311 training observations). Physical performance data were collected using STATSports GPS APEX, while RPE and sRPE were collected using the PM Reporter Pro smartphone application. Associations between RPE, sRPE, and selected independent variables were investigated using LMM analysis.

Results: The final model demonstrated accurate predictive performance of RPE and sRPE, with coefficient of determination (R^2) values of 0.65 and 0.66, root mean squared error (RMSE) of 1.16 and 142.2 and mean absolute error (MAE) 0.93 and 104.9, respectively. Key predictors included session duration, total distance (TD), high-speed running distance (HSRD), sprint distance (SpD), peak speed ($Peak_{speed}$), and player status (starter or substitute). The high intraclass correlation coefficients (ICC) indicated that a considerable proportion of the total variability in sRPE and RPE responses could be attributed to individual player differences.

Conclusion: This study highlights the potential of using GPS-derived data to predict RPE and sRPE values in female football players, providing practitioners with valuable information to tailor individual training sessions and balance training intensity and recovery.

1 Introduction

In recent years advancements in sports science and technology have led to an extensive collection of medical, training, and performance data in professional football (Bullock et al., 2022). With increased data collection and competition demands (FIFA, 2022; Randell et al., 2021), coaching practices must evolve accordingly (Brocherie & Beard, 2020). With numerous factors influencing the daily planning process, the need for tools to help elicit the necessary training load is ever-present.

Research has suggested using predictions to evaluate medical, training and performance data in football (Bullock et al., 2022; McCall et al., 2017; Rico-González et al., 2023; Seshadri et al., 2021). Prediction models can be used to assist practitioners with clinical decision-making; they incorporate data from multiple predictor variables measured at a point in time to estimate an individual's probability of health- or performance-related outcome being present at the time of measurement (diagnosis) or if it will occur in the future (prognosis) (Bullock et al., 2022; Collins et al., 2015; McCall et al., 2017; van Royen et al., 2022).

However, datasets are often complicated and characterised by multiple dependent observations across training and matches and imbalanced data due to injuries, illness and team selection (Newans et al., 2022). In the context of football, it is important to consider not only the team-level variability but also the inter-individual variability when analysing workload data, as exercise responses may vary not only between players but also within the same players, creating multilevel hierarchical data (Gelman & Hill, 2006; Iannaccone et al., 2021; Impellizzeri et al., 2004; Nakagawa & Schielzeth, 2013; Newans et al., 2022).

Training load quantification is typically categorised into internal and external components (Impellizzeri et al., 2019). Internal training load (ITL) represents an individual's psycho-physiological response to a specific workload and is frequently assessed by collecting players' ratings of perceived exertion (RPE) and session rating of perceived exertion (sRPE) (Impellizzeri et al., 2019; Impellizzeri et al., 2004; Thorpe et al., 2015). RPE and sRPE are cost-effective, easily administered, and have demonstrated correlations with other physiological measures, such as heart rate and blood lactate (Foster et al., 2021; Impellizzeri et al., 2004). External training load (ETL) is commonly monitored using global positioning systems (GPS), proven valid and reliable for assessing distance and determining peak speed

(Beato et al., 2018; Beato & de Keijzer, 2019; Gualtieri et al., 2023; Hennessy & Jeffreys, 2018).

Recent studies have employed machine learning (ML) and other statistical methods to predict RPE from GPS-derived workload data, demonstrating accurate predictions for RPE (Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022). This research holds significant implications, as coaches could use past external load data to predict internal load responses, leading to more precise training methods and effectively balancing training intensity, duration, and recovery periods to ensure players are prepared for match demands without incurring excessive fatigue (Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022).

Choosing the appropriate statistical method is essential and requires careful consideration of the trade-off between statistical accuracy and model complexity. Prediction models must be accurate and practical simultaneously (van Royen et al., 2022). While more sophisticated approaches such as ML can improve models' accuracy, researchers have suggested that adopting linear mixed models (LMM) when analysing longitudinal sports-science datasets offers a balanced approach by capturing team-level and inter-individual variability without being too complex, making it practical to use (Bullock et al., 2022; Iannaccone et al., 2021; Newans et al., 2022).

To our knowledge, no previous study has predicted RPE and sRPE from external load measures in elite female football players using LMM. The primary aims of this study are to **(1)** investigate whether external load indicators, such as total distance (TD), high-speed running distance (HSRD), sprint distance (SpD), and peak speed ($Peak_{speed}$), can accurately predict RPE and sRPE for female football players using LMM; **(2)** identify which external load indicator contributes the most to RPE and sRPE; and **(3)** provide practitioners with a simple yet effective method for predicting sRPE based on these relationships.

2 Methods

2.1 Data collection

With ethical institutional approval from the Norwegian Centre for Research Data (reference number: 296155) and written informed consent from the participants, 58 female football players (22 ± 4 years of age) from two top-level Norwegian clubs were included in the study.

A thorough description of the data collection can be found in Winther et al. (2022) and Baptista et al. (2022) and will, for practical purposes, be reiterated here. Locomotor data from the two clubs' training sessions in the 2020 and 2021 seasons were collected using GPS APEX (STATSports), with a sampling frequency of 10 Hz. The validity and levels of accuracy (bias <5%) of this tracking system has been previously presented (Beato et al., 2018). During training and matches, each player wore a tight vest with the GPS unit on the back of their upper body between the scapula as described by the manufacturer. To minimise inter-device error, each player used the same GPS unit during the entire collection period (Beato et al., 2018).

Doppler-derived speed data was exported from manufacturer software (STATSport Sonra 2.1.4) into Python 3.7.6 for processing (linearly interpolating any missing raw data) and to derive metrics. Raw acceleration was then calculated for 0.6 s. After deriving all the metrics, the data were transferred to R (R.4.0.5, R Core Team, 2021) for statistical analysis.

Approximately 30 minutes post-training session, players reported RPE values, following Foster et al. (2001) recommendations to minimise the effect of immediate fatigue after training and reduce the influence of peer pressure (Malone et al., 2015). The scores were recorded individually using the RPE CR-10 scale (Foster et al., 2001) via the PM Reporter Pro smartphone application. Each player was fully familiarised with the scale before reporting. After submitting their responses, data were automatically uploaded to cloud-based software.

2.2 Physical performance variables

The physical parameters analysed included total distance (TD), high-speed running distance (HSRD) (19-23 km/h), sprint distance (SpD) (>23 km/h), and peak speed ($Peak_{speed}$). The speed thresholds were chosen according to previous research (Park et al., 2019) and are the same thresholds used by FIFA (Bradley & Scott, 2020). Acceleration and deceleration were not included in our analysis, as we aimed to use relatively standard variables across GPS systems. This decision was made because various methods exist for calculating acceleration and deceleration, which could introduce inconsistencies and limit the generalisability of our findings.

Table 1: *Variable description.*

	Variable	Type	Units
Dependent	Total distance (TD)	Continuous	Complete load, Meters (m)
	High-speed running distance (HSRD)	Continuous	19-23 km/h
	Sprint distance (SpD)	Continuous	>23 km/h
	Peak speed ($Peak_{speed}$)	Continuous	Highest speed achieved during training
Fixed	Position	Nominal	Central defender (CD), Wide defender (WD), Central midfielder (CM), Wide midfielder (WM), Striker (S)
	In match squad	Nominal	Playing status (starter, sub, unused sub)
	Match	Nominal	1 = Match, 0 =Training
	Minutes played Last match	Continuous	Number of minutes played last match
	Session duration	Continuous	The length of the training session (min)
Random	Player ID	Nominal	Unique ID

2.3 Data cleaning

Data collected with fewer than 12 satellites, HDOP higher than 5, session duration over 210 and less than 1 minute, and sprint distance greater than 4000 meters were treated as missing. These thresholds were chosen based on visually inspecting the histograms of each variable for inaccurate and unlikely data. In addition, RPE values equal to 0 were treated as missing since PMsys currently does not have a dedicated missing value. Positions were grouped into central defender (CD), wide defender (WD), central midfielder (CM), wide midfielder (WM) and striker (S). Goalkeepers were excluded from the analysis.

Following data cleaning, 61.22% of reported duration, RPE, and sRPE values were missing, along with 62.09% of average absolute acceleration, the mean number of satellites, mean HDOP, and session duration values. Additionally, 62.2% of SpD, HSRD, TD, and $Peak_{speed}$ values were missing. These missing values were removed using complete case analysis (CCA), leaving a final sample of 4311 training observations from 58 athletes.

2.4 Statistical analysis

An LMM analysis was used to investigate associations between RPE and sRPE and selected independent variables. After the data cleaning process was completed, the dataset was divided into a training set (70% of sessions) and a holdout set (30% of sessions). The full model consisted of session duration, TD, HSRD, SpD, $Peak_{speed}$, squad status (starter, substitute, or unused in current or last match), playing position (in current or last match), and minutes played in current or last match. Backward elimination with a significance level of 0.05 was used to remove non-significant predictors. The final model consisted of session duration, TD, HSRD, SpD, $Peak_{speed}$, starters, substitutes and unused substitutes.

As RPE and GPS data were nested within players (multiple training sessions for each player), and players were nested within teams, the data are hierarchical. This creates dependent RPE and sRPE scores within players. The chosen method can handle such multilevel data. RMSE and MAE were computed for the observed RPE and sRPE values and values predicted by the final model to evaluate predictive accuracy. An "explained" variance measure, a conditional R^2, was computed for the same observed and predicted values (Nakagawa & Schielzeth, 2013). Moreover, the intraclass correlation (ICC) was used to quantify the proportion of variance in RPE and sRPE data that can be attributed to differences between individual players (Nakagawa et al., 2017). Additionally, RMSE, MAE, ICC and R^2 were computed on the holdout set to evaluate how the model performed on unseen data.

3 Results

The models accurately predicted the RPE and sRPE from the training set, with RMSE, MAE, and R^2 values shown in Table 1. These models accounted for 65% (RPE) and 66% (sRPE) of the variability in the training set. Predictive performance was slightly lower on the holdout set (Table 1), with R2 values of 53% (RPE) and 60% (sRPE).

The final model indicated that session duration, TD, HSRD, SpD, $Peak_{speed}$, starters, and substitute variables were significantly associated with sRPE. For the RPE model, all variables mentioned, except peak speed, showed significant associations. Additionally, the unused substitute variable was significantly related to RPE.

The predicted sRPE values deviate by 60.70 AU within individual players (residual) and 143.56 AU between different players. For the RPE model, the predicted RPE values deviate by 0.72 AU within individual players (residual) and by 1.20 AU between different players.

Between-player variance suggested that players might respond differently to the same training stimuli. The ICC revealed that 84.79% of the total variance in sRPE responses could be attributed to between-player differences, while 15.21% was due to within-player differences or variability between training sessions. The high ICC indicates that a substantial proportion of the total variability in sRPE responses is attributed to player differences. Moreover, the ICC for the RPE model was approximately 73.44%, suggesting moderate to high consistency in players' RPE responses across training sessions.

Table 2: *The performance of the training and holdout set for both RPE and sRPE models. Abbreviations: RMSE = Root mean squared error, MAE = Mean absolute error,* R^2 *= Coefficient of determination.*

	RMSE	**MAE**	$\mathbf{R^2}$
sRPE Train	142.2	104.9	0.66
sRPE Holdout	158.7	112.9	0.60
RPE Train	1.18	0.93	0.65
RPE Holdout	1.29	0.98	0.53

Table 3: *Full sRPE model coefficients. Abbreviations: std. error = standard errors, df = degrees of freedom, T = T-values, P = significance level, Std. dev = standard deviations, ICC = intraclass correlation coefficient.*

Fixed effect	**Estimate ± std. error**	**df**	***T***	***P***
Intercept	236.2 ± 36.65	1.447	6.443	< 0.001
Session duration	-55.83 ± 23.11	2.121	2.416	0.0158
TD	-735.4 ± 23.015	2.121	23.015	< 0.001
HSRD	-22.9 ± 5.960	2.121	5.960	< 0.001
SpD	10.19 ± 6.039	2.091	-6.039	< 0.001
$Peak_{speed}$	-29.24 ± 5.184	2.117	-5.640	< 0.001
Starter	81.18 ± 14.57	2.120	-5.573	< 0.001
Substitute	27.82 ± 9.552	1.592	2.913	0.0036
Unused substitute	26.97 ± 17.86	2.088	1.510	0.1312
Random effects	**Variance component**	**Std. dev**	**ICC**	
Residual	3695	60.79		
Between-players	20609	143.56	0.8479	

Table 4: *Full RPE model coefficients. Abbreviations: std. error = standard errors, df = degrees of freedom, T = T-values, P = significance level, Std. dev = standard deviations, ICC = intraclass correlation coefficient.*

Fixed effect	**Estimate**	**df**	***t***	***P***
Intercept	3.842 ± 0.304	811	12.620	< 0.001
TD	0.0004765 ± 0.00001651	2.109	28.869	< 0.001
HSRD	0.002056 ± 0.000318	2.114	6.464	< 0.001
SpD	0.0008421 ± 0.0001406	2.090	-5.990	< 0.001
$\mathbf{Peak_{speed}}$	-0.1287 ± 0.0434	2.109	2.966	0.0030
Starter	0.6371 ± 0.0825	2.104	7.723	< 0.001
substitute	0.4925 ± 0.0815	2.121	6.046	< 0.001
Unused substitute	0.4789 ± 0.1503	1.984	3.186	0.0015
Random effects	**Variance component**	**Std. dev**	**ICC**	
Residual	0.5196	0.7208		
Between-players	1.4374	1.1989	0.7344	

4 Discussion

This study aimed to investigate the relationships between external load indicators (TD, HSRD, SpD, and $Peak_{speed}$) and RPE and sRPE for female football players using LMM, identify which external load indicator contributes the most to RPE and sRPE, and provide practitioners with a simple yet effective method for predicting RPE and sRPE. Our findings revealed that the final LMM demonstrated accurate values of R^2 (0.65 and 0.66), RMSE (1.18 and 142.2) and MAE (0.93 and 104.9) for RPE and sRPE, respectively (Table 1). This suggests that a few external load variables can accurately predict RPE and sRPE for female players.

Our results align with previous research on predicting RPE for male football players (Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022). However, comparing findings is difficult because GPS devices from different manufacturers are used in other studies. Furthermore, differences in training content and statistical approaches may explain differences between studies.

The aforementioned studies have used more advanced analytical approaches (Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022) and many external load variables to improve model accuracy (Geurkink et al., 2019; Jaspers et al., 2018). While incorporating more variables might lead to increased accuracy, it is essential to consider the practicality and interpretability of these models, ensuring that practitioners can effectively use these tools alongside their expert knowledge to make informed decisions (Malone et al., 2020; van Royen et al., 2022).

In contrast to Geurkink et al. (2019) and Jaspers et al. (2018), we argue that the number of variables included is less relevant and that models should be created using variables that are both predictive and actionable in a practical context, as others have proposed (Akenhead & Nassis, 2016; Marynowicz et al., 2022; van Royen et al., 2022). While perhaps not as sophisticated, our model still demonstrated arguably acceptable predictive ability sacrificing some accuracy in return for a more practical approach, which might provide practitioners with a simple tool to understand the dose-response relationship between external load measures and RPE and sRPE.

However, these variables were prescreened and not included in the full model due to the mentioned limitation regarding acceleration and deceleration. This contrasts with previous research (Askow et al., 2021; Douchet et al., 2021; Gaudino et al., 2015; Geurkink et al., 2019; Jaspers et al., 2018; Marynowicz et al., 2022) and should be interpreted with caution. Acceleration and deceleration are known to impose considerable physiological and mechanical loading demands, which are important factors when monitoring players' training load (Bloomfield et al., 2007; Dalen et al., 2021; Dello Iacono et al., 2023; Douchet et al., 2021; Osgnach et al., 2010; Verheul et al., 2021).

The high ICC value for the sRPE model (84.79%) indicates that a substantial proportion of the total variability in sRPE responses can be attributed to player differences. Similarly, the ICC value for the RPE model (73.44%) suggests a moderate to high level of variability between players' RPE responses across training sessions. This highlights the importance of considering individual differences when interpreting and applying RPE and sRPE data, which is consistent with the theoretical model of Impellizzeri et al. (2005) and prior prediction research on male football players (Geurkink et al., 2019; Marynowicz et al., 2022). However, this finding contrasts with Jaspers et al. (2018), which found group models to have better predictive performance than individual models.

4.1 Limitations

Our study has some limitations that should be acknowledged. Firstly, a significant proportion of the data was missing (RPE and sPRE 61.22%, session duration 62.09%, HSRD, SpD, TD and $Peak_{speed}$ 62.09%). The high number of missing values can be attributed to a lack of GPS devices, injuries, illness, and days off granted to players. Moreover, there were more observations in the 2021 season compared to the 2020 season due to the COVID-19 pandemic.

We used a CCA to deal with the missing data, which has notable limitations. First, it can lead to a substantial loss of statistical power and reduced generalisability if a large proportion of cases are removed due to missing data (Graham, 2009). This may result in less precise estimates and decreased model prediction reliability. Second, if data are not MCAR but instead are MAR, CCA may introduce bias in the parameter estimates, as the observed data no longer accurately represent the underlying population (Graham, 2009; Schafer & Graham, 2002).

Due to the amount of missing data and the use of CCA, the sample size for prediction was reduced. Consequently, the model demonstrates overfitting, as evidenced by the diminished performance on the holdout set compared to the training set. Therefore, the prediction model is likely to have limited generalisability and perform inadequately on individuals not present in the dataset used for model development (Newans et al., 2022). Further research could improve missing data treatment by using imputation techniques, which allows for unbiased and statistical valid missing data handling (Bhaskaran & Smeeth, 2014; Greenland & Finkle, 1995; Sterne et al., 2009).

In addition, the uneven distribution of player positions in the dataset, particularly the larger number of midfielders, may introduce bias in the predictive accuracy of the models, as they may be more accurate in estimating RPE and sRPE for midfielders than other positions.

While our model provides accurate estimations of RPE and sRPE (Tables 3 and 4), it does not capture the full complexity of the relationship between external and internal loads. Incorporating acceleration and deceleration metrics could have improved the model's accuracy, offering a more comprehensive understanding of the relationship between external and internal training loads.

However, acceleration and deceleration were not included due to the limited consistency in calculating these metrics across different GPS systems. Future research may benefit from exploring the trade-off between model simplicity and accuracy and the potential effects of including additional external load variables on prediction outcomes.

Lastly, future research should investigate disparities in perceived training load among different positions in each training session, considering the activities and drills performed. Additionally, incorporating wellness questionnaires can improve insights into within-player variability, providing practitioners with a more comprehensive understanding of how athletes' physiological and psychological states influence perceived training load (Gallo et al., 2016; Saw et al., 2016).

4.2 Practical implications

The prediction model developed in our study provides a parsimonious yet accurate approach to predicting RPE and sRPE values based on GPS-derived data. The model could provide practitioners with information on individual training responses and thus help tailor individual training sessions, balancing training intensity and recovery. By monitoring training load and perceived exertion, practitioners can make more informed decisions about training programs and recovery strategies.

Furthermore, our model assists in identifying potential risk factors for injury or overtraining by detecting anomalies in RPE and sRPE, allowing proactive intervention to mitigate these risks. This methodology helps understand which practices generate higher exertion levels, how players adapt to the proposed loads, and potentially reduces the risk of injuries. The practicality and accuracy of our approach make it a valuable tool for managing athletes' internal load and enhancing their performance and well-being.

5 Conclusion

To the best of our knowledge, this is the first study to predict RPE and sRPE values using GPS-derived data in female football players. The potential for predicting RPE and sRPE values through GPS-derived data shows promising results, although future studies should include more extensive data collection, longitudinal monitoring, and better missing data treatment. In conclusion, in this specific context, the accurate prediction of RPE and sRPE values using GPS-derived data could contribute to more effective training load management in female football players.

References

Akenhead, R., & Nassis, G. P. (2016). Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *International Journal of Sports Physiology and Performance*, *11*(5), 587-593. https://doi.org/10.1123/ijspp.2015-0331

Askow, A. T., Lobato, A. L., Arndts, D. J., Jennings, W., Kreutzer, A., Erickson, J. L., Esposito, P. E., Oliver, J. M., Foster, C., & Jagim, A. R. (2021). Session Rating of Perceived Exertion (sRPE) Load and Training Impulse Are Strongly Correlated to GPS-Derived Measures of External Load in NCAA Division I Women's Soccer Athletes. *J Funct Morphol Kinesiol*, *6*(4). https://doi.org/10.3390/jfmk6040090

Baptista, I., Winther, A. K., Johansen, D., Randers, M. B., Pedersen, S., & Pettersen, S. A. (2022). The variability of physical match demands in elite women’s football. *Science and Medicine in Football*, 1-7. https://doi.org/10.1080/24733938.2022.2027999

Beato, M., Coratella, G., Stiff, A., & Iacono, A. D. (2018). The Validity and Between-Unit Variability of GNSS Units (STATSports Apex 10 and 18 Hz) for Measuring Distance and Peak Speed in Team Sports [Original Research]. *Frontiers in Physiology*, *9*(1288). https://doi.org/10.3389/fphys.2018.01288

Beato, M., & de Keijzer, K. L. (2019). The inter-unit and inter-model reliability of GNSS STATSports Apex and Viper units in measuring peak speed over 5, 10, 15, 20 and 30 meters. *Biology of sport*, *36*(4), 317-321. https://doi.org/10.5114/biolsport.2019.88754

Bhaskaran, K., & Smeeth, L. (2014). What is the difference between missing completely at random and missing at random? *Int J Epidemiol*, *43*(4), 1336-1339. https://doi.org/10.1093/ije/dyu080

Bloomfield, J., Polman, R., & O'Donoghue, P. (2007). Physical demands of different positions in FA Premier League soccer. *Journal of sports science & medicine*, *6*(1), 63.

Bradley, P., & Scott, D. (2020, 6. july 2020). *Physical abalysis of France 2019 shows increase in speed and intensity* Fédération Internationale de Football Association (FIFA) https://www.fifa.com/tournaments/womens/womensworldcup/france2019/news/physical-analysis-of-france-2019-shows-increase-in-speed-and-intensity

Brocherie, F., & Beard, A. (2020). All Alone We Go Faster, Together We Go Further: The Necessary Evolution of Professional and Elite Sporting Environment to Bridge the Gap Between Research and Practice. *Front Sports Act Living*, *2*, 631147. https://doi.org/10.3389/fspor.2020.631147

Bullock, G. S., Hughes, T., Arundale, A. H., Ward, P., Collins, G. S., & Kluzek, S. (2022). Black Box Prediction Methods in Sports Medicine Deserve a Red Card for Reckless Practice: A Change of Tactics is Needed to Advance Athlete Care. *Sports Med*, *52*(8), 1729-1735. https://doi.org/10.1007/s40279-022-01655-6

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine*, *13*(1), 1. https://doi.org/10.1186/s12916-014-0241-z

Dalen, T., Sandmæl, S., Stevens, T. G. A., Hjelde, G. H., Kjøsnes, T. N., & Wisløff, U. (2021). Differences in Acceleration and High-Intensity Activities Between Small-Sided Games and Peak Periods of Official Matches in Elite Soccer Players. *The Journal of Strength & Conditioning Research*, *35*(7), 2018-2024. https://doi.org/10.1519/jsc.0000000000003081

Dello Iacono, A., McLaren, S. J., Macpherson, T. W., Beato, M., Weston, M., Unnithan, V. B., & Shushan, T. (2023). Quantifying Exposure and Intra-Individual Reliability of

High-Speed and Sprint Running During Sided-Games Training in Soccer Players: A Systematic Review and Meta-analysis. *Sports Medicine*, *53*(2), 371-413. https://doi.org/10.1007/s40279-022-01773-1
Douchet, T., Humbertclaude, A., Cometti, C., Paizis, C., & Babault, N. (2021). Quantifying Accelerations and Decelerations in Elite Women Soccer Players during Regular In-Season Training as an Index of Training Load. *Sports*, *9*(8), 109. https://www.mdpi.com/2075-4663/9/8/109
FIFA. (2022, 16. december). *Women's Football: Accelerate the growth of women's football* https://www.fifa.com/womens-football
Foster, C., Boullosa, D., McGuigan, M., Fusco, A., Cortis, C., Arney, B. E., Orton, B., Dodge, C., Jaime, S., Radtke, K., van Erp, T., de Koning, J. J., Bok, D., Rodriguez-Marroyo, J. A., & Porcari, J. P. (2021). 25 Years of Session Rating of Perceived Exertion: Historical Perspective and Development. *International Journal of Sports Physiology and Performance*, *16*(5), 612-621. https://doi.org/10.1123/ijspp.2020-0599
Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A new approach to monitoring exercise training. *J Strength Cond Res*, *15*(1), 109-115.
Gallo, T. F., Cormack, S. J., Gabbett, T. J., & Lorenzen, C. H. (2016). Pre-training perceived wellness impacts training output in Australian football players. *J Sports Sci*, *34*(15), 1445-1451. https://doi.org/10.1080/02640414.2015.1119295
Gaudino, P., Iaia, F. M., Strudwick, A. J., Hawkins, R. D., Alberti, G., Atkinson, G., & Gregson, W. (2015). Factors influencing perception of effort (session rating of perceived exertion) during elite soccer training. *Int J Sports Physiol Perform*, *10*(7), 860-864. https://doi.org/10.1123/ijspp.2014-0518
Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
Geurkink, Y., Vandewiele, G., Lievens, M., de Turck, F., Ongenae, F., Matthys, S. P. J., Boone, J., & Bourgois, J. G. (2019). Modeling the Prediction of the Session Rating of Perceived Exertion in Soccer: Unraveling the Puzzle of Predictive Indicators. *International Journal of Sports Physiology and Performance*, *14*(6), 841-846. https://doi.org/10.1123/ijspp.2018-0698
Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annu Rev Psychol*, *60*, 549-576. https://doi.org/10.1146/annurev.psych.58.110405.085530
Greenland, S., & Finkle, W. D. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol*, *142*(12), 1255-1264. https://doi.org/10.1093/oxfordjournals.aje.a117592
Gualtieri, A., Rampinini, E., Dello Iacono, A., & Beato, M. (2023). High-speed running and sprinting in professional adult soccer: Current thresholds definition, match demands and training strategies. A systematic review [Review]. *Frontiers in Sports and Active Living*, *5*. https://doi.org/10.3389/fspor.2023.1116293
Hennessy, L., & Jeffreys, I. (2018). The Current Use of GPS, Its Potential, and Limitations in Soccer. *Strength & Conditioning Journal*, *40*(3), 83-94. https://doi.org/10.1519/ssc.0000000000000386
Iannaccone, A., Conte, D., Cortis, C., & Fusco, A. (2021). Usefulness of Linear Mixed-Effects Models to Assess the Relationship between Objective and Subjective Internal Load in Team Sports. *Int J Environ Res Public Health*, *18*(2). https://doi.org/10.3390/ijerph18020392
Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2019). Internal and External Training Load: 15 Years On. *Int J Sports Physiol Perform*, *14*(2), 270-273. https://doi.org/10.1123/ijspp.2018-0935

Impellizzeri, F. M., Rampinini, E., Coutts, A. J., Sassi, A., & Marcora, S. M. (2004). Use of RPE-based training load in soccer. *Med Sci Sports Exerc*, *36*(6), 1042-1047. https://doi.org/10.1249/01.mss.0000128199.23901.2f

Impellizzeri, F. M., Rampinini, E., & Marcora, S. M. (2005). Physiological assessment of aerobic training in soccer. *J Sports Sci*, *23*(6), 583-592. https://doi.org/10.1080/02640410400021278

Jaspers, A., De Beéck, T. O., Brink, M. S., Frencken, W. G. P., Staes, F., Davis, J. J., & Helsen, W. F. (2018). Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn From Machine Learning? *International Journal of Sports Physiology and Performance*, *13*(5), 625-630. https://doi.org/10.1123/ijspp.2017-0299

Malone, J. J., Barrett, S., Barnes, C., Twist, C., & Drust, B. (2020). To infinity and beyond: the use of GPS devices within the football codes. *Science and Medicine in Football*, *4*(1), 82-84. https://doi.org/10.1080/24733938.2019.1679871

Malone, J. J., Di Michele, R., Morgans, R., Burgess, D., Morton, J. P., & Drust, B. (2015). Seasonal training-load quantification in elite English premier league soccer players. *Int J Sports Physiol Perform*, *10*(4), 489-497. https://doi.org/10.1123/ijspp.2014-0352

Marynowicz, J., Lango, M., Horna, D., Kikut, K., & Andrzejewski, M. (2022). Predicting ratings of perceived exertion in youth soccer using decision tree models [journal article]. *Biology of sport*, *39*(2), 245-252. https://doi.org/10.5114/biolsport.2022.103723

McCall, A., Fanchini, M., & Coutts, A. J. (2017). Prediction: The Modern-Day Sport-Science and Sports-Medicine "Quest for the Holy Grail". *Int J Sports Physiol Perform*, *12*(5), 704-706. https://doi.org/10.1123/ijspp.2017-0137

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R(2) and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J R Soc Interface*, *14*(134). https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142. https://doi.org/https://doi.org/10.1111/j.2041-210x.2012.00261.x

Newans, T., Bellinger, P., Drovandi, C., Buxton, S., & Minahan, C. (2022). The Utility of Mixed Models in Sport Science: A Call for Further Adoption in Longitudinal Data Sets. *International Journal of Sports Physiology and Performance*, 1-7. https://doi.org/10.1123/ijspp.2021-0496

Osgnach, C., Poser, S., Bernardini, R., Rinaldo, R., & di Prampero, P. E. (2010). Energy cost and metabolic power in elite soccer: a new match analysis approach. *Med Sci Sports Exerc*, *42*(1), 170-178. https://doi.org/10.1249/MSS.0b013e3181ae5cfd

Park, L. A., Scott, D., & Lovell, R. (2019). Velocity zone classification in elite women's football: where do we draw the lines? *Science and Medicine in Football*, *3*(1), 21-28.

Randell, R. K., Clifford, T., Drust, B., Moss, S. L., Unnithan, V. B., De Ste Croix, M. B. A., Datson, N., Martin, D., Mayho, H., Carter, J. M., & Rollo, I. (2021). Physiological Characteristics of Female Soccer Players and Health and Performance Considerations: A Narrative Review. *Sports Med*, *51*(7), 1377-1399. https://doi.org/10.1007/s40279-021-01458-1

Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F. M., & Baca, A. (2023). Machine learning application in soccer: a systematic review. *Biology of sport*, *40*(1), 249-263. https://doi.org/10.5114/biolsport.2023.112970

Saw, A. E., Main, L. C., & Gastin, P. B. (2016). Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a

systematic review. *Br J Sports Med*, *50*(5), 281-291. https://doi.org/10.1136/bjsports-2015-094758
Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychol Methods*, *7*(2), 147-177.
Seshadri, D. R., Thom, M. L., Harlow, E. R., Gabbett, T. J., Geletka, B. J., Hsu, J. J., Drummond, C. K., Phelan, D. M., & Voos, J. E. (2021). Wearable Technology and Analytics as a Complementary Toolkit to Optimize Workload and to Reduce Injury Burden [Review]. *Frontiers in Sports and Active Living*, *2*. https://doi.org/10.3389/fspor.2020.630576
Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj*, *338*, b2393. https://doi.org/10.1136/bmj.b2393
Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., & Gregson, W. (2015). Monitoring Fatigue During the In-Season Competitive Phase in Elite Soccer Players. *Int J Sports Physiol Perform*, *10*(8), 958-964. https://doi.org/10.1123/ijspp.2015-0004
van Royen, F. S., Moons, K. G. M., Geersing, G. J., & van Smeden, M. (2022). Developing, validating, updating and judging the impact of prognostic models for respiratory diseases. *Eur Respir J*, *60*(3). https://doi.org/10.1183/13993003.00250-2022
Verheul, J., Nedergaard, N. J., Pogson, M., Lisboa, P., Gregson, W., Vanrenterghem, J., & Robinson, M. A. (2021). Biomechanical loading during running: can a two mass-spring-damper model be used to evaluate ground reaction forces for high-intensity tasks? *Sports biomechanics*, *20*(5), 571-582.
Winther, A. K., Baptista, I., Pedersen, S., Randers, M. B., Johansen, D., Krustrup, P., & Pettersen, S. A. (2022). Position specific physical performance and running intensity fluctuations in elite women's football. *Scand J Med Sci Sports*, *32 Suppl 1*, 105-114. https://doi.org/10.1111/sms.14105