MASTERARBEIT | MASTER'S THESIS

Titel | Title

Preventable Defensive Dangerous Accessible Space - Advanced Metric for Defensive Performance in Football

verfasst von | submitted by

Maximilian Nicolas Hartmann

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 926
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Wirtschaftsinformatik
Betreut von | Supervisor:	ao. Univ.-Prof. i.R. tit. Univ.-Prof. Dipl.-Ing. Dr. Erich Schikuta

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Jonas Bischofberger, MSc, for his continuous support, insightful feedback and encouragement throughout the entire process.

I would also like to thank Univ.-Prof. Dipl.-Ing. Dr. Erich Schikuta for the opportunity to write this thesis and for his valuable guidance.

Furthermore, I would like to thank my family and friends for their unwavering support and motivation. I am also grateful to my colleagues for their understanding and encouragement.

Finally, I would like to give a very special thanks to my father, Dr. Marco Hartmann-Rüppel, for the time and effort in proofreading this thesis.

Abstract

Defensive performance in football is inherently difficult to quantify, as successful defending is often characterized by positioning, anticipation and collective behavior rather than observable actions. Traditional defensive metrics therefore tend to misrepresent a player's true defensive contribution. This thesis proposes a novel framework for the evaluation of defensive performance based on tracking data. For each defensive frame, player positions are discretely optimized under realistic spatial constraints, such as limited displacement and minimum distance to teammates, in order to minimize the opponent's Dangerous Accessible Space (DAS). The difference between the observed and optimized DAS values is used to derive a performance metric that quantifies positional efficiency, the Preventable Defensive DAS (PDD). The framework is applied to a set of professional match data from the German *Bundesliga*. Empirical results show that the proposed metric captures substantial variation in defensive performance across players and positions and reveals positional inefficiencies that are not reflected in conventional defensive statistics. On average, optimized positioning reduced the opponent's DAS by 6.88%, indicating measurable positional inefficiencies in observed defensive behavior. Overall, this thesis contributes to football analytics by introducing an advanced, data-driven measure of defensive performance that provides a more nuanced assessment of player contributions.

Kurzfassung

Die Verteidigungsleistung im Fußball lässt sich nur sehr schwer bemessen, da sich ein erfolgreiches defensives Spiel oftmals durch eine gute Positionierung, das Vorausahnen von Spielsituationen und kollektives Verhalten auszeichnet und weniger durch einzelne und klar definierte Aktionen am Ball. Klassische Metriken haben daher die Tendenz, den tatsächlichen Anteil eines Spielers am defensiven Ergebnis falsch einzuschätzen. Diese Arbeit stellt ein neues, auf Tracking-Daten basierendes Framework für die Bewertung der Defensivleistung vor. Für jeden Frame, in dem ein Verteidiger aktiv ist, wird seine Positionierung mit dem Ziel optimiert, das Dangerous Accessible Space (DAS) des Gegners zu minimieren. Mögliche optimierte Positionen werden dabei in einem räumlichen Kontext eingeschränkt, der realistische Spielerbewegungen sicherstellt. Der Unterschied zwischen den beobachteten und den optimierten DAS-Werten ist die Grundlage für die abgeleitete Metrik, Preventable Defensive DAS (PDD), welche die Qualität des Positionsspiels abbildet. Das Framework wird auf ein Datenset aus Spielen der deutschen *Bundesliga* angewendet. Die empirischen Ergebnisse zeigen, dass die Metrik substantielle Unterschiede zwischen Spielern und Positionen offenlegt und Aspekte des Positionsspiels erfasst, welche von bestehenden Metriken nicht erfasst werden. Durchschnittlich lässt sich das DAS des Gegners um 6,88 % reduzieren, woraus sich deutliche Potentiale in der Verbesserung des Positionsspiels ableiten lassen. Zusammenfassend trägt diese Arbeit eine neue datengetriebene Metrik bei, die die ganzheitliche Betrachtung und Beurteilung defensiver Performance fördert.

Contents

List of Tables

List of Figures

List of Algorithms

1. Introduction

Traditionally, football performance evaluation relied largely on subjective observation and expert judgment, both on and off the pitch. Coaches, scouts and analysts assessed players and tactical systems primarily through personal experience and qualitative judgment. While this approach allowed for rich contextual understanding, it lacked consistency, scalability and reproducibility (Rein and Memmert, 2016). Football nowadays has evolved from a cultural phenomenon into a global industry generating revenues in the multi-billion-dollar range (Fan et al., 2023). This implies the need for more systematic and objective performance evaluation methods for tactical as well as for squad planning purposes.

Over the past two decades, football has undergone a profound transformation driven by advances in data collection and computational analysis. Performance analysis is no longer limited to post-match observation but is embedded in tactical preparation, player development, recruitment and strategic decision-making. The availability of tracking data, which captures the precise field positions of every player throughout a match, has fundamentally changed the analytical possibilities in football. These developments have enabled the design of increasingly sophisticated methodologies and performance metrics to evaluate performance in football (Ehrenberg-Silies, 2024).

Within this evolving field of football analytics, **offensive performance** has received disproportionate attention. Metrics such as expected Goals (xG) or expected Threat (xT) (Singh, 2024) are now widely used in both academic research and professional practice to quantify attacking contributions. These models have significantly improved the understanding of how goals and promising attempts are created and which actions contribute most to scoring opportunities. In contrast, **defensive performance** metrics remain comparatively underexplored, despite empirical evidence highlighting the critical role of a strong defense for sustained success in football (Freitas et al., 2023).

This imbalance is particularly striking given the nature of the game. Football is inherently a low-scoring sport compared to other team sports, like basketball or American football, and individual goals can have a decisive impact on match outcomes. Preventing goals is therefore as important as scoring them, if not more important. Several studies show that the number of goals conceded and the effectiveness of defensive transitions are strongly associated with team success (Delgado-Bordonau et al., 2013; Kempe et al., 2016; Winter and Pfeiffer, 2016). Nevertheless, defensive contributions are often harder to identify, interpret and quantify than offensive actions.

One explanation for this discrepancy lies in the inherent complexity of defensive behavior. Unlike offensive actions such as shots or passes, many defensive contributions occur off-the-ball and aim to prevent potential actions rather than produce observable events. Effective defensive positioning may discourage a forward pass or delay an attack without producing a measurable action. As a consequence, an important part of defensive behavior remains invisible to traditional event-based data and metrics (Freitas et al., 2023; Santos and Lago-Penas, 2019; Wu and Swartz, 2023).

To address these challenges, recent research has explored more advanced modeling techniques, including machine-learning-based approaches. Supervised learning methods such as deep neural networks and gradient boosting have been applied to football data, predominantly event-based data, to predict outcomes or estimate action values (Baladram et al., 2020; Decroos et al., 2019; Merhej et al., 2024). While these approaches have demonstrated strong performance in offensive modeling, their application to defensive analysis remains limited. Defensive off-the-ball behavior lacks clearly defined labels and machine-learning models often function as black boxes, making their outputs difficult to interpret in tactical contexts.

At the same time, the use of tracking data introduces substantial computational challenges due to the high temporal resolution and dimensionality of the data. Consequently, recent work has focused on improving the computational efficiency by using algorithms such as LightGBM (Ke et al., 2017), which have proven to be promising in this domain with large-scale datasets. (Wu and Swartz, 2023).

The thesis addresses these limitations by adopting a different modeling perspective. Rather than predicting defensive actions or movement, the focus shifts to evaluating defensive positioning of football players. In this regard, defensive performance is conceptualized as a **spatial optimization problem** and based on tracking data. The central idea is to evaluate the quality of positioning of a defender relative to an optimal positioning in a given game situation. To this end, the offensive metric Dangerous Accessible Space (DAS), introduced by Bischofberger and Baca (2025), is extended and applied in a defensive context. DAS provides a spatial representation of the danger accessible to the attacking team by combining a position-based non-shot xG value field with the space that is accessible via pass for the attacking team.

By simulating alternative defensive positions and evaluating their impact on DAS, it becomes possible to determine an optimal defensive position for an individual player in a specific situation. The relative deviation between the player's actual position and this optimal position then serves as the basis for a novel defensive metric that captures off-the-ball behavior, i.e. the quality of defensive positioning. This thesis makes three main contributions: (i) the **design of a tracking-data-based defensive metric** that evaluates individual defensive positioning, (ii) the **implementation** of this metric in a fully reproducible and interpretable framework and (iii) its **empirical validation** using professional football tracking data. The proposed approach does not rely on labeled training data and remains easily interpretable from a tactical perspective, making it suitable for both academic and practical purposes.

The thesis is structured as follows. Chapter 2 provides a detailed overview of the theoretical background and related work in football analytics, with a particular focus on defensive performance evaluation and current gaps in research. Chapter 3 introduces the conceptual design and specification of the proposed defensive DAS optimization model, including data representation and the introduction of the resulting metric. Chapter 4 provides the implementation details, alternative design decisions and the evaluation framework of the model. In chapter 5, the results both from the metric itself as well as from the evaluation are presented, followed by the interpretation and integration in the context of the domain of football analytics, as well as a discussion of limitations and practical implications in chapter 6. Finally, chapter 7 concludes the thesis and outlines directions for future research.

2. Theoretical Background and Related Work

2.1. Football Analytics and Data Foundations

For a long time and still until now, football in Europe lagged behind other major sports such as baseball or American football with respect to data-driven performance analysis. One key reason for this delay was the limited availability of structured and publicly accessible data. In 2010, the *New York Times* referred to football as the "least statistical" of all major sports (Müller et al., 2017). Although major data providers such as *Opta* have since collected increasingly large datasets, data-driven decision-making has historically been underutilized by many professional football clubs, both in day-to-day tactical preparation and at the strategic management level, for example in player scouting or squad planning (Chatziparaskevas et al., 2024).

Another fundamental challenge arises from the nature of football itself. Unlike sports such as baseball or American football, which are characterized by clearly separated and well-defined plays, football is a continuous game with fluent transitions between different phases. Actions are often interdependent and difficult to isolate, making it inherently harder to quantify individual contributions (Ehrenberg-Silies, 2024). Furthermore, football is a highly interactive team sport, where the performance of a single player is strongly influenced by the behavior and positioning of teammates and opponents. This interdependence complicates the evaluation of individual performance compared to sports such as tennis, where actions can be attributed directly to individual players (Müller et al., 2017).

Despite the challenges, the adoption of data-driven approaches has increased substantially over the past decade. In particular, clubs in the English Premier League such as *Brighton & Hove Albion FC*, *Brentford FC* and *Liverpool FC* have demonstrated how systematic data analysis can support competitive advantages in recruitment, tactical preparation and match analysis (Chatziparaskevas et al., 2024). The development has coincided with major advances in data collection technologies, enabling increasingly detailed representations of match behavior (Ehrenberg-Silies, 2024).

From a tactical perspective, football can be divided into five distinct moments of play: **established attack**, **established defense**, **offensive transition**, **defensive**

transition and **set pieces** (Hewitt et al., 2016). These moments reflect the continuous flow of the game while allowing for a structured analytical framework. Among them, established defense and defensive transition are particularly relevant for defensive analysis, as they describe situations in which the opposing team is in possession of the ball. During these phases, the primary objectives of the defending team are to regain possession and to prevent the opponent from creating goal-scoring opportunities. Figure 2.1 shows the possible transitions between those phases, highlighting the unique application of set pieces, which cannot be classified as either offensive or defensive phase.

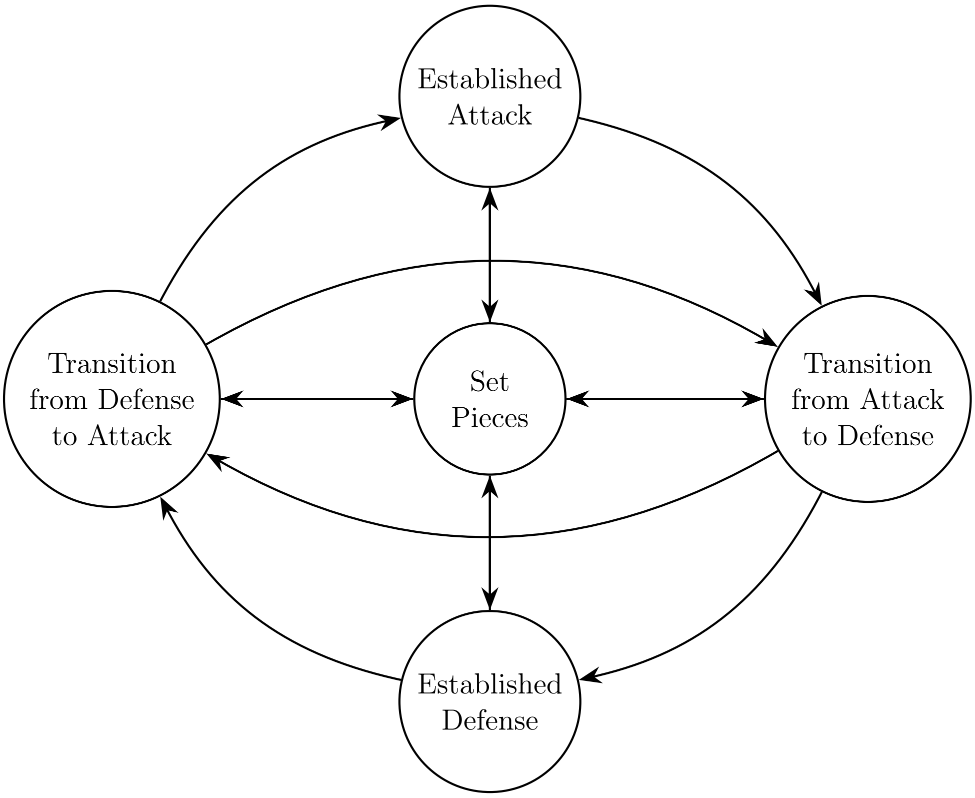

Figure 2.1.: Five moments of play

Given that football is inherently a low-scoring game compared to sports like American football or basketball, each goal scored or conceded can have a profound impact on a team's success or failure (Singh, 2024; Eigenrauch et al., 2024). This implies that it is difficult to score a goal and easier to simply prevent the opponent from scoring. Consequently, strong defensive performance is a critical determinant of a game's outcome. Empirical studies support this observation, showing that the number of goals conceded is strongly associated with team success (Kempe et al., 2016). Additional research highlights the critical role of defensive transitions and defensive stability, as well as the relevance of metrics such as shots on target conceded, in differentiating successful from unsuccessful teams (Delgado-Bordonau et al., 2013; Winter and Pfeiffer, 2016).

Despite this undeniable importance, defensive performance analysis remains underrepresented in both academic research and practical applications. One explanation for this imbalance lies in the inherent complexity of defensive behavior. In contrast to offensive actions such as shots, passes or dribbles, many defensive contributions occur off the ball and aim to prevent events rather than to produce observable outcomes. As a result, defensive actions are often less tangible and more difficult to measure (Wright et al., 2014). Moreover, offensive aspects of the game tend to receive greater public attention, as attacking players and goal-scoring actions are generally perceived as more entertaining. This bias is reflected, for example, in individual awards such as the Ballon d'Or, possibly the most important individual award in football, which has rarely been awarded to defensive players (UEFA.com, 2025; Freitas et al., 2023).

From a data perspective, football analytics predominantly relies on two types of data: **event data** and **tracking data** (Renkin et al., 2022). Event data consists of discrete, manually or semi-automatically annotated actions such as passes, tackles, interceptions or shots, each associated with a timestamp and contextual information. Due to its accessibility and relatively low computational requirements, the use of event data is well established and has laid the foundation of football analytics. However, a key limitation of event data is its inability to capture off-the-ball behavior and spatial relationships between players. Many essential defensive behaviors, such as positions to block passing lanes or maintaining team compactness, remain unmeasured (Wu and Swartz, 2023).

In contrast, tracking data captures the exact two-dimensional positions of all players and the ball at high temporal resolution (usually 25 frames per second). These data are obtained using advanced optical tracking systems or multi-camera setups (Piersma et al., 2020; Wu and Swartz, 2023; Bischofberger and Baca, 2025). Tracking data enables the analysis of movement dynamics, spatial relationships, velocities, accelerations, pitch control and the geometric structure of team formations. This richness makes tracking data particularly valuable for the analysis of defensive behavior and especially for off-the-ball behavior that cannot be captured by event-based approaches. At the same time, the use of tracking data introduces some significant challenges. The high dimensionality of the data, large file sizes, complex preprocessing requirements and substantial computational demands provide obstacles for both modeling and evaluation. Nevertheless, for research questions concerned with defensive positioning and spatial control, tracking data is not only advantageous but essential (Eigenrauch et al., 2024; Bischofberger et al., 2024).

2.2. Traditional Defensive Performance Metrics

Traditional approaches for the evaluation of defensive performance in football are primarily based on event data. Commonly used defensive metrics include tackles, interceptions, blocks, clearances and defensive duels won. These metrics are intuitive and easy to interpret, which explains their widespread use in both academic research and applied performance analysis (Merhej et al., 2024). They provide a first approximation of defensive involvement and allow for basic comparisons between players or teams.

Despite their practicality, traditional defensive metrics come with several well-documented limitations restricting their informational value and usefulness for a comprehensive evaluation of defensive performance. The most prominent issue with the absolute number of defensive events is the **possession bias**. Teams that tend to have a larger proportion of the game without ball possession are naturally exposed to more defensive situations and hence, accumulate a higher number of defensive actions. As a result, a higher number of tackles or interceptions may only reflect more defensive situations rather than superior defending quality (Piersma et al., 2020). However, some approaches have been proposed to mitigate possession bias, most notably by normalizing defensive actions with respect to ball possession or defensive playing time (Santolini, 2022). Such adjustments can improve comparability between teams but keep several further limitations. One of those limitations lies in the lack of contextual information. Traditional metrics typically treat all defensive actions equally, regardless of their location on the pitch, the game situation or the potential danger they prevent. For example, an interception near or in one of the penalty areas can have a significantly higher tactical impact than an interception in a midfield zone. However, in traditional metrics both events are counted equally. This absence of spatial and situational context makes it difficult to assess the true defensive contribution of individual actions (Merhej et al., 2024). Furthermore, traditional defensive metrics are inherently outcome-oriented. They only capture situations in which a defensive action results in an observable event. A very important aspect of defensive behavior, however, is to prevent dangerous situations from occurring in the first place and hence, not produce any measurable event. Examples for this include maintaining appropriate distance between defenders, closing passing lanes through intelligent positioning, applying pressure on the player in possession or forcing opponents into less dangerous areas of the pitch. These actions have a big impact on the opponents' decision-making but often do not produce an event. Consequently, they remain invisible in event-based datasets. This issue is particularly relevant for the evaluation of defensive positioning. Effective positioning often manifests through the absence of events rather than their presence. A well-positioned defender may discourage a forward pass, delay an attack or guide an opponent towards a less threatening option without directly interacting with the ball. Traditional metrics are structurally incapable of capturing such preventive contribution (Merhej et al., 2024; Piersma et al., 2020).

Other traditional defensive metrics include statistics such as conceded goals, conceded shots on target or off target and defensive effectiveness, which represents the percentage of goals conceded relative to the total shots faced. Although these metrics have been shown to be critical for optimizing success, as highlighted in prior studies, they do not provide deeper insights into the underlying defensive dynamics or the quality of specific defensive actions (Delgado-Bordonau et al., 2013; Kempe et al., 2016; Merhej et al., 2024).

In summary, while traditional defensive metrics provide a useful starting point, they are insufficient for a detailed and context-aware evaluation of defensive performance. Their reliance on event data, lack of spatial information and focus on observable outcomes limit their ability to capture off-the-ball behavior and positional defending. These limitations motivate the development of more advanced approaches that integrate spatial and situational context, continuous movement data and a deeper understanding of defensive decision-making. The following sections therefore introduce advanced approaches to address these challenges.

2.3. Advanced Performance Metrics and Modeling Paradigms

The limitations of traditional, event-based metrics have motivated the development of more advanced performance evaluation models in football analytics. A wide range of methods has emerged that aim to capture the value of actions in regard to contextual information and spatial representation of the pitch. These methods use various methodological foundations but share the goal of providing a more comprehensive view than traditional metrics.

One prominent paradigm of advanced performance metrics is based on the concept of **expected value**. Metrics that belong to this category are expected Goals (xG), expected Assists (xA), expected Threat (xT) or Expected Possession Value (EPV). The basic idea behind those metrics is to calculate the likelihood that a given action or game situation results in a favorable outcome, typically a goal (Singh, 2024; Bischofberger and Baca, 2025). These models have significantly improved the evaluation of offensive performance by assigning value to passes, carries and field position and not only shots. By modeling the game as a sequence of states with associated expected outcomes, they allow for a more nuanced understanding of attacking contribution (Mead et al., 2023). Despite their success in offensive analytics, those expected-value-based models are less adopted for measuring defensive performance. Some models implicitly capture defensive effects, for example by a reduced xG if a shot is taken under pressure. However, this does not allow to directly derive defensive contribution. This limits the ability to evaluate the contribution of individual defenders but also of the defensive team performance (Piersma et al., 2020).

In parallel to expected-value models, **machine-learning-based** approaches are becoming more popular in football analytics. Supervised learning techniques such as logistic regression, gradient boosting decision trees and neural networks are commonly used to predict outcomes like shot success, pass completion or action value. These methods are particularly effective for nonlinear relationships and a large number of contextual features (Jiang, 2021; Baladram et al., 2020). As a result, machine learning has become a standard tool for offensive modeling and action valuation. In the context of defensive analytics, however, machine-learning-based approaches face several challenges. They typically rely on labeled data, which is difficult to obtain for off-the-ball behavior because, unlike shots or passes, defensive positioning does not have a clearly observable outcome that could serve as a training label. Secondly, machine-learning models are normally black boxes, making the results difficult to interpret in tactical terms. This can be problematic in applied settings with analysts, scouts and coaches, which often require transparent explanations. Nonetheless, some promising approaches exist to make use of neural networks (Merhej et al., 2024) or gradient boosting trees (Decroos et al., 2019) for measuring defensive performance, which will be introduced in chapter 2.4.

A third paradigm focuses on **spatial and tracking-data-based** models. These methods leverage the continuous positional information provided by tracking data to analyze space occupation, dynamics of movement and the interaction between players. Examples for this are pitch control, spatial dominance metrics or evaluations of defensive behavior. By explicitly modeling space and movement, these approaches are better suited to capture off-the-ball behavior and defensive positioning (Wu and Swartz, 2023). Nevertheless, as all approaches mentioned, spatial models involve some trade-offs as well. They require strong modeling assumptions and computational resources. Moreover, many existing models are rather descriptive. They describe how space is controlled, but they do not address explicitly whether a defender is in an optimal position or how the position could be improved.

Taken together, the literature on advanced performance metrics reveals three broad modeling paradigms: event-based value models, machine-learning-based predictive approaches and tracking-data-based spatial models. While these categories in practice often overlap, for example many expected value models rely on machine learning techniques to estimate a value, they are distinguished by their primary focus and underlying data requirements. Each paradigm can contribute valuable insights, yet none fully resolves the challenge of evaluating defensive positioning. This observations motivates the need for alternative approaches that combine the strengths of more than one paradigm. The following section therefore reviews state-of-the-art defensive metrics in detail and highlights the approaches to address these challenges, setting the stage for the framework proposed in this thesis.

2.4. State of the Art in Defensive Analytics

This section reviews state-of-the-art approaches for advanced metrics in football in general and defensive performance in particular. Building upon the limitations of traditional metrics established in section 2.2 and modeling paradigms discussed in section 2.3, the focus lies on methods that aim to measure the defensive behavior. The approaches reviewed differ in terms of data requirements and usage, methodological design and the aspect of defensive performance they capture. For each method, both its contribution and its limitations are discussed in order to position the proposed approach of this thesis within the existing literature.

2.4.1. Common Advanced Metrics in Practice

While the metrics described in section 2.2 are relatively easy to calculate and limited in their significance, the evaluation of football performance has significantly improved by the development of advanced performance metrics, particularly on the offensive side. One of the most widely adopted metrics is the expected Goals mentioned in section 2.3. The metric assigns a probability between zero and one to each shot, representing the likelihood that the shot results in a goal, where 0 indicates no chance of scoring and 1 would be a guaranteed goal (Mead et al., 2023). Due to its intuitive interpretation and straightforward application, this metric has been rapidly adopted by professional football clubs, analysts and fans alike (Chatziparaskevas et al., 2024; Mead et al., 2023).

However, xG has notable limitations, most prominently that the metric only credits the player taking the shot. This limitation led to the introduction of expected Assists, which credits the player providing the final pass before a shot occurs (Singh, 2024). More comprehensive approaches are grouped under the term of Expected Possession Value models, with expected Threat being the most prominent example. These metrics assign values to field positions based on the probability that possession from those locations in combination with the positioning of all players will eventually lead to a goal (Bischofberger and Baca, 2025; Singh, 2024).

While those metrics have significant advancements compared to the traditional metrics for offensive performance evaluation some of them can also be adopted as defensive metrics. Metrics such as xG conceded or xT conceded are also very common. However, the primary focus of these metrics remains offensive. But the main weakness of those metrics in defensive analytics is that those values cannot be broken down to individual players. Therefore the defensive adoption of those metrics can only be used for defensive performance on team level.

The general lack of advanced defensive metrics in football is also highlighted by a meta-study from Freitas et al. (2023), which reveals the imbalance between evaluation metrics for offensive and defensive analytics in football. An example from practice further demonstrating this imbalance can be found at the German Football League (DFL), which is the organization behind the highest two divisions in German football. They developed a set of advanced metrics, so-called *Match Facts*. Among these, only one out of fifteen metrics, the *Ball Recovery Time*, can be classified as a defensive metric for field players, while seven are explicitly offensive metrics (Bundesliga, 2024).

2.4.2. Event-based Action Valuation Models

In response to the limitation of traditional defensive metrics and the dominance of offensive-focused value models, researchers have proposed methods that aim to quantify the value of defensive actions and behavior more directly. One type of approaches focuses on quantifying the value of defensive actions.

Defensive Action Expected Threat (DAxT) is one of those approaches introduced by Merhej et al. (2024). The central idea behind this approach is to estimate the attacking threat value that was prevented by a defensive action such as a tackle or an interception. To this end, the model aims to predict the hypothetical field position that the attacking team would have reached if the defensive team had not successfully intervened using a neural network. The value of the defensive action is then computed as the xT associated with the predicted field position. In this way, DAxT can effectively quantify the impact of defensive interventions.

Another noteworthy approach in the field of valuing defensive actions is the **Defensive Contribution Rating**. Decroos et al. (2019) propose an approach that measures the overall value of both defensive and offensive actions in a game. Therefore, they use the gradient boosting algorithm **CatBoost** (Prokhorenkova et al., 2019) to estimate the probability of scoring and conceding goals before and after each action in a match. The difference between these probabilities is interpreted as the value of the action, with positive contributions assigned when the probability of conceding a goal decreases or the probability of scoring a goal increases. By aggregating these action values, the model produces both offensive and defensive player ratings. It is therefore not an approach that primarily aims to measure defensive performance but contributes a valuable metric to defensive analytics in football.

Both models combine the paradigms of expected-value-based and machine-learning-based approaches and come with similar and notable limitations. First, both models rely exclusively on event data and do not incorporate tracking data. As a result, the spatial context of defensive situations is only captured indirectly. The value of the situations may not capture the specific context of certain situations, such as counterattacks. Second, both models only evaluate on-the-ball actions and successful defensive actions. Defensive positioning that discourages passes or limits attacking options without resulting in an explicit observable event, as well as unsuccessful tackles, remain unaccounted for (Merhej et al., 2024; Decroos et al., 2019). Another limitation in the approach from Decroos et al. (2019) for defensive evaluations is the imbalance between offensive and defensive action types. Only three, tackles, interceptions and clearances, out of 21 action types are classified as defensive actions for outfield players. Taken together, these limitations restrict the applicability of both approaches for evaluating defensive performance and defensive positioning in particular.

2.4.3. Pressure Metrics

In addition to event-based valuation models, another line of research in defensive analytics focuses on **pressure- and pressing-based metrics**. These approaches aim to describe, formalize and quantify the pressure that is applied to the player in possession by the defenders and the constraints the defenders set on the decision-making options of this player. Those metrics are mostly independent from concrete events and therefore address the limitation of event data-based models by incorporating aspects of off-the-ball behavior.

Early work has focused on the characteristics of pressure situations using tracking data. For instance, Andrienko et al. (2017) propose a visual analytics framework to explore defensive pressure by analyzing the proximity between players, movement directions and temporal dynamics around the ball. Their approach allows them to identify pressing zones and pressure patterns across the pitch. This enables qualitative insights into collective defensive behavior. However, this framework remains rather exploratory and does not define a scalar defensive performance metric that allows for systematic comparisons across players.

More recent work moves towards the detection and classification of pressing behavior and situations. Bauer and Anzer (2021) follow a machine-learning-based method to detect counter-pressing situations following ball losses. By modeling collective defensive movement patterns from tracking data, their approach allows for analysis of pressing frequency and structure at a large scale. While this represents a significant methodological advancement, the focus remains on the classification of defensive transition phases rather than the evaluation of defensive performance.

Other studies aim to quantify the **effectiveness of pressing**. One approach introduced by Merckx et al. (2021) is to model pressing as a trade-off between risk and reward. They combine tracking and event data to evaluate whether pressing leads to ball recoveries or exposes the defense to an increased threat. Their framework introduces an explicit measure for pressing effectiveness, capturing the outcome of such situations in probabilistic terms. Nevertheless, the evaluation is inherently ball-centric and limited to situations in which pressing is explicitly initiated. While this provides some valuable insights in the behavior of defenders that does not result in observable actions, it also excludes broader aspects of defensive positioning.

An important extension of pressure-based analysis was introduced with the **exPress** framework (Lee et al., 2025). The approach shifts the focus from the team level to individual player contribution within pressing situations by estimating the probability of successful ball recovery and assessing how player positioning influences this probability. By incorporating position adjustments, exPress provides valuable insights into the role of an individual defender during pressing situations. However, similar to the approach by Merckx et al., the evaluation remains restricted to situations with pressing contexts and does not allow for general evaluation of defensive positioning.

Taken together, pressure metrics constitute a substantial advancement in defensive analytics. They shift the analysis from discrete events to spatial relations between players and incorporate tracking data to capture pressing situations. However, the focus of those approaches is always ball-centric and narrowed to certain defensive situations. Most of those metrics focus on the pressure on team level and do not allow for the evaluation of individual contribution. As a result, while pressure metrics complement existing methods, they leave the gap of evaluating the individual defensive positioning in a holistic and situation-independent manner.

2.4.4. Off-the-Ball Evaluation

Besides pressure-based approaches, several studies aim to evaluate defensive performance by explicitly focusing on off-the-ball behavior. These approaches attempt to quantify defensive impact without relying on discrete events or certain situations. An important step towards incorporating off-the-ball defensive behavior in defensive analytics is presented by Wu and Swartz (2023), who propose a tracking-data-based approach for evaluating defensive movement. This approach shifts the focus towards the evaluation of continuous player movement using positional data, addressing a significant gap by the metrics presented above.

This approach utilizes the **LightGBM** (Ke et al., 2017), which is a faster and more efficient variant of the gradient boosting algorithm utilized by Decroos et al. (2019) mentioned above, to predict a two-dimensional velocity vector representing the average movement of a defensive player in a given situation. This predicted vector serves as a benchmark against which the actual movement of a player, derived from the tracking data, is compared. Players are rewarded for moving faster in the predicted direction and penalized for deviations from the predicted direction or a deficit in speed (Wu and Swartz, 2023). In contrast to DAxT and the defensive contribution rating, this method utilizes the third modeling paradigm of spatial and tracking-based data and combines that with a machine-learning-based approach.

While this approach represents a significant innovation by explicitly addressing off-the-ball behavior, it also introduces new challenges and limitations. Most notably, the predicted average movement is treated as optimal behavior. As a consequence, unconventional or context-specific defensive decisions that deviate from the average movement may be penalized, even if they are tactically effective. For example, a player with more speed might be able to choose a more effective direction than the average player but gets penalized for that deviation. Another limitation is that the importance of that movement is not considered. The resulting value is independent from the field position or game-specific contexts. Additionally, the reliance on supervised learning requires large labeled datasets and limits the interpretability of the outputs (Wu and Swartz, 2023; Jiang, 2021).

A complementary perspective is introduced by who propose a hybrid framework combining tracking and event data with an EPV model to identify spatial advantages created by off-the-ball positioning. Their approach evaluates whether players occupy positions that would increase the expected value of a possession if a pass were played and attributes defensive responsibility for allowing such advantageous situations. Although this method explicitly links positioning to value creation, defensive quality is assessed only in relation to realized passes that exploit spatial weaknesses. It remains the open question of how individual defensive positioning quality can be assessed in a holistic manner.

2.4.5. Dangerous Accessible Space

The **Dangerous Accessible Space (DAS)** framework, introduced by Bischofberger and Baca (2025), combines event-based and tracking data to evaluate football performance from a spatial perspective. The core idea of DAS is to simulate passes in all directions from a given game situation and to determine the area of the pitch that is accessible to the attacking team via pass. A location is considered accessible if a pass is more likely to reach an attacking player than to be intercepted by a defender or leave the field out of play.

To simulate the passing outcomes, the model builds upon the work of Spearman et al. (2017), treating every pass as a simple Markov process. The resulting accessible space from the simulated passes is then combined with a non-shot-xG metric, which is an enhanced xG model capturing also situations where no actual shot has occurred, for those field locations. By aggregating the non-shot-xG values of the accessible space, DAS quantifies the danger associated with a given game situation.

DAS provides an intuitive and theoretically grounded representation of spatial danger and has primarily been applied to offensive analysis. It allows for the evaluation of actions such as passes or carries by comparing DAS values before and after the action to derive a $\text{DAS}^{\text{gained}}$ (Bischofberger and Baca, 2025), which describes the increase in DAS that is achieved by a pass. By leveraging the combination of the paradigm of spatial and tracking-based modeling with the paradigm of expected-value-based modeling, this metric provides an innovative approach for evaluating off-the-ball behavior in general.

However, also the DAS framework provides some limitations. So far, it does only consider low passes. That means the accessible area might be inaccurate to some extent if it would only be accessible via high passes. Additionally it is computationally expensive, which makes it difficult to apply in large scale in practice (Bischofberger and Baca, 2025). Focusing on defensive evaluation, DAS can provide valuable insights into the defensive performance on the team level. Nevertheless, it does not offer granular analysis of individual player contribution. Hence, the model does not allow for a direct evaluation of individual defensive positioning.

This limitation highlights a key opportunity for further research. While the metric captures the spatial structure of attacking danger, it does not address the question of how defenders should position themselves to minimize that danger. This thesis builds upon this framework and extends it by introducing a defensive optimization perspective, which is described in detail in the following chapter.

2.5. Research Gap and Positioning of this Thesis

The preceding sections have demonstrated that, despite significant advances in football analytics, the evaluation of defensive performance, particularly with respect to off-the-ball behavior, remains an open research challenge. While traditional event-based metrics and modern machine-learning approaches have substantially improved the understanding of offensive performance, defensive analytics continues to lag behind both methodologically and conceptually.

Event-based defensive metrics, such as tackles or interceptions, provide only a partial and often biased view of defensive performance. As discussed in section 2.2, these metrics inherently focus on observable actions and therefore fail to capture defensive behavior that prevents dangerous situations from emerging in the first place. Since many high-quality defensive actions are positional rather than reactive, their contribution remains largely invisible in event-based datasets.

More advanced modeling approaches, including machine-learning-based methods, attempt to overcome some of these limitations. Metrics such as DAxT or Defensive Contribution Ranking improve upon traditional counting statistics by incorporating contextual information and probabilistic modeling. However, these approaches still rely predominantly on event data and thus remain limited to on-the-ball defensive actions. As a consequence, they are unable to assess defensive positioning that successfully discourages passes, closes space or forces attackers into suboptimal decisions without producing an explicit event.

Tracking-data-based approaches mark an important step towards addressing off-the-ball defensive behavior. The work of Wu and Swartz (2023), for example, uses machine learning to evaluate defensive movement patterns by comparing observed player trajectories to predicted average behavior. While this approach provides valuable insights into defensive motion, it introduces new challenges. In particular, the assumption that average behavior constitutes optimal behavior may penalize context-specific or unconventional defensive decisions. Moreover, the reliance on supervised learning requires large amounts of labeled data and results in models that are often difficult to interpret in tactical contexts.

Across the existing literature, a common limitation emerges: current defensive metrics either focus on discrete actions or aim to predict average behavior, but they do not explicitly address the question of **optimal defensive positioning**. In other words, while previous work evaluates what defenders did or what defenders typically do, it remains largely unexplored how defenders should position themselves in a given situation to minimize the opponent's attacking potential.

This thesis addresses this gap by adopting a fundamentally different modeling perspective. Instead of predicting defensive actions or movements, defensive performance is framed as a **spatial optimization problem**. Building upon the DAS framework, which integrates spatial accessibility with the expected threat values, this work extends an originally offensive model into a defensive context. The core idea is to determine, for a given game situation, the position of an individual defender that minimizes the dangerous space accessible to the opponent.

By simulating alternative defensive positions using tracking data and evaluating their impact on DAS, an explicit benchmark for optimal defensive positioning can be established. The deviation between a player's actual position and this optimized position then serves as the basis for a novel defensive metric. This approach offers several advantages over existing methods: it captures off-the-ball defensive behavior, does not require labeled training data and remains fully interpretable from a tactical perspective.

In summary, this thesis positions itself at the intersection of spatial modeling and defensive analytics. It complements existing event-based and machine-learning-based approaches by introducing an optimization-driven framework that directly quantifies the quality of defensive positioning. The following chapter builds upon this conceptual foundation and presents the design and implementation of the proposed defensive DAS optimization model.

3. Model Design

Building upon the theoretical background and the identified research gap discussed in the previous chapter, this chapter introduces the conceptual design and methodological framework of the proposed defensive performance metric. The primary objective of this work is to develop a tracking-data-based approach that allows the evaluation of individual defensive positioning by explicitly modeling optimal behavior. In contrast to existing event-based or machine-learning-driven defensive metrics, this approach frames defensive performance as a spatial optimization problem. Instead of analyzing defensive actions, the model evaluates how defenders should position themselves in a given situation in order to minimize the attacking potential for the opponent. This perspective enables the establishment of an explicit benchmark for optimal defensive positioning, which can be used to assess the quality of a player's actual positioning.

The chapter focuses on the design and specification of the defensive optimization model, including the underlying concept, data representation and metric definition. In doing so, section 3.1 outlines the research design and conceptual considerations that guided the development of the model. It discusses the overall objective, the scope of the analysis and alternative concepts that were considered. Section 3.2 describes the data basis and preprocessing steps required to apply the model, while section 3.3 presents the model itself in detail. And section 3.4 introduces the resulting metric based on the optimization model.

3.1. Research Design

The goal of this thesis is the development of an advanced metric for defensive performance in football. The metric should capture the quality of individual player positioning based on tracking data. In contrast to traditional defensive metrics that focus primarily on on-the-ball actions, the proposed approach aims to evaluate defensive contributions that comes from off-the-ball behavior and spatial control. These contributions manifest through the prevention of dangerous situations rather than through explicit defensive events. As a result, they are underrepresented in current event-based metrics.

To address this limitation, defensive positioning is conceptualized as a spatial optimization problem. The basic idea is to compare the observed position of a defender in a given situation with the optimal position for the defender. This requires the definition of an explicit benchmark for the optimal defensive positioning. For this purpose the DAS framework (Bischofberger and Baca, 2025) provides a suitable basis as it quantifies the danger associated with a given game situation by combining the spatial accessibility with a value of the opponent's attacking potential.

The analysis in this thesis is conducted at the level of individual players and individual game situations. Each situation is represented by a single frame of tracking data, containing the positional data of all players and the ball. This frame-based perspective allows for a precise spatial representation of the game state and enables the evaluation of defensive positioning independently of subsequent actions or outcomes. The focus on individual frames also ensures that the proposed method remains computationally tractable and interpretable.

Other modeling approaches have been considered during the conceptual development of the defensive metric for this thesis. Two promising approaches are:

- **Expected Interceptions**: This approach combines tracking and event data and compares the outcome of actual passes to the expected pass completion by the DAS-model. This provides a value that can be interpreted as an expected interception value. A higher value for expected interceptions may indicate better positioning. The expected interception rate could be compared with the actual interception rate to derive a second value that measures interception abilities.

- **Team Preventable Defensive DAS**: This approach is similar to the one pursued in this thesis. The idea is to measure the difference between the observed positioning of the defending team and the optimal positioning. The difference to the approach in this thesis is that only the positions of the attacking teams are locked in and the goal is to find the optimal positioning and formation for the whole defensive team.

While the first approach could deliver a comprehensive view of defensive actions, the focus would shift slightly away from off-the-ball behavior and defensive positioning. By definition, the metric could only include situations in which the attacking team has played a pass. Hence, situations in which the defending team might have prevented a pass by excellent positioning would be excluded. The second approach offers a very interesting angle on defensive play and tactics on a team level but at the same time provides a very complex problem to solve from a computational perspective. It is an optimization problem with multiple variables and even heuristic approaches would need too much time. Therefore, this remains an approach that might be considered in future work as further development.

The approach selected for this thesis focuses on the optimization of the position on the individual defender level. For a given frame, alternative defensive positions are simulated to find the optimal one, i.e. the position with minimal DAS-value for the opposing team.

By comparing the DAS value of the opposing team with the actual positioning of the player and the DAS value of the opposing team with the optimal positioning of a player in a given situation, the quality of position can be assessed. This value can be interpreted as **Preventable Defensive DAS (PDD)** by a player. This modeling choice represents a deliberate trad-off between informational value from the metric and feasibility from a computational perspective. By focusing on individual player positioning at the frame level, the metric is interpretable and captures relevant aspects of defensive behavior.

3.2. Data Basis and Preprocessing

As outlined in section 2.1, two types of data are generally available for football games. These are event-based data and tracking data. For the proposed defensive optimization model, the primarily relevant and essential data is tracking data. Since the quality of positioning always depends on the positioning of teammates, opponents and the ball, this information is needed. Event data, e.g. the moments of actual passes, can be used to determine which frames of the tracking data are going to be analyzed but is not mandatory. For the parameters used in this thesis the event data will not be utilized.

The dataset used for the model in this thesis is available as open-source data provided by *Sportec Solution*, a subsidiary of the German Football League (DFL) (Bassek et al., 2025), and made available for use in Python by the library *DataBallPy* (Oonk, 2025). This data contains tracking, event-based and meta-data of seven official matches from the first and second divisions in Germany in the 2022/23 season. The data itself is consistent and in the same format for all seven games, which makes it convenient for data processing. While these attributes make the dataset very suitable for the purpose of this thesis, the dataset is relatively small. However, not many open-source datasets are available that contain the required tracking data. This small sample size is one key limitation of this work and will be further discussed in the following chapters. The tracking dataset itself consists of positional and contextual information for all players and the ball at a sampling rate of 25 frames per second. Table 3.1 provides an overview of the columns contained in the dataset.

This raw data provided by the DFL and parsed by *DataBallPy* then needs to be enhanced to be usable for the DAS-model. Those enhanced columns are derived from the positional data and include the speed of the players and the player that is in possession of the ball.

However, the calculation of DAS is computational expensive. Therefore the dataset is reduced upfront. The sampling rate is set from 25 frames per second to 1 frame every 5 seconds. This sample rate is chosen because it makes the computation feasible in an acceptable time frame, while preserving enough information to capture the continuous positioning of the player. A further reduction of the sample would potentially lead to the

Column Name	Description	Datentyp	Einheit
frame	number of the frame (unique)	int	
datetime	timestamp of the frame	time	hh:mm:ss.ms
td_gametime	timestamp of the gametime	time	mm:ss
<player>_x, <player>_y	x- and y-coordinate for all players (two columns each)	float	meters
ball_x, ball_y	x- and y-coordinate the ball	float	meters
team_possession	team in possession	string	
ball status	ball in or out of the game	string	
<player>_vx, <player>_vy, <player>_velocity	velocity for all players (three columns each)	float	m/s
ball_vx, ball_vy, ball_velocity	velocity for the ball	float	m/s
player_possession	player with ball possession	string	

Table 3.1.: Overview Tracking Data

lack of important situation throughout the game. Other researchers use the same sample rate for their studies, e.g. Baouan (2025), indicating that this rate is a good compromise. In order to reduce the calculation of frames that cannot provide a valuable insight the dataset is further filtered to cover only those frames for which a single player in possession of the ball can be assigned for. This excludes frames, in which the ball is out of play or traveling in the air or on the ground as consequence of a pass or shot. The reasoning behind that is, that the DAS only considers the space which is accessible via pass and in those situations no player could play a pass to access the dangerous space. Therefore, the DAS value in those situations is not very meaningful. Another idea for the reduction of the dataset was to keep only those frames in which a pass actually happened but this was rejected for the same reasons already discussed in section 3.1. It might exclude frames, in which excellent positioning of the defender has prevented the occurrence of the pass. Hence, the metric would have a negative bias towards situations that offer good passing opportunities. An additional and obvious constraint for the frames that can be considered for each player, is that the opposing team needs to be in possession of the ball.

To make use of the DAS-model the datasets needs to be in a long-format, which means it consists of one row per frame and player. In contrast, the dataset provided by DataBallPy and used in the previous steps of the preprocessing pipeline, was in a wide format. This means it consists of one row per frame and one column with coordinates for each player. For this transformation the function provided directly in the DAS-library is used (Bischofberger, 2025). Tables 3.2 and 3.3 show the difference between the two formats and the result of the transformation.

Frame	A_x	A_y	B_x	B_y	C_x	C_y
1	10.2	5.0	15.8	4.7	11.8	10.4
2	10.6	4.3	16.1	4.0	12.8	10.3
3	11.0	3.5	16.4	3.0	9.8	7.4

Table 3.2.: Mock tracking data in wide format

Frame	Player	x-Position	y-Position
1	A	10.2	5.0
1	B	15.8	4.7
1	C	11.8	10.4
2	A	10.6	4.3
2	B	16.1	4.0
2	C	12.8	10.3
3	A	11.0	3.5
3	B	16.4	3.0
3	C	9.8	7.4

Table 3.3.: Mock tracking data in long format

With that data the DAS for the original positioning can be calculated and is added as a column to the corresponding frames. Having the DAS values for the original dataset, the dataframe is further filtered to only incorporate frames with a certain DAS threshold. This is done because the resulting metric is computed as a relative value and frames with very low DAS could destabilize the metric and are not very meaningful. At the same time, the restriction of frames enhances computational efficiency. After that step the resulting dataframe is passed to the defensive DAS model. Additional parameters that are required are the player to be optimized (future work may enhance the model to optimize several players at a time), a list of the frames that are passed with the Dataframe, a dataframe that includes the positions for the players 1 second before each frame and parameters used for fine-tuning of the optimization. The target variable is the new DAS value, which is ideally the optimal, hence minimal, one that the respective player could achieve with optimal positioning and the new coordinates of that optimal position.

3.3. Model: Defensive Dangerous Accessible Space

3.3.1. Problem Definition

The objective of the defensive optimization model is to evaluate the quality of individual defensive positioning by explicitly defining an optimal position in a given game situation. To this end, defensive positioning is formulated as a spatial optimization problem based on tracking data and the DAS framework.

A game situation is represented by a single frame of tracking data which contains the two-dimensional positions of all players and the ball at a specific point in time. The analysis is restricted to defensive situations, defined as frames in which the opposing team is in possession of the ball. For each such frame, the perspective of the model is limited to a single defending player whose positioning is to be evaluated. All other players, including teammates and opponents are treated as fixed within the frame and remain unchanged during the optimization process. The central assumption underlying this approach is that defensive quality can be expressed through the impact of a defender's position on the attacking potential of the opponent. Instead of focusing on discrete defensive actions, the model evaluates how the positioning of a defender influences the amount of dangerous space that is accessible to the attacking team. This perspective aligns with the preventive nature of many defensive contributions, which aim to reduce threat and prevent attacking actions before they occur.

Formally, the defender's position is treated as the decision variable of the optimization problem. For a given frame, the position of the player one second before that frame is determined. The corresponding frame is the **pre-frame** and the corresponding position is the **original position**. Alternative positions that are within a constrained neighborhood of the original position are considered as candidate positions. The original position is used to account for the fact that the player needs time to move to the correct position. For each candidate position, the DAS of the attacking team is recalculated while all other aspects of the game situation remain unchanged. The objective is to find the position that minimizes the opponent's DAS value. Conceptually, the optimization problem can be expressed as the minimization of a scalar-valued objective function:

$$\min_{(x,y)\in\Omega} \mathrm{DAS}_{\mathrm{opp}}(x, y)$$

In the formula (x,y) stands for the candidate position of the defending player and Ω represents the feasible set of positions resulting from the defined constraints. The resulting minimizer is interpreted as the optimal defensive position for the analyzed player in the given game situation.

The difference between the player's observed position and the optimized position provides the basis for evaluating defensive positioning quality. While the optimization does not promise to identify the globally optimal defensive strategy, it establishes a situation-specific benchmark that reflects how a defender could improve his individual positioning within realistic movement constraints and within the team's tactical defensive formation. The model avoids the complexity of team-level optimization by framing defensive positioning as a problem on the individual player level. This design choice ensures interpretability and computational feasibility. The following subsection introduces the parameters and constraints chosen for the optimization process.

3.3.2. Parameterization and Constraints

The optimization problem defined in the previous subsection requires the definition of a feasible search space and a set of constraints to ensure both computational tractability and tactical plausibility. Without the definition of appropriate parameters, the optimization could produce unrealistic and unreachable defensive positions or lead to extensive computational costs. The parameters in this section therefore serve two purposes. First, they restrict the optimization to a space that limits the candidate positions to tactically reasonable positions. Second, they keep the number of simulations in a range that stays feasible from a computational perspective.

A first group of parameters controls the **spatial context of the optimization**. This group of parameters ensures both, reasonable candidate positions from a physical and tactical perspective and feasible computational costs. Candidate positions are only searched within a defined **search radius** around the original position. This reflects the assumption that the player can only move a limited distance in such a time frame. Hence, limiting the search radius ensures that the optimal position remains physically reachable and consistent with defensive behavior.

The discretization of the search space is defined by the **search step size**. This step size is used to generate candidate positions. Smaller step sizes result in a finer spatial resolution, potentially leading to more accurate approximations of the optimal position. But they also increase the number of potential candidates that need to be evaluated and thus the computational costs of the optimization. Conversely, larger step sizes reduce the runtime at the cost of reduced spatial precision. The selected step size always represents a trade-off between accuracy and efficiency.

To avoid implausible defensive formations, a **minimal distance between defenders** ensures that candidate positions do not overlap with the positions of other players in the defending team. This constraint ensures that candidate positions remain tactically reasonable and a defensive structure is preserved. Potential candidate positions that violate this constraint are excluded from the optimization.

The second group of parameters controls **spatial-independent** aspects of the optimization. One parameter that was already introduced in section 3.2 is the temporal resolution.

The temporal resolution for this model is set to one frame every five seconds to reduce the computational costs. This parameter does not have an impact on the optimization itself on the frame level but does limit the overall runtime by reducing the number of frames. Consequently, a higher temporal resolution would come with higher computational costs but more accuracy on aggregated values for games as a whole.

A further parameter, which has been introduced in the previous chapter as well, sets a **minimum DAS threshold** for each frame. By applying this threshold to the dataset before the optimal positions are evaluated, the total number of frames that is optimized decreases, thereby decreasing the overall runtime. At the same time, this parameter ensures that there are no frames with insignificant baseline threats destabilizing the metric.

All parameters introduced in this section are selected heuristically based on empirical considerations and domain knowledge. The goal is not to reproduce exact player movement dynamics but to approximate realistic defensive adjustments within a simplified and, more importantly, interpretable framework. By constraining the optimization to a local, discrete and tactically plausible search space, the model achieves a balance between realism, interpretability and computational efficiency. The impact of the chosen parameter values on both optimization results and runtime performance is discussed in section 4.2. This ensures that model specification remains concise, while design choices and their implications are presented transparently.

3.3.3. Optimization Algorithm

The optimization problem defined in subsections 3.3.1 and 3.3.2 is solved using a deterministic, simulation-based search procedure. The algorithm operates at the level of individual frames and individual players. It evaluates a set of candidate positions within a constrained spatial neighborhood around the defender's original position. The algorithm explicitly enumerates and evaluates feasible alternatives instead of relying on gradient-based optimization or learning-based methods. This ensures transparency and interpretability of the optimized positions.

For a given player, the algorithm works sequentially through the selected frames, in which the player is in the defending team, i.e. a player from the opposing team is in possession of the ball. For each defensive frame, the corresponding pre-frame one second prior is determined. The original position is then the center for the candidate positions. Candidate positions are generated for each frame. These positions are relative to the original pre-frame location by combining horizontal and vertical displacement within the search radius specified during parameterization. The result is a two-dimensional grid around the original position and represents the discrete search space of the optimization.

In the next step, the generated candidate positions are filtered to ensure the minimum distance to other defenders introduced in subsection 3.3.2. Candidate positions that violate the minimum distance constraint to teammates are removed. This ensures that all remaining candidates correspond to tactically plausible defensive formation and prevents unrealistic overlap between players.

For each of the remaining candidate positions, a simulated game situation is constructed. This is achieved by replacing the original position of the analyzed defender with a candidate position. The other players and the ball remain unchanged, while the frame column gets adjusted so that each candidate position has a unique frame value for all entries, i.e. rows for each other player, the ball and the adjusted defending position. Thus, each candidate position is represented by a hypothetical frame with an otherwise identical game situation. The unique frames for each candidate position are necessary to use the DAS framework for all frames and corresponding candidate positions in one run. Due to the advanced parallelization of the framework, this is the most efficient approach from a computational perspective.

Therefore, all simulated frames are simultaneously evaluated using the DAS framework. For each candidate position, the DAS value of the attacking team is computed, resulting in a scalar measure of the spatial danger accessible to the opponent under the simulated defensive configuration. This constitutes the core of the optimization process, as it directly links the defensive positioning to an interpretable measure of attacking threat.

After all candidate positions have been evaluated, the algorithm selects the position that minimizes the opponent's DAS for each original frame. This position is interpreted as the optimal defensive position for the analyzed player in that frame, subject to the imposed constraints and discretization of the search space. The optimization is therefore local in nature, as it identifies the best position within a limited neighborhood rather than a global optimum over the entire pitch. The optimization procedure described above is summarized in Algorithm 1 for clarity.

This algorithmic formulation highlights the deterministic and local nature of the optimization and provides a transparent link between defensive positioning and the resulting DAS values. The described procedure is applied independently to each frame and each analyzed player. As a result, the optimization does not have temporal dependencies between frames and does not assume coordination between multiple optimized players. This design choice simplifies the algorithmic structure and ensures that the optimization remains computationally feasible. Overall, the algorithm provides a transparent and straightforward mechanism for approximating optimal defensive positioning. By explicitly simulating alternative positions and evaluating their impact on DAS, the algorithm establishes a clear connection between spatial decisions and defensive effectiveness. The following section formalizes the objective function underlying this evaluation and further discusses its interpretation.

Data: Tracking Data T;
Pre-frame data $T - pre$;
Player d;
Frame Set F;
Config parameters (R_{max}, Δ, d_{min});
Result: Optimized positions p^*_f for all defensive Frames f;
$F_{def} \leftarrow \{f \in F \mid team_possession(f) \neq team(d)\}$;
$\Delta_{set} \leftarrow \{(dx, dy) \mid dx, dy \in [-R_{max}, R_{max}], \sqrt{dx^2, dy^2} \leq R_{max}$, step size $\Delta\}$;
$S \leftarrow \emptyset$;
foreach $f \in F_{def}$ **do**
  $p_{center} \leftarrow$ position of $d$at$(f - frame_rate)$;
  $P_{valid} \leftarrow \emptyset$;
  foreach $(dy, dy) \in \Delta_{set}$ **do**
    if $\forall$ *teammate* t*:* $\|(p_{center} + (dx, dy)) - p_t(f)\| \geq d_{min}$ **then**
      add (dx, dy) to P_{valid};
    end
  end
  ensure $(0, 0) \in P_{valid}$;
  foreach $(dx, dy) \in P_{valid}$ **do**
    $p \leftarrow p_{center} + (dx, dy)$;
    Construct simulated frame f' with position p;
    add f' to S;
  end
end
Compute $DAS_{new} \leftarrow$ `get_dangerous_accessible_space`(S);
foreach $f \in F_{def}$ **do**
  $p^*_f \leftarrow \arg\min DAS_{new}(f')$;
end
Return$\{p^*_f\}_{f \in F}$

Algorithm 1: Discrete Defensive Position Optimization

3.3.4. Objective Function

As described in the subsections above, in this thesis, the Dangerous Accessible Space (Bischofberger and Baca, 2025) is used as the objective function and serves as the central criterion for evaluating the quality of defensive positioning. DAS provides a scalar value that quantifies the attacking potential of the opposing team for a given game situation. The scalar nature of this function makes it very suitable for the purpose of this thesis.

DAS measures the amount of spatial danger available to the attacking team by combining the accessibility of space through potential passes and the non-shot-xG associated with the accessible pitch locations (Singh, 2024). A location is considered accessible if a pass to that location is more likely to reach a teammate of the player who plays the pass than it is to be intercepted by a defender or to leave the field out of play. The accessible area is then weighted by non-shot-xG values, resulting in a continuous representation of attacking danger across the pitch.

Within the proposed framework, DAS is interpreted from a defensive perspective. A high DAS value indicates a promising situation for the attacking team, whereas a low DAS value implies effective defensive control and good positioning. By minimizing the opponent's DAS value, the optimization explicitly aims to identify defensive positions that reduce both the quantity and the quality of space available to the attacking team.

DAS is chosen as the objective function for several reasons. First, DAS is inherently tracking-data-based and therefore able to capture off-the-ball behavior, interaction and spatial relationships that are central to defensive positioning. Second, the framework combines spatial accessibility with the modeling of expected values, allowing defensive positioning to be evaluated in terms of its impact on potential future outcomes. Third, DAS provides an interpretable and continuous measure that can be directly compared across different simulated positions within the same game situation and the outcomes are also understandable for practitioners.

Unlike event-based objective functions, the use of DAS does not depend on the occurrence of specific actions such as tackles or interceptions. Instead, defensive effectiveness is assessed purely through spatial structure and threat reduction. This aligns with the conceptual focus of this thesis on preventive defensive behavior and the evaluation of positioning quality independently of subsequent events.

3.3.5. Model Characteristics and Limitations

Several characteristic properties follow directly from the conceptual design and algorithmic formulation of the defensive optimization model. First, the model is **deterministic**. For a given frame, defender and set of parameters, the optimization always yields the same result. This property enhances reproducibility and facilitates systematic evaluation and comparison across players and situations.

Second, as already mentioned in subsection 3.3.2, the optimization is **local** in nature. Candidate positions are generated within a discrete and limited neighborhood around the defender's original position. The algorithm identifies the position that minimizes the objective function within this constrained search space. While this approach does not guarantee a global optimum across the entire pitch, it reflects realistic defensive adjustment and avoids the combinatorial complexity associated with unrestricted spatial optimization.

Third, the model operates **frame-wise**. Each game situation is evaluated in isolation, without temporal dependencies to the evaluation of other frames. This design choice simplifies the optimization process and ensures computational feasibility when applied to large datasets. At the same time, it implies that longer-term defensive dynamics, such as anticipation or coordinated pressing sequences, are not explicitly modeled.

The framework offers several advantages. It provides a transparent and interpretable measure of defensive positioning, does not require labeled training data and directly incorporates tracking data to capture off-the-ball behavior. Moreover, the model establishes an explicit benchmark for optimal positioning while staying computationally feasible. However, the model also entails important limitations. The focus on individual defenders neglects interactions and coordination effects at the team level. Additionally, the discretization of the search space and the imposed constraints introduce approximation errors and may affect the precision of the identified optimal position. Furthermore, the computational cost of the optimization increases with finer spatial resolution, necessitating trade-offs between accuracy and runtime. Finally, another limitation is directly inherited from DAS as an objective function. As for now, DAS does not consider a z-coordinate in the simulation of passes, which imposes a further factor for inaccuracy of the simulated result. This inaccuracy is inherited by the model proposed in this thesis.

These limitations are the result of deliberate design decisions aiming at balancing realism, interpretability and feasibility within the scope of this thesis. They also point to several directions for future research, such as team-level optimization, adaptive parameterization strategies or development of the DAS framework. The implications of these characteristics and limitations are further discussed in the context of the empirical evaluation in the following chapters.

3.4. Metric Definition

The optimization framework introduced in section 3.3 yields an optimal defensive position for a given player and game situation. In order to transform this positional benchmark into a quantitative measure of defensive performance, a defensive metric is defined that evaluates the player's observed positioning relative to the optimized reference. This section introduces the **Preventable Defensive DAS (PDD)**, a relative metric that quantifies the proportion of attacking danger that could have been prevented through improved defensive positioning.

The metric, like the model itself, is defined at the level of individual frames. For a given defensive frame f, the DAS of the attacking team is first evaluated at the player's observed position, resulting in a baseline value denoted as $DAS_{observed}(f)$. This value is computed once per frame and remains constant throughout the optimization process, as the original game situation is not altered. The optimization procedure then identifies a locally optimal defensive position for the analyzed player and yields a corresponding DAS value $DAS^{*}(f)$, which by construction satisfies $DAS^{*}(f) \leq DAS_{observed}(f)$, because if no position can be found that constitutes an improvement for the defender, the original position is used as the optimal position.

For all remaining frames, the PDD is defined as the relative reduction in attacking danger achieved by the optimized defensive position, expressed as a percentage:

$$PDD_f(d) = \max(0, \frac{DAS_{observed}(f) - DAS^*(f)}{DAS_{observed}(f)}) \times 100$$

By construction, the metric has a lower boundary of zero. A value of $PDD_f(d) = 0$ indicates that no improvement of the defensive positioning can be found within the defined search space, implying that the player's observed position is already locally optimal. Higher values reflect increasing potential for defensive improvement, with larger percentages indicating that a greater share of the opponent's DAS could have been reduced through alternative positioning of the observed player.

The use of relative, percentage-based formulation serves several purposes. First, it should enable comparisons across game situations with different baseline danger levels, such as situations close to or far from the goal. Second, it aims to allow for more meaningful comparison between players occupying different positions on the pitch, for example defenders and attackers involved in defensive transitions. Finally, the metric retains a direct and intuitive interpretation. It quantifies how much of the opponent's attacking danger was preventable through improved positioning.

To obtain a player-level defensive score, the frame-level metric is aggregated across all valid frames associated with the player. But rather than averaging frame-level ratios, the aggregation weights frames implicitly by their observed DAS by following a ratio-of-sums formulation, thereby emphasizing defensive positioning in situations of higher offensive threat. The player-level PDD score is defined as:

$$PDD(d) = \frac{\sum_{f \in \mathcal{F}_d} (DAS_{observed}(f) - DAS^*(f))}{\sum_{f \in \mathcal{F}_d} DAS_{observed}(f)} \times 100$$

where $\mathcal{F}_d$ denotes the set of frames in which player d is evaluated and for which $DAS_{observed}(f) \geq DAS_{min}$. Alternative aggregation strategies, such as the use of medians or stratification by tactical context, have also been explored but are not considered for the metric in this thesis. They represent potential extensions for future work.

The Preventable Defensive DAS metric provides a direct, interpretable and context-aware measure of defensive positioning quality. By grounding the metric in a spatial optimization framework and expressing performance relative to an explicit benchmark, the proposed approach bridges the gap between positional analysis and quantitative defensive evaluation.

4. Model Optimization

After the introduction of the model design and the resulting metric in the previous chapter, this chapter focuses on implementation details, alternative implementation approaches and the evaluation of the final metric. First, section 4.1 will provide detailed information about the implementation, for example used languages and libraries. In section 4.2, alternative optimization approaches and parameter selections are introduced together with the reasoning for the final setup. Finally, section 4.3 reports the evaluation framework and methods that are used for validation of the final metric.

4.1. Implementation Details

The proposed optimization framework and the resulting metric computation are implemented in Python and integrated into a modular processing pipeline that follows the conceptual separation introduced in this chapter: (i) preparing frame-based inputs from tracking data, (ii) running defensive optimization for selected players and frames and (iii) computing and aggregating metric values.

Software environment and libraries

All experiments are conducted in a Python-based environment. Game data loading and basic enhancements, such as player velocity, are performed using the *DataBallPy* library (Oonk, 2025). Numerical operations and data manipulation are implemented with standard scientific Python tools like NumPy and pandas. The computation of Dangerous Accessible Space relies on the implementation by Bischofberger (2025), which is treated as a black-box function that maps a frame representation to a scalar value for the attacking team.

Pipeline structure

The pipeline structure is divided into three main components:

1. **Preprocessing and frame selection**

Raw tracking data is converted into a frame-oriented representation. Frames are filtered to include only defensive situations, which means the opponent is in possession of the ball, and to exclude non-representative game states such as the ball being out of play or set piece phases. To keep runtime manageable, frames are sampled down at a fixed interval, producing a reduced but representative set of defensive situations for each game. Furthermore, after computing the DAS values for the original positions, the $DAS_{threshold}$ is used to exclude frames with an insignificant baseline danger.

2. **Defensive Optimization**
   For each evaluated frame for the selected player, the optimization routine generates candidate positions in a local neighborhood around the observed position, filters candidates according to defined constraints and evaluates the opponent's DAS for each simulated position. The candidate with the minimum DAS is selected as the local optimum, yielding $DAS^*(f)$ and the corresponding optimal position p^*.

3. **Metric computation and aggregation**
   Based on the original frame DAS value $DAS_{observed}(f)$ and the optimized value $DAS^*(f)$, the frame-level **Preventable Defensive DAS** is computed as a relative percentage reduction. Player-level scores are then obtained by aggregating across frames using the ratio-of-sums formulation defined in section 3.4.

Caching and intermediate storage

Computing DAS repeatedly is the dominant cost of the pipeline. To avoid recomputation and to support reproducibility, intermediate results are cached. In particular, original frame-level DAS values as well as optimization outputs are stored persistently per game and sampling configuration. This design allows experiments to be rerun with different aggregation routines or evaluation splits without repeating the full optimization step. Cached objects are stored in a structured directory layout and serialized using Python's native persistence mechanisms (e.g., pickle).

Computational considerations

Runtime is primarily driven by the number of evaluated frames and the number of candidate positions per frame. Consequently, the implementation is designed to (i) reduce the number of frames by systematic downsampling, (ii) restrict the search space using a local movement radius and a discrete step size and (iii) discard infeasible candidates early using constraint checks. These measures reduce unnecessary DAS evaluations and provide a practical trade-off between optimization precision and computational cost. The reasoning behind the parameter decision is reported in chapter 4.2.

Reproducibility

The optimization procedure is deterministic given a fixed parameter configuration and fixed input frames. All parameters used for candidate generation, feasibility checks and frame selection are stored alongside the outputs to ensure that results are exactly reproducible. The separation of preprocessing, optimization and aggregation steps further supports transparent auditing of the pipeline.

The model specification, metric definition and implementation details presented in this chapter establish a complete and reproducible framework for evaluating defensive positioning. The following chapter builds upon this framework and introduces the evaluation design used to assess the behavior and robustness of the proposed Preventable Defensive DAS metric.

4.2. Design Decisions and Alternative Optimization Strategies

The optimization framework that has been introduced in this chapter involves several design choices that directly affect the quality of the results, the computational feasibility and the interpretability. While the previous section describes the final model and implementation, this section discusses alternative optimization strategies and parameter choices that were considered during development but ultimately not adopted. The objective of this section is to provide transparency regarding the decision-making process and to justify the selected configuration as a balanced trade-off between effectiveness and efficiency.

4.2.1. Alternative Optimization Strategies

The final optimization approach uses a deterministic, exhaustive search over all feasible candidate positions within a constrained neighborhood. This choice was made after evaluating alternative strategies that aim either to reduce computational cost by limiting the number of evaluated candidates or to increase the accuracy without posing too many additional computational costs.

A first alternative strategy considered during the development was **random sampling** within the feasible search space. Instead of evaluating all candidates within the search space, a fixed number of candidates would be drawn at random and evaluated using the DAS objective function. In the conducted experiments, the number of sampled candidates was limited to values of $n = 20$ and $n = 50$, compared to a maximum of 81 feasible positions in the deterministic search.

From a purely computational perspective, random sampling achieved notable runtime improvements. Reducing the number of evaluated candidates to 20 resulted in an approximate reduction of runtime to one quarter of the deterministic baseline, while sampling 50 candidates reduced runtime to roughly one half to two thirds. This reduction in runtime is the logical consequence since reducing the number of candidates leads to a decrease in runtime per frame in approximately the same proportion. Hence, the absolute runtime decreases as well. However, given that the absolute runtime of the chosen deterministic approach is already manageable, these improvements were not critical for the practical applicability.

In terms of optimization quality, the random sampling approach produced slightly worse results on average, as expected due to the reduced coverage of the search space. Importantly, however, the observed degradation in optimization quality did not scale proportionally with the reduction in the number of evaluated candidates. Instead, result quality varied noticeably across runs, even when the same number of samples was used. This variability highlights the primary limitation of the random sampling approach in the context of this thesis. This approach has a **lack of stability and reproducibility**. Different random samples potentially lead to different locally optimal positions and DAS reductions. Therefore, the resulting metric values would be sensitive to the random seed and it cannot be guaranteed to produce consistent results across multiple runs with the same parameterization. This stochastic behavior complicates both interpretation and evaluation, as differences in metric values may arise from sampling variability rather than from meaningful differences in defensive positioning.

Given that one central objective of this thesis is to define a transparent and interpretable benchmark metric for defensive positioning, this lack of reproducibility was considered a decisive drawback. Consequently, despite offering notable runtime savings, the random sampling approach was not adopted in favor of the deterministic evaluation of all feasible candidate positions.

A second alternative strategy explored during the development was an **iterative coarse-to-fine optimization**. The idea is to explore the search space with a coarse spatial resolution first in order to identify the most promising regions. In the second step, those regions are explored again with a smaller step size. Conceptually, this approach aims to approximate the deterministic solution more efficiently by reducing the number of fine-grained evaluations.

While this strategy produced slightly improved optimization results compared to the deterministic baseline, the additional refinement step significantly increased computational complexity and runtime. Using the coarse-to-fine approach with an initial step size of 1.5, the average relative reduction in DAS between the observed and optimized position increased by approximately 6% compared to the deterministic baseline with a step size of 1. At the same time, runtime increased by about 13%. Given the limited gain in optimization quality relative to the higher computational cost and implementation complexity, the iterative approach was not adopted for the final model.

Based on these considerations, the **deterministic evaluation of all feasible candidate positions** was selected as the final optimization strategy. Despite its computational cost, this approach offers several advantages that are aligned with the objectives of this thesis. It provides stable and reproducible results, ensures that the local optimum within the defined search space is consistently identified and maintains a clear and interpretable relationship between candidate positions and the resulting DAS values.

4.2.2. Parameter Selection and Trade-offs

Besides the choice of the fundamental optimization strategy, several parameters were explored to balance the optimization accuracy and computational feasibility. These include the maximum movement radius, the spatial step size used for candidate generation and the minimum DAS threshold required for the metric computation.

The values of these parameters were selected based on a combination of football domain knowledge, empirical experimentation and computational considerations. The movement radius was chosen to reflect realistic short-term defensive adjustments and has a significant impact on the runtime. The step size controls the spatial resolution of the optimization and directly influences the runtime as well. The step size was chosen as a trade-off between runtime and optimization quality. In this thesis the parameters were set to $searchRadius = 5$ and $stepSize = 1$. This combination keeps the computational cost on a manageable level while ensuring consistent and good results. A reduction of spatial resolution to a step size of 1.5 already delivers worse results than the random approach, introduced in the previous subsection, with step size 1. The minimum distance to teammates does not have a significant impact on the runtime and has been chosen solely by considerations of realistic game situations.

The minimum DAS threshold for the original game situation was introduced to avoid unstable relative improvements in game situations with a very low threat for the defensive team. At the same time, this threshold lowers the number of frames to be evaluated and optimized and hence, reduces the runtime. From this perspective, one idea was to use a relative threshold for the model, for example the 25th percentile as a cutoff. However, depending on the game, this would include situations with very little threat that might destabilize the metric or exclude situations with a high threat that should be included. Therefore, a fixed value of $thresholdDAS = 0.1$ was chosen.

Overall the approach was to select a pragmatic compromise that yields stable optimization behavior, interpretable metric values and manageable computational cost across the evaluated dataset rather than optimizing the parameters exhaustively.

4.3. Evaluation Methodology

In addition to defining and implementing the proposed optimization framework, a structured methodology is required to examine the behavior and interpretability of the resulting metric PDD. Since it is derived from repeated observations across games, players and frames, its usefulness does not only depend on individual values but also on its consistency across different contexts and relationship to external reference measures.

This section outlines the methodological approach to evaluate consistency, stability and significance of the PDD metric. For this objective, three complementary perspectives are considered. First, **inter-match stability** is assessed to examine if player-level PDD values remain consistent across different games. Second, **intra-match stability** is analyzed to quantify the variation within a single match across multiple defensive situations. And third, an **external validation** approach is introduced by comparing PDD values to established player ratings.

All empirical results obtained using this methodology are reported in section 5.4, while their implications are discussed in chapter 6.

4.3.1. Inter-Match Stability

To assess the consistency of the PDD metric across different matches, inter-match stability is evaluated using the Intraclass Correlation Coefficient (ICC). The ICC is a statistical measure commonly used to quantify the reliability of repeated measurements of the same entity under varying conditions (Koo and Li, 2016). In the context of this thesis, it captures to which extent player-level PDD values remain stable when computed across multiple games.

The ICC is well suited for this setting for two reasons. First, it explicitly models measurements grouped by entity. In this case, repeated PDD values are associated with the same player across different games. Second, unlike rank-based or pairwise correlation measures, the ICC directly quantifies the proportion of total variance that can be attributed to differences between players instead of variability across matches (Shrout and Fleiss, 1979).

The inter-match stability is evaluated for two different aggregations. Player-level PDD values are aggregated for full matches and for half-periods. The reason for the second approach is the relatively small data basis. By using first and second half periods instead of full matches, the sample size is artificially enlarged. To avoid effects of differences in between both half-periods, full matches are analyzed as well. Furthermore, only players with a minimum number of matches are included to ensure meaningful estimation. Different inclusion thresholds are considered in order to examine how stability estimates change as the number of observations per player increases.

Two specific forms of the ICC are used in this work. The first follows a one-way random-effects model with single measurements, commonly denoted as ICC(1,1). It can be used with different games for the various players. That means there is no need for common games between the players. This is particularly useful in this thesis because of the relatively small data set (Shrout and Fleiss, 1979). The second follows a two-way mixed model with single measurements, commonly denoted as ICC(3,1). For this, all players considered are rated for all games considered. This lowers the sample size but the value is inherently more significant (Shrout and Fleiss, 1979).

By applying the ICC in this manner, inter-match stability is assessed in terms of relative consistency rather than absolute agreement. This aligns with the intended use of PDD as a comparative metric.

4.3.2. Intra-Match Stability

Intra-match stability examines the consistency of the PDD metric **within a single game** across multiple defensive situations. Since defensive positioning is highly dependent on game context, variation at the frame level is expected.

Two complementary perspectives are adopted to assess the intra-match stability. First, **half-period aggregation** is used to compare player-level PDD values between the two halves of a match. The consistency between halves is assessed using Spearman and Pearson correlations, allowing for the evaluation of both linear and rank-based associations (Pearson, 1895; Spearman, 1904).

Second, **frame-level variation** is analyzed to capture fine-grained fluctuations of PDD within a match. For this, the ICC measures introduced in subsection 4.3.1 are used. In this case, the frames are the raters, while the players remain the observed objects.

By combining half-level correlation analysis with frame-level dispersion measures, intra-match stability is evaluated across multiple temporal scales. This approach allows the analysis to distinguish between expected situational variability and broader patterns of consistency in defensive positioning behavior.

4.3.3. External Validation

To complement the internal consistency analysis, an external validation approach is adopted. Some challenges arise with the comparison to external ratings since established measures of defensive positioning are missing. Since the data basis for this thesis consists primarily of single games, the comparison with general ratings of players is not very meaningful as well.

Therefore, the external validation is based on player ratings for the corresponding games publicly available on online platforms. Those are subjective ratings, measuring the overall performance of players, not the positioning in particular. Since those player ratings are typically ordinal in nature and may not exhibit linear scaling properties, rank-based correlation measures are used as the primary evaluation criterion. Specifically, Spearman's rank correlation coefficient is employed to assess monotonic relationships between PDD and external ratings (Spearman, 1904). Pearson's correlation coefficient is additionally reported as a supplementary measure to capture potential linear associations.

The analysis is performed at the player level and restricted to matching positional groups to ensure contextual consistency. The objective is to indicate whether the proposed metric exhibits systematic alignment with established evaluation frameworks.

5. Evaluation

This chapter presents the results that were obtained using the final configuration of the proposed optimization framework and the resulting Preventable Defensive DAS metric. The focus of this chapter lies on describing the observed metric values, their distributions and their variation across players and position groups. In contrast to chapters 3 and 4, which detailed the methodological design and implementation choices, including the reasoning behind them, this chapter aims to present the result purely descriptively.

The results shown in this chapter are based on the final optimization strategy and parameter configuration introduced in the previous chapter. Alternative optimization methods or parameter settings are not considered unless they are used as a comparison when necessary for interpretation. The objective is to establish a clear empirical picture of the behavior of the PDD metric when it is applied to real tracking data.

The chapter is structured as follows. Section 5.1 provides an overview of the resulting PDD values and introduces their basic properties and value ranges. Section 5.2 analyzes the distribution of PDD values across positional groups and introduces player rankings. Section 5.3 shows selected situational examples of the metric to strengthen an intuitive understanding. Finally, section 5.4 provides the results of the evaluation methods.

5.1. Overview of the Resulting PDD Metric

This section provides a high-level overview of the PDD metric and the values obtained from the application of the proposed framework. The aim is to introduce the basic characteristics of the metric and establish a reference for the more detailed analyses presented in subsequent sections.

Figure 5.1 illustrates the distribution of **player-level PDD values** across all evaluated players with at least 10 frames. By definition, PDD is expressed as a percentage and quantifies the proportion of the opponent's DAS that could have been prevented through an optimal defensive positioning by the evaluated players. Player-level values are computed using the ratio-of-sums aggregation defined in section 3.4.

The observed distribution is continuous and concentrated within a moderate value range. Most player-level PDD values lie below 15%, with a gradual decline in frequency

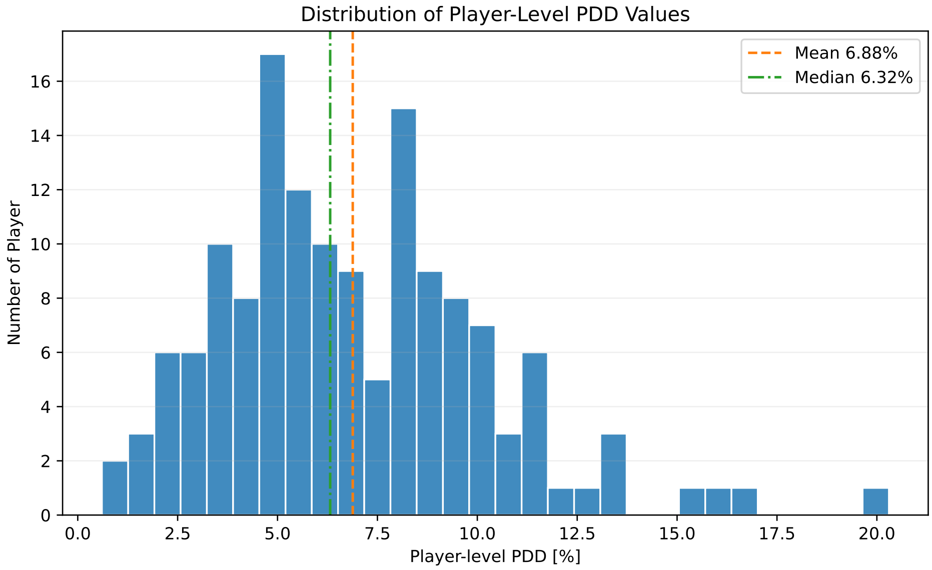

Figure 5.1.: Distribution of Player-Level PDD-Values

towards higher values. The mean PDD value across all players is 6.88%, while the median is slightly lower at 6.32%, indicating a mildly right-skewed distribution. This suggests that, for the majority of players, only a limited proportion of the opponent's DAS is preventable through an alternative positioning within the defined search space.

Importantly, the distribution does not exhibit pronounced clustering at extreme values. Very low PDD values correspond to players whose observed positioning is frequently close to the locally discovered optimum, whereas higher values indicate increasing potential for positional improvement. This overview establishes the general scale and distributional properties of the PDD metric. The following sections build on this foundation by analyzing how PDD values differ across positional groups and how they vary across games.

5.2. Distribution of PDD Values Across Players and Positions

This section presents aggregated results of the PDD metric at the player level and with a focus on the distribution across positional groups. As established in the previous section, PDD values are relative, situation-dependent measures and therefore require careful contextualization when comparing across players. Defensive responsibilities and situational contexts vary substantially between tactical roles. Therefore, aggregated results are analyzed with respect to positional groupings.

To establish an understanding of the context, the results are first compared at the

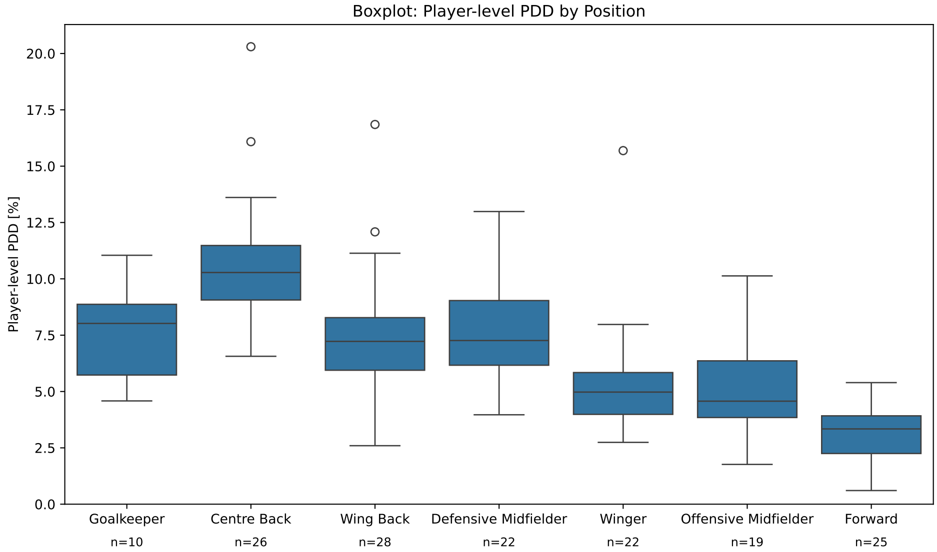

Figure 5.2.: PDD by Position Group

level of positional groups before individual player rankings are introduced. This reflects the fact that absolute PDD values vary significantly between different roles and that meaningful comparisons at the player level should be performed within homogeneous role groups, even though it is defined as a relative measure as stated in section 3.4. In addition, players that appear in multiple positions across different games are treated separately for each role, such that the role-specific behavior is preserved in aggregation.

5.2.1. Comparison of PDD Distribution Across Positions

Figure 5.2 illustrates the distribution of PDD at the player level across the seven positional groups: goalkeepers, centre backs, wing backs, defensive midfielders, wingers, offensive midfielders and forwards. The number of players considered for each positional group is indicated below each boxplot.

Across the positions, the metric exhibits clear differences in both central tendency and dispersion. Defensive roles, particularly centre backs but wing backs and defensive midfielders as well, show higher typical PDD values compared to the offensive positions. In addition, these defensive groups tend to display a wider spread of values, indicating that the potential for preventable dangerous space varies more for players in these roles.

In contrast, the primarily offensive positions, such as offensive midfielders, wingers and forwards, generally show lower typical PDD values and tend to have a more compact distribution. Goalkeepers form a distinct group and display higher typical values but a compact distribution. This is consistent with their constrained positional freedom but defensive positioning on the pitch in open play. Goalkeepers are excluded from further analysis due to the distinct nature and requirements of the position.

Importantly, these differences across positions should be interpreted as descriptive characteristics of the metric rather than direct indicators of defensive ability. PDD quantifies relative, situation-dependent potential for positional improvements and is influenced by role-specific spatial contexts and defensive responsibilities. The positional distributions presented in this section therefore serve primarily to contextualize the scale and behavior of the metric across tactical roles.

5.2.2. Player Rankings within Positional Groups

Based on the positional comparison presented in the previous subsection, player-level rankings are computed separately within each positional group. This proceeding ensures that the comparison is meaningful and players are measured against other players with similar defensive roles and responsibilities during the match. Consequently, no global ranking across positions is reported.

For the defensive positions, centre backs, wing backs and defensive midfielders are shown in tables 5.1, 5.2 and 5.3. These are the positions that are most directly involved in defensive positioning and have the most significant impact on the DAS value of the opponent. For each position, the five players with the lowest and with the highest PDD values are reported. The rankings for the remaining positions are provided in the appendix.

Rank	Player	PDD(%)	Number of Frames
1	Erhan Mašović	6.56%	82
2	Ivan Ordets	7.80%	82
3	Adam Dźwigała	7.89%	73
4	James Alexander Lawrence	7.93%	144
5	Eric Anders Smith	8.49%	73
22	Matthijs de Ligt	13.15%	91
23	André Hoffmann	13.56%	64
24	Rick van Drongelen	13.61%	90
25	Kevin Kraus	16.09%	78
26	Edmond Fayçal Tapsoba	20.30%	97

Table 5.1.: Top and bottom five Centre Backs by PDD

Each table presents the aggregated PDD value for the player-position combination together with the number of frames contributing to the aggregation. To ensure a certain

Rank	Player	PDD(%)	Number of Frames
1	Luca-Milan Zander	2.60%	12
2	Kingsley Schindler	2.82%	32
3	Kristian Majdahl Pedersen	2.98%	19
4	Manolis Saliakas	4.56%	61
5	Leon Guwara	5.49%	67
24	Keven Schlotterbeck	9.67%	28
25	Mitchel Bakker	9.71%	97
26	Scott Fitzgerald Kennedy	11.14%	19
27	Joao Pedro Cavaco Cancelo	12.08%	30
28	Saidy Janko	16.85%	82

Table 5.2.: Top and bottom five Wing Backs by PDD

Rank	Player	PDD(%)	Number of Frames
1	Kevin Stöger	3.97%	79
2	Patrick Osterhage	4.59%	28
3	Ao Tanaka	4.69%	132
4	Maximilian Thalhammer	5.06%	86
5	Anthony Losilla	5.26%	82
18	Dennis Dressel	9.63%	90
19	Nadiem Amiri	9.97%	84
20	Michal Karbownik	10.30%	124
21	Hikmet Çiftçi	11.73%	104
22	Sadik Fofana	12.99%	22

Table 5.3.: Top and bottom five Defensive Midfielders by PDD

robustness, only combinations with a minimum of 10 frames are included in the rankings.

As already indicated by the boxplot in 5.2, the centre backs have the highest PDD values in both top and bottom five compared to the players in the other two positional groups. This is most probably a consequence of their positioning in that area of the field that potentially yields the highest non-shot-xG values. Despite that, some names within the rankings may appear counterintuitive when compared to general player or team reputation. For example, *Edmond Tapsoba* is positioned at the lower end of the centre backs. This result is influenced by match-specific contexts included in the dataset, such as unfavorable game states. This will be further discussed in chapter 6.

5.2.3. Players Appearing in Multiple Positions

As already mentioned, some players in the dataset have played in different positional groups between the covered games. They are treated separately for each role since the role has a direct impact on the PDD metric values.

Table 5.4 shows all players in the datasets for which this scenario applies. Notably, both players with defensive positions - *Matthias Zimmermann* and *Michal Karbownik* - have a significant difference between the values for the different positions. This highlights the need to differentiate the rankings for each group and make sure comparisons are made within homogeneous groups.

Player	Position 1	PDD(%) 1	Position 2	PDD(%) 2
Dawid Igor Kownacki	Forward	3.78%	Offensive Midfielder	3.20%
Elione Fernandes Neto	Offensive Midfielder	1.76%	Winger	3.32%
Emmanuel Iyoha	Forward	3.74%	Winger	4.37%
Kwadwo Kyeremeh Baah	Forward	2.39%	Winger	3.77%
Matthias Jürgen Zimmermann	Centre Back	11.09%	Wing Back	7.14%
Michal Karbownik	Defensive Midfielder	10.30%	Wing Back	7.96%
Shinta Karl Appelkamp	Offensive Midfielder	3.86%	Winger	2.74%

Table 5.4.: Players with Multiple Positional Groups

5.3. Case Studies and Qualitative Examples

To complement the aggregated results presented in the previous section, this section provides qualitative, frame-level examples illustrating how the PDD metric manifests in concrete game situations. The objective is to visualize the relationship between defensive positioning, spatial control and the resulting change in DAS. This supports an intuitive understanding of the metric.

For each example, the original defensive position and the optimized position identified by the optimization model are shown in comparison using subfigures. All other players and the ball remain logically unchanged, since the change in DAS is solely attributable to the analyzed player. All examples are taken from the game *Fortuna Düsseldorf - 1. FC Nürnberg (15th Gameday, 2022/23 Season, 2. Bundesliga)*. In each of the figures, the left side displays the DAS of the original frame, with the player marked in blue. On the right-hand side, the DAS of the optimized frame is displayed with the player marked in green.

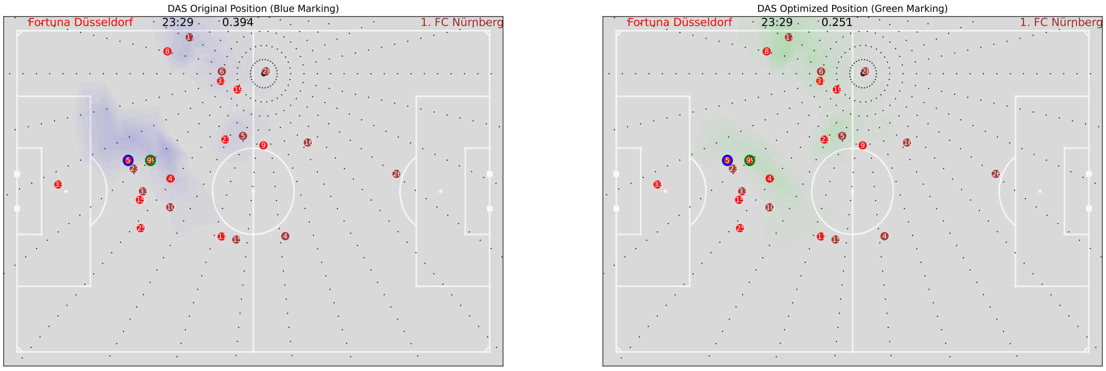

Figure 5.3.: Frame Centre Back

Example 1: Centre Back

Figure 5.3 illustrates a defensive situation involving a centre back in the last line of defense. In the original frame, the defender keeps the attacking player with the number 23 onside, allowing a lot of space to be accessible for that player in a pitch area that inherits high xT values. The total DAS for that position is 0.394.

In the optimized frame, the defender is repositioned slightly in front of the attacking player, putting him offside. This attacker is now not reachable via pass and hence, the accessible space, particularly directly at the penalty area, is visibly reduced. The new DAS is 0.253, leading to a PDD score for that situation and player of 36.23%.

Example 2: Wing Back

In figure 5.4, the positioning of a right back is illustrated, where the defensive line stands high, almost at the middle line. The right back is standing close to the passer, covering the angle to the right sideline at the middle line. In his back the field is not covered, leaving space for a through ball. The total DAS in this situation is 1.193.

In the optimized frame, the defender opens the space to the right sideline to cover more area behind. This allows new pass options for the attacking player, but the accessible space is less dangerous. The optimal positioning yields a PDD of 53.50% and a new DAS of 0.555.

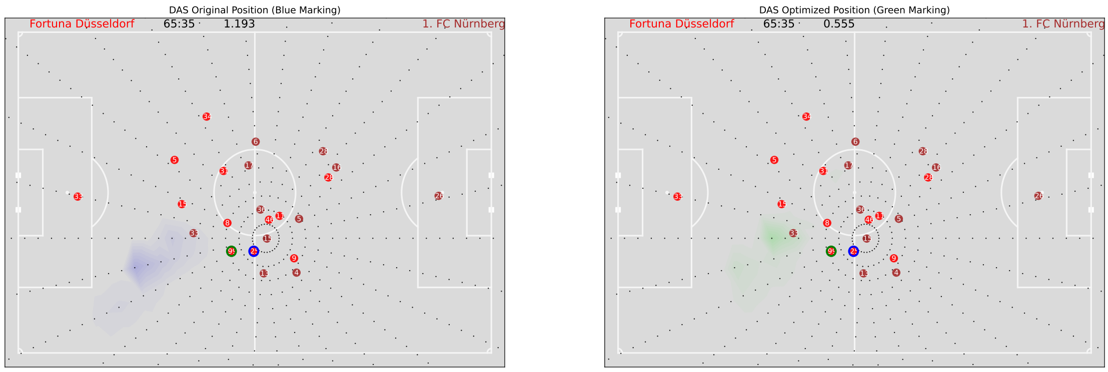

Figure 5.4.: Frame Wing Back

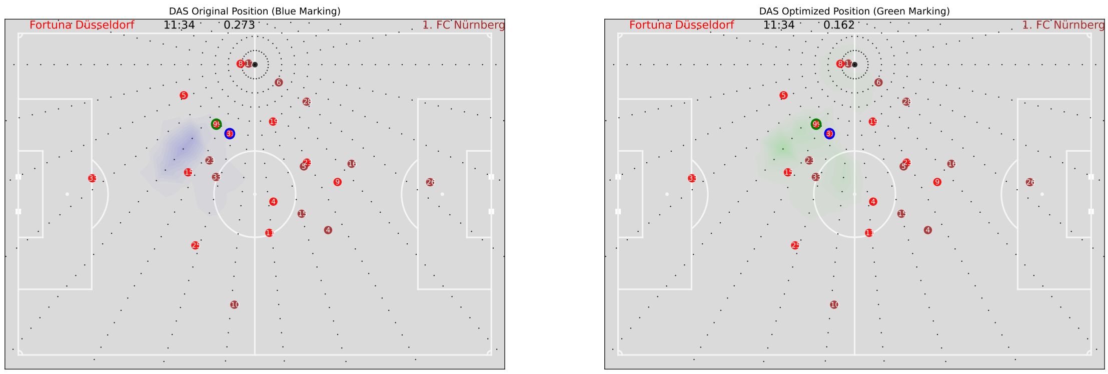

Figure 5.5.: Frame Defensive Midfielder

Example 3: Defensive Midfielder

Figure 5.5 presents a situation for a defensive midfielder where the ball is in a central position on the field. The player in possession has a good passing opportunity for a through ball to the attacker with number 23 in between the two central defenders. The defensive midfielder covers the pass in the direction of the middle of the field, leaving the through pass option uncovered. The resulting DAS is 0.273.

The optimized frame shows the impact of a slight change in position, preventing the deep pass and allowing more accessible space with a less dangerous field position in the middle of the pitch. This lowers the overall to DAS 0.162. The PDD in this scene is 40.49%.

This last example, however, shows also that defensive positioning is always dependent on the teammates. When the space between the two central defenders would be smaller, the defensive midfielder might already be standing optimally. This indicates a possible starting point for future work.

5.4. Validation Results of the Final Configuration

This section presents results related to the consistency and robustness of the PDD metric and its relationship to external evaluations. The objective is to report observable patterns in the data that provide a basis for the subsequent discussion of the metric stability and interpretability. At this stage, the focus remains on describing the results obtained from the dataset, without drawing conclusions regarding the validity or reliability of the metric.

This section is divided into three parts. First, the inter-match consistency is examined by comparing the values at the player level across different games. Second, intra-match stability is analyzed, assessing how PDD values vary over the duration of a single game. And third, the results of the metric are compared against external player ratings.

5.4.1. Inter-Match Stability

In this subsection, the empirical results of the inter-match stability of the PDD metric are reported. As introduced in section 4.3, the inter-match stability assesses the extent to which player-level PDD values are consistent across different games and half-periods treated as separate entities. The stability is evaluated using the ICC.

Table 5.5 summarizes the resulting ICC values with increasing minimum numbers of matches for a player to be included (*match-count*). For players with at least two matches, an ICC value of 0.51 is observed based on 18 players. Restricting the analysis to players with at least three matches results in ICC of 0.49 (15 players). For the subset of players with at least five matches, the ICC increases to 0.76 for ICC(1,1) and 0.77 for ICC(3,1), based on 8 players.

ICC Type	Min Match-Count	Number of Player	ICC
ICC(1,1)	2	18	0.505
ICC(1,1)	3	15	0.492
ICC(1,1)	4	13	0.626
ICC(1,1)	5	8	0.763
ICC(3,1)	5	8	0.765

Table 5.5.: ICCs for Inter-Match Stability (Full Games)

To increase the number of games artificially, an alternative analysis has been conducted. For this, each halftime period is treated as an independent observation. This approach increases the number of measurements per player while preserving the temporal structure. The resulting values from this halftime period aggregation can be seen in table 5.6. For players with at least 4 halftime periods, the ICC has a value of 0.61 with 14 players observed. For a minimum of 8 periods, the value increases to 0.71 with 11 observed players. And for 10 periods and 3 observed players the ICC(1,1) and ICC(3,1) yield a value of 0.79.

ICC Type	Min Halftime-Count	Number of Player	ICC
ICC(1,1)	4	14	0.612
ICC(1,1)	6	13	0.615
ICC(1,1)	8	11	0.717
ICC(1,1)	10	3	0.792
ICC(3,1)	10	3	0.787

Table 5.6.: ICCs for Inter-Match Stability (Halftime Periods)

Across both aggregation schemes, the ICC values vary depending on the minimum number of observations per player. Higher stability estimates are generally observed for subsets of players with a larger number of repeated measurements. The reported results provide a descriptive overview of how PDD behaves across matches under different inclusion criteria. A detailed interpretation of these findings and their implications for the robustness of the metric is deferred to chapter 6.

5.4.2. Intra-Match Stability

After presenting the results of the inter-match stability analysis, this subsection reports the results of the intra-match stability analysis. The intra-match stability examines how the PDD values vary for individual players within a single game across multiple situations. As outlined in section 4.3, again two complementary perspectives are considered. Consistency between two halftime periods within the same match and a fine-grained variation at the frame level.

First, the intra-match consistency is analyzed at the level of halftime periods within a match. Player-level PDD values are computed and aggregated for each half independently. To assess the stability across these halftime periods, correlation-based measures are used. Across 150 player-match combinations, a Pearson correlation (Pearson, 1895) of 0.71 and a Spearman rank correlation of 0.72 are observed between first-half and second-half PDD values. These correlations indicate a noticeable degree of association between halves.

Second, frame-level variation is analyzed to capture if there are relevant differences between players that can be observed consistently over the duration of one game in varying situations. For each combination of player and match, frame-level PDD values are used to compute the ICC values. Across all seven games in the dataset, the ICC is between 0.09 and 0.19, which is a very low number for the ICC.

This analysis highlights the extent to which PDD values fluctuate across different defensive contexts encountered during a single match. Taken together, both perspectives of the analysis show consistency of PDD values for a player across longer time periods, which potentially cover a range of different game situations, while the observations at the frame level indicate a high short-term variability. This should not be surprising given the nature of defensive positioning and the dependence on situational contexts. A further discussion and interpretation of these results is provided in chapter 6.

5.4.3. External Validation

In the third part of the evaluation, the PDD is compared to external rating sources. As outlined in section 4.3, external ratings are not treated as ground truth for defensive performance but are the best source to validate not only the consistency but also the results themselves to be significant and coherent with general domain perception.

External validation is performed at the player level by correlating aggregated PDD with player ratings from publicly available online platforms, i.e. SofaScore and WhoScored. To ensure contextual consistency, the analysis is restricted to defensive positional groups only. For more offensive players it is expected that ratings are rather based on other aspects of the game. The results for offensive players can be found in the appendix but are not further analyzed for that reason. Correlation-based measures are used to quantify the association between PDD and external ratings.

Given the ordinal nature of player ratings and the potential presence of nonlinear relationships, Spearman's rank correlation coefficient is used as the primary evaluation measure. For centre backs the correlation with SofaScore ratings is -0.16 and with WhoScored -0.12. Wing backs have a correlation of -0.03 with SofaScore and 0.06 with WhoScored. The highest correlation can be observed for defensive midfielders. The correlation with SofaScore ratings is -0.26 and with WhoScored -0.11. Since low PDD values mean a good performance and high ratings mean a good performance, a negative correlation coefficient can be interpreted as a positive correlation between the PDD values and the ratings.

For centre backs and defensive midfielders a low correlation can be observed, which is already noticeable since currently those ratings are not directly influenced by defensive positioning measures. For wing backs no correlation can be observed, indicating that the focus of those ratings for these positional groups lies on other aspects of the game.

As for the two other parts of the evaluation in this chapter, a further analysis of the results and their implications on the PDD metric is conducted in chapter 6.

6. Discussion of Results

In chapter 5, the empirical results obtained from the application of the proposed optimization framework and the resulting PDD metric are presented. These results describe the distribution of PDD values, their variation across players and positional groups, qualitative frame-level examples and multiple consistency analyses. In this chapter, those findings are discussed and placed in a broader methodological and football-analytical context.

The discussion focuses on interpreting what the PDD metric captures in practice, how its observed behavior should be understood and which conclusions can and cannot be drawn from the reported results. Particular attention is also paid to the interpretation of evaluation outcomes, including stability analyses and comparisons with external player ratings. In addition, methodological limitations and practical applications of the proposed approach are presented.

6.1. Interpretation of the PDD Metric

This section discusses the results of the PDD metric presented in the previous chapter and how the metric should be interpreted. The objective is to clarify what the metric is able to capture in practice, how the observed patterns across players and positions arise and which conclusions can and cannot be drawn from PDD values.

At its core, PDD quantifies the relative proportion of DAS that could have been prevented by a player through alternative, respectively optimal, positioning in a given situation. The metric does not evaluate any defensive actions or outcomes but focuses on the potential of game situations. On the level of a single frame, PDD is very context dependent and the values should be interpreted carefully but at the aggregated player level the metric can give an indication about the quality of defensive position for that player. A low value in a given situation implies that the player leaves few dangerous space for the opponent that he could have prevented. Conversely, a high value means the player could have prevented more DAS in the given situation. Consequently, at the aggregated player level, a low PDD average value for a game means that the player is frequently observed close to the optimal position while a high value indicates a lot of preventable space that was accessible for the opponent.

A central characteristic of PDD is its strong dependence on role and context. As shown in section 5.2, defensive positions such as centre backs, wing backs and defensive midfielders exhibit higher typical PDD values and greater variance than the offensive positions. This pattern makes sense since defensive players find themselves more often in situations where a small adjustment can substantially alter the DAS. Therefore, comparisons between players should be limited to positional groups to provide a meaningful basis.

An example that illustrates this limitation, is that across the whole dataset the three players with the lowest PDD values are *Eric-Maxim Choupo-Moting* (PDD 0.61%), *Tim Lemperle*(1.16%) and *Jan Uwe Thielmann* (1.49%), all of whom are forwards. Conversely, the three players with the highest PDD values include two centre backs - *Edmond Tapsoba* (20.30%) and *Kevin Kraus* (16.09%) - and one wing back - *Saidy Janko* (16.85%). While the PDD metric is normalized as a relative percentage reduction in DAS and therefore enables comparisons across defensive situations of varying baseline danger, it does not account for systematic differences in defensive roles and exposure across playing positions. Forwards are typically involved in fewer and less dangerous defensive situations, whereas centre backs are consistently exposed to high-risk defensive contexts. Consequently, even a relative, percentage-based metric can produce implausible rankings when applied across heterogeneous positional roles. Given that the dataset further contains only a limited number of matches for most players, these extreme differences are unlikely to reflect true disparities in defensive positioning quality. These observations reinforces the conclusion that PDD values should be interpreted within positional groups rather than used for position-independent player comparisons.

However, within a positional group, PDD values can provide a meaningful indication of the performance, especially for defensive positional groupings. This is supported by the evaluation results presented in section 5.4. Player performances remain consistent across different games and halftime periods, as the ICC values for the inter-match stability and the correlations for the intra-match stability show. They also correlate, even though the correlation is low, with external ratings for centre backs and defensive midfielders. Nevertheless, one aspect that has to be highlighted is that the player rankings presented in section 5.2 cannot be seen as a ranking of the general abilities of the players but only evaluate the performance in the matches of the dataset. Due to the small sample size those values cannot be generalized. This directly provides an explanation for some players that may not seem to be ranked intuitively. While *Ersan Masovic* and *Ivan Ordets* from the *VfL Bochum* do probably not belong to the elite defenders in the Bundesliga, *Edmond Tapsoba* from *Bayer 04 Leverkusen* might be one of them. This statement is not backed by the provided PDD values but can be explained by looking at the game that is the basis for this. The *VfL Bochum* won the game with 3-0 and played more than 80 minutes with one player more on the field due to an early red card for Leverkusen's *Amine Adli.*

Overall, the presented results indicate that PDD could provide an accurate measure for defensive positioning and that it can be a good complement for existing defensive metrics since it differs conceptually from most of them. While metrics for observable defensive actions such as tackles or interceptions are outcome focused, PDD provides a measure that focuses on preventable danger and the potential of situations, which adds to a holistic approach for understanding the game.

6.2. Interpretation of the Evaluation Results

This section supplements the interpretation from the previous section, focuses on the results of the evaluation analysis presented in section 5.4 and interprets their implications for the proposed PDD metric. The three parts of the evaluation, inter-match stability, intra-match stability and external validation, will be analyzed and put together to provide insights into the reliability, sensitivity and practical interpretability of the metric. In addition, this section starts with a discussion of contextual effects that emerge from the empirical results, which are relevant for the interpretation of the proposed PDD metric.

Beyond individual positioning tendencies, the evaluation results suggest that PDD values are influenced by the team-level defensive context. Players embedded in compact and structurally stable defensive systems are exposed to fewer extreme defensive situations and individual positioning errors are more likely to be compensated by teammates. As a consequence, the observed preventable DAS tends to be lower for players from well-organized and dominant teams. This effect becomes apparent when comparing matches with different team dynamics, such as the example of *VfL Bochum* versus *Bayer 04 Leverkusen* discussed in section 6.1. While PDD is computed at the individual player level, the underlying DAS values are inherently shaped by collective defensive behavior and game context. Compact spacing, coordinated shifting and effective pressure reduce the overall accessibility of dangerous space, limiting the potential impact of suboptimal individual positioning. As a result, PDD may systematically favor players from teams with strong collective defensive structures, while players in less compact or more frequently exposed teams are confronted with a higher baseline of defensive risk. This observation does not contradict the design of the metric but highlights that individual defensive positioning is always embedded within a team context. Consequently, PDD values should be interpreted with awareness of tactical and structural team characteristics, particularly when comparing players across different teams.

Inter-Match Stability

Looking at the ICC values for both full games and halftime periods as independent entities, the aggregated PDD values can exhibit meaningful consistency across matches, particularly when a sufficient number of observations per player is available. As shown by the increase in ICC values for players with more observations, stability improves as random situational effects are averaged out. This is consistent with the interpretation in the previous section of PDD as a metric that captures recurring positional tendencies rather than isolated events.

At the same time, the observed variation in ICC values for smaller sample sizes highlights an important limitation. PDD values derived from a small number of matches should be interpreted with caution. Match-specific factors can strongly impact the resulting PDD values as we have potentially seen in the example of *Edmond Tapsoba*. The inter-match results, therefore, support the use of PDD when leveraging a dataset of sufficient size.

Intra-Match Stability

The intra-match stability analysis reveals a different but complementary aspect of the metric's behavior. While PDD values show a noticeable degree of consistency when aggregated at halftime period level, frame-level values exhibit substantial variation within individual matches. This pattern is not unexpected, as defensive positioning is inherently context-dependent. But it shows that at the frame level, the situation itself has more impact on the resulting PDD value than the player for which it is computed.

This might seem to indicate instability, but from a practical perspective, it is logical that any metric that aims to evaluate defensive positioning throughout a game is sensitive to situational changes. Taken together, both perspectives on intra-march stability indicate that PDD captures consistent and stable results over time, although frame-level values require careful interpretation and contextualization.

External Validation

The external validation is the third and last part of the evaluation and shifts the focus more to the domain of football, analyzing the relationship to established metrics. In general, it is difficult to find appropriate benchmarks for defensive positioning. As shown in section 2.5, distinct metrics for this are missing and as stated in sections 4.3 and 6.1 the general abilities of a player are not reflected in the PDD values with the underlying small dataset. Therefore, external validation is done with player ratings for the corresponding games, which are not solely based on defensive positioning and the contribution of defensive positioning to the overall ratings remains opaque and likely inconsistent. Consequently, the resulting correlations have to be analyzed carefully and can only be meaningful if defensive positioning is an important part of the role. Hence, in this thesis, the correlations are only considered for defensive roles.

Centre backs and defensive midfielders have a low correlation between PDD values and both external rating sources. This is an indication that the results of the metric align with general perception of player performances in the observed games. Especially for those central defensive positions, it is very plausible that the positioning influences the ratings without any explicit metric to be based on. For wing backs no correlation could have been observed. This might be explainable with the fact that the role of a wing back varies a lot and has commonly more differences in tactical instructions than the role of centre backs or defensive midfielders. Another reason that might add to this lack of correlation is the expected offensive outcome of wing backs. They are often more involved in offensive play, potentially leading to less focus on defensive positioning in the rating.

Summary

Taken together, the evaluation results indicate that PDD exhibits stability and consistency across multiple games and is sensitive for situational contexts during a game. A limited alignment with external rating sources highlights that PDD measures a distinct aspect of defensive performance. These findings support the interpretation of PDD as complementary metric that can enrich evaluation frameworks and complement to a wholistic view on defensive play. The evaluation also suggest that further analysis with larger datasets might be insightful.

6.3. Methodological Limitations

While the proposed PDD metric provides a novel perspective on defensive positioning, several methodological limitations must be acknowledged. These limitations arise from modeling choices, data constraints and the conceptual scope of the approach and should be considered when interpreting the results.

First, the optimization framework operates heuristically to find a local optimum. The defensive position is optimized within a restricted neighborhood and discrete search space. As a result, the model identifies locally optimal alternative positions. While this design ensures computational feasibility and preserves interpretability, it does not necessarily find the optimal position for a defender on the pitch.

Second, the optimization focuses on individual players only, without modeling team-level coordination. Interaction between defenders, such as collective shifting of defensive lines or compensatory movements by teammates, are not incorporated. While this aligns with the objective of isolating player-specific positioning potential, it necessarily abstracts from collective team behavior.

Another important limitation concerns the dependency on the objective function DAS. Since PDD is derived directly from DAS values, its behavior and interpretation are inherently tied to the assumptions and parameterization of the DAS model, for example the fact that DAS does not account for high passes in its current version. Alternative definitions of dangerous space could lead to different optimization outcomes and PDD distributions.

Furthermore, the metric is sensitive to game context, such as tactical setup and match state. As seen in the example in section 6.1, one red card can have a huge impact on the PDD values for that particular game. Therefore, a sufficient dataset is needed to avoid such outliers when trying to evaluate the general player ability in defensive positioning. At a fine-grained level, on a frame basis, PDD values can be very difficult to interpret in isolation. Since the value for a single frame depends heavily on the context, this level is often not suitable for statements regarding player quality in defensive positioning.

In addition, the robustness and stability of the metric depends on a very small sample size, yet. The observed values might not hold for larger sample sizes. However, the observed results show increasing robustness with increasing observation so far. The external validation is subject to limitations of the reference measures themselves. Player ratings are often influenced by single outstanding situations and subjective assessments. Consequently, the correlations should be interpreted as indicators rather than definitive validation.

Importantly, these limitations reflect deliberate modeling choices and trade-offs as well as potential work in the future rather than deficiencies of the approach. The proposed framework is designed to complement existing defensive metrics by adding the new perspective of preventable danger.

6.4. Practical Implications and Use Cases

Despite its methodological constraints, the proposed PDD metric offers several practical applications within football analytics, particularly in contexts where spatial positioning is of central interest. Rather than serving as a standalone performance indicator, PDD is best understood as a complementary analytical tool that provides additional insight into defensive positioning behavior under specific game conditions.

One primary use case of PDD might be in **player profiling and scouting**. By quantifying the relative amount of preventable dangerous space associated with a player's positioning, PDD can help identify systematic tendencies. When applied within positional groups, the metric allows the comparison of players with similar tactical roles and to highlight individual players whose positioning is consistently superior across situations. Another focus might be on specific game contexts at the frame level that can be identified as potential weakness and to find players that are particularly strong in those situations.

A second application concerns player development and coaching feedback. A frame-level analysis and aggregated PDD values can be used to illustrate how small positional adjustments may reduce opponent access to dangerous space. This makes the metric suitable for instructional settings, where the focus lies on spatial awareness and decision-making.

A further application might be opponent scouting. By aggregation PDD values across matches or phases of play, analysts can identify patterns how defensive structures respond to different attacking setups. While the current framework does not model team-level coordination, individual weakness might be identified by this approach.

Finally, the proposed framework offers potential for integration into larger analytical pipelines. PDD values can serve as input for higher-level models, such as player similarity analyses. Future extensions could incorporate team-level interactions or contextual adjustments, further enhancing the practical relevance of the approach.

In summary, the practical value of PDD lies in its ability to translate complex spatial information into an interpretable, role-specific indicator of positional improvement potential. When used alongside established metrics and qualitative analysis, it can support more nuanced assessments of defensive behavior in professional football.

7. Summary and Future Work

This thesis introduces a novel framework for defensive positioning in football by focusing on preventable dangerous space instead of observable defensive actions. Building on tracking data and the DAS model, an optimization-based approach is proposed to estimate how much of the opponent's Dangerous Accessible Space could have been reduced through improved positioning of a particular defender. The Preventable Defensive DAS (PDD) metric provides a role-specific, context-aware perspective on defensive behavior.

The methodological contribution of this work lies in the combination of spatial modeling and constrained optimization at the frame level. By evaluating locally optimal positioning alternatives under fixed game conditions, the framework isolates positional improvement potential in a transparent and interpretable manner. The proposed aggregation strategy enables meaningful player-level comparisons while accounting for differences in baseline danger across situations.

Empirical results demonstrate that PDD values exhibit structured variation across positional groups, reflecting differences in defensive roles and spatial responsibilities. Qualitative frame-level examples illustrate how small positional adjustments can lead to measurable changes in DAS, supporting the interpretability of the metric. Stability analyses show that aggregated PDD values can exhibit meaningful consistency across matches given sufficient data, while intra-match variation reflects the situational nature of defensive positioning. Comparisons with external player ratings indicate that PDD captures aspects of defensive behavior that are not fully represented in traditional evaluation systems.

Several limitations of the proposed approach are discussed, including the reliance on local optimality assumptions and the absence of explicit team-level coordination modeling. These limitations highlight the trade-offs inherent in designing an interpretable and computationally feasible metric but do not diminish the contribution of the framework as a complementary analytical tool.

Future work could extend the proposed approach in multiple directions. From a methodological perspective, incorporating team-level optimization or coordinated defensive moments could provide a more holistic representation of defensive structures. Temporal extensions that account for sequential decision-making or anticipation may further enhance the realism of the model. In addition, alternative definitions of dangerous space could be explored to assess the robustness of PDD under different spatial modeling assumptions.

From an applied perspective, future research may investigate the integration of PDD into a broader analytical pipeline, such as a player development tracking or a tactical scouting system. Combining the metric with contextual information, including match state or opponent characteristics, may further improve its interpretability and practical relevance.

In conclusion, this thesis demonstrates that optimization-based analysis of defensive positioning offers a viable and insightful complement to existing football analytics methodologies. By shifting the focus from realized actions to the potential threat of situations, the proposed framework contributes a more nuanced understanding of defensive performance in modern football.

Bibliography

Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., Von Landesberger, T., Weber, H.: Visual analysis of pressure in football. Data Mining and Knowledge Discovery **31**(6), 1793–1839 (2017) `https://doi.org/10.1007/s10618-017-0513-2` . Accessed 2026-02-01

Bauer, P., Anzer, G.: Data-driven detection of counterpressing in professional football: A supervised machine learning task based on synchronized positional and event data with expert-based feature extraction. Data Mining and Knowledge Discovery **35**(5), 2009–2049 (2021) `https://doi.org/10.1007/s10618-021-00763-7` . Accessed 2026-02-01

Baouan, A.: Statistical analysis of team formation and player roles in football. arXiv. arXiv:2502.03342 [stat] (2025). `https://doi.org/10.48550/arXiv.2502.03342` . `http://arxiv.org/abs/2502.03342` Accessed 2025-11-02

Bischofberger, J., Baca, A.: Dangerous Accessible Space: A Unified Model of Space and Value in Team Sports. In Review (2025). `https://doi.org/10.21203/rs.3.rs-6932689/v1` . `https://www.researchsquare.com/article/rs-6932689/v1` Accessed 2026-01-25

Bischofberger, J., Baca, A., Schikuta, E.: Event detection in football: Improving the reliability of match analysis. PLOS ONE **19**(4), 0298107 (2024) `https://doi.org/10.1371/journal.pone.0298107` . Accessed 2024-11-20

Bischofberger, J.: Github jonas-bischofberger/accessible-space (2025). `https://github.com/jonas-bischofberger/accessible-space` Accessed 2025-11-02

Baladram, M.S., Koike, A., Yamada, K.D.: Introduction to Supervised Machine Learning for Data Science. Interdisciplinary Information Sciences **26**(1), 87–121 (2020) `https://doi.org/10.4036/iis.2020.A.03`

Bassek, M., Rein, R., Weber, H., Memmert, D.: An integrated dataset of spatiotemporal and event data in elite soccer. Scientific Data **12**(1), 195 (2025) `https://doi.org/10.1038/s41597-025-04505-y` . Accessed 2025-10-15

Bundesliga: Bundesliga (2024). `https://aws.amazon.com/de/sports/bundesliga/` Accessed 2024-11-08

Chatziparaskevas, P., Saprikis, V., Antoniadis, I.: The impact of information systems and data science on management in modern professional football: Moneyball theory and the development model of Brentford FC, Aizuwakamatsu, Japan, p. 050011 (2024). https://doi.org/10.1063/5.0237053 . https://pubs.aip.org/aip/acp/article-lookup/doi/10.1063/5.0237053 Accessed 2024-11-06

Delgado-Bordonau, J., Doménech-Monforte, C., Guzmán, J., Mendez-Villanueva, A.: Offensive and defensive team performance: Relation to successful and unsuccessful participation in the 2010 Soccer World Cup. Journal of Human Sport & Exercise **8**, (2013) https://doi.org/10.4100/jhse.2013.84.02

Decroos, T., Bransen, L., Van Haaren, J., Davis, J.: Actions Speak Louder than Goals: Valuing Player Actions in Soccer. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1851–1861. ACM, Anchorage AK USA (2019). https://doi.org/10.1145/3292500.3330758 . https://dl.acm.org/doi/10.1145/3292500.3330758 Accessed 2024-11-13

Eigenrauch, S., Bischofberger, J., Schikuta, E.: A Data Science Approach for Predicting Soccer Passes Using Positional Data (2024)

Ehrenberg-Silies, S.: Big Data und KI im Fußball (2024)

Fan, M., Liu, F., Huang, D., Zhang, H.: Determinants of international football performance: Empirical evidence from the 1994–2022 FIFA World Cup. Heliyon **9**(10), 20252 (2023) https://doi.org/10.1016/j.heliyon.2023.e20252 . Accessed 2026-01-25

Freitas, R., Volossovitch, A., Almeida, C.H., Vleck, V.: Elite-level defensive performance in football: a systematic review. German Journal of Exercise and Sport Research **53**(4), 458–470 (2023) https://doi.org/10.1007/s12662-023-00900-y . Accessed 2024-11-06

Hewitt, A., Greenham, G., Norton, K.: Game style in soccer: what is it and can we quantify it? International Journal of Performance Analysis in Sport **16**(1), 355–372 (2016) https://doi.org/10.1080/24748668.2016.11868892 . Accessed 2024-11-10

Jiang, H.: Machine Learning Fundamentals: A Concise Introduction, 1st edn. Cambridge University Press, Cambridge (2021). https://www.cambridge.org/core/product/identifier/9781108938051/type/book Accessed 2024-11-15

Koo, T.K., Li, M.Y.: A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine **15**(2), 155–163 (2016) https://doi.org/10.1016/j.jcm.2016.02.012 . Accessed 2026-01-13

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Advances in Neural Information Processing Systems, vol. 30, pp. 3149–3157. Curran Associates, Inc., Long Beach, California, USA (2017)

Kempe, M., Vogelbein, M., Nopp, S.: The Cream of the Crop: Analysing FIFA World Cup 2014 and Germany's Title Run. Journal of Human Sport and Exercise **11**(1) (2016) `https://doi.org/10.14198/jhse.2016.111.04` . Accessed 2024-11-10

Lee, M., Jo, G., Hong, M., Bauer, P., Ko, S.-K.: exPress: Contextual Valuation of Individual Players Within Pressing Situations in Soccere. (2025)

Merhej, C., Beal, R., Ramchurn, S., Matthews, T.: What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data | Enhanced Reader (2024) `https://doi.org/10.48550/arxiv.2106.01786`

Mead, J., O'Hare, A., McMenemy, P.: Expected goals in football: Improving model performance and demonstrating value. PLOS ONE **18**(4), 0282295 (2023) `https://doi.org/10.1371/journal.pone.0282295` . Accessed 2024-11-12

Merckx, S., Robberechts, P., Euvrard, Y., Davis, J.: Measuring the Effectiveness of Pressing in Soccer (2021)

Müller, O., Simons, A., Weinmann, M.: Beyond crowd judgments: Data-driven estimation of market value in association football. European Journal of Operational Research **263**(2), 611–624 (2017) `https://doi.org/10.1016/j.ejor.2017.05.005` . Accessed 2025-11-17

Oonk, A.: Welcome to DataBallPy! — DataBallPy Documentation (2025). `https://databallpy.readthedocs.io/en/latest/index.html` Accessed 2025-10-31

Pearson, K.: VII. Note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London **58**(347-352), 240–242 (1895) `https://doi.org/10.1098/rspl.1895.0041` . Accessed 2026-01-13

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: CatBoost: unbiased boosting with categorical features. arXiv. arXiv:1706.09516 (2019). `http://arxiv.org/abs/1706.09516` Accessed 2024-11-13

Piersma, J.P.T., Koning, D.A.J., Supervisor, S., Supervisor, S., Bransen, D.L.: Valuing Defensive Performances of Football Players (2020)

Renkin, M., Bischofberger, J., Schikuta, E., Baca, A.: Validation and Optimisation of Player Motion Models in Football. In: Groen, D., De Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds.) Computational Science – ICCS 2022 vol. 13351, pp. 26–32. Springer, Cham (2022). https://link.springer.com/10.1007/978-3-031-08754-7 Accessed 2024-11-20

Rein, R., Memmert, D.: Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. SpringerPlus **5**(1), 1410 (2016) `https://doi.org/10.1186/s40064-016-3108-2` . Accessed 2024-11-08

Santolini, N.: What are we doing with defensive stats? (2022). https://medium.com/@nicola.santolini.94/what-are-we-doing-with-defensive-stats-29fabb130c87 Accessed 2026-01-25

Spearman, W., Basye, A., Dick, G., Hotovy, R., Pop, P.: Physics-Based Modeling of Pass Probabilities In Soccer, (2017)

Shrout, P.E., Fleiss, J.L.: Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin **86**(2), 420–428 (1979) https://doi.org/10.1037/0033-2909.86.2.420 . Accessed 2026-01-13

Singh, K.: Introducing Expected Threat (xT) (2024). https://karun.in/blog/expected-threat.html Accessed 2024-11-06

Santos, P.M., Lago-Penas, C.: Defensive positioning on the pitch in relation with situational variables of a professional football team during regaining possession. Human Movement **20**(2), 50–56 (2019) https://doi.org/10.5114/hm.2019.81019 . Accessed 2024-11-06

Spearman, C.: The Proof and Measurement of Association between Two Things. The American Journal of Psychology **15**(1), 72 (1904) https://doi.org/10.2307/1412159 . Accessed 2026-01-13

UEFA.com: Ballon d'Or laureates: Who has won football's most prestigious award? | Ballon d'Or 2024/25. Section: Football (2025). https://www.uefa.com/ballondor/news/0292-1c27056169a7-bc88d1fdba43-1000--ballon-d-or-laureates-who-has-won-football-s-most-prestig/ Accessed 2025-11-17

Wright, C., Carling, C., Collins, D.: The wider context of performance analysis and it application in the football coaching process. International Journal of Performance Analysis in Sport **14**(3), 709–733 (2014) https://doi.org/10.1080/24748668.2014.11868753 . Accessed 2024-11-10

Winter, C., Pfeiffer, M.: Tactical metrics that discriminate winning, drawing and losing teams in UEFA Euro 2012®. Journal of Sports Sciences **34**(6), 486–492 (2016) https://doi.org/10.1080/02640414.2015.1099714 . Accessed 2024-11-10

Wu, Y., Swartz, T.: Evaluation of off-the-ball actions in soccer. Statistica Applicata - Italian Journal of Applied Statistics (2) (2023) https://doi.org/10.26398/IJAS.0035-008 . Accessed 2024-11-06

Acronyms

DAS Dangerous Accessible Space. iii, v, 2, 3, 15, 17, 20–24, 26–31, 33–37, 41, 42, 44, 46–48, 53–55, 57, 61

DAxT Defensive Action Expected Threat. 12, 14, 16

DFL German Football League. 11, 21

EPV Expected Possession Value. 9, 11, 14

ICC Intraclass Correlation Coefficient. 38, 39, 49–51, 54–56

PDD Preventable Defensive DAS. iii, v, 21, 30, 31, 34, 35, 38–51, 53–59, 61, 62

xA expected Assists. 9, 11

xG expected Goals. 1, 2, 9, 11, 15, 28, 29, 45

xT expected Threat. 1, 9, 11, 12, 47

A. Appendix

A.1. Full Positional Rankings by PDD

Rank	Player	PDD(%)	Number of Frames
1	Erhan Mašović	6.56%	82
2	Ivan Ordets	7.80%	82
3	Adam Dźwigała	7.89%	73
4	James Alexander Lawrence	7.93%	144
5	Eric Anders Smith	8.49%	73
6	Kangni Frederic Ananou	8.98%	84
7	Steve Breitkreuz	9.03%	86
8	Lukas Thomas Fröde	9.17%	90
9	Timo Bernd Hübers	9.20%	127
10	Julian Jeffrey Gaston Chabot	9.70%	127
11	Betim Fazliji	9.81%	51
12	Jordy Adriana Jozefina de Wijs	10.00%	79
13	Dayotchanculle Oswald Upamecano	10.28%	91
14	Tim Christopher Oberdorf	10.28%	328
15	Boris Tomiak	10.38%	182
16	Christoph Klarer	11.06%	345
17	Matthias Jürgen Zimmermann	11.09%	124
18	Christopher Wolfgang Georg August Schindler	11.37%	144
19	Jonathan Glao Tah	11.46%	97
20	Robin Bormuth	11.49%	182
21	Jan Elvedi	11.53%	86
22	Matthijs de Ligt	13.15%	91
23	André Hoffmann	13.56%	64
24	Rick van Drongelen	13.61%	90
25	Kevin Kraus	16.09%	78
26	Edmond Fayçal Tapsoba	20.30%	97

Table A.1.: Centre Backs by PDD

Rank	Player	PDD(%)	Number of Frames
1	Luca-Milan Zander	2.60%	12
2	Kingsley Schindler	2.82%	32
3	Kristian Majdahl Pedersen	2.98%	19
4	Manolis Saliakas	4.56%	61
5	Leon Guwara	5.49%	67
6	Noussair Mazraoui	5.53%	85
7	Nico Neidhart	5.76%	90
8	Fabian Nürnberger	6.01%	110
9	Jeremie Agyekum Frimpong	6.07%	97
10	Leart Paqarada	6.10%	73
11	Jonas Armin Hector	6.17%	127
12	Konrad Faber	6.28%	86
13	Benno Schmitz	6.32%	95
14	Matthias Jürgen Zimmermann	7.14%	283
15	Dominik Friedrich Schad	7.31%	104
16	Benjamin Jacques Marcel Pavard	7.50%	91
17	Nicolas Gavory	7.60%	91
18	Michal Karbownik	7.96%	204
19	Anderson-Lenda Lucoqui	8.08%	90
20	Dominique Heintz	8.13%	54
21	Enrico Valentini	8.22%	22
22	Jan Kwasi Gyamerah	8.44%	122
23	Erik Durm	9.06%	182
24	Keven Schlotterbeck	9.67%	28
25	Mitchel Bakker	9.71%	97
26	Scott Fitzgerald Kennedy	11.14%	19
27	Joao Pedro Cavaco Cancelo	12.08%	30
28	Saidy Janko	16.85%	82

Table A.2.: Wing Backs by PDD

Rank	Player	PDD(%)	Number of Frames
1	Kevin Stöger	3.97%	79
2	Patrick Osterhage	4.59%	28
3	Ao Tanaka	4.69%	132
4	Maximilian Thalhammer	5.06%	86
5	Anthony Losilla	5.26%	82
6	Marcel Sobottka	6.12%	283
7	Afeez Aremu Olalekan	6.32%	61
8	Ryan Jiro Gravenberch	6.53%	76
9	Eric Martel	6.84%	127
10	Benedikt Steffen Gimber	6.86%	86
11	Ellyes Joris Skhiri	6.91%	127
12	Johannes Geis	7.62%	122
13	Marlon Maurice Ritter	8.28%	182
14	Adam Bodzek	8.47%	25
15	Jorrit Petrus Carolina Hendrix	8.58%	281
16	Joshua Walter Kimmich	8.63%	91
17	Exequiel Alejandro Palacios	9.17%	97
18	Dennis Dressel	9.63%	90
19	Nadiem Amiri	9.97%	84
20	Michal Karbownik	10.30%	124
21	Hikmet Çiftçi	11.73%	104
22	Sadik Fofana	12.99%	22

Table A.3.: Defensive Midfielders by PDD

Rank	Player	PDD(%)	Number of Frames
1	Elione Fernandes Neto	1.76%	12
2	Denis Huseinbašić	2.56%	32
3	Andreas Albers Nielsen	2.73%	84
4	Dawid Igor Kownacki	3.20%	75
5	Florian Kainz	3.83%	95
6	Shinta Karl Appelkamp	3.86%	213
7	Thomas Müller	3.96%	61
8	Philipp Förster	4.27%	54
9	Dong-gyeong Lee	4.54%	43
10	Marcel Hartel	4.57%	65
11	Manuel Paul Wintzheimer	4.74%	34
12	Mike Wunderlich	5.37%	151
13	Jackson Alexander Irvine	5.66%	73
14	Florian Richard Wirtz	5.91%	97
15	Svante Ulf Ingel Ingelsson	6.82%	53
16	Philipp Klement	7.86%	31
17	Lino Tempelmann	8.13%	144
18	Sebastien Thill	9.17%	47
19	Adam Hložek	10.13%	13

Table A.4.: Offensive Midfielders by PDD

Rank	Player	PDD(%)	Number of Frames
1	Shinta Karl Appelkamp	2.74%	22
2	Elione Fernandes Neto	3.32%	13
3	Leroy Aziz Sané	3.67%	91
4	Kwadwo Kyeremeh Baah	3.77%	11
5	Pascal Breier	3.83%	37
6	Dejan Ljubicic	3.86%	108
7	Linton Maina	4.36%	108
8	Emmanuel Iyoha	4.37%	225
9	Connor Isaac Metcalfe	4.57%	12
10	Mats Møller Dæhli	4.66%	102
11	Moussa Diaby	4.95%	96
12	Kenny Prince Nebil Hussein Redondo	4.99%	172
13	Felix Klaus	5.27%	255
14	Nicklas Shipnoski	5.35%	78
15	Jens Castrop	5.48%	144
16	Joshua Mees	5.64%	84
17	Kristoffer Paul Peterson	5.91%	276
18	Aaron Opoku-Tiawiah	6.77%	78
19	Jean Zimmer	7.06%	182
20	Erik Wekesser	7.84%	42
21	Kingsley Junior Coman	7.98%	76
22	Daniel Patrick Hanslik	15.69%	10

Table A.5.: Wingers by PDD

Rank	Player	PDD(%)	Number of Frames
1	Eric Maxim Choupo-Moting	0.61%	15
2	Tim Lemperle	1.16%	19
3	Jan Uwe Thielmann	1.49%	32
4	Daniel Ginczek	1.81%	75
5	David Otto	1.89%	12
6	Philipp Hofmann	2.13%	74
7	Aygün Yildirim	2.25%	78
8	Kwadwo Kyeremeh Baah	2.39%	38
9	John Verhoek	2.42%	84
10	Terrence Anthony Boyd	2.58%	165
11	Serge David Gnabry	2.99%	91
12	Davie Selke	3.29%	95
13	Etienne Amenyido	3.34%	65
14	Takuma Asano	3.37%	74
15	Kwadwo Antwi Duah	3.41%	102
16	Emmanuel Iyoha	3.74%	25
17	Dawid Igor Kownacki	3.78%	265
18	Rouwen Hennings	3.89%	191
19	Kai Pröger	3.92%	84
20	Lukas Stephan Horst Daschner	4.72%	61
21	Christoph Daferner	4.92%	144
22	Christopher Antwi-Adjei	5.08%	82
23	Lukas Schleimer	5.11%	42
24	Sardar Azmoun	5.17%	97
25	Lex-Tyger Lobinger	5.39%	17

Table A.6.: Forwards by PDD