Degree in Mathematics

Degree in Engineering Physics

Bachelor's Degree Thesis

Data-driven identification of optimal substitutions in soccer

Joan Hernanz i Ibáñez

Director (KU Leuven): Jesse Davis

Tutor (UPC): Enrique Romero Merino

May 2023

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Centre de Formació Interdisciplinària Superior

KU LEUVEN

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona

telecos BCN

Abstract

Data-driven football analytics is a rising field, and clubs are spending more and more resources on gaining a competitive edge through data-driven techniques. Substitutions are the main tool a coach has to intervene in the course of the game, and their limitation and relevance have attracted interest to their study. In this thesis, we seek a data-informed approach to identifying and predicting optimal substitutions.

With the recent change in football legislation, now each team is permitted up to five substitutions per match. We compare the new paradigm with the prior one and look for optimal substitutions with the use of data-based models. With machine learning classifiers and a substitution sensible in-game win probability model, player changes have been assessed. The same models have been used for simulating alternative types of substitution and timing, in order to make data-driven approaches that increase the chances of success.

The addition of extra substitutions has resulted in an increase in tactical interventions by coaches, but the match dynamics have remained the same. Machine learning models obtain good results in the prediction of substitution assessment. The win probability model is sensible to substitutions, which have generally a better effect on the substituting team. Offensive substitutions generally increase the winning probabilities, especially for losing teams. No timing is observed to be significantly better for doing substitutions, but in particular cases can be very relevant.

Keywords: Football Analytics, Football, Big Data, Substitutions, Machine Learning, Win Probability Model, Bayesian Model, Data-Driven Decisions.

American Mathematical Society 2020 Mathematics Subject Classification: 62-07, 60E99

Resum

L'anàlisi basada en dades del futbol és un camp en auge, i els clubs dediquen cada vegada més recursos a obtenir un avantatge competitiu mitjançant tècniques basades en dades. Les substitucions són la principal eina de la qual disposa un entrenador per a intervenir en el transcurs del joc, i la seva limitació i rellevància han despertat l'interès pel seu estudi. En aquesta tesi busquem un enfocament analític per a la identificació i predicció de substitucions òptimes.

Amb el recent canvi en la legislació futbolística, ara es permeten a cada equip fins a cinc substitucions per partit. Comparem el nou paradigma amb l'anterior i busquem les substitucions òptimes amb l'ús de models basats en dades. Amb classificadors d'aprenentatge automàtic i un model de probabilitat de victòria en directe sensible a les substitucions, s'han avaluat els canvis de jugadors. Els mateixos models s'han utilitzat per a simular substitucions de diferent tipus i en moments alternatius, amb la finalitat de realitzar plantejaments basats en dades que augmentin les probabilitats d'èxit.

L'addició de canvis addicionals ha provocat un augment de les intervencions tàctiques dels entrenadors, però la dinàmica del partit es manté igual. Els models d'aprenentatge automàtic obtenen bons resultats en la predicció de la valoració de les substitucions. El model de probabilitat de victòria és sensible a les substitucions, que solen tenir un millor efecte per a l'equip que les realitza. Les substitucions ofensives solen augmentar les probabilitats de victòria, especialment per als equips que van perdent. No s'observa que cap moment sigui significativament millor per a fer el canvi, però en casos particulars pot ser rellevant.

Paraules clau: Analítica de Futbol, Futbol, Big Data, Aprenetatge Automàtic, Model de Probabilititat de Victoria, Model Bayesià, Decisions Basades en Dades.

Resumen

El análisis basado en datos del fútbol es un campo en auge, y los clubes dedican cada vez más recursos a obtener una ventaja competitiva mediante técnicas basadas en datos. Las sustituciones son la principal herramienta de la que dispone un entrenador para intervenir en el transcurso del juego, y su limitación y relevancia han despertado el interés por su estudio. En esta tesis buscamos un enfoque analítico para la identificación y predicción de sustituciones óptimas.

Con el reciente cambio en la legislación futbolística, ahora se permiten a cada equipo hasta cinco sustituciones por partido. Comparamos el nuevo paradigma con el anterior y buscamos las sustituciones óptimas con el uso de modelos basados en datos. Con clasificadores de aprendizaje automático y un modelo de probabilidad de victoria en directo sensible a las sustituciones, se han evaluado los cambios de jugadores. Los mismos modelos se han utilizado para simular sustituciones de distinto tipo y en momentos alternativos, con el fin de realizar planteamientos basados en datos que aumenten las probabilidades de éxito.

La adición de cambios adicionales ha provocado un aumento de las intervenciones tácticas de los entrenadores, pero la dinámica del partido se mantiene igual. Los modelos de aprendizaje automático obtienen buenos resultados en la predicción de la valoración de las sustituciones. El modelo de probabilidad de victoria es sensible a las sustituciones, que suelen tener un mejor efecto para el equipo que las realiza. Las sustituciones ofensivas suelen aumentar las probabilidades de victoria, especialmente para los equipos que van perdiendo. No se observa que ningún momento sea significativamente mejor para hacer el cambio, pero en casos particulares puede ser relevante.

Palabras clave: Analítica de Fútbol, Fútbol, Big Data, Arendizaje Automático, Modelo de Probabilidad de Victoria, Modelo Bayesiano, Decisiones Basadas en Datos.

Acknowledgements

First of all, I would like to thank Jesse Davis for his support, for always finding a moment to talk and guide me with his ideas and overall clear picture, and Maaike Van Roy, for the weekly feedback and always encouraging my work. I have very much enjoyed working in the football analytics topic with such a world-leading group. Thanks to KU Leuven for offering all of the services of one of the most prestigious universities in mainland Europe.

Gracias a Enrique Romero, mi tutor por correspondencia electrónica, por toda la ayuda, correciones y ánimo a distancia. Gràcies en especial al CFIS, per la confiança i l'ajuda econòmica durant tota la carrera, i per oferir-me l'oportunitat de passar un any on volia investigant el que m'agrada.

Gràcies als que heu fet que la meva estància a Bèlgica valgui molt la pena: Leuven Team, la Resistencia i els Earls of Leuven. I als que m'heu acompanyat tota la carrera: els matfis, i la gent de la FME i EF i la gent del Thau. Una abraçada molt forta pel Marc, que sé que estas molt orgullós de tots nosaltres. Gràcies a l'Andreu pels memes, i pare, mare i familia en general per les visites i videotrucades absurdes. Gràcies, en especial, a la Martina, que hi cap en totes les categories i que es mereix molt més que aquesta frase.

Joan, Leuven, May 2023

Contents

Chapter 1

Introduction

1.1 Problem Description

Football is the most important sport in the world, with fans all over the globe tuning in every day to follow games, and shows discussing everything related to those games. With the increasing amount of data being generated in football, the industry has started promoting data-driven and data-informed [13] techniques to optimize each and every task.

In recent years, football analytics has emerged as a powerful tool for understanding the game and gaining a competitive advantage [30]. While the use of analytics in sports is not new, it has been most prominently associated with isolated events and higher-scoring games (such as baseball or basketball). Data analysis has been used for decades in these sports to evaluate players and teams, inform strategy, and make informed decisions [20]. However, the traditional mindset of football coaches and fans, combined with the naturally random spirit of football due to its low-scoring nature, has often made football resistant to the use of data and statistics in the sport. In contrast, many football analysts and experts believe that the use of analytics is essential for understanding the game and improving performance. The success of teams that have embraced analytics in all areas provides compelling evidence for the value of this approach.

Substitutions, in soccer, are important resources, because they alter the ability to change tactics [23], which can often directly influence the final outcome of the game. Through substitutions, the coach explicitly defines what his intention is in relation to the game, and can show their ability to maneuver mid-match.

Nowadays, coaches can perform up to five substitutions per match, in a total of three windows (and possibly at half-time) [19]. Coaches have to use these substitutions for possible injuries, fatigued players, protection for booked players and, potentially, change the course of a match through a tactical or player change. In this thesis, we will focus on the latter. Changes in formation, playing style and even a team's mentality can be accomplished without substituting a player. We are interested in identifying optimal substitutions and knowing the effectiveness of substitutions, according to the progress of the match.

The type of substitution, whether it must change the team formation of style, picking the correct player and being spot on with timing are very important for a coach to accomplish their objective. Accomplishing this can result in altering the outcome if they are losing, or maintaining the victory if they are already winning. Multiple strategies can be followed, since tactical decisions are simultaneously factors for the team stabilization and the opposition's destabilization ([12], [4], [28], [14]).

1.2 Research Questions

Three substitutions per team and per match were the rule until a few years ago. With the COVID break, many leagues accepted that teams could do up to five substitutions per match, due to schedule overload. Even though some major leagues went back to the three substitutions per match, from this season (2022/23) all major leagues in Europe and continental competitions have opted to allow five changes per match [21]. Since there is no apparent extra match load, we pose the following question.

Research Question 1 *Has the new 5-substitutions per match rule made that more substitutions become relevant? Do coaches have more ways to maneuver?*

We will focus on this question from an observational point of view, which will be addressed in section 3.3, by comparing of the last complete season previous to COVID with the seasons that have followed that allowed five substitutions per match.

Next, we will focus on assessing substitutions. This key part of the investigation is the base for the whole Thesis, as investigating optimal substitutions depends on how you define them. We pose the following research question.

Research Question 2 *How can we determine if a substitution is useful? What is the performance of the subbed player in comparison to the starting one? Do substitutions have an effect on a team's win probability?*

The question will be addressed in chapter 4. Different evaluation methods will be discussed and analyzed, including changes in result, match momentum or in substitute performance. Machine learning models will be trained for the prediction of such optimal substitutions. A model based on [29] will be developed, using substitution information to calculate the win probability during the match and discuss how sensible the model is to such features. The models developed in this chapter will be used during the course of this report.

The next topic that will be addressed is the selection of the adequate player to come in. Given the moment at which a coach decides to introduce some kind of modification to the team's players, we can use data to respond to the following question.

Research Question 3 *How can we select the optimal substitute? Which is the right type of substitution? Which bench player is expected to produce a better performance?*

To address the problem, in chapter 5 we will study what can be expected from the players on the bench, with the intention to identify the optimal pick from those the coach can choose. The type of substitution and which particular bench player are the two main decisions that a data-informed approach can better the teams' chances.

Last but not least, we will focus on timing. Timing is of the utmost importance since a correct change at the correct moment can modify the course of the match. Thus, we will formulate the last research question.

Research Question 4 *What is the optimal moment for a substitution?*

This last question will be addressed in chapter 6, and we will use the models developed previously in this report. We can focus on what is the time distribution of the more beneficial substitutions, or given the game circumstances, what do the models predict as a better time frame for doing a substitution. We will make simulations on the win probability model to find the minute here a substitution gives the team the biggest advantage.

As of the rest of the thesis, chapter 2 is an overview of publications in the substitution problem, as well as a review of relevant results in football analytics. In chapter 3 the data used for the models is discussed, with analysis at the player and team level, and also the study of Research question 1. Finally, chapter 7 discusses the results obtained through the report.

Chapter 2

Background

2.1 Introduction to Football Analytics

Over the last decades, the growing interest in football has produced an increasing amount of research in football analytics. Trying to see more in-depth what simple statistics such as shots and shots on target could offer, a new metric was developed, expected goals (xG) [15]. Expected goals assign the probability a shot has to become a goal, based on the location of the shooter, body part, goalkeeper positioning and many other variables. xG models have been built and renewed by many companies, in private and public, and continue to be upgraded continuously [35] as the quantity and quality of data increases. Thus, by assessing the number of expected goals, instead of observed ones, analysts have been able to better predict future performances. More recently, xG has broken into the public scene, being now a common term between football pundits on TV shows and during matches. Expected assists (xA) measures the quantity of assists a player should expect from the quality of their passes, assuming an average goal conversion from their teammates.

To value not only goalscorers but the rest of the players, many metrics have been developed. Karun Singh [32] introduced the concept of expected threat (xT), which assigns to every zone of the pitch a probability that the attack ends successfully in a goal. Thus, by seeing where the players moved the ball, analysts can see which impact they actually had in the building of an offensive play. Other metrics, such as VAEP [11], measure all actions on the ball and give them a value that relates to the goal probability increase, and takes into account the probability of the opposing team scoring. This allows analysts to value every on-ball action of a player, including defensive ones and therefore gives a more complete model to rate players in terms of their impact in a game.

Football is a very complex game, with multiple situations happening at the same time, not all captured by the data or the camera, and a very big amount of factors that impact the game. This is why it has been discussed [13] that a data-driven approach to football might not be sufficient. Instead, a data-informed one where all the metrics and results from big-data trained machine learning models are used together with professional expertise, would be useful to increase performance.

2.2 The Substitution Problem

Substitution analysis is emerging, but there is limited literature on the field [16]. Many papers addressing this problem centered on the physiological aspect of football players. This thesis will try to focus on the tactical approach. The influence of substitutions in the development of matches has been proven ([1], [23]) and how the game state gives preference to certain kinds of substitutions ([12], [4], [28], [14]). Machine learning techniques have been applied for the identification of optimal substitutions ([7]) and optimal timing both by modeling football matches ([17], [3]) and by identifying successful patterns of substitutions ([26], [31]). In this section, we will describe in detail the results obtained by the different studies and review their techniques and results.

Game modeling for optimal substitution identification

Timing on substitutions was first addressed by Hirotsu and Wright [17], using a Markov model to model a match, and studying how different players affected the transition probabilities between states. Therefore, they used dynamic programming to find the optimal substitution time for a certain substitution depending on the match scoreline.

On a similar line, Beal et al. [3], modeled a football match in two distinct ways. First, a pre-match Bayesian game, where a game consists of two teams selecting tactics with a certain probability and a payoff function. Second, an in-game stochastic model which defines a game as two teams, with the corresponding set of strategies, the game states, the transition probabilities between those states and a payoff function. A set of 760 matches is used to train the transition probabilities, and substitutions are modeled as different strategies which change such probabilities, therefore trying to select strategies to obtain a result depending on the approach a team wants to have.

Substitutions' effect on result

When trying to assess the impact of substitutions in the final match result, Myers [26] presented a substitution rule, known as the 58-73-79 rule. The results, which made it to the mainstream media, said that losing teams should do the first substitution before the 59$^{\text{th}}$ minute, the second before the 73$^{\text{rd}}$ and the last before the 79$^{\text{th}}$. This result is based on the number of times that teams following this rule were able to reduce the goal difference. It does not give any rule for winning teams or drawn matches.

As Silva and Swartz discussed in [31], the paper by Myers has some flaws such as not considering the opposition strategy. With a different approach, by using Bayesian logistic regression, they showed that there is no apparent optimal time to do the substitutions, since the coefficients of such regression do not show any significant variation with time. Substitutions have been observed to change the goal intensity or the probability of goal scoring [1]. The first and second own substitutions increase the probability of scoring a goal, but the opposition's third diminishes it.

It is clear that substitutions affect the result, so clinical timing is important. Rey, Lago-Ballesteros and Padrón-Cabo [28], proved that the timing of substitutions is related to the current scoreline of the match. With their analysis of Champions League matches, observations are that winning teams tend to substitute later, and they claim that it should be suggested to coaches that reverting losing scenarios requires changes in tactics early in the match. For the first substitution, it has been observed that the most important factor is the scoreline as it stands before the time of the player substitution, as shown by Del Corral, Barros and Prieto-Rodriguez [12], through their study of the substitution patterns in the *Primera División de España* 04-05, where they also conclude that defensive substitutions happen later in matches than offensive ones.

Most of the subs are made either at half-time or between the 60^{th} and 85^{th} ([4], [14]). When studying which players are involved in changes, it is observed that the majority of substitutions involve midfielders or attacking players, and substitutions do depend on contextual-related variables of the match [14]. All studies agree that coaches are able to alter the final outcome of the match and the playing tactics. When the scoreline is negative, teams make changes more rapidly.

Prediction of substitutions

Machine learning techniques have been used to predict the efficiency of substitutions. Brutti, Duarte and Del Bianco studied the Brazilian Serie A (2015-2018) [7] and used techniques such as k-Nearest Neighbours, Decision Trees, Random Forests and Support Vector Machines to predict the usefulness of substitutions based on match features such as goals scored, offensive or defensive substitutions, time, team strength and scoreline. The assessment is based on the result of the match. The obtained results were reasonably good and suggested that this could be a promising analysis of the best moment to replace a player.

Different types of substitutions are also a field where prediction is also relevant. According to player roles, Lorenzo-Martinez et al. [23] define defensive, neutral and offensive substitutions and discuss the differences between them. Their results showed that there exist discernible results between the three kinds of substitutions, including offensive substitutions to move the teams' centroid forward and defensive substitutions reducing the teams' stretch. Thus, concluding that substitutions have a palpable effect on tactical terms.

Physical performance

From a physical point of view, and concerning the performance of players when substituted as opposed to when they are starting, Bradley et al. [4] observed that substitute players covered more distance and ran more distance at high intensity, with similar results to those found by other studies ([25], [8], [9], [5], [36]). Thus, the conclusion is clear and, from a physical point of view, the match performance of substitutes is higher than starting players in the ending part of the match.

The inclusion of two extra possible substitutions due to COVID has multiplied the options that coaches have. When, in the nineties, a third substitution was included, this led to an increase in tactical substitutions [34]. In the last years, it had been discussed the options of adding extra substitutions, not just in overtime but in normal time too [22]. Additional substitutions have been observed to reduce significantly player load through one match, and less so through the season, in competitions where the change rule was already implemented [24].

2.3 Related Work

Research in football analytics is much broader than analyzing successful substitutions. Following are interesting papers with ideas that will be brought up during the thesis.

Win-probability models are a popular tool to assess the state of the game. Robberechts, Van Haaren, Davis [29] developed an in-game Bayesian win probability model. The model tries to predict, given a game state, the probability of goal scoring and therefore, through a binomial distribution, predict the score and the win probability. The Bayesian approach, where the effects of a given state are dependent on time and related to close time frames, provides the model with a way of finding evolution through time of the effect of a certain feature.

Harmony between the substituted player and their mates is important for the change to be successful. Bransen and Van Haaren [6] provide a new approach to studying chemistry among players in a team, using the VAEP ([11]) of consecutive actions between a pair of players to determine how much joint hazard they produced. It also measures defensive performance based on the expected indicators from the opposing attacking players in their supposed position.

Knowing the difference in strength between teams is relevant when studying substitutions. Elo rating systems, while originally introduced in chess, have been proven to have predictive power in football [18]. Thus, we can use Elo systems as information about the team's relative quality. Elo ratings for team sports include a Home Field Advantage (HFA), which measures the effect that playing in home ground has on the relative quality difference. The web page clubelo.com uses a system that takes into account different home field advantages, international matches for relative ratings between leagues and the number of goals scored to continuously update their Elo ratings and is well considered by the football industry as the reference provider of Elo ratings.

Chapter 3

Data Overview and Analysis

3.1 Data Description

3.1.1 Data providers

The main source of data is fbref.com. This football-related website includes data from data-provider Opta. The majority of data has been scrapped using the `worldfootballR` package for R [10]. All data used in this thesis is public and accessible.

`fbref.com` match information

With the help of the R package, we can access, for each match, a shooting sheet of information that `fbref` provides. These sheets include for each shot, minute, player, expected goals (xG), the players involved in the two Shot Creating Actions (SCA), and some additional metrics and information. We can also access a lineup sheet with the players involved in the match, both on the pitch and on the bench. Finally, *fbref.com* provides us with a match report, including all bookings, substitutions and goals in a match. Figure 3.1 is a diagram of all the information we can get from a single match.

Definition 3.1 ***Expected Goals (xG)*** *[15] is a metric that measures, for each shot, the probability that such a shot ends in goal. xG are accumulated during the match and are an indicator of the number of goals a team should have scored with an average finishing rate.*

Definition 3.2 *A* ***Shot Creating Action (SCA)*** *is any of the two actions performed by a teammate prior to a shot. It can be a pass, cross, dribble, received fault or a previous shot.*

Note that SCA do not have to be performed by different players of the attacking team. This is, one player can make a dribble, take a shot, and then shoot again from the rebound. They can also make a pass, receive the ball back and then shoot. And a player can be responsible for two SCA for the same shot they didn't take.

But SCA are important because they provide information on the game previous to a goal chance which is just a shot.

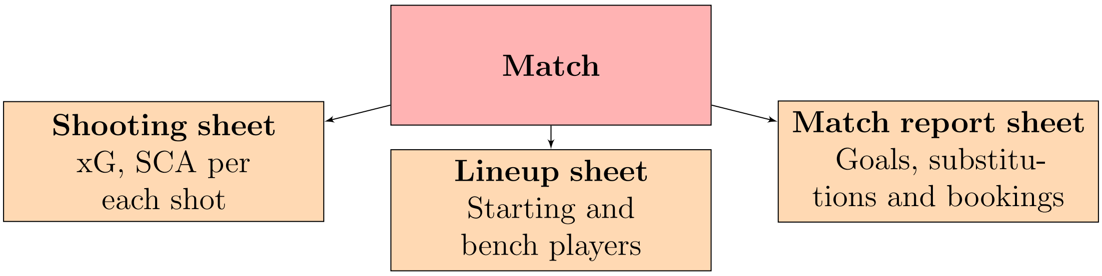


Figure 3.1: Diagram of information obtained from a match

The shooting sheet is of the utmost importance since it gives xG data for each shot. This recent feature, that `fbref.com` has included due to a recent change in data-providers[1], enables us to do a much more in-depth analysis. The low-scoring nature of football means that goals are a rarity, so xG is a weighted way to study the teams' performance. Since the sheet provides the SCA, we can analyze the influence of players at the shot level, not only in shots but also in the actions leading to them, opening the doors to many analyses.

Following the definition of xG, even though the value assigned to a shot might differ depending on the provider of the data, a series of metrics [15] and terminology is derived from it, which are directly provided by the sheets or easily computed.

Definition 3.3 ***Expected Goals Against (xGA)*** *is the amount of xG amassed by the opposition during a period of time (period of a match, complete match or season).*

Definition 3.4 ***Expected Goals Difference (xGD)*** *is the difference in the xG amassed by a team and the opposition during a period of time (period of a match, complete match or season).*

Definition 3.5 *To the player that does the pass previous to the shot, we assign an* ***Expected Assists (xA)*** *value equal to the xG assigned to the shot.*

Shot-by-shot data gives this report a much more profound scope. This falls short of the event-by-event and tracking data that clubs and industry leaders are currently working on, which gives the opportunity for a more in-depth analysis, which ultimately can lead to a competitive advantage [27].

Next on, we have access to the lineups of both teams. Further analysis of this data will be conducted in section 5.1. Lineups include the position the player has played during the match, including multiple positions if the coach has considered tactical changes throughout the match. For unused bench players, position data is void,

[1] https://www.sports-reference.com/blog/2022/11/fbref-shot-level-xg-now-on-match-reports-for-20-competitions/

since they did occupy any position. Nonetheless, we can study their positions during the season or get access to the players' transfermarkt.com page, where their usual position is available.

Finally, from the match report we get the commonly known match features: goals, yellow and red cards, and substitutions. This information gives us an overview of the match state. Such data has been available for much more time than the new in-depth spatiotemporal data, but this report shows that important analysis can be done with such simple, and widely available, data points.

One missing feature from `fbref.com`'s data is the finishing times of both first-halves and matches. Such information can be relevant, in particular when large injury time is introduced, but since we have the minute of all shots, bookings and substitutions, it will be possible to assess the final minute of playing time with sufficient accuracy.

`clubelo.com` Elo ratings

Elo ratings are obtained from `clubelo.com` page, a reference page for European Football Club Rankings based on the Elo system. We can access the complete rankings on any given day, enabling us to obtain the Elo ratings of the teams on every match day. These rankings are comparable through different leagues since they are weighted based on international matches. We do not have the particular Home Field Advantage (HFA) of each competition, so we must add a general HFA on all computations.

3.1.2 Training and test data

As training data, we have selected to study the top leagues in European men's football. We will divide between data for the study of player performance and the data for the study of substitutions.

To study players' productivity, we have studied the shot-by-shot xG of four complete seasons, from 2018/19 up to 2021/22, in the following leagues: the English Premier League, the Spanish *LaLiga*, the Italian *Serie A*, the German *Bundesliga* and the French *Ligue 1*. These five competitions are known as the 5 big leagues and will be referred to as so during this Thesis. This sums up to a total of 7203 matches and 177980 shots taken into account for the training data.

Unfortunately, we cannot use all this data when training models that evaluate substitutions, since the 5-changes per match rule was implemented after the COVID stop. This was through the 2019/20 season. The last matches of such had 5 substitutions, but the change was just implemented, so coaches might not have it used to fully profit from such a rule. Thus, we are counting only the 2020/21 and 2021/22 seasons for substitution model training. The English Premier League decided to go back to the 3-subs per match ruling, so we cannot include it in the training. To minimize the effect of the reduction on matches, we have added these seasons of the Dutch *Eredivisie* and the Portuguese *Primeira Liga*, the next leagues according to the UEFA coefficient[2]. Thus we are evaluating 4116 matches and 34368 substitutions.

The number of matches per league is 380 for a 20-team competition, which are the top-tier leagues in England, Spain, Italy and France, and 306 matches for the 18-teams leagues, such as the top-tier German, Dutch and Portuguese leagues. The French first division, during the COVID break, had only 279 matches played, which explains the odd number of matches in the corresponding dataset.

The test data are all matches played this season (2022/23) at the male European top 5 leagues up to the FIFA match day of finals of March since the English league has used 5 substitutions per match this year [21]. We consider coaches are used to this ruling due to being in the majority of Europe and international competitions. Thus, our test datasets include $n = 1305$ matches played in the first divisions of England, Spain, Italy, Germany and France since the start of the season up to the 20$^{\text{th}}$ of March. Table 3.1 provides a summary of the datasets used in this Thesis, and the relevant number of data points each of them offers.

Seasons	**Data**	**Matches**	**Data points**
Top5 18/22	Training Data-Players	7203	177980 shots
ES, IT, DE, FR, NL, PT 20/22	Training Data-Subs	4116	34368 subs
Top5 22/23 (up to 20$^{\text{th}}$ March)	Test Data	1305	11377 subs 32461 shots

Table 3.1: Summary of the data sets used for training and testing.

We have discarded international competitions such as the Champions League since its knockout stage and short group phase can produce situations where the objective of the team is not to win (depending on results from previous matches) or that reserve players are used (teams already classified or eliminated). In a similar way, we have not taken into account any of the relegation or title playoffs in Germany, France, the Netherlands or Portugal. Thus, we are considering only regular league games, where these situations can happen in a much more sporadic way.

3.2 Data Analysis

In a normal match, during the first-half, lineups are what coaches decided previous to the encounter, and most of the tactics are previously decided and given to the players. in the second-half, players might receive detailed instructions at half-time on how to attack the opposition's weaknesses, and new players are introduced to potentially modify the tactics of a team. Thus, we can distinguish between first-halves, where the initial approach plays a major role, and second-halves where adaptability becomes a main asset.

[2]https://www.uefa.com/nationalassociations/uefarankings/country/

3.2.1 Substitutions and goal scoring

Comparing the first and second-halves can provide us with an insight into the effect of substitutions in a game. Although it is true that second-halves are usually longer, it is assumed that the injury time is intended to correct for lost time, so we can assume that the effective playing time is similar between the two halves. But the facts are that 55.6% of the xG and 56.1% of the goals happened in second-halves, which is a significant increase.

When there were three substitutions, goal-scoring intensity increased after the first and second substitutions, but was reduced after the opposition's third [1]. When comparing this to last season, 21/22, with five substitutions through 3 substitution windows we obtain similar results. In Figure 3.2 we see how the xG generated evolves after each substitution window, normalized by minutes. xG intensity, and consequently goal probability, increases after the first and second coach interventions, due to the fresh legs and new tactical information brought into the pitch. It also decreases after the third, which is most probably due to the overlapping with the opposition's substitutions, which decreases goal-scoring probability [1].

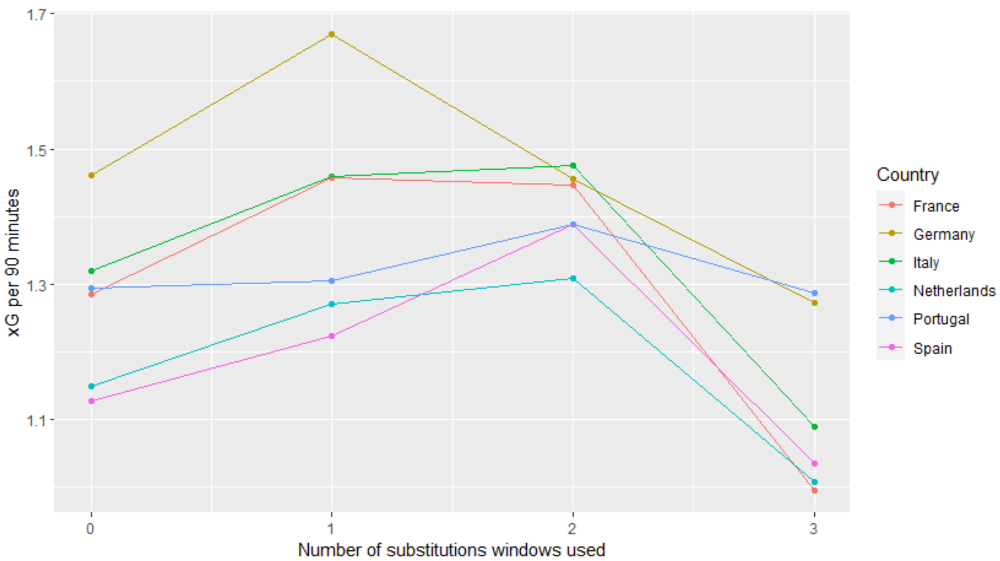


Figure 3.2: Average xG generated after each substitution window, grouped by leagues.

Matches do tend to be more open in the second-half, as most of the goals and goal opportunities happen there. Coaches can introduce tactical variations and therefore increase their attack production. But at the end of the match, tactical decisions oppose themselves and the goal-scoring intensity decreases a little bit, even though the number of minutes after the last window is smaller and might be less significant.

3.2.2 Players' performance through time

The main metric for assessing the actions of players will be the newly defined *expected Value.*

Definition 3.6 *We define a new metric* ***Expected Value (xV)*** *as the amount of xG amassed by a player in all of their shots and SCA.*

We are aware of the shortcomings of xV. Metrics such as expected threat [32] or VAEP [11] give value to all actions, and are more sensitive to the amount of real hazard the action posed on the defending team. We are not valuing all the actions, just shots and the two previous teammates' actions; and we are giving actions the value of the shot they helped produce, which may not be directly related to the quality of the actions.

Extreme examples can also happen, such as a player doing a dribble, getting fouled and taking the penalty. Since the penalty has an xG of 0.76, and the player produced the two SCA, they would amass 2.18 xV, which is much higher than the value of the player in that circumstance, but such cases are only theoretical and in practice are treated as outliers. Knowing the faults this metric can have, it gives much more information than just xG or even xG+xA, since it takes into account many more types of actions, and since we do not work with event-by-event data, it gives us the broader view possible.

Expected Value will be used, through the Thesis, when talking about players, while expected Goals refers to the team since otherwise we would be double, or even triple, counting each goal-scoring chance.

Substitutions can have multiple reasons, and fatigue is usually indicated as the cause of many of the substitutes. By having the xG data of every shot, we can compute the xV that players generate through time. On Figure 3.3 we can compare the Predicted xV of an average top 5 leagues player, depending on whether they start on the pitch or the bench.

Definition 3.7 ***Predicted xV*** *is the amount of xV expected to generate in the number of minutes played, by multiplying the rate of xV per minute by the number of minutes.*

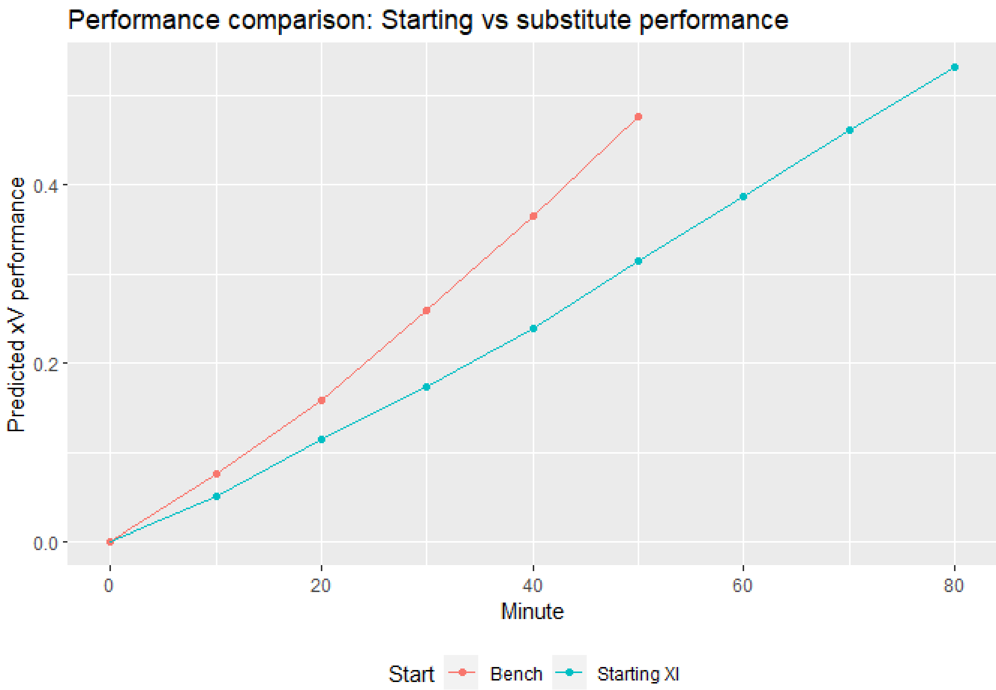


Figure 3.3: xV generated through a match at the average player rate of xV per minute. For bench players, in red, minutes are counted from the moment they enter the pitch

There are two main observations to be made. On the one hand, we can see that the rate of xV generated is almost constant through time. This happens for both starting players and bench players. This is, the rate of attack generated by a starting player does not decay, which would indicate that fatigue does not really apply, or is compensated by the increase in goal opportunities overall.

On the other hand, substitute players produce a higher amount of xV. This is a bit of a counter-intuitive fact since one would expect that the best attacking players generate most of the occasions, and that the best players are starters for their teams. This happens due to the overlapping of two phenomenons: an increase in overall generated xG and fatigue of starting players. Matches tend to become more open and have more xG when the match is ending, and starting players produce fewer xV in relation to the total generated. Thus, substitute players usually play in more open matches and can produce more attacking opportunities, independently of their quality in comparison to starters.

3.3 Comparison Between 3-subs and 5-subs Eras

In Research question 1 we wondered about the effect of the new rule that allows five substitutions per team per match. This rule, adopted as an emergency at the restart after COVID lockdown, seems to establish the norm for the future. Substitutions are the main tool coaches have to change the course of the match. Since they usually happen in the second-half, we can focus on sign changes as a measure of coaches' influence.

We found that over the last compete season before COVID, 2018/19, a 42.6% of the matches changed their result from half-time to the end of the match, and also that 61.7% of the matches drawn after the first-half ended up being a victory for either team. When we check the two following complete seasons with 5 substitutions, we observe that the percentage of matches that changed their result were 39.3% in 20/21 and 41.4% in 21/22, very similar numbers and if something, smaller. Similarly, a 60.9% of half-time drawn matches ended up not being a draw in 20/21 and 61.5% in 21/22, which are the same percentages with a bit of variation.

Players distribute themselves through the pitch, with different roles depending on their position [23]. For each position, the data providers define, we have assigned a value, the *offensiveness*, related to their proximity to the opposition's goal. This way, we can assign an offensive value to the player depending on the positions they have occupied during a match. If they have occupied multiple positions, we compute the average of the offensiveness of all the positions filled. In Table 3.2 a relation between positions and their offensiveness value can be found.

Position	**Abbreviation**	**Offensiveness**
Goalkeeper	GK	0
Centre, Left or Right back, Defender	CB, LB, RB, DF	1
Defensive Midfielder, Wingback	DM, WB	2
Central, Left or Right Midfielder	CM, LM, RM, MF	3
Attacking Midfielder, Left or Right Winger	AM, LW, RW	4
Forward, Second Striker	FW, SS	5

Table 3.2: Different positions, with the abbreviations used by the data providers and the relative offensiveness assigned.

Definition 3.8 *We define a player change as an **offensive substitution** if the offensiveness of the player coming in is more than 0.5 units higher than the player going out. Similarly, a **defensive substitution** is where the positions occupied by the new player are less offensive by at least 0.5 units. All other substitutions are considered **neutral**.*

There are multiple reasons to substitute a player: protecting a booked player, injuries, fatigue, or tactical substitutions. According to the types of substitution, all offensive and defensive substitutions carry tactical adjustments. A neutral substitution can also be tactical since different players have different profiles, but they might also be just replacements due to physical conditions or the context of the match. Due to our usage of widely available data, we cannot distinguish between these two cases. Thus, we will compare neutral and non-neutral substitutions.

The percentage of non-neutral substitutions was 24.6% during season 18/19, and continued to be 24.3% in 19/20 up to the COVID lockdown. When leagues resumed, allowing five substitutions, only 18.0% of substitutions were non-neutral. The same rate continued in 20/21 with a 17.2%. The number of non-neutral substitutions per team per match remained the same, close to 0.7. This means that coaches made a similar number of tactical changes, keeping their usual number of tactical alterations through a match, but increased non-tactical ones, probably due to fatigue, COVID sequels and a more compact schedule. A summary of these rates is in Table 3.3. These percentages may vary depending on the definition of offensive and defensive substitutions [7].

On season 21/22, with coaches having time to adapt, the percentage of non-neutral subs raised back to 24.4% of the total ones, and coaches did 1.1 non-neutral substitutions per team per match. This trend seems to be contradicted by the results of 22/23 season so far, going back to 17.8% of totals. This might be due to the World Cup being held in December, which has made teams play more frequently and players play more matches, thus increasing the need for physical substitutions.

Season	**Non-neutral subs percentage**	**Non-neutral subs per match**	**Subs per match**
18/19	24.6%	0.72	2.93
19/20 pre-COVID	24.3%	0.70	2.87
19/20 post-COVID	18.0%	0.77	4.29
20/21	17.2%	0.72	4.17
21/22	24.4%	1.1	4.33
22/23	17.8%	0.77	4.35

Table 3.3: Summary of the non-neutral substitutions per season, both as percentage of total substitutions and per match, and also the number of substitutions per team per game.

Finally, we would like to see if teams are adapting to the new rule, so we look at the number of overall substitutions made by the teams over time. After performing a Welch Two Sample t-test on the substitutions per team per match through seasons 20/21 and 21/22 we obtain that the mean of the latter, 4.33 subs per team with a standard deviation of 0.88, is statistically significant (p-value $4.328e-15$) bigger than the 4.17 of the first season, with 0.92 deviation. The comparison with the ongoing 22/23 season shows that the value remains stable at around 4.35 subs per match, and with a similar standard deviation.

Including a new substitution is a very relevant change in the rules [34]. We have observed that even though coaches have more maneuverability, the change in dynamics in second-halves remains very similar. A possible explanation is a Red Queen effect [33], where both teams evolve so that the coaches' effect cancels itself. When the new rule was introduced, the tactical substitutions per match remains the same, but the percentage decreased. With the adaptability of the teams' staff, which can be seen in the increase of overall substitutions a team uses, the amount of non-neutral substitutions has increased.

Chapter 4

Substitution Evaluation

This section focus on Research question 2 from different perspectives. First, we will try to define what is a useful or positive substitution according to different criteria. We will also try to compute the winning probability and thus derive the impact substitutions have.

Football's low-scoring nature generates multiple situations where better-playing teams do not necessarily convert into a positive result. Thus, if a substitution has an impact on the game that is not reflected in the data we have available, we might not be able to evaluate it properly. In this chapter, we use the available information to try and assess substitutions in the best possible way.

4.1 Pre-Study Analysis

Evaluating a substitution can be a difficult task, and more so not having event-by-event data, so we cannot get valuations on the effect of a particular player's actions, such as xT [32] or VAEP [11]. What we do have is shot-by-shot data, and therefore we can compute the expected Value generated by a player with their shots and Shot Creating Actions. We are going to pack substitutions into windows, as the effect of two simultaneous changes can not be discerned, so in this section the term *substitution* will be used also to refer either to a single sub or a window of multiple player swaps, according to context.

The analysis in this section is new, with rules developed especially for this thesis. Even though some other studies use similar assessment rules ([26], [31], [7]), we introduce a wider range of ways to study and evaluate substitutions.

4.1.1 Assessment rules definition

For those windows including an offensive or defensive substitution, we are going to define simple rules: an offensive change is *successful* if a goal is scored afterwards; a defensive one is *successful* if no goal is conceded from that moment to the end of the match. This problem can be thought of as a two-class problem, where substitutions are either *positive* or *negative*.

We are going to define three rules where substitutions are assessed based on the scoreline and its change, the xG generated by both teams, and the xV generated by the players coming in and out.

Definition 4.1 *2-Class* ***Scoreline Evaluation Rule****. A substitution is tagged as* ***positive*** *if one of the following cases is true: the team is winning or drawing having an inferior Elo rating, and the result remains the same; the team was losing by multiple goals and reduces it to one, or the team betters the amount of points scored in that match. In all other cases, it is considered* ***negative***

This long case-by-case definition tags as *positive* those substitutions that better your result or maintain it when it is already bad, and tags it as *negative* otherwise.

We can also use more *advanced* metrics, based on xG, to better assess the team's performance past their finishing rates. We use xGD to value a team's dominance over time, so we can separate substitutions depending on an xGD increase or decrease. By analyzing chances weighed by their threat, we can measure a shift in match momentum even if the finishing is unequal.

Definition 4.2 *2-Class* ***xGD Shift Rule****. A substitution is* ***positive*** *when such is the variation in xGD, $\Delta xGD = xGD_{post} - xGD_{prev}$. Otherwise, it is* ***negative****.*

Similarly, we can evaluate the performance of the players coming in and out by looking at their generated xV. This time we have to account for time since with more time you can generate more xV. With xGD we do not need to account for time because it measures both teams, and both had the same time to produce scoring opportunities.

Definition 4.3 *2-Class* ***xV Change Rule****. A substitution is considered* ***positive*** *if the xV per minute of the incoming players is higher than those that have been brought off, and* ***negative*** *otherwise.*

These three rules can also be defined for a three-class problem, where we include the possibility of *inconsequential* substitutions (neutral substitutions refer to non-offensive nor defensive ones). For the scoreline-based rule, we expect the change in scoreline to be significant, while in the xGD and xV-based rules, we define a threshold from which we consider the change sufficiently relevant.

Definition 4.4 *3-Class* ***Scoreline Evaluation Rule****. A substitution is tagged as* ***positive*** *if the amount of points increases, as* ***negative*** *if it decreases and* ***inconsequential*** *if it remains the same.*

Definition 4.5 *3-Class* ***xGD Shift Rule****. A substitution is* ***positive*** *when $\Delta xGD > 0.5$,* ***negative*** *when $\Delta xGD < -0.5$, and* ***inconsequential*** *otherwise.*

Definition 4.6 *3-Class* ***xV Change Rule****. A substitution is* ***positive*** *if the rate of xV per minute increases but 20% or more,* ***negative*** *the rate decreases by a 20% or more, and* ***inconsequential*** *otherwise.*

4.1.2 Rules correlation

Previously, we have offered various definitions of what a helpful substitution might be, depending on the scoreline, the team's or the player's performance. It is a logical question to study how they relate to themselves. As discussed previously, better playing teams might not convert it into a better scoreline. This is why we would like to compare the xG dynamics, our way to assess team performance, with the 3-Class Scoreline Evaluation Rule. Figure 4.1 compares, for each of the classes of such rule, the changes in expected goals difference. As we can see, there exists a correlation, and substitutions that ended in a better scoreline were accompanied by the team bettering their xGD. Of course, between the distribution curves we see some overlapping; and similar changes in xGD can result in better, same or worse scorelines.

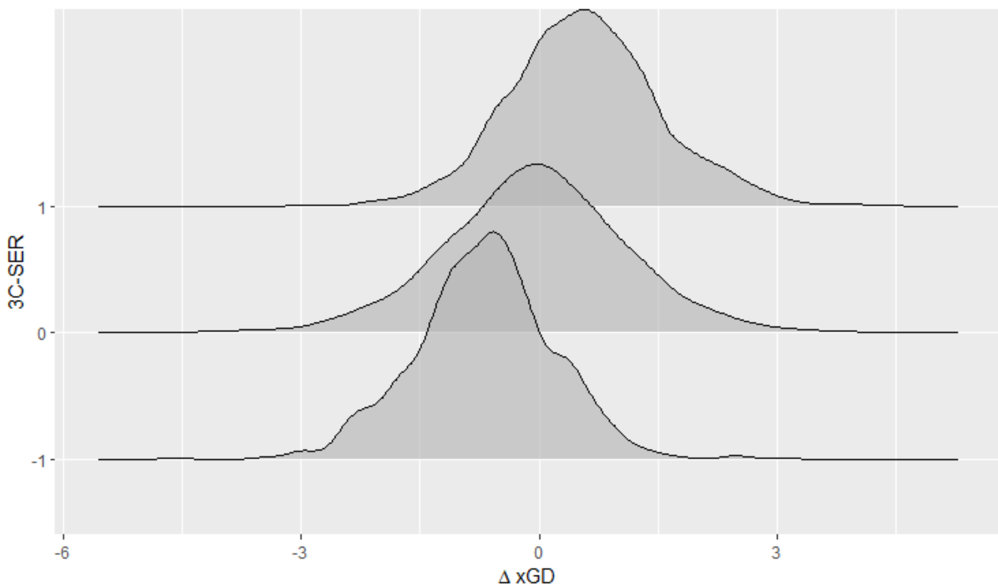


Figure 4.1: Distribution of ΔxGD divided by the 3-Class Scoreline Evaluation Rule categories, where 1 represents *positive* substitutions; 0, *inconsequential* ones and -1, *negative* modifications.

Substitutions with variation in the player's performance are also correlated with the change in the team's expected goals. The substitution windows classified as *positive* in the 2-Class xV Change Rule, result in a positive ΔxGD in 64.3% of the cases, while the *negative* xV subs, the team only betters their performance a 43.0% of the times. Of course, a team can produce more attacking quality chances but the player in particular not be involved in it. Or, reversely, a player can have a better performance than the teammate they substituted but the overall team's execution is worse. Even though these cases do happen, we see a correlation between all the rules, independently of their approach to assessment.

4.2 Classification Models

As done in the study by Brutti, Duarte and Dal Bianco [7], we want to use machine learning models to try and predict if a future substitution is successful. Their study was on the Brazilian Tournament First Division Championship (2015-2018), and used k-Nearest Neighbors (kNN), Decision Tree, Random Forest (RF) and Support Vector Machines (SVM). We will use these algorithms, but not the Decision Tree since it is a simpler version of Random Forest. The total number of substitutions evaluated is 23797, a smaller number than the over 34 thousand that we explained in Table 3.1 since we group transfers by windows.

Brutti et al. studied substitutions made by the visiting team, and defined them as effective if the team scored, or if the team maintained a clean sheet after the substitution and the substitution was not offensive. Our aim is to evaluate all kinds of substitutions, home or away and in any order. The features used to codify substitutions we used are similar to the ones used by Brutti et al. The main difference is that Brutti et al. used the information of each moment a substitution had been made to predict future substitutions, while we only use the information at the moment of the substitution.

- `elo_difference`: A float that measures the relative difference in quality between both teams [18]. Brutti et al. used 4 variables to codify the offensive and defensive force for both teams, and two extra variables to indicate which was favorite, and two more variables to measure the relative force between teams. Elo ratings allow us to use only one variable, but we might not capture teams with exceptionally good attack or defense.
- `is_home`: a Boolean variable, which is `TRUE` if the substitution is done by the local team, and `FALSE` for the away team. It also works as an indicator of a Home Field Advantage.
- `minute`: the minute at which the substitution is done. Brutti et al. only codified the 15-minute period in which the substitution happened.
- `goal_difference`: Goal difference for the team doing the substitution. Is negative if the team is losing. Also used by Brutti et al.
- `total_subs, off_subs` and `def_subs`: the number of total, offensive and defensive substitutions done in a particular window. This could not be used in the Brazilian league where only 3 substitutions per team were allowed. If the substitution was offensive, neutral or defensive was codified.
- `xGD_prev`: xGD in favor of the team doing the substitution, from the start of the match to the moment of the substitution. Brutti et al. did not use xG data in their analysis.

To evaluate the performance of the predictors we will use the **accuracy**, which is the percentage of correct predictions. The number of substitutions with each tag is evenly distributed in all the rules previously defined, so we can take accuracy as a global assessment of the machine learning model. Even though other metrics such as F1, Precision or Recall are important when predicting Boolean problems since we are working with different algorithms, different assessment rules, and both two-class and three-class problems... We believe that four different scoring metrics would make the thesis illegible. Even though only accuracy is shown in the text, for tuning and assessment of the models we have used the appropriate metrics.

4.2.1 Offensive and defensive substitutions

We start by predicting the effectiveness of offensive and defensive substitutions. By taking into account only windows with such substitutions, we train the models to predict if they are *successful* of not, as previously defined. For the optimization of hyperparameters, we use Python's library `scikit-learn` and the function `GridSearchCV`. By using a 10-fold Cross-validation, using F1 score as the primary metric for hyperparameter tuning because the target variable is Boolean and we want to balance precision and recall in our predictions. These are the best hyperparameters for each model when trying to predict successful offensive substitutions.

- Model: RF
  - `n_estimators`: 500
  - `max_depth`: 10
  - `min_samples_leaf`: 2
  - `max_features`: sqrt
- Model: SVM
  - `kernel`: rbf
  - `C`: 1
  - `gamma`: scale
- Model: kNN
  - `k`: 5
  - `weights`: uniform
  - `algorithm`: auto

The hyperparameters when tuning for the defensive substitutions are very similar, since we are working with similar dataset. k in KNN and `min_samples_leaf` and `max_depth`. The accuracies obtained stand between 68% and 73% for offensive substitutions and 70% and 75% for the defensive ones. In both cases, kNN performs the best out of the three algorithms and Support Vector Machines work out the worse.

This is the most comparable to the work by Brutti, Duarte and Dal Bianco [7]. Their evaluation is based on whether a goal is scored, for all substitutions, and if a clean sheet is maintained for defensive substitutions. They used two datasets: the first one was for predicting the effectiveness of the second substitution, while the second dataset was to predict the third and last substitution. Results for the first dataset vary between 70% and 78% accuracy, depending on the algorithm, while on the second dataset they get results between 78% and 87% accuracy. Our results are apparently worse than theirs, but no direct comparison can be established due to the difference in procedure.

4.2.2 Scoreline conditioned evaluation

Valuing substitutions depending on the evolution of the scoreline is the easier and most usual ([26], [31]) way to assess substitutions. For our analysis, we employ

Definition 4.1 and Definition 4.4 of the Scoreline Evaluation Rules. These definitions classify substitutions as *positive* when a team performs well in terms of goal scoring and *negative* when they do not. In addition, we also introduce the concept of *inconsequential* substitutions for situations where the result of the game remains unchanged. This approach allows us to consider the various outcomes of substitutions and identify their impact on the game.

Since we want to compare how the models predict the Scoreline Evaluation Rule, both for the two-class and three-class problems, we are not using the Support Vector Machines. On Table 4.1 and Table 4.2 we can see the confusion matrices obtained by the k-Nearest Neighbors predictor, with $k = 7$, the smallest k where the accuracy gets constant. Accuracy is of 74.7% for the two-class problem and 80.4% for the three-class. The difference is mostly explained by the accumulation of results with no direct influence on the scoreline, and thus the model tends to predict such results. The results for the Random Forest predictor were similar but a bit worse, with accuracies of 70.5% for the 2-class problem and 76.5% for the three-class.

		Actual	
		0	1
Prediction	0	9127	2851
	1	3174	8645

Table 4.1: Confusion Matrix for the kNN 2C-SER

		Actual		
		-1	0	1
Prediction	-1	683	820	186
	0	271	16075	412
	1	8	2951	2391

Table 4.2: Confusion matrix kNN 3C SER

The confusion matrices are with the training data. By using Cross Validation, we tried to train models in a way that they are not overfitted, so the differences between the validation error and the test error are minimal. After selecting the hyperparameters with the training data, we evaluated the test dataset, obtaining very similar results. These results are of the order of the ones Brutti et al. obtained with the simple dataset, which are a bit worse than the more informative dataset. But results are not directly comparable due to the different definitions of both features and effectiveness. Moreover, this is a complex problem where high accuracies cannot be expected, since randomness is a big part of the game.

We also trained a kNN classifier to predict the 3-Class Scoreline Evaluation Rule but ignored the xGD_{prev} feature. This is, not include information on how the team was performing prior to the substitution being made. The results reveal that the accuracy is marginally lower, so we cannot conclude that previous team performance is an important feature for the prediction of the effectiveness of a substitution.

4.2.3 Advanced metrics evaluation

Besides studying the effect of substitutions in the scoreboard, we have defined the xGD Shift Rule and the xV Change Rule to assess substitutions based on more advanced metrics. We will now focus on evaluations using those metrics.

For the xGD Shift Rule, we are in the exact same problem as we were for the Scoreline Evaluation Rule since the features used and the categories of the substitution assessment are the same. Hyperparameter tuning was, once again, done by 10 fold cross-validation. The models were fit, and when validation data that the models had never seen, averaged by the 10 folds of cross-validation, we obtained a 70.2% accuracy for the RF and 72.8% for kNN on the two-class problem. The increase for the three-class problem was not as big as happened with the Scoreline Evaluation Rule. That increase came mostly from having a majority of the substitutions be *inconsequential*, i.e. the result did not change. For the three-class xGD Shift Rule, we defined cut-offs so that the amount of *inconsequential* substitutions was not so big, thus the not-so-big increase in accuracy.

When predicting the xV changes, we included an extra feature relative to the player's performance.

- `player_out_avg_xv`: The average xV per 90 minutes generated by the player going out.
- `player_in_avg_xv`: The same metric for the player coming in

These are the average xV per 90 minutes the players have amassed during the last season. Including the actual season could be a significant improvement, especially on the latter part of the season. We can include information about the player coming in since we assume that the coach knows which player they are going to introduce. With this new information, the accuracy goes up to 78% for both algorithms. These features allow us to know the quality of the players involved in the substitution, which adds to the match dynamics and team information we already had, which allows for the model to produce better predictions.

4.3 Win Probability Model

This section is based on the in-game win probability model by Robberechts, Van Haaren and Davis [29]. The main difference is we have included substitution features and made some variations on what the model tries to predict.

4.3.1 Model definition and validation

We define the game state, for a number of remaining minutes t and for a team $i \in \{\text{home,away}\}$ as $x_{t,i} = (t, i, \tau_{t,i}, s_{t,i})$. For a particular i, we have the team features coded in $\tau_{t,i}$ which include: team strength difference based on Elo ratings [18], goal difference, player difference, and yellow cards. $s_{t,i}$ includes the information of substitutions of both teams: number of offensive, defensive and total substitutions.

The model is trained to predict $\theta_{t,i}$, the probability that a goal is scored at any of the remaining minutes. The probabilities for training are obtained as a division of the goals that were scored in those minutes. Thus, with a Binomial distribution, we can calculate the number of goals i is expected to score in the last t minutes as

$g_i \sim B(t, \theta_{t,i})$. By adding that to the already scored goals, $G_{t,i}$, we can simulate the final result.

Different approaches were taken in the way to predict $\theta_{t,i}$. The intuitive idea is that the goal-scoring probability at the next minute increases with time, but different features can affect it in different ways. It has been shown that the relative importance of variables evolves non-linearly of time [29], but also intuitively close time frames should have a close relationship between the variables. The four algorithms were:

- Multiple Logistic Regressions (mLR)

  A logistic regression is calculated for each time remaining t. We use this approach and not a general logistic regressor because the importance of certain features varies with time.

$$\theta_{t,i} = \frac{e^{w_t x_{t,i}}}{1 + e^{w_t x_{t,i}}} \tag{4.1}$$

- Random Forest (RF)

  A Random Forest regressor. Random forest is able to deal with non-linear interactions between the features.

- Multiple Random Forests (mRF)

  A different Random Forest for each time frame, but using also close-time values also as training to prevent a certain time frame to be an outlier. This approach is basically the same as a simple random forest but with a very much increased number of trees.

- Gaussian Walk (GW)

  This Bayesian approach also uses an inverse logistic function with a series of weights $\alpha_{t,i}$ which take as a prior the weights of the preceding time frame, when the minutes remaining are t.

$$\theta_{t,i} = \text{invlogit}(\alpha_t \cdot x_{t,i} + \beta) \qquad \alpha_t \sim N(\alpha_{t+1}, 2) \qquad \beta \sim N(0, 10) \tag{4.2}$$

Comparing this particular model to the one developed by Robberechts, Van Haaren and Davis [29], we see a few important differences. Since we want to look at substitutions, we only study the probabilities for the second-half of matches. We work with the number of minutes remaining t, while Robberechts et al. divide the match into 100 time frames. Our approach gives each minute the same importance, even if the match counted with a big amount of stoppage time. But the model faces the same problem if it was to be implemented in real life since injury time is unknown. For the test data, we assume the 96$^{\text{th}}$ minute to be the last, since it is the mode, and we believe this model should be evaluated without knowing the real last minute. The second main difference is we predict $\theta_{t,i}$ and then simulate the match, while Robberechts et al. implement it in their model and directly predict the probabilities of each result.

In order to validate the performance of a win probability model, we need to look at the whole picture. If the model gives a team an 8% of win probability on a particular match, that result is either going to happen or not, because the match only happens one time. Thus, we group all the predictions of an 8% and we expect that such a percentage of the cases actually happen. We call *predicted probability*, to the probability the model predicts for a certain result, and *actual probability* the percentage of times such result comes true. In Figure 4.2 we see the graphs for the four algorithms and their performances on the three different results: home win, away win and draw.

The size of the point represents the number of such predictions that were made. This is a difference with respect to the model of Robberechts et al. Since we calculate probabilities as a result of the simulation, given the $\alpha_{t,i}$, and not as a direct prediction of the model, we get a much more even distribution of predictions. This is, for the Random Forest or multiple Logistic Regressor, they got a non-uniform distribution of predictions, having few predictions with high probabilities. When implementing the Bayesian approach, that problem was partially solved. In our case, for all models we have a much more uniform distribution, even though predictions of 0-2% are a bit more typical.

4.3.2 Effect of substitutions

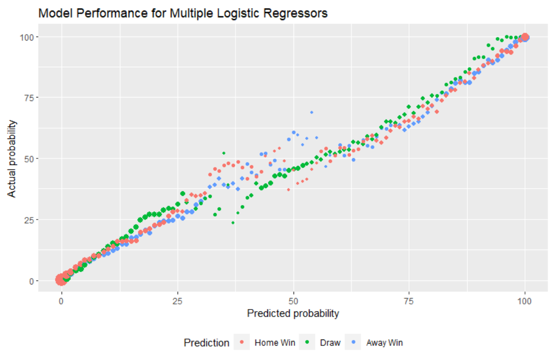
A team might take multiple approaches for a match, such as maximizing their win probability or minimizing the losing one [3]. These decisions might depend on many factors: league positions, relative strength, psychological momentum, etc... In this thesis, we will focus on regular league games, and based on that 3-point per win, 1-point per draw rule, we will measure expected points as a unified approach.

Definition 4.7 ***Expected Points (EP)*** *is the expected value of points for a team given their winning, drawing and losing probabilities in a game.*

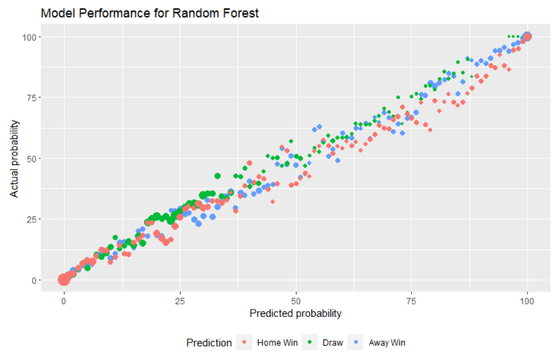
It is logical that a team would like to maximize the points they obtained, since it may give them better long-term results. Thus, we will use EP as a min-max-style approach, and we will assess possible strategies as a result of their effect on EP. This metric weights the draw and win probabilities in the same way the league scoring system does, and therefore we believe it can be accepted as a general strategy for all teams.

We compare the winning probability after the substitution to the minute before it happened. Consequently, when a goal and a change happen at the same minute, so it reports `fbref.com`. But we cannot discern what happened before. This produces a massive change in win probability, and a very high ΔEP, which is due to the goal and not the substitution, so we discard those substitutions from the analysis.

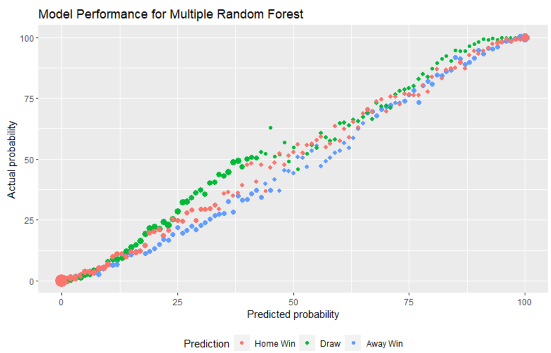
To ensure that the substitution features were relevant we trained the same model but



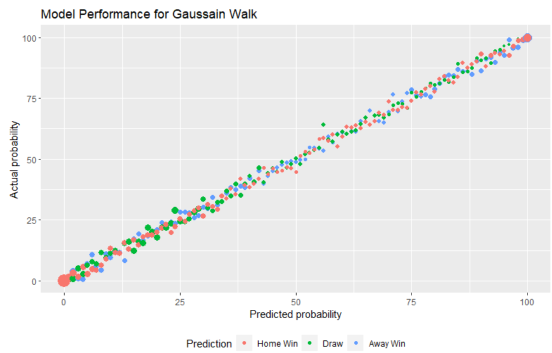
(a) Multiple Logistic Regressors



(b) Random Forest



(c) Multiple Random Forest



(d) Gaussian Walk

Figure 4.2: Evaluation of different techniques to calculate $\theta_{t,i}$ and the final performance of the in-game win probability model, comparing predicted probability to actual probability for each type of prediction. Size indicates the number of predictions.

without those features. This is, predict $\theta_{t,i}$ from $x_{t,i} = (t, i, \tau_{t,i})$. We shall call this model *simple* as opposed to the complete model which we will refer to as *complex*. In Figure 4.3 we see the distribution of the variation of expected points when a substitution is made for the two models. The *simple* model is oblivious to the fact that the substitutions have been made, so the distribution should be as if we picked random moments in matches' second-halves and calculated their ΔEP.

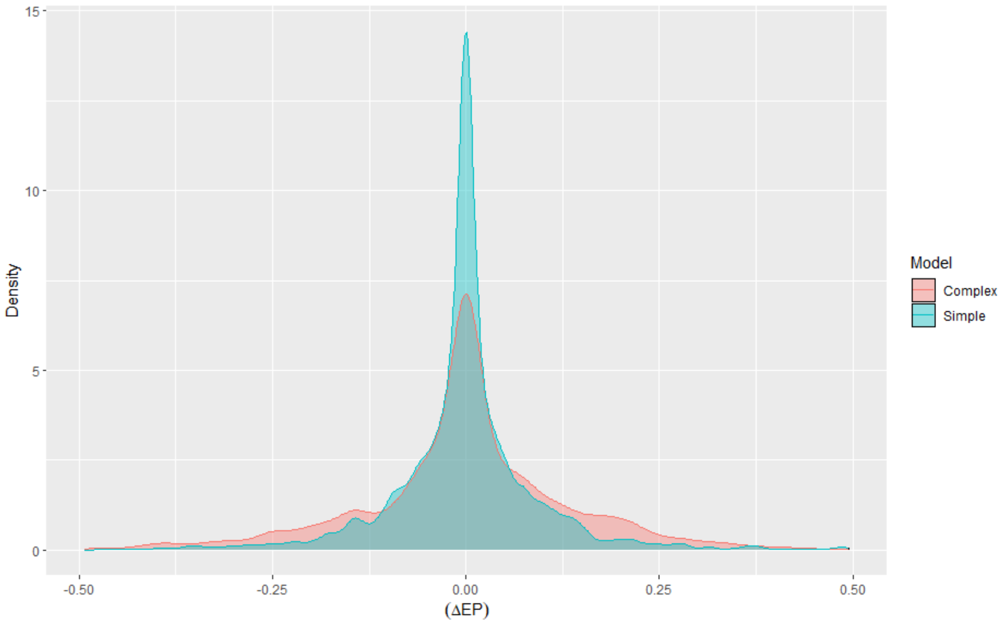


Figure 4.3: Distribution of the variation of EP after a substitution. Comparison between the *simple* and the *complex* model.

Instead, in the distribution of the *complex* model we see how the number of substitutions of none or very little impact in EP is much lower, and that the substitutions with an increase or decrease in EP grow, as it can be seen in both tails of the complex distribution curve. If we separate the substitutions in positive and negatives, as in Figure 4.4, we can focus on the difference between the positive and negative tail. In order to better visualize these tails in smaller numbers, we apply the transformation to the x-axis and expose $\sqrt{\Delta EP}$.

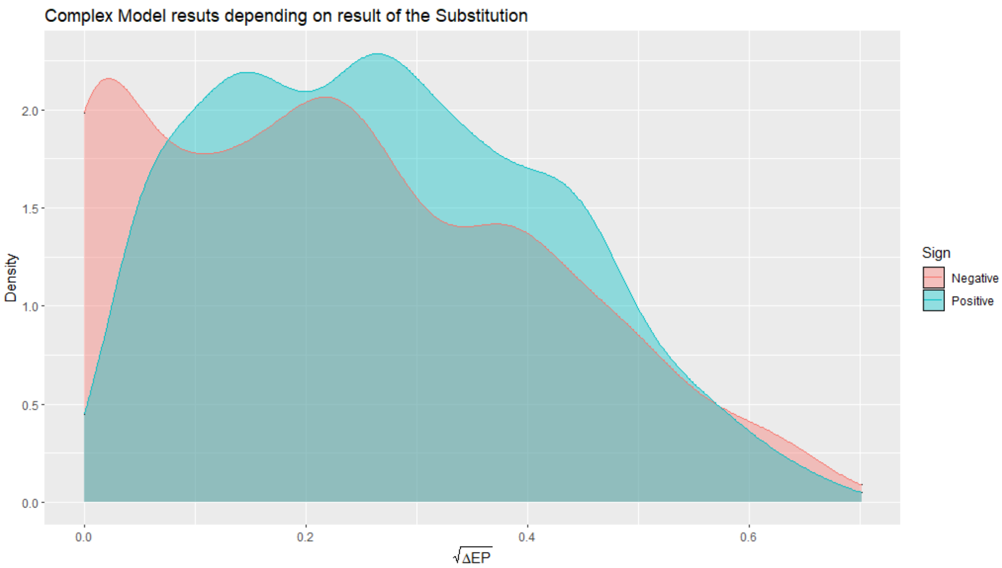


Figure 4.4: Distribution of $\sqrt{\Delta EP}$ for positive and negative substitutions in the *complex* model.

Here we can see how the two curves are not symmetric, which means substitutions are not just random moments according to the model. We can also see how the density is bigger for the substitutions with a positive impact on the match. One of the conclusions is that the majority of substitutions have a positive effect on the team's goal-scoring probability, while not diminishing the defense so match, so the win and draw probability increase and so do the expected points.

4.4 Results and Discussion

In this chapter, we wanted to study the problem of properly evaluating the effectiveness of a substitution. This is a difficult thing to do with the available data since many of the player's actions go unnoticed. Different approaches have been taken for the assessment, and even though they are correlated, in many cases a single player's performance, or even the whole team, might not translate into a palpable result in the form of goals.

We trained many models to predict the success of the substitutions. In a similar study on the Brazilian League [7], they obtained model performances of 75-85%. Our results oscillate between the high sixties and high seventies. As we described, methods in both studies differ, and this might explain why a more detailed dataset gives worse predictions. It is true, though, that Brutti et al. focus much on the team's strength, while we only counted for Elo ratings. A more team-based approach might be necessary to improve predictions on substitutions. We observed that player-specific information did better the results of the models predicting xV Change Rule.

The second approach consisted in calculating the importance of substitutions as a byproduct of the variation in win probability. Our in-game win probability model, based on the one developed by Robberechts, Van Haaren and Davis [29], proved to be sensitive to substitutions, as seen in Figure 4.3. Our change in the model, trying to predict future goals and then simulating the match, instead of predicting result percentages, turned out to be effective in having predictions of all possible percentages for all results. Similarly, as in Robberechts et al., the Bayesian approach with the Gaussian Walk has resulted as the most effective, having a high relation between predicted and actual probability.

One aspect to discuss about the win probability model is the effect that is shown that substitutions have. Most of the substitutions have a small impact, both for positive and negative outcomes. In actual matches, a substitution can have a much bigger effect, since the player that comes in might be especially inspired and have an above-average finishing, or on the contrary have a bad day. The model averages the effect of the substitutions given a game state. It does not account for other contextual features that might have an impact on the development of the game, such as the psychological aspects of a substitution. In general, though, the impact of a player replacement is usually beneficial, as seen from Figure 4.4.

Chapter 5

Substitute Selection

A natural question a coach may come up with is which player should be introduced. Research question 3 is the formulation of such doubt and this chapter aims to provide an answer. With the new regulations, in many of the leagues, the number of players on the bench has now increased, though it is different for each country. Deciding which player comes into the match depends mostly on the available resources.

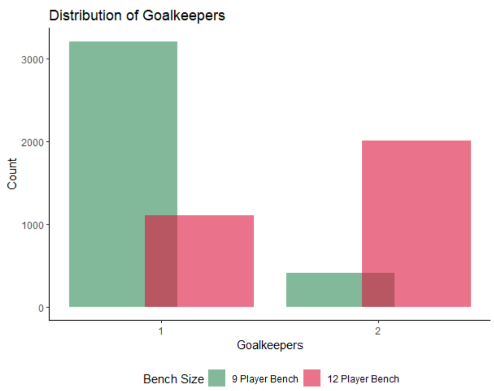
Two players are involved in a substitution: the player on the field that is removed, and a bench player who is introduced. Deciding on which player should leave the pitch is a very interesting topic. Unfortunately, we cannot address it since we are working with public and widely available data, that does not include event-by-event, therefore, a detailed analysis of the player's performance through the game cannot be made, making it difficult to properly assess which player is having a below par match. Moreover, many other factors, such as physical form or the tactical needs of the team need to be taken into account.

Of course, some of these problems arise as well when selecting the bench player that will participate in the match. Physical problems are not a concern for substitutes, professionals can play a match as bench players, and have been proven to cover more spaces and do more sprints than starting players [4]. We give a data-based approach to substitute selection, even though tactical knowledge and specific on-match situations cannot be captured into data are still important.

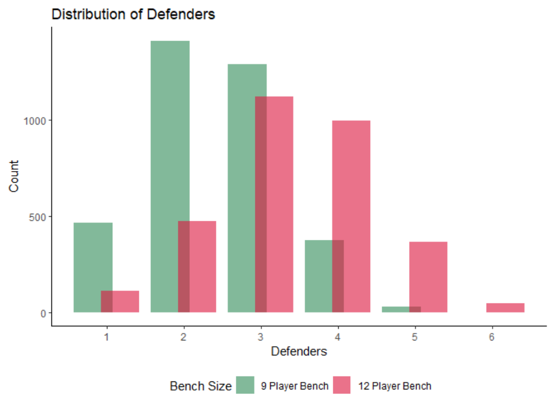
5.1 Bench Analysis

For the purpose of simulations and proper substitute selection, we need to know which players are on the bench. On a general basis, if we want to decide between various possible types of substitutions, knowing a team's formation and the available players is very important. In this section, we aim to provide some ground rules for simulation.

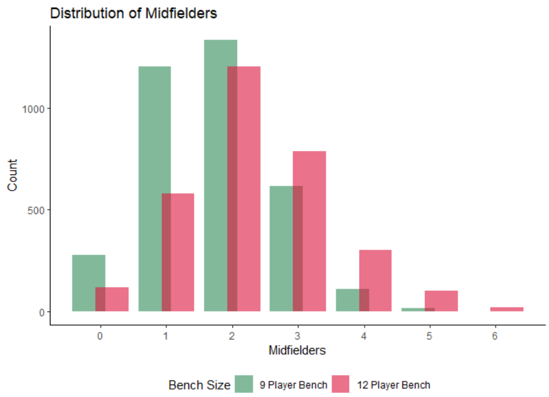
In the 2020/21 and 2021/22 seasons, teams could have up to 12 players on the bench for the Spanish LaLiga, the Italian Serie A and the Dutch Eredivisie; but only up to



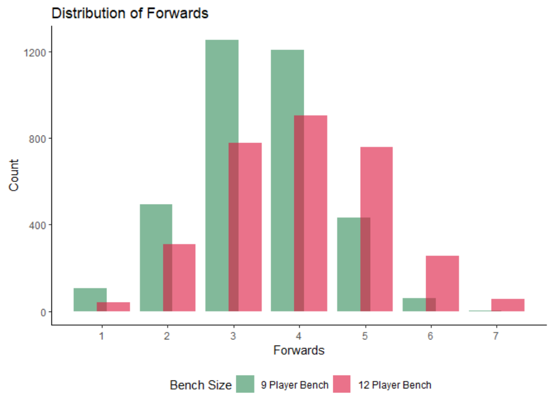
(a) Goalkeepers Histogram



(b) Defenders Histogram



(c) Midfielders Histogram



(d) Forwards Histogram

Figure 5.1: Histograms of the distribution of each position, separated by maximum bench size.

9 in the German Bundesliga, the French Ligue 1 and the Portuguese Primeira Liga. Thus, our analysis will have to be dual. We first start by seeing which is the typical distribution of a bench to see which assets a coach has at their disposal. Figure 5.1 shows the distribution of the main positions a player can be: goalkeeper, defender, midfielder or forward.

A few observations about the distribution of players. In Figure 5.1a we see that, while in smaller benches teams mainly have just one bench goalkeeper, when twelve spots are available, usually two goalkeepers stand at the bench. Substitute goalkeepers rarely come into play (roughly 0.3%), and needing two extra goalkeepers is unheard-of, but teams rather have them both on the bench, possibly due to not needing 11 substitute field players. In Figure 5.1c we observe that some benches do not have midfielders at all. While this sounds odd, some teams might play their midfielders in other positions, since they are the most versatile, or plan on playing in midfield a player which the data providers tag as either defender or forward. Additionally, through the season squads get affected by injuries and the number of players initially planned could not be available in certain matches

Selecting adequate bench players is key to the coach's success because it defines their tools during the match, the pieces they can fit into the playing XI to change the game dynamics. Beal et al. [3], develop a formal model for a pre-Match Bayesian game. Teams have different tactics, which are the formations that they use, and a multi-class classification deep neural network is trained to learn the payoffs of the different strategies. In such a way, they try to predict the opposition's strategy and suggested an optimal tactic. The results showed teams with tactics similar to the suggested ones achieved better results.

The pre-match modeling of the game is an area where bench management could be an important asset. Although we were unable to explore the idea in this report, we believe it is an important area for future research. Including the information on the opposition's bench could improve the model by giving more valuable information. In this analysis, we aim to give an approximate idea of the bench composition for accurate and real simulations.

	12-Player Bench				9-Player Bench			
Formation	Gk	Def	Mid	For	Gk	Def	Mid	For
4-5-1	2	3	3	4	1	3	2	3
4-4-2	2	4	2	4	1	3	2	3
4-3-3	2	3	3	4	1	3	2	3
3-5-2	2	4	3	3	1	2	2	4
3-4-3	2	4	2	4	1	2	2	4
5-3-2	2	3	3	4	1	2	2	4

Table 5.1: Number of players per position in the bench, separated by the two bench sizes.

In Table 5.1 we have summed up the bench that we will be working on during the simulations, based on the averages of players from such position and such that they are a full-sized bench. The types of formations are ordered by appearances on our dataset, and all types of formations that appear more than 50 times can be included in one such formation. Formations are defined by the number of defenders, midfielders and forwards separated by a dash. This is a simplification of the previously defined positions in Table 3.2, so multiple tactical approaches are lost. For example, both the 4-2-3-1 and 4-1-4-1, two widely used formations, are packed under 4-5-1, even though they might have more players defined as forwards.

Following the example of [2], we will assume that on the pitch we can have a minimum of 3 and a maximum of 5 defenders, same with midfielders, and 1 up to 3 attackers. Even though the number of forwards on the pitch is smaller, we know that most of the substitutions involve attacking players [4] so coaches want to have more options in these positions. These constraints will be applied for the simulation of full matches.

5.2 Type of Substitution

We have classified substitutions as *offensive, defensive* or *neutral*, according to Definition 3.8, depending if there is a reasonable increase or decrease in the *offensiveness* on the position of the incoming player. We can keep the moment a coach decides to make a substitution constant. For this, we will assume that no substitutions are due to an injury, and hence require a similar player as a replacement. During chapter 4 we studied different ways to evaluate the effectiveness of a substitution. We are going to use those same models to evaluate if another type of substitution could have been better according to the models.

For each window substitution, we have simulated the expected result if those substitutions had had another nature. For offensive substitutions, we simulate them being neutral or defensive, and evaluate what the performance would have been. Similarly, we have done the same for neutral and defensive substitutions. The wide spectrum of targets defined in subsection 4.1.1, multiplies the opportunities we have here, so we are just going to explore the options we find more interesting.

A real-life application of this kind of model could be for coaches, to decide the type of substitution they have to do or to be data-informed as to which is more likely to accomplish their goals. Additional models, that not only differentiate between offensive and defensive substitutions but also incorporate other factors, have the potential to equip technical staff with a broader range of tools for conducting simulations.

5.2.1 Win probability simulations

Simulations with the in-game win probability model, detailed in section 4.3, not only allow us to see if certain simulated substitutions would be successful but since we work with Expected Points (Definition 4.7), we are provided with a numerical weighing of the effect of changing the type of player swap.

For each substitution, we have simulated that modification in an offensive, neutral and defensive way. The Bayesian win probability model is applied to each generated game state. This is, given a game state $x_{t,i} = (t, i, \tau_{t,i}, s_{t,i})$ we modify $s_{t,i}$, so we have $s_{t,i,\text{off}}$, $s_{t,i,\text{neu}}$ and $s_{t,i,\text{def}}$. Similarly, for that same t, we modify the game state of the opposing team i^* into $x_{t,i^*} = (t, i^*, \tau_{t,i^*}, s_{t,i^*,\text{off/neu/def}})$, depending on the case. Thus, for every case, we get three pairs for each case: $(\theta_{t,i,\text{off}}, \theta_{t,i^*,\text{off}})$ for the offensive substitution state, $(\theta_{t,i,\text{neu}}, \theta_{t,i^*,\text{neu}})$ for the neutral one and $(\theta_{t,i,\text{def}}, \theta_{t,i^*,\text{def}})$ for defensive. With the definition of $g_{t,i} \sim B(t, \theta_{t,i})$ and adding the current scoreline $(G_{t,i}, G_{t,i^*})$, we simulate and get the home win, draw and away win probabilities. Note that this analysis is independent of which of the nature of the substitutions. For either offensive, defensive or neutral substitution, we simulate the three states.

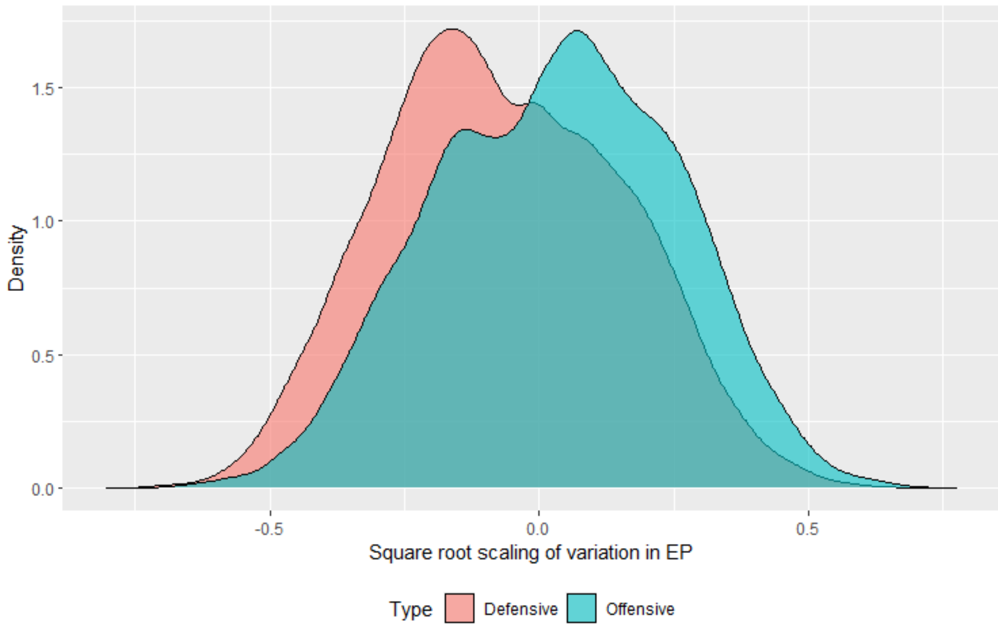


Figure 5.2: Variation in the amount of Expected Points. x-axis is square root scaled. Separation between the comparison of offensive substitutions with respect to neutral ones (blue) and negative ones with respect to neutral (red).

The graphical representation of the simulation results can be observed in Figure 5.2. From the previous simulations we obtain EP_{off}, EP_{neu} and EP_{def}. In blue, we have the distribution of $\Delta EP_{\text{off}} = EP_{\text{off}} - EP_{\text{neu}}$, and in red we have $\Delta EP_{\text{def}} = EP_{\text{def}} - EP_{\text{neu}}$. The x-Axis has been rescaled by a square root transformation for better visualization, but maintaining the sign of the change, so the amount plotted is actually $\text{sgn}(EP)\sqrt{\Delta EP}$. If we plot simply ΔEP the curve is centered and spiked. This is due to 60.3% of the substitutions having a ΔEP of less than 0.03, which corresponds to less than 1% change in EP since we work with a 3-point system.

This is important since we see that almost two of each five substitutions could have a significant displacement in terms of win probability if the proper type of substitution is taken. In Figure 5.2 we see that, on the one hand, the blue curve is displaced to the right, which means that, in general, an offensive substitution gives you better odds of winning than a neutral one. On the other hand, the red curve is slightly displaced to the left, giving you worse win probability percentages. This does not mean it is always the case since we see a big part of the ΔEP_{off} curve is negative, giving you worse chances than a neutral sub, and similarly with the ΔEP_{def} in the positive changes.

Looking at the numbers, we see that over 20% of the substitutions could have very significant, more than 0.3 expected points, swing in expected points. This means that, by properly selecting the substitution type, one of every five substitutions could have a real impact on the match and the winning probability.

When analyzing the substitutions of a match as a whole, we must include the restrictions of section 5.1. As an example, we will examine a particular match: on Figure 5.3 shows the evolution of the Expected Points through the game for the clash between Levante and Real Madrid at the start of the 2021/22 Spanish *LaLiga*. In that match, both teams were using a 4-3-3 formation. Observe that the expected points in the last minute for Madrid are 1.25, because with $\theta_{\text{away},1}$, the model considered the away team had a 25% winning chance, even at the last minute, due to play against a weaker team with a red card.

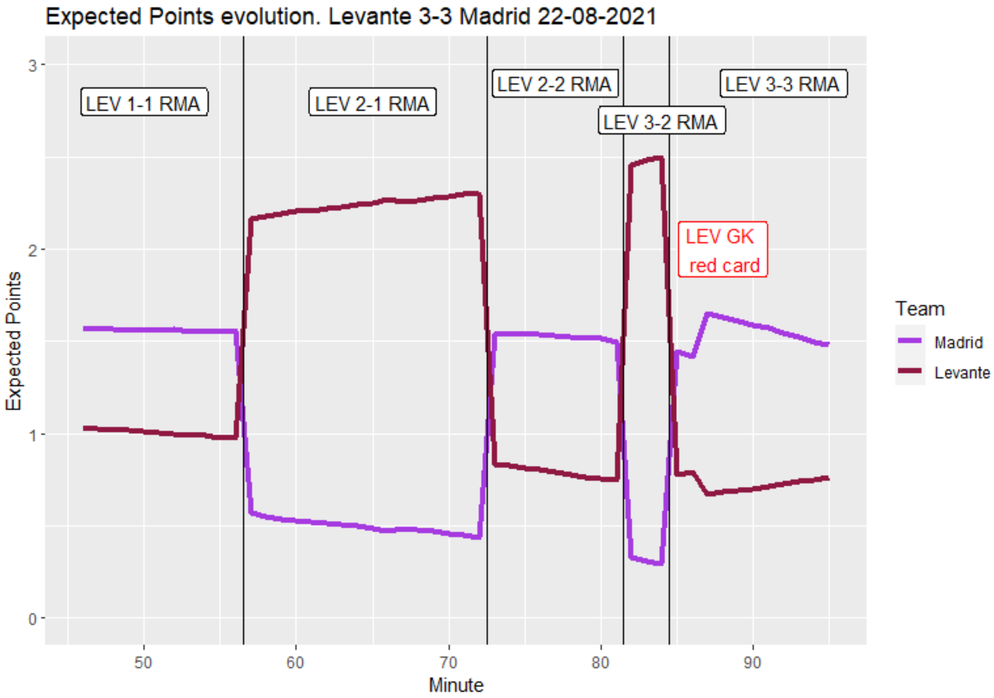


Figure 5.3: Expected Points evolution of the Levante 3-3 Real Madrid second-half on 22nd August 2021.

Due to their formation having already 3-forwards, Madrid did 5 neutral substitutions throughout the match, one of them being playing a forward, Luka Jovic, as a right midfielder. Our model suggests that the best substitution would have been if, at minute 65, an offensive substitution would have been made, resulting in 0.55 EP instead of 0.49 obtained by the real substitution. Madrid could only do one offensive substitution, and of the windows they used, the second one, at the 65th minute, was the better one. Defensive substitutions significantly lowered their EP.

For Levante, the model suggests that their 67th minute subs, being defensive, were the correct choice, but on the 78th minute, with the score being 2-2, they did two neutral substitutions, and the model gives them 0.78 EP. If they did those substitutions defensively, adopting a more defensive approach, the model would output 0.80 EP. Being a weaker team in Elo rating, the model suggests it is better to diminish the opposition's scoring probability than to raise your own at the expense of the other team's attack.

5.3 Player Selector. Theoretical Framework

This section has not been developed in practice due to time constraints, and just the theoretical framework is described. Nonetheless, the section has been included as a starting point for future research. As such, while this section may not be as comprehensive as the others, it represents a foundation for future work on the substitution problem.

The main question to be answered in this section is *from which player can I expect a better performance as a substitute*. In the previous analysis done in subsection 3.2.2, we observed that bench players produce a higher amount of xV. This counter-intuitive fact is due to them playing in more open matches where teams go more directly for goals and where the initial caution is a bit lost.

Figure 3.3 showed that the rate of expected value per unit of time is maintained, both for starting and bench players. Therefore, through all of our analysis and modeling, we will assume that this rate remains constant over time. Many players have a very small amount of time played as incoming bench players because they are usual starters. And even if they have played multiple matches from the bench, with the small number of minutes they amass each time, any sporadic big contribution in xV might bias the analysis. Thus, for a single player, we work with their xV as the amount counting all the minutes. When predicting their predicted xV, just a small correction for the difference between bench and starting players is necessary.

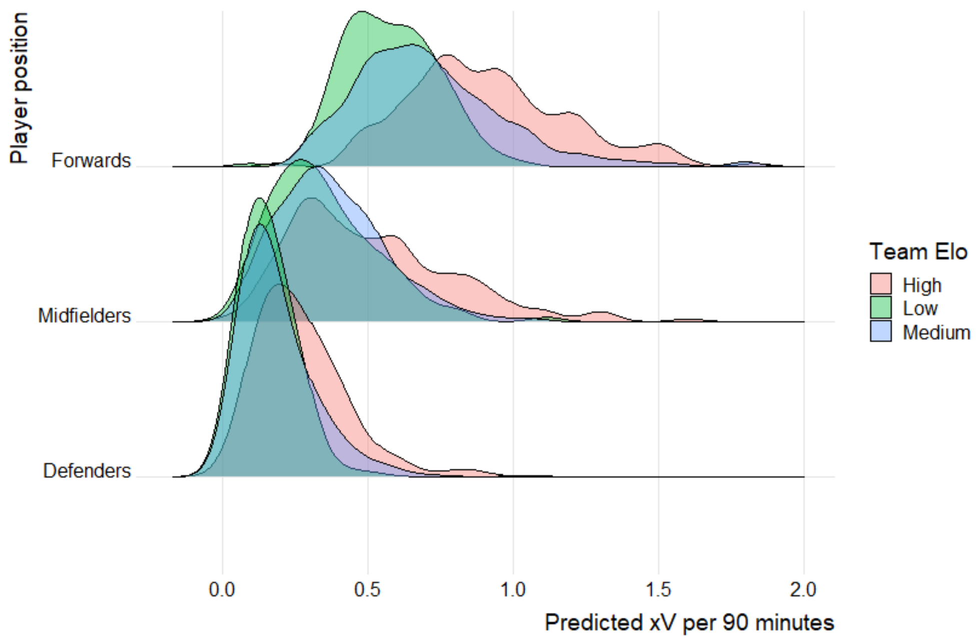


Figure 5.4: Distribution of the predicted xV per 90 minutes for the players depending on their position and the strength of their team

Many factors can influence the rate of xV generated. In Figure 5.4 we see the big differences between positions, and the variation depending on the strength of the team, taking into account the Elo ratings relative to their leagues. The main idea is to take all these factors to have a probability distribution of the player's expected performance and use it as a prior distribution. Additional information about the player such as past seasons' performance, the current season or their performance as bench players should be treated as posteriors.

The idea of chemistry between players was developed by Bransen and Van Haaren [6]. By analyzing pairs of players and the value of their consecutive actions, and also the actions of the player they should be defending, they analyze and then predict the chemistry between pairs of players. They also develop this into a starting XI selector that maximizes the chemistry. This idea should be integrated into further work on the substitution problem, where more detailed data is available. By looking at the chemistry between the players on the bench, and those in the field, trying to maximize their value would be a very important factor when selecting the optimal substitute. Chemistry between a bench player and the rest of their mates is a very important feature that could be implemented into the probability distribution of expected.

Chapter 6

Optimal Timing

In this chapter, we will address Research question 4. Optimal timing can be very relevant since having the right players at the right time might be clinical. We will focus on the two principal kinds of models described in section 4.2, classifiers, and section 4.3, an in-game win probability model. For both cases, we will both analyze the timing of actual successful substitutions and use the model to search for better timing of the real substitutions that were done in the match.

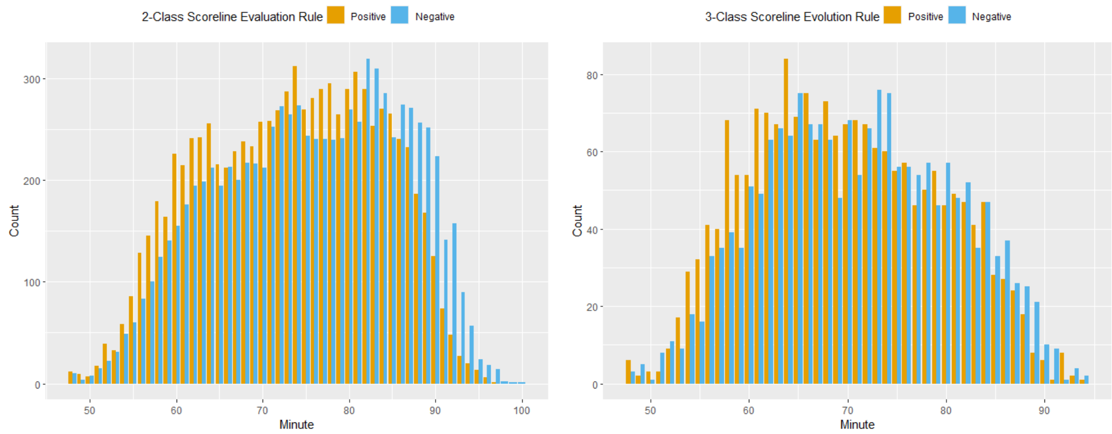
6.1 Timing Analysis of Positive Substitutions

6.1.1 Substitution classification

When assessing substitutions depending on the evolution of the match scoreline, we use the Scoreline Evaluation Rules, Definition 4.1 and Definition 4.4, which basically tag as *positive* substitutions in which the team has a good performance goal wise, and *negative* otherwise. In the three-class problem, we introduce the concept of *inconsequential* subs, for those cases in which the result of the game does not change. The results when assessed by xGD shift are very similar, so we will focus on the scoreline in this section.

In Figure 6.1 we can see how those two variables were distributed through time. While in Figure 6.1a all substitutions are taken into account, we have not shown the inconsequential substitutions in Figure 6.1b, and hence the difference in the y-axis. For both distributions, we observe a similar pattern, where the orange bars, substitutions that accomplish a better result or hold on to a good one, are a bit higher for an early substitution, while late player modifications seem to have a higher failure rate. During the last minutes of the game, teams with a disadvantage change more players, which usually means the result is not modified, and thus that substitution is tagged as negative. On the other hand, when the team is content with the result, most of the times will not require further substitutions.

Focusing on the three-class problem in Figure 6.1b, we see that the amount of either positive or negative subs decreases with time. From the start of the second-half, and up to the 65th minute, the percentage of inconsequential substitutions remains similar, because there is enough time to turn the tables around. After that, the number of substitutions with no effect on the scoreline increases, and thus we have fewer positive or negative substitutions. This also supports the claim that the high number of negative substitutions, in the two-class problem, at the end of the match are non-changing results considered bad for the team.



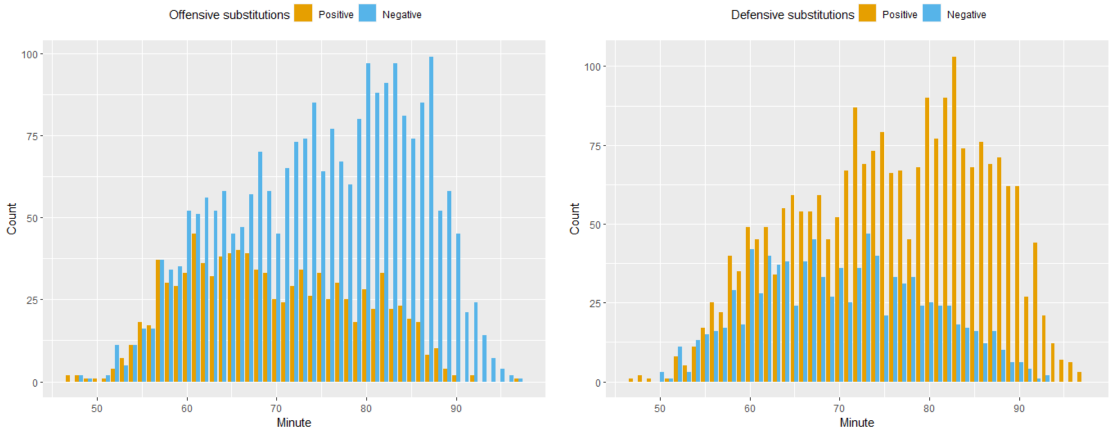
(a) 2-Class Scoreline Evaluation Rule distribution (b) 3-Class Scoreline Evaluation Rule distribution

Figure 6.1: Distribution through time for both the 2-Class and 3-Class Scoreline Evaluation Rule, comparing positive and negative substitutions. Inconsequential substitutions are not shown for the three-class problem.

On these two figures, substitutions made at halftime are not included, since the 46th minute is the one with the most substitutions and it distorts the graph. In the first minutes of the second-half there are usually no substitutions since there is no point in wasting a window when being able to use half-time. The decrease in substitutions towards the end of the match is due to different injury-time and thus different ending times for each match.

Half-time substitutions are relevant, though, and we see half of them being positive and the other half negative in the 3-Class Scoreline Evaluation Rule. For the 3-Class problem, we get similar results, with the same amount of substitutions being positive than negative, and a big majority being inconsequential.

We also studied the distribution of offensive and defensive substitutions, which is shown in Figure 6.2. These are the substitutions with a significant change in the players' position in terms of offensiveness. We defined the success of an offensive substitution if a goal is scored afterward. Similarly, a defensive substitution is considered positive if no goal is allowed from that moment until the end of the match. According to those definitions, the logical consequence would be that late offensive substitutions are not so successful, while late defensive changes tend to be more effective.



(a) Result of offensive substitutions through time (b) Result of defensive substitutions through time

Figure 6.2: Distribution through time for both the 2-Class and 3-Class Scoreline Evaluation Rule, comparing positive and negative substitutions. Inconsequential substitutions are not shown.

The result of offensive substitutions is in the same line as Figure 3.2, where we saw that the intensity of goal-scoring chances decreases after the last substitution. Those substitutions made at the last minutes of the match have a lower xG per minute, and together with the less time to score, results in a very low amount of successful late offensive substitutions. In the study by Amez et al. [1] after the third substitution, the probability of an opposition goal decreases. Our results are according, we see that late defensive substitutions, which usually are the third window, have a very high rate of success, given by their clean sheets.

6.1.2 Win probability model

Our in-game win probability model is based on the simulation of matches given the goal-scoring probability. This is added to the current scoreline for the simulated results. The effect of scoring a goal gets much bigger with time. While a goal in the first-half gives the other team plenty of time to turn it around, the closer the goal is to the final whistle, the higher the expected points swing. In this regard, the model successfully portrays intuitive ideas. The idea here is to study in which part of the match substitutions have a bigger effect.

In Figure 6.3 we have a box plot showing, for each minute, the swing in EP due to the substitutions in that minute. The plot shows three zones: from half-time through the 66$^{\text{th}}$ minute, the ΔEP are widely distributed, the outliers are rare, and the Q1-Q3 range is big; from the 66$^{\text{th}}$ to the 76$^{\text{th}}$ minute, the interquartile range gets smaller, but swings in EP are still wide; after the 76$^{\text{th}}$ minute until the end of the match, very few substitutions have a significant change in expected points and are classified as outliers. We can observe too that the box plot for each minute is mainly centered, showing that there is no apparent better timing to do a substitution to obtain a win probability advantage. We do notice that the model is more sensitive to substitutions early in the second-half.

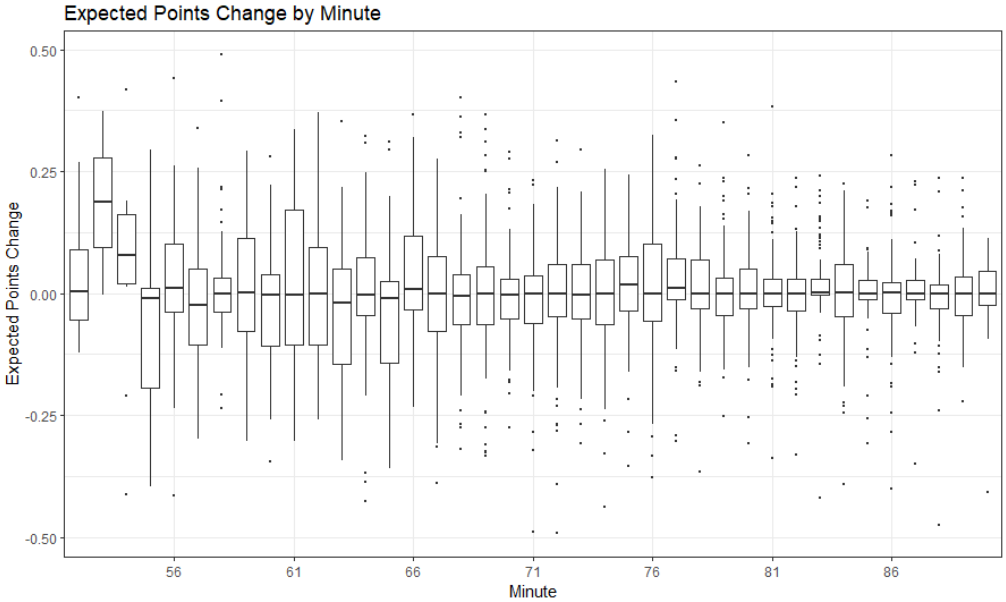


Figure 6.3: Distribution of the variation of EP after a substitution. Distribution at each minute.

6.2 Simulation of Alternative Timing

In this section we will study the alternative timing for the substitutions coaches did during the studied matches. This is, for each substitution that had been made, we are simulating moving it forward and backward through time. We have respected the coaches' choices in terms of the type of substitution and the different substitutions windows used, so if a team made the first two changes at the same time and then two isolated ones, we have also done so in the simulations, while also including a five-minute minimum space between the simulated substitutions, since it is no real scenario wasting two windows in two consecutive player changes, except for extraordinary cases such as an injury, a red card, or trying to lose time.

Simulations have followed the same structure for our two types of models. Knowing how the substitutions were packed into windows and their offensive or defensive nature, we were able to generate new game states $x_{t,i}$ by varying the substitution information. This applied both to the classification models, which had the updated state input, and the win probability model, in a similar way to the simulations previously explained in this thesis.

6.2.1 Substitution classification models

After generating the adequate game states for simulation, we applied the trained substitution classification models. We grouped substitutions according to their proximity to the real match case, and the results are shown in Table 6.1. Time groupings are not homogeneous because we want to remark the minutes closer to the real substitution. We show the change in the percentage of positive or successful substitutions. This is, offensive substitutions, when simulated from 6 to 10 minutes prior to their real minute, obtained better predictions 3.6% of the time.

	Change in positive/successful substitutions (%)					
Time to real sub	[-20,-11]	[-10,-6]	[-5,-1]	[1,5]	[6,10]	[11,20]
Offensive Subs	+8.2	+3.6	+0.4	-0.2	-1.3	-6.5
Defensive Subs	-7.6	-3.4	-1.2	+0.9	+2.5	+7.3
2C-SER	+3.2	+2.5	+0.4	+0.2	-0.1	-0.4
3C-SER	+1.5	+0.7	0.0	-0.2	-0.7	-1.1
xV-Change Rule	+10.2	+6.1	+1.4	-2.1	-4.2	-8.3

Table 6.1: Results of timing simulations with the classification models. Results grouped by different assessment techniques and time with respect to the real substitution. Numbers represent the variation in the percentage of positive or successful substitutions. Negative numbers represent a decrease in the number of successful outcomes

Results in Table 6.1 are in line with most of the other results. Offensive substitutions have a higher probability of success if done earlier, while defensive ones behave the other way around: success increases as they are done later in the match. This agrees with the results in Figure 6.2, and the main reasons are the same: time left to achieve the goal. The Scoreline Evaluation Rule, both for the two and three-class problems, seems to be a bit better when substitutions are done earlier than in real life, but the numbers are very small and not very significant. The biggest results are for the xV Change Rule, which compared the incoming player performance to the outgoing one. We see that having more time as a substitute is correlated with a higher probability of a successful substitution.

6.2.2 Win probability model

From Figure 6.3, we observed that there is no apparent general better moment to do a substitution, but on a single match we have observed significant changes in the win probabilities. Thus we simulate for each substitution an alternative moment. The results, in general, agree that substitutions should be done, if anything, previously to the moment when they happen in reality. In Figure 6.4 we have plotted the distribution of ΔEP depending on the time previous to which the substitution took place. The biggest changes are when substitutions are made 30 minutes prior to their real-time, which practically could never happen.

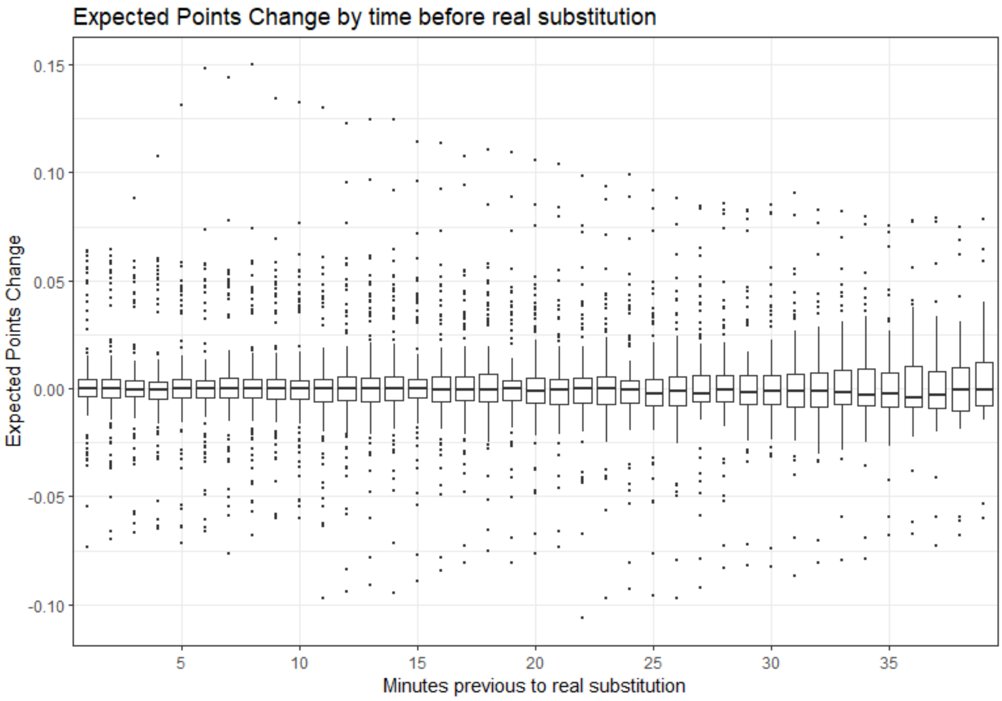


Figure 6.4: Distribution of the variation of EP depending on the time previous to the real substitution

Again, substitutions in a match are related, so considering their relations is relevant. If we apply it to the Levante 3-3 Real Madrid match we discussed earlier, whose expected points evolution is in Figure 5.3, we can see that Madrid could have started their comeback before, according to the model. This is due to Madrid having a higher Elo rating, so the model gives a higher goal-scoring probability by doing the substitutions, so it suggests doing it before and thus enhancing the winning probabilities. On the other hand, for Levante, it shows very little change when just changing the timing. Rey et al. [28] showed that losing teams substitute before while leading teams tend to delay changes, and our analysis suggests it is the optimal strategy.

While the previous match may have been more frenetic and high-scoring, we can focus also on a quieter and more measured game, which now presents an equally valuable opportunity for analysis. In fact, by studying the strategies and tactics used in a more low-key setting, we may be able to uncover new insights and approaches that could prove useful in future matches. This is the case for Osasuna 0-0 Espanyol, played also at the start if the 2021/22 *LaLiga*. Both teams had very similar Elo ratings, and it was a close goal-less match. Espanyol made a half-time sub, and then a double substitution at the 63$^{\text{rd}}$ minute, after which they had 1.07 EP. Had they done the substitution before, the EP after the substitution would have been 0.02 EP above the value they got in the real match. When arrived to the 63$^{\text{rd}}$ minute, the simulation gives the same value, since the game state $x_{t,i}$ is the same. The fact is, during the minutes the substitution had been done, Espanyol had a higher win probability. When we make this double substitution later in the match, Espanyol's EP goes lower while Osasuna's goes up.

In that match, Osasuna just made three changes. In this particular case, the variation with time of those substitutions has almost no effect, less than 0.01 EP on the winning probabilities. What simulations do say is that, if those changes had been offensive, a real change could have been seen, as we discussed in section 5.2. Had Osasuna's coach decided to introduce a fourth and fifth bench player into the match, their EP would have gone down by close to 0.02 EP, so just staying with the players on the field is the decision the model would have suggested.

6.3 Results and Discussion

Timing of substitutions is key. Most of the substitutions occur between the 60^{th} and the 85^{th} minute [4], in our case we found that 65.3% of the substitutions were on that time period. Early substitutions showed a higher tendency to turn the result into a desired one, while late substitutions have a higher percentage of failure. By bringing fresh legs and new tactics onto the field, coaches have the opportunity to inject energy and creativity into their team's performance, potentially shifting the momentum in their favor. Even though generated chances, measured by xG, and actual goal scoring are not directly related, the analysis for both cases is very similar. The simulations with alternative timing on the Scoreline Evaluation Rule resulted in no significant change.

The distribution in time of success for offensive and defensive substitutions is the logical consequence of their definition. Since we need a goal to consider the offensive substitution as accomplished, the probability is higher if the change is done earlier. Similarly, the probability of keeping a clean sheet after a defensive change increases the less time the opposition has for goal-scoring. Moreover, the results align with the variation of goal-scoring intensity by Amez et al. [1]. Similarly, the number of *inconsequential* substitutions, those that do not vary the result, increases as the final whistle gets closer. Simulations with alternative timing gave according to results, both for offensive and defensive substitutions.

Offensive and defensive substitutions, which we defined as a change in the position occupied by the players has an effect on the centroid of the whole team shifting [23]. In subsection 3.2.2, especially captured in Figure 3.3, we discussed how substitutes tend to produce more expected value, which makes a logical decision to advance an offensive substitution if a player is performing poorer. While defensive substitutions have higher chances of being successful if done later in the match, it does not mean that the optimal strategy is such. Teams that are winning, though, have been observed to tend to delay their substitutions [28].

The in-game win probability model seems to have a higher sensibility towards early substitutions. This is logical due to the effect of the substitutions being able to have repercussions in the game, both for good and bad substitutions. This result, together with the analysis of the different assessments, points towards the same direction: early substitutions have a higher impact on the game.

The higher rate of xV per minute of substitutes is one of the reasons we obtain a tendency to better results for earlier substitutions. In the first stage of the investigation, no constraints were included, and the win probability model suggested the winning combination was to burn all substitutions as soon as possible. This, with the proper amount of information, does not make sense, since playing over 30 minutes with no substitutions left can be harmful to injury and fatigue wise. Since teams never use this strategy, and the last substitutions are done very late into the game ([4], [28], [14]), the model has not learned that it could be a problem. Adding fresh players does increase your goal-scoring probabilities, so the model suggests adding even more fresh players into the pitch. When doing time simulations with the xV Change Rule we obtained that doing earlier the substitutions resulted in more success. This means the incoming players are expected to outperform their colleagues if enough time is given to them because bench players outperform starters on average.

Simulations on complete matches exposed the problem of the model suggesting too early substitutions. We could not find any evidence of Myers' substitution rule [26] as an optimal strategy. Even though we have now five substitutions and study a different dataset of matches, the problems discussed by Silva and Swartz [31] still apply: similar strategies are equally useful and some *bad* patterns can bias the analysis.

Chapter 7

Conclusions

Substitutions are the main asset a coach has to directly influence the game. By substituting players on the field with fresher bench players, clubs' technical staff are able to introduce new tactics, make more adjustments and possibly change momentum. The objective of this thesis is to study optimal substitutions through the use of the data. In order to do so, we studied the substitution problem from different points of view.

Following the COVID pandemic, the majority of leagues introduced a fourth and fifth substitution for each team, as a temporary measure for the big amount of matches played in a small period of time. Some leagues decided to keep these extra substitutions, and this now seems to be the new rule of international top-tier football. Through the analysis of the European male top leagues, we responded to Research question 1, in which way did the introduction of the additional substitutions affect the coaches' maneuverability and the relative importance of substitutions.

The results show how clubs throughout Europe have needed a period of adjustment. The current season, 2022/23, is the third complete season where the 5 substitutions per match rule is valid. The number of actual player changes done by a team has increased significantly speaking since this rule was introduced and now seems to be stable, with 4.35 substitutions per team per game in average. Even though substitutions have increased, the percentage of matches that changed results after the second-half, which is the one affected by substitutions, has remained the same to pre-COVID era. The percentage of matches that are drawn at half-time and finish with a winner has also remained very similar. This is the result of teams evolving to continue to be competitive with each other, and we cannot see an effect of the increase of substitutions in match scorelines.

We defined substitutions as offensive or defensive depending on the positions the incoming and outgoing players occupy. If the positions are distinct, it means that the substitution has a tactical component, and the behavior of the team is changed [23]. Of course, same position substitutions can be due to tactical reasons or simply fatigue, injuries or bookings, but due to the data we are working with, we have no way to know. When the new 5-subs rule was introduced, coaches maintained their number of non-neutral substitutions per match in absolute value, which by the second season had been raised, equaling the percentage of almost 1 of every 4 substitutions being non-neutral. This is further proof of the adaptation process teams have taken since the COVID outbreak, and how the new 5-subs rule has changed the tactical landscape, with non-neutral substitutions increasing by a 50% in absolute numbers.

During the whole thesis, we have worked with public data available for many leagues, not only European male top-tier competitions, so that this analysis could be assimilated into other leagues. The counterpart is the lack of data, which in this case means many substitutions have a tactical effect we cannot see through the data. Having a better description of the substitutions would be a great asset to study how the change between 3-subs per match and 5 has changed the coaches' tactics. In particular, having event-by-event data is the main point of improvement of this thesis and throughout the conclusions will be mentioned multiple times.

Moving forward, we explored Research question 2, where we focus on the evaluation of substitutions. The aim of the thesis is to identify optimal substitutions, so first we needed to state what we define as a successful one. We based our assessment on three points of view: the scoreboard and its evolution, the goal-scoring chances and the players' performance. All three ways of assessing substitutions are correlated, but on a particular game or substitution the assessments can be distinct. Also, for all the perspectives we looked at the problem as a two-class problem, where substitutions are either positive or negative, and a three-class problem where we introduced *inconsequential* tags to the substitution without a sufficiently relevant effect.

Once we had the rules where we defined what we consider as *successful*, we trained different machine learning models, mainly k Nearest Neighbors, Support Vector Machines and Random Forest classifiers. The results were close, but relatively worse, than the ones obtained by Brutti et al. in a similar study [7], even though methodologies are slightly different and use distinct data. The best predictions were obtained for the three-class problems, due to accuracy being higher for the amount of inconsequential substitutions. We did not find any significant difference between the algorithms used to predict the classification of substitutions.

The second way in which we assessed substitutions was as a byproduct of an in-game win probability model, based on the idea by Robberechts et al. [29]. Our model predicts, given a game state, the probability of scoring a goal and simulates the rest of the match. Of the different techniques tried for the prediction of goal scoring, the one that resulted in more accurate predictions was the Gaussian Walk, as the Bayesian approach really transfers information between time frames. The validation curves shown by the other methods, the Multiple Logistic Regressors, Random

Forest and Multiple Random Forest were worse than the Bayesian method but still achieved a reasonable level of accurate predictions.

From this model, we studied how the win, tie and loss probabilities changed with the substitutions. While comparing to a model without the substitution information, we proved that player modifications are relevant to predicted goal scoring and, therefore, winning probability. We also observed that substitutions tend, in a majority, to give a boost to the teams' performance, given that the majority of the changes produced a positive swing in expected points.

These two main models, the classifiers and the win probability, are the base of the thesis. Classification models could be improved with more information about teams, such as their offensive and defensive performances, or a more detailed game state. The win probability model has a main point of improvement, which is that due to the observed higher performance of substitutes, it usually suggests more substitutions to be done earlier. Teams always wait until the end of the match to finish their substitution possibilities, for eventual red cards or injuries. The model has not learned the consequences of having the substitutions done so early, because there are no such cases in the training set.

Turning to Research question 3, we investigated the problem of knowing the expected performance of the bench players. Before proceeding to the simulations, we did an analysis of the distributions of the bench, depending on the number of allowed players and the starting formation. This allows us to make a general analysis of the substitutions without depending on each particular bench from each match.

We have worked with offensive, neutral and defensive substitutions for the type of substitution in the simulations, which are proven to modify the team's tactics [23]. For every substitution that happened in a match, we simulated as if it was any of the three options, using the win probability model to obtain a numeric value of the impact of the substitution. Having access to more detailed data, such as the event-by-event or tracking data, would allow these same models to introduce more features and different kinds of substitutions. If a similar model were to be implemented for real use, coaches would like a wider range for the types of available substitutions, with more tactical nuances than a simple offensive-defensive choice.

The results of the simulations were interesting. First, we saw that the change in expected points for offensive substitutions with respect to neutral was in majority positive. Instead, doing the substitution defensively gives, on average, a worse EP value than neutral. This means that the model tends to better value the offensive substitutions. But we saw that defensive changes can actually induce a positive swing in EP. It was the case of the Levante - Real Madrid match, where a lower Elo rated team, Levante, with a favorable result, the model suggested that a defensive player swap was the correct option. Overall, the results of the simulations showed that a reasonable amount of substitutions can have a significant change in their expected points value depending on the type of substitution.

The selection of a substitute has many improvement ways. The expected performance of a player is an analysis of its own, and in this thesis could only be theorized and very small worked with. Many metrics from extended seasons can be taken into account to properly assess the player, and football clubs could even include training data which is evidently never public. Moreover, studying relationships between teammates on the football pitch [6] can also be a powerful asset to identify optimal substitutions. The creation of a Bayesian model, where every piece of information about a player can be inputted for a better performance prediction, is a promising way to start the investigation on bench players' achievement forecast.

Finally, we addressed Research question 4 and dealt with the timing problem. We moved the substitutions through the time, respecting the windows on which they were made, and applying reasonable constraints for the simulations. The generated game states were fed into the models, both the classifiers and the win probability.

The result for offensive substitutions was that it was better to make them as soon as possible, which is logical since it gives the team more time to score a goal. In a similar way, when assessing substitutions by xV of the player, early substitutions are much more effective that the end of the match. Reversely, the defensive substitutions are proven more effective in the latest part of the match, since the opposition has less time to score. These findings are consistent with other analyses of goal-scoring frequencies and their correlation with substitutions [1]. Obviously, the results are a consequence of the definition. While bench players generate more xV and is logical to substitute a player performing poorly, it is not clear that delaying defensive substitutions is the optimal strategy, even though is what teams do [28].

When studying the substitutions with respect to the Scoreline Evaluation Rules, we see that positive subs happen before, in general, than negative ones. But when doing the simulations, we did not find a very significant change in the amount of positive and negative substitutions, neither for the two-class or three-class problem. Similarly, the simulations with the win probability model did not show much variation in time. We did observe that the substitutions that have a higher effect on the winning probabilities are the ones made early in the second-half and at half-time. These results align with the study by Silva and Swartz [31], where they found no apparent better moment to make the substitution, contradicting Myers' [26] timing rule.

While with the type of substitution we obtained a high proportion of substitutions with a relevant change in expected points, with the timing the results are much more restrained. In some cases, we obtain that doing the substitution earlier can provide the team with a higher winning probability during a few minutes, but the increments in EP are lower than with the type of substitute.

During the thesis we have worked separately on two different parts of the substitution problem: the incoming player and the time at which they do. A complete system of data-driven forecast of substitutions, as future work on this problem, should include a coordination of those two problems, and also the decision of the outgoing player. With access to event-by-event or tracking data, a more profound analysis can be done of a player's actions ([32], [11]) and their effect on the match, while joining with physical information through GPS [36] or multi-camera tracking systems ([8], [5], [4]), the system would be able to provide with a more clear view of the under-performers of the game.

Such a complete model could implement various ideas: trying to predict the decline in physical performance, identifying defensive under-performance from an opposition high value. Such information, adequately combined with a timing analysis and a proper substitute selection, could provide coaches and their staff with a powerful data-informed tool to gain a competitive advantage and to know the best data-based decisions. The key idea is to weave all the analysis for more real information about the problem. In this thesis, we simulated alternative substitute selection and alternative timing separately. Being able to introduce simulations of the outgoing player and combining the three factors could lead to a well-functioning complete model for forecasting and identification of optimal substitutions in soccer.

In this thesis we developed two types of models: classification models and a substitute sensible win probability model. We have used them to study the most beneficial substitutions and simulate alternative settings for better performance. We saw that coaches can definitely use substitutions to change the course of matches and that by selecting mostly the appropriate type of substitution can obtain a big increase in their win probabilities. Timing is less important numerically wise, but can play a relevant role in some cases.

In conclusion, this thesis has provided a comprehensive analysis of substitutions in soccer, exploring various approaches and methods for evaluating their impact on the game. Through our investigation, we have studied some factors that influence the effectiveness of substitutions, including the timing of the substitution and the tactical approach of the team from the type of substitution they choose. Our findings provide insights that can inform decision-making processes for coaches and analysts alike. Moving forward, it is clear that further research is needed to fully understand the complexities of substitutions in soccer and to develop more sophisticated methods for evaluating their impact. Nonetheless, we believe that this thesis represents a valuable contribution to the field of football analytics, and offers a solid foundation for future work on the substitution problem.

Bibliography

[1] Simon Amez et al. "The right man in the right place? Substitutions and goal-scoring in soccer". In: *Psychology of Sport and Exercise* 54 (2021), p. 101898.

[2] Ryan Beal et al. "Learning the value of teamwork to form efficient teams". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 05. 2020, pp. 7063–7070.

[3] Ryan Beal et al. "Optimising game tactics for football". In: *arXiv preprint arXiv:2003.10294* (2020).

[4] Paul S Bradley, Carlos Lago-Peñas, and Ezequiel Rey. "Evaluation of the match performances of substitution players in elite soccer". In: *International journal of sports physiology and performance* 9.3 (2014), pp. 415–424.

[5] Paul S Bradley and Timothy D Noakes. "Match running performance fluctuations in elite soccer: indicative of fatigue, pacing or situational influences?" In: *Journal of sports sciences* 31.15 (2013), pp. 1627–1638.

[6] Lotte Bransen and Jan Van Haaren. "Player chemistry: Striving for a perfectly balanced soccer team". In: *arXiv preprint arXiv:2003.01712* (2020).

[7] Nicholas Sangoi Brutti, Denio Duarte, and Guilherme Dal Bianco. "Predição da efetividade da substituição no futebol: caso campeonato Brasileiro da Série A". In: *Revista Brasileira de Computação Aplicada* 13.1 (2021), pp. 42–52.

[8] Christopher Carling et al. "Work-rate of substitutes in elite soccer: A preliminary study". In: *Journal of science and medicine in sport* 13.2 (2010), pp. 253–255.

[9] Daniel Barbosa Coelho et al. "Effect of player substitutions on the intensity of second-half soccer match play". In: *Revista Brasileira de Cineantropometria & Desempenho Humano* 14 (2012), pp. 183–191.

[10] James Curley. *worldfootballR: R Interface to the World Football Database API*. R package version 0.1.0. 2021. URL: https://github.com/jalapic/worldfootballR.

[11] Tom Decroos et al. "Actions speak louder than goals: Valuing player actions in soccer". In: (2019), pp. 1851–1861.

[12] Julio Del Corral, Carlos Pestana Barros, and Juan Prieto-Rodriguez. "The determinants of soccer player substitutions: A survival analysis of the Spanish soccer league". In: *Journal of Sports Economics* 9.2 (2008), pp. 160–172.

[13] Paul Gamble, Lionel Chia, and Sian Allen. "The illogic of being data-driven: reasserting control and restoring balance in our relationship with data and technology in football". In: *Science and medicine in football* 4.4 (2020), pp. 338–341.

[14] Miguel-Angel Gomez, Carlos Lago-Peñas, and L Adam Owen. "The influence of substitutions on elite soccer teams' performance". In: *International Journal of Performance Analysis in Sport* 16.2 (2016), pp. 553–568.

[15] Sam Green. *Assessing the Performance of Premier League Goalscorers*. Blog post. Apr. 2012. URL: https://www.statsperform.com/resource/assessing-the-performance-of-premier-league-goalscorers/.

[16] Samuel P Hills et al. "Profiling the responses of soccer substitutes: A review of current literature". In: *Sports Medicine* 48 (2018), pp. 2255–2269.

[17] Nobuyoshi Hirotsu and Michael Wright. "Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions". In: *Journal of the Operational Research Society* 53.1 (2002), pp. 88–96.

[18] Lars Magnus Hvattum and Halvard Arntzen. "Using ELO ratings for match result prediction in association football". In: *International Journal of forecasting* 26.3 (2010), pp. 460–470.

[19] International Football Association Board. *Laws of the Game 2020/21*. Retrieved on April 24, 2023. 2020. URL: https://digitalhub.fifa.com/m/1cf301829f1cf996/original/ifab-laws-of-the-game-2020-21.pdf.

[20] Kaan Koseler and Matthew Stephan. "Machine learning applications in baseball: A systematic literature review". In: *Applied Artificial Intelligence* 31.9-10 (2017), pp. 745–763.

[21] Premier League. *What's new in 2022/23: Five substitutes*. Retrieved on April 24, 2023. 2021. URL: https://www.premierleague.com/news/2669074.

[22] Verlaine Leite and Roberto Figueredo. "Is there a need to increase the number of substitutions in modern professional football?" In: *Fizička kultura* 74.1 (2020), pp. 5–18.

[23] Miguel Lorenzo-Martınez et al. "The effect of substitutions on team tactical behavior in professional soccer". In: *Research quarterly for exercise and sport* 93.2 (2022), pp. 301–309.

[24] Johannes Meyer and Stefanie Klatt. "Impact of One Additional Substitution on Player Load and Coaching Tactics in Elite Football". In: *Applied Sciences* 11.16 (2021), p. 7676.

[25] Magni Mohr, Peter Krustrup, and Jens Bangsbo. "Match performance of high-standard soccer players with special reference to development of fatigue". In: *Journal of sports sciences* 21.7 (2003), pp. 519–528.

[26] Bret R Myers. "A proposed decision rule for the timing of soccer substitutions". In: *Journal of Quantitative Analysis in Sports* 8.1 (2012).

[27] Robert Rein and Daniel Memmert. “Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science”. In: *SpringerPlus* 5.1 (2016), pp. 1–13.

[28] Ezequiel Rey, Joaqun Lago-Ballesteros, and Alexis Padrón-Cabo. “Timing and tactical analysis of player substitutions in the UEFA Champions League”. In: *International Journal of Performance Analysis in Sport* 15.3 (2015), pp. 840–850.

[29] Pieter Robberechts, Jan Van Haaren, and Jesse Davis. “Who will win it? An in-game win probability model for football”. In: *arXiv preprint arXiv:1906.05029* (2019).

[30] SciSports. *State of the Football Analytics Industry in 2021*. Retrieved on April 24, 2023. 2021. URL: https://www.scisports.com/state-of-the-football-analytics-industry-in-2021/.

[31] Rajitha M Silva and Tim B Swartz. “Analysis of substitution times in soccer”. In: *Journal of Quantitative Analysis in Sports* 12.3 (2016), pp. 113–122.

[32] Karun Singh. *Introducing Expected Threat (xT)*. Blog post. Accessed: March 7, 2023. 2019. URL: https://karun.in/blog/expected-threat.html.

[33] Leigh Van Valen. “The red queen”. In: *The American Naturalist* 111.980 (1977), pp. 809–810.

[34] Carlos Varela-Quintana, Julio del Corral Cuervo, and Juan Prieto-Rodriguez. “The effect of an additional substitution in association football. Evidence from the Italian Serie A”. In: *Revista de psicologıa del deporte* 25.1 (2016), pp. 101–105.

[35] Dinesh Vatvani. *Upgrading Expected Goals*. Blog post. May 2022. URL: https://statsbomb.com/articles/soccer/upgrading-expected-goals/.

[36] Jason D Vescovi and Terence G Favero. “Motion characteristics of women’s college soccer matches: Female Athletes in Motion (FAiM) study”. In: *International Journal of Sports Physiology and Performance* 9.3 (2014), pp. 405–414.