

Institute of Exercise Training and Sport Informatics

German Sport University Cologne

**The Importance of Context in Football Match Analysis
Research: A Statistical Modeling Perspective**

Doctoral thesis accepted for the degree

Doktor der Sportwissenschaft

by

Maximilian Klemp-Weins

from

Bonn

Cologne 2025

First reviewer: Univ.-Prof. Dr. Daniel Memmert

Second reviewer: Prof. Dr. Paul Bradley

Chair of the doctoral committee: Univ.-Prof. Dr. Mario Thevis

Thesis defended on: August 13th 2025

This dissertation was written and submitted for examination in accordance with the doctoral regulations of February 20th, 2013, and was revised and published in accordance with the doctoral regulations of July 9th, 2024.

Affidavits following §7 section 2 No. 4 and 5 of the doctoral regulations from the German Sport University Cologne, February 20th 2013 as well §7 section 2 No. 9 of the doctoral regulations from the German Sport University Cologne, July 9th 2024:

Hereby I declare:

The work presented in this thesis is the original work of the author except where acknowledged in the text. This material has not been submitted either in whole or in part for a degree at this or any other institution. Those parts or single sentences, which have been taken verbatim from other sources, are identified as citations.

I further declare that I complied with the actual “guidelines of qualified scientific work” of the German Sport University Cologne.

Cologne, 22.09.2025

Date, Signature

A handwritten signature in black ink, appearing to read 'M. Klaf', with a stylized, flowing script.

"All models are wrong, some are useful"

— **George E. P. Box, 1976**

Acknowledgements

First of all, I want to thank Daniel for giving me the opportunity to conduct my Ph.D. studies with such great flexibility and freedom, while still always keeping an eye on me not getting caught up in too many side projects and perfectionism. I am deeply grateful for the trust you placed in me over the years, and for the opportunities you gave me to grow as a scientist.

Also, I want to thank Paul Bradley who (unknowingly) sparked my interest in the study of physical performance but, more importantly, taught me very early in my scientific career that “Context is King”.

A huge thanks is due to Robert who showed me that stats was fun, coding was fun, and taught me most of what I know about either of them today. It is safe to say that I would be nowhere near where I am today without you. Thank you for your invaluable guidance and mentoring.

A big part of why I enjoyed this Ph.D. journey so much has always been the amazing atmosphere in our lab. I want to thank all the colleagues that I have worked with over the last 6 years, with a special mention to my fellow Ph.D. students. Starting with the people who introduced me so kindly to the Institute and the infamous room 113: Ann-Kathrin, Dennis, Giulia, Dennis, Niels, Michel, Marius, Ben, Fabian, and Anne, followed by the next generation of Ph.D. students to whom I owe unforgettable memories: Ash, Manu, Saumya, Henrik, Dominik, Justus, and Albert. You all made this work not feel like work, through professional support, inspiring discussions, and a lot of fun we had inside and outside of the lab. I want to devote a special thanks to Marc for all the (more or less professional) discussions at the kitchen table and the mutual support during the regular crises. Gràcies germà. Also, I want to thank all students, interns, and assistants who have supported me in various ways in conducting my projects. And, of course, a big thank you to all my friends who have helped me to disconnect and recharge in order to keep my motivation and energy high for all the years.

It is hard to put into words how grateful I am to my family who has supported me not only over the course of my Ph.D. studies but throughout my entire life. To my Mom who has worked incredibly hard for me to pursue my goals while providing me

with loving support. To Phil who has always been there for me as my brother and friend (and advanced Algebra tutor). And to Geli and Josef who have supported me in any possible way at any time. I want to dedicate this work to my Dad who can unfortunately not be here to experience this moment with us.

And, last but not least, I want to express my endless gratitude for having you, Hannah, in my life. Thank you for your endless, but most importantly, unconditional support. You have given me the strength and belief that I can do this, while letting me know that I didn't have to. Thank you for your patience and your ability to make me laugh in an instant.

Abstract

Match Analysis Research (MAR) in football aims to enhance the understanding of team and player performances by linking specific *Performance Indicators* to successful outcomes. Despite rapid advancements in data availability and analytical techniques, inconsistencies in research findings remain prevalent, undermining the reliability and applicability of *MAR* results. These inconsistencies frequently stem from methodological limitations and a lack of contextualization, where crucial *Contextual Factors* influencing performance are not adequately accounted for in study designs and analyses. The primary motivation for this dissertation emerges from the necessity to address these methodological shortcomings, to understand their effects on study results, and to reinforce the scientific rigor and reliability within the *MAR* domain.

This cumulative doctoral thesis addresses the critical need to comprehensively incorporate and evaluate *Contextual Factors* within *MAR*, applying statistical modeling and causal inference approaches. The thesis systematically investigates how context shapes football team behavior, influences performance indicators, and impacts match outcomes, providing insights that clarify previously reported contradictions and enable the formulation of recommendations for study design in *MAR*. This is achieved by formulating and pursuing five objectives, which are resolved in the course of this thesis.

The first objective of this dissertation was to offer a detailed, formal description of the *MAR* problem from both conceptual and statistical perspectives. To this end, the history and evolution of performance analysis, specifically *MAR* in football, are described, followed by a review of the statistical modeling strategies applied in the research field. The statistical learning problem related to *MAR* is formally described alongside a formulation in the language of causal inference.

The second objective involved systematically assessing the influence of contextual variables on team behaviors and performance. A critical distinction is made between match-level context and dynamic in-match context, elucidating how these factors differentially shape the tactical, technical, and physical behaviors of teams.

The third objective explicitly aimed at providing coherent explanations for controversial and contradictory results frequently reported in previous *MAR* literature.

The impact of *Contextual Factors*, as well as the effects of study design choices (specifically data aggregation strategies), are discussed in the light of the causal inference framework, proposing an explanation for the emergence of confounder bias in *MAR* studies.

The fourth objective further expands this methodological discourse, exploring the interrelations between study design choices, contextual variables, and biases within *MAR*. Leveraging the theoretical framework introduced over the course of the thesis, specifically the synopsis of the dissertation studies, presents empirical evidence for the potentially confounding effect of context. The effect of different data aggregation strategies on study findings is demonstrated empirically.

The final objective of the dissertation moves beyond the identification and description of these methodological challenges toward providing tangible methodological solutions. First, a methodology for investigating the risk of confounding due to *Contextual Factors* using the statistical notion of *collapsibility* is proposed. Next, the *segmented match analysis paradigm* is suggested as a principle for study design that provides robustness against confounder bias due to choosing the appropriate level of aggregation and allowing the integration of *Contextual Factors*.

Collectively, this thesis highlights the critical importance of context in football match dynamics and outcomes. It points out limitations of traditional methods that overlook the relationship between context, study design, and potential bias. By combining statistical methods and the causal inference framework with insights into football performance, the thesis advances *MAR* methodology significantly. The findings provide researchers with robust tools for context-aware analyses. Additionally, the practical insights offer valuable guidance for practitioners aiming to enhance team performance analysis and decision-making in football.

Zusammenfassung

Die *Spielanalyseforschung* (engl. *Match Analysis Research, MAR*) im Fußball zielt darauf ab, das Verständnis von Team- und Spielerleistungen zu verbessern, indem der Zusammenhang zwischen Leistungsindikatoren und Erfolg untersucht wird. Trotz regelmäßiger Innovationen in der Datenerhebung und analytischen Verfahren bleiben widersprüchliche Forschungsergebnisse ein Problem, was die Glaubwürdigkeit und Anwendbarkeit von *MAR*-Erkenntnissen beeinträchtigt. Diese Widersprüche sind oftmals auf methodische Inkonsistenzen und einen Mangel an Kontextualisierung zurückzuführen. Zentrale Kontextfaktoren werden in Studiendesigns nicht angemessen berücksichtigt, obwohl sie starken Einfluss auf Leistung und Erfolg nehmen. Die vorliegende Dissertation ist motiviert durch die Notwendigkeit, diese methodischen Defizite zu adressieren, ihre Auswirkungen auf Studienergebnisse zu verstehen und die wissenschaftliche Stringenz sowie Aussagekraft innerhalb des *MAR*-Forschungsfelds zu stärken.

Diese kumulative Dissertation greift die dringende Notwendigkeit auf, Kontextfaktoren systematisch in die *MAR* zu integrieren und zu evaluieren. Hierzu werden statistische Modellierungsansätze und Konzepte der kausalen Inferenz angewandt. Die Arbeit untersucht systematisch, wie kontextuelle Einflüsse das Verhalten von Fußballmannschaften formen, Leistungsindikatoren beeinflussen und Spielausgänge prägen. Ziel ist es, bestehende Widersprüche in der Literatur zu klären und Empfehlungen für zukünftige Studienkonzeptionen abzuleiten. Die Umsetzung dieser Zielsetzung erfolgt anhand von fünf Zielen, die im Rahmen der Dissertation bearbeitet werden.

Das erste Ziel bestand in der konzeptionellen und statistischen Formalisierung des *MAR*-Problems. Hierzu wird die historische Entwicklung der Leistungsanalyse im Sport nachgezeichnet sowie der Ursprung der *MAR* untersucht. Daran anschließend wird ein Überblick über die in der *MAR* verwendeten Modellierungsansätze gegeben. Zudem wird das statistische Lernproblem im Kontext von *MAR* formal dargestellt und in die Terminologie der kausalen Inferenz übersetzt.

Das zweite Ziel war die systematische Untersuchung des Spielkontexts sowie dessen Einfluss auf das Verhalten und die Leistung von Fußballmannschaften. Eine zentrale Unterscheidung wird zwischen kontextuellen Merkmalen auf Spielebene und sich während des Spiels verändernden Kontextfaktoren getroffen, um deren unterschiedliche Auswirkungen auf das Verhalten von Mannschaften zu beleuchten.

Das dritte Ziel bestand darin, kohärente Erklärungen für widersprüchliche Ergebnisse in der bestehenden *MAR*-Literatur zu liefern. In diesem Zusammenhang wird der Zusammenhang zwischen Kontextvariablen, der Wahl des Studiendesigns sowie den beobachteten Effekten im Rahmen des Kausalitätsparadigmas diskutiert. Daraus wird eine Erklärung für das Auftreten von Konfundierungseffekten in *MAR*-Studien abgeleitet.

Das vierte Ziel sah eine Erweiterung dieser methodologischen Diskussion durch eine empirische Untersuchung der Zusammenhänge zwischen Studiendesign, Kontextfaktoren und potenziellen Verzerrungen in *MAR* vor. Unter dem theoretischen Rahmen, der im Verlauf der Arbeit entwickelt wurde, wird insbesondere anhand der in der Synopse dargestellten empirischen Studien gezeigt, dass Kontextfaktoren potenziell konfundierende Effekte auslösen können. Dazu werden vor allem die Auswirkungen unterschiedlicher Datenaggregierungsstrategien auf Studienergebnisse empirisch untersucht.

Das fünfte und letzte Ziel dieser Dissertation geht über die Beschreibung methodologischer Herausforderungen hinaus und fokussiert auf konkrete Lösungen. Zunächst wird eine Methodik zur Bewertung des Risikos für konfundierende Effekte anhand des statistischen Konzepts der *Collapsibility* vorgestellt. Anschließend wird das Paradigma der *segmentierten Spielanalyse* als konzeptionelle Grundlage für robuste Studiendesigns vorgeschlagen, welches ein angemessenes Aggregationslevel erlaubt und die Integration relevanter Kontextfaktoren ermöglicht.

Zusammenfassend illustriert diese Dissertation die entscheidende Bedeutung von Kontextfaktoren in der Analyse von Fußballspielen. Sie zeigt die Grenzen der üblichen Methodologie auf, die den Zusammenhang zwischen Kontext, Studiendesign und potenziellen Verzerrungen vernachlässigt. Durch die Integration statistischer Modelle und kausaler Konzepte mit sportwissenschaftlicher Expertise leistet die Arbeit einen zentralen Beitrag zur Methodologie der Spielanalyse. Die Erkenntnisse bieten agierenden Personen aus Wissenschaft wie Praxis gleichermaßen fundierte Grundlagen für kontextsensitive Analysen und datenbasierte Entscheidungsprozesse im modernen Fußball.

List of Publications

Peer-review publications

The following articles have been published (one in submission) as part of this cumulative dissertation. These articles include four first-authorships (one shared first-authorship) as well as three co-authorships.

- I Klemp, M., Wunderlich, F., & Memmert, D. (2021).** In-play forecasting in football using event and positional data. *Scientific Reports*, 11(1). [IF 2023: 7.5, **Q1**]
- II Klemp, M., Memmert, D., & Rein, R. (2021).** The influence of running performance on scoring the first goal in a soccer match. *International Journal of Sports Science & Coaching*, 17 (3), 558–567. [IF 2023: 3.5, **Q1** Social Sciences (miscellaneous)]
- III Klemp, M., Memmert, D., & Rein, R. (In preparation).** The Role of Contextual Factors in Match Analysis Research and their Potential for Confounding Effects: A Case for a Segmented Match Analysis Paradigm.
- IV Memmert, D., Klemp, M. (shared first-authorship), Schwab, S., & Low, B. (2023).** Individual attention capacity enhances in-field group performances in soccer. *International Journal of Sport and Exercise Psychology*, 1–18. [IF 2023: 7.3, **Q1** Social Psychology]
- V Brinkschulte, M., Furley, P., Klemp, M., & Memmert, D. (2021).** English goalkeepers are not responsible for England's poor performance in penalty shootouts in the past. *Scientific Reports*, 11(1). [IF 2023: 7.5, **Q1**]
- VI Guerrero-Calderón, B., Klemp, M., Castillo-Rodriguez, A., Morcillo, J. A., & Memmert, D. (2020).** A new approach for training-load quantification in elite-level soccer: Contextual factors. *International Journal of Sports Medicine*, 42(08), 716–723. [IF 2023: 4.8, **Q1** Physical Therapy, Sports Therapy and Rehabilitation]
- VII Guerrero-Calderón, B., Klemp, M., Morcillo, J. A., & Memmert, D. (2021).** How does the workload applied during the training week and the contextual factors affect the physical responses of professional soccer players in the match? *International Journal of Sports Science & Coaching*, 16(4), 994–1003. [IF 2023: 3.5, **Q1** Social Sciences (miscellaneous)]

Table of Contents

Acknowledgements	III
Abstract	V
Zusammenfassung	VII
List of Publications	IX
List of Figures	X
1 Introduction	1
1.1 Motivation	3
1.2 Objectives	5
1.3 Structure of the Thesis	6
2 Literature Review	7
2.1 Performance Analysis in Sports	7
2.1.1 Endurance Performance	9
2.1.2 Strength and Power Performance	10
2.1.3 Physical Performance Analysis in Football	12
2.1.4 Deterministic Models in Performance Analysis	14
2.2 Match Analysis Research	15
2.2.1 Contextual Factors in <i>MAR</i>	19
2.2.2 Level of Observation in <i>MAR</i>	24
2.2.3 Randomness and Uncertainty in Football	26
2.2.4 Types of Data Utilized in <i>MAR</i>	30
2.3 Statistical Learning	33
2.3.1 Formalization of the Statistical Learning Problem in <i>MAR</i>	36
2.3.2 Statistical Modeling Approaches in <i>MAR</i>	38
2.4 Causal Inference	42
2.4.1 <i>MAR</i> as a Case of Observational Research	42
2.4.2 The Counterfactual Model of Causation	44
2.4.3 Confounding in Causal Inference	46
2.4.4 Confounding in <i>MAR</i>	48
2.4.5 The Experimental Match Analysis Paradigm	50
3 Synopsis	52
3.1 Study 1: In-play forecasting in football using event and positional data	52
In-play forecasting in football using event and positional data	54
3.2 Study 2: The influence of running performance on scoring the first	
goal in a soccer match	55
The influence of running performance on scoring the first goal in a soccer	
match	57

3.3	Study 3: The Role of Contextual Factors in Match Analysis Research and Their Potential for Confounding Effects: A Case for a Segmented Match Analysis Paradigm	58
4	Discussion and Conclusion	60
4.1	Summary	60
4.2	Main Findings	61
4.3	Practical Implications	62
4.3.1	Implications for Research	62
4.3.2	Implications for Sports Practice	64
4.4	Contributions	65
4.5	Limitations and Scope	67
4.6	Future Directions	68
	References	70
A	Appendix	95
	Study 4	96
	Study 5	97
	Study 6	98
	Study 7	99

List of Figures

- 2.1 Causal diagram of a confounder C (Hernán, Clayton, and Keiding 2011). 47

1 Introduction

Association football has fascinated people all over the world for more than a century and attracts vast numbers of active participants as well as spectators. In 2024, an estimated 250 million active members were organized in 211 national football associations under the umbrella of the World Federation *FIFA* (Weil and Britannica 2024), and the latest men’s football World Cup final in 2022 attracted 1.5 billion spectators worldwide (Jones 2023). Similarly, 260 million people watched the women’s football World Cup final in 2019 (Glass 2019).

Since its early days, the sport’s popularity has motivated fans, practitioners, and scholars to think and talk about players’ or teams’ performances and ways to improve the latter. While Thomas Reilly began to investigate the physical demands of the game in the 70s (Reilly and Thomas 1976), the study of tactical aspects dates back to the seminal work of Charles Reep (Reep and Benjamin 1968), identifying mathematical laws to drive the occurrence of the number of goals or the length of passing sequences.

These early works can be considered the foundations of the study of *Performance Analysis* (Franks and Goodman 1986) in team sports, which has again led to the more specific study of *Match Analysis* (Memmert 2021), which describes the analysis of performance in competition.

In the decades following these pioneering studies, football practitioners have made use of match analysis in order to prepare their teams for upcoming matches (Mehta et al. 2023). At the same time, academic scholars have tried to advance understanding of the dynamics of the game by uncovering quantitative principles driving the behavior of teams and the results of matches (McGarry 2017). Knowledge of such dynamics has been utilized to guide the design of training regimens (Paul S. Bradley 2020), understand the influence of a game’s circumstances (Guerrero-Calderón et al. 2021), or to forecast the outcomes of future matches (Wunderlich and Memmert 2020). An overarching theme of most endeavors in the field of *Match Analysis Research (MAR)* has been the identification of *Performance Indicators* that are positively linked to successful outcomes (Hughes and Bartlett 2002).

The first systematic approaches in *MAR* mainly utilized notational analysis, gathered by human observers either witnessing the performance live or watching video recordings (Franks and Goodman 1986). While early approaches to this task relied

on pen-and-paper work (Reep, Pollard, and Benjamin 1971), advances in technology, particularly the advent of the Personal Computer, introduced computer-aided systems to the field (Nevill, Atkinson, and Hughes 2008; Hughes 1984).

This first era of *MAR*, relying on computer-aided video analysis and notational data, revealed the first insights into successful tactics (A. Tenga et al. 2017) or playing styles (Fernandez-Navarro et al. 2016), while also shedding light on more specific phenomena like defensive transitions (Vogelbein, Nopp, and Hökelmann 2014) or creative behavior (Kempe and Memmert 2018).

The beginning of the 21st century brought two main innovations to the field of *MAR*. First, technological innovations with respect to data collection techniques introduced new types of data and made this data available for a broader number of matches. In particular, player-tracking technologies enabled the generation of *position data*, locating every player and the ball on the pitch up to 25 times per second (Manaffard, Ebadi, and Moghaddam 2017). At the same time, the original coding systems used to create notational data were further developed and complemented with semi-automated data generation techniques. This resulted in the advent of *event data*, logging every event in a match accompanied by its timestamp, location, and further features describing the specifics of the action (Garnica-Caparrós 2024). Second, mathematical advances from the field of statistics and machine learning began to gain traction in various research areas, including sports science (Rein and Memmert 2016; Memmert, Lemmink, and Sampaio 2016). Methods of data mining and machine learning enabled researchers as well as practitioners to handle the increasing volume and complexity of data that was now available due to the previously mentioned data collection techniques.

These developments caused a leap in the field of *MAR* with respect to sample sizes and complexity of analyses. Specifically, the traditional statistical modeling approaches were complemented and replaced mainly by machine learning techniques (Bunker, Yeung, and Fujii 2023; Herold et al. 2019). One major contribution of the application of machine learning to football data was the development of the so-called *expected metrics*, which estimate the success probabilities of certain actions or sequences based on context and other predictors. In this strand of research, the most prominent metric is probably *expected goals* (Macdonald 2012; Mead, O’Hare, and McMenemy 2023), with other aspects of match-play such as passes (Anzer and Bauer 2022) or counterattacks (Biermann et al. 2023) also being investigated.

Potentially the most generic and advanced type of expected metric is the *expected possession value*, which estimates a team’s probability of scoring from a given possession. Expected possession value models have been proposed utilizing machine learning techniques for event data (Decroos et al. 2019) or position data (Fernández,

Bornn, and Cervone 2019) as well as Markov chain models (Van Roy et al. 2020). Models of this type allow for the quantification of the value or contribution of any action (Decroos and Davis 2020) or movement (Fernandez and Bornn 2018; Llana et al. 2022; Gregory et al. 2024) during a match.

While technological advances have changed the way in which football matches are being analyzed, the motivation behind these studies has remained largely the same since the very beginnings of *MAR* in the 1960s: identifying *Performance Indicators* that describe teams' behavior and performance as closely as possible while providing some prognostic value regarding match outcomes (McGarry 2017). Regardless of the methodology used, deriving causal conclusions from this type of data presents unique challenges and risks that will be discussed in detail in the present thesis.

1.1 Motivation

Match Analysis Research (MAR) in football has produced a vast body of studies, incorporating increasingly sophisticated metrics and models. Over the years, various approaches have emerged to address specific analytical challenges, such as computer vision for the extraction of tracking data (Thomas et al. 2019), expected metrics for action valuation (Van Roy et al. 2020), and space control models to predict player movement (Dick, Tavakol, and Brefeld 2021). Despite these advancements, a significant portion of the literature has continued to focus on the classical *MAR* problem: linking *Performance Indicators* to *Success Indicators* (Floris R. Goes, Kempe, and Lemmink 2019; Lepschy, Wäsche, and Woll 2018).

As is the case in any emerging research field, controversies and potential shortcomings in *MAR* have been pointed out over time. These shortcomings, characterized mainly by methodological inconsistencies, might cause heterogeneity in study outcomes and have the potential to undermine the trustworthiness of results obtained in the field. It is therefore mandatory to identify such potential sources of bias and to investigate how to avoid them.

The first concerns about the methodological approach of *MAR* have been raised through two articles published in the early 2010s. A review paper by Mackenzie and Cushion (2012) brought to attention the lack of contextualization in *MAR* studies. The authors point out that only a minority of included studies contextualized the analyzed performances with respect to the opponent. Furthermore, out of 42 studies in the review, only ten accounted for whether games were played at home or away. Another systematic review by Sarmiento and colleagues (2014) similarly highlighted the need for contextualization while also pointing out the need for predictive instead of comparative studies.

The authors of both reviews indicate that these methodological issues might contribute to the production of contradictory or paradoxical results. While their works have triggered the increased inclusion of contextual information into *MAR* studies (e.g., Paul Simon Bradley et al. 2014), controversies persist, most likely due to issues with respect to data aggregation strategies.

Indeed, contradictions within the *MAR* literature are evident, with studies reporting conflicting results on similar research questions. For instance, the relationship between physical performance and success has been found to be positive in some cases (Modric et al. 2019; Schauburger, Groll, and Tutz 2017; Ingebrigtsen et al. 2012), negative in others (Di Salvo et al. 2009), and sometimes nonexistent (Hoppe et al. 2014; Lepschy, Wäsche, and Woll 2020). Researchers have suggested that these inconsistencies may stem from variations in study design (Phatak et al. 2022), demonstrating that differences in data aggregation, covariate selection, or normalization can influence the direction of reported effects (Phatak, Rein, and Memmert 2021).

Importantly, even the advances in modeling techniques and the broad availability of complex data are not generally a remedy for these risks of bias through study design. Even with the use of machine learning models, which require almost no assumptions about the data, it is possible to experience issues of confounding through lurking variables and the inherent causal structure of research problems (Davis and Robberechts 2024). On the contrary, the use of more sophisticated methods might even bring new challenges while potentially conveying a false sense of certainty (Hernán, Hsu, and Healy 2019; Shah, Steyerberg, and Kent 2018; Christodoulou et al. 2019).

The contradictory findings in the field, along with the methodological concerns raised by multiple researchers, pose a considerable threat to the credibility of *MAR* studies. Moreover, the increasing complexity and volume of new data types have introduced additional challenges. In general, *MAR* suffers from vulnerability to biases associated with observational research, where no controlled experiments are possible (Greenland, Pearl, and Robins 1999). This vulnerability is likely aggravated by the heterogeneity in study design, specifically data aggregation strategies and the subsequent inclusion of contextual information.

The present doctoral thesis seeks to contribute to resolving these challenges by investigating the influence of contextual factors on team behavior in football and identifying potential sources of bias in *MAR* studies. By addressing these methodological concerns, this work aims to enhance the robustness and credibility of football *Match Analysis Research*.

1.2 Objectives

In the previous sections, the current state of the *MAR* field has been delineated very briefly, pointing out controversial findings and potential shortcomings due to methodological inconsistencies. Furthermore, the lack of consideration of contextual information seems to play a pivotal role in the emergence of such issues.

The efforts presented in this doctoral thesis are targeted at improving upon this situation by investigating the specific causes and mechanisms of controversies and by proposing methods to handle the complexities of the field of study.

In order to identify overarching problems and causes of potential bias, before finally proposing possible solutions, a clear picture of the *MAR* problem has to be drawn. Therefore, this thesis pursues multiple objectives: first, formalizing certain aspects of the *MAR* problem, then identifying potential issues, and finally proposing ways to handle these problems.

The objectives of the present thesis are the following:

Objective 1

Provide a formal description of the *MAR* problem in a conceptual and statistical sense.

Objective 2

Describe the effect of contextual variables on team behavior and performance.

Objective 3

Provide an explanation for the controversial results in the previous literature under a common framework.

Objective 4

Describe the relationships between contextual variables, study design, and potential bias in *MAR* studies.

Objective 5

Propose methods to assess the risk of bias due to study design and to avoid such bias.

1.3 Structure of the Thesis

The present doctoral thesis is structured in the following way.

In Chapter 2, a review of the literature relevant to the thesis is conducted. This literature review begins with the history and current state of the research field of *Performance Analysis*, which is the broader area from which *Match Analysis Research (MAR)* has originated. From there, the emergence of *MAR* is described, along with the most recent developments in the field, including the study of *Contextual Factors*, common choices in study designs, and the various statistical approaches to the problem. Lastly, the *causal inference* framework is introduced as a useful vehicle to view the *MAR* problem in a theoretical way and to derive potential sources of bias.

Chapter 3 contains the synopsis of this thesis, thereby outlining its main contributions through the articles forming the core of this cumulative dissertation. The chapter introduces these research articles by describing their backgrounds and objectives. The articles are included in the chapter as well in the format of publication or submission, and can be read independently from the main body of the thesis.

The thesis closes with Chapter 4, providing a summary of its contributions as well as a general conclusion. The chapter also discusses limitations of the research presented in this thesis and outlines possible future directions.

2 Literature Review

The main topic of this dissertation is the research area of *Match Analysis Research* in football and, more specifically, its methodological basis. To understand how the field of *Match Analysis Research* emerged in its current form, it is mandatory to review the history of *Performance Analysis* and the concept of a *Performance Indicator*.

Since the origins of *Performance Analysis* in sports lie in the field of exercise physiology, leading to the study of physical performance in team sports, these origins will be reviewed in order to explain the foundations of modern *Match Analysis Research*. The study of the physiological basis of sporting performance will be discussed, as well as the use of deterministic models in biomechanics, which has likely been the inspiration for the endeavour of seeking *Performance Indicators* related to successful outcomes.

This discussion will, therefore, lead to the study of *Match Analysis Research (MAR)*, which is essentially a case of *Performance Analysis* in competition. The modern approach to *Match Analysis Research*, utilizing statistical methods and machine learning, is discussed, and a formalization of the *MAR* problem in statistical terms is derived. It will then be shown that *MAR* is a case of observational research, bearing certain limitations and potential for bias. This issue is viewed in the light of Causal Inference, a theoretical framework that deals explicitly with drawing causal conclusions from (potentially) non-experimental data. The framework is introduced, along with its concepts and methods, and it is shown how the *MAR* problem can be modeled as a case of Causal Inference. From this analogy, it is stated how specific study designs commonly deployed in *MAR* might be prone to biases if not careful consideration is given to the causal structure of the problem.

2.1 Performance Analysis in Sports

Performance analysis in sports is concerned with advancing the understanding of sporting behavior (McGarry 2017) to inform decision-making in an attempt to enhance sports performance (O'Donoghue 2009). It may focus on the study of physical, technical, or tactical capabilities of athletes within a given sport. Performance Analysis generally aims to assess performance, understand how performance emerges

in various areas, and, consequently, to utilize this knowledge to improve performance (Franks and Goodman 1986).

Traditionally, Performance Analysis Research has attempted to study skills and traits that are thought to be related to sporting performance in competition. One approach to this problem has focused on the physiological basis of sporting performance, identifying organs and subsystems responsible for various aspects of physical performance in the human organism (Robergs 2003). These fundamental insights are then used to identify measurable equivalents of athletic performance, which can subsequently serve as targets for training and testing.

The physiological approach to human performance has its roots in the 19th century (Flint 1878), while the most significant advancements in understanding how physiological systems contribute to athletic performance emerged in the mid-20th century. Archibald V. Hill described the mechanism of muscle contraction for the first time in 1922 (A. V. Hill 1922). For this contribution, he was awarded the Nobel Prize for Physiology or Medicine jointly with Otto Meyerhof in 1922 (Bassett 2002). This work laid the foundation for an investigation of the physiological basis of human performance, specifically of muscular exercise. The following research in the field of performance analysis in exercise physiology has generated insights into the facets that comprise human performance and the markers of physiological capacity that underlie this performance.

The distinction between different aspects of physical performance was driven mainly by the various metabolic systems providing energy for performance on varying timescales (Baker, McCormick, and Robergs 2010) and can, on a high level, be made between endurance and strength performance¹. Endurance performance is characterized by a person's ability to sustain a particular mode of movement for a prolonged time, with duration and intensity varying according to the respective sporting discipline or training regime (Michael J. Joyner and Coyle 2008). Strength performance represents the ability to exert force on an external object or resistance (Siff 2000). Depending on the sport or event, this external resistance may be gravity, as when moving one's own body mass (Suchomel, Nimphius, and Stone 2016). Performance analysis research in the realm of exercise physiology in the 20th and early 21st century aimed to identify appropriate targets for training and testing

¹Of course the actual taxonomy of human physical performance is much more complex, spanning more nuanced distinctions within endurance and strength as well as performance types that are made up of a mixture of strength and endurance capacities. It is also acknowledged that a differentiation can and should be made with respect to the tissue that is involved with the generation of human performance such as the central and peripheral nervous system or cell metabolism. However, since exercise physiology is not a major aspect of this thesis, a more detailed differentiation is beyond the scope of this work.

endurance or strength performance capacities. The most important results from this strand of research are illustrated in the following.

2.1.1 Endurance Performance

A multitude of measures exist that aim to describe the physiological basis of endurance performance. One of the most prominent models of the physiology of endurance performance was proposed by Pate and Kriska (1984), stating three primary factors of aerobic endurance performance: VO_2max , Lactate Threshold (LT), and work economy (C). The model is supported by several authors and publications (e.g., Conley and Krahenbuhl 1980; Farrell et al. 1993; P. E. di Prampero et al. 1986) and can, therefore, be considered an agreed-upon framework of endurance performance capabilities.

Among these three physiological constructs, probably the most crucial physiological predictor of endurance performance is VO_2max . Hill and Lupton (1923) first described VO_2max as the maximal ability of the body to utilize oxygen during exercise. It is defined as the highest rate at which an individual can consume oxygen during intense exercise and is measured during an all-out running or cycling test by means of spirometry. VO_2max is commonly expressed in either absolute terms as liters per minute ($l \cdot min^{-1}$) or relative to body mass as milliliters per kilogram per minute ($ml \cdot kg^{-1} \cdot min^{-1}$), with the latter allowing for better comparisons between individuals of different body sizes (Michael J. Joyner and Coyle 2008).

Per-Olof Åstrand and Bengt Saltin, pioneers of exercise physiology, examined VO_2max kinetics during various types of exercise (Åstrand and Saltin 1961a), during the temporal course of exercise (Åstrand and Saltin 1961b), and measured VO_2max in elite athletes (Saltin and Åstrand 1967). Foster (1983) demonstrated that VO_2max is a powerful predictor in race performance across endurance events of different lengths, which was supported by later findings (Pietro Enrico di Prampero 2003; M. J. Joyner 1991).

As described by Bassett and Howley (2000), VO_2max is limited by the ability of the cardiorespiratory system to transport O_2 to the muscle, which is driven by cardiac output. Midgley and colleagues (2006) review the literature on VO_2max training and summarize that training-induced improvements can largely be attributed to changes in *stroke volume*, which is again determined by the strength of the *myocardium*.

It should be mentioned that there exists considerable debate about whether scaling VO_2max with respect to body mass linearly is a valid method (Ulrik Wisløff, Helgerud, and Hoff 1998) and that approaches of allometric scaling have been proposed to better account for the nonlinear relationship between body mass and exercise

oxygen cost (Bergh et al. 1991). However, further discussion of this issue is beyond the scope of this thesis.

The next factor in the model is LT or rather the highest possible exercise intensity at LT . The Lactate Threshold is defined as the highest work intensity that can be sustained without causing a gradual increase in the blood lactate concentration (Brooks 1985). In any real-world endurance performance, a combination of aerobic and anaerobic metabolism is responsible for meeting the energetic needs of the exercising organism (Baker, McCormick, and Robergs 2010; Wahl, Bloch, and Mester 2009), and so the lactate threshold (the highest intensity at which exercise can be carried out without increasingly accumulating lactate in a muscle cell) is an important predictor of long-term endurance performance. The running velocity at the lactate threshold was shown to be related to 3 km running performance (Grant et al. 1997). For highly trained individuals who do not exhibit significant differences with respect to VO_{2max} , velocity at the lactate threshold has been demonstrated to explain residual variance in performance during prolonged endurance exercise (Coyle et al. 1988). As an alternative to the labor-intensive lactate diagnostic, Wasserman and McIlroy (1964) introduced the concept of ventilatory threshold as a non-invasive indicator of LT .

Lastly, C describes the relationship between exercise intensity and energy expenditure of an organism. More specifically, in most applications, it is defined as the ratio between work output and oxygen cost (Conley and Krahenbuhl 1980). Helgerud demonstrated interindividual differences in C among athletes with similar endurance performance (Helgerud 1994).

2.1.2 Strength and Power Performance

Strength and power are essential for athletic performance. As described above, strength represents the ability to exert force on an external object or resistance. Power means the rate at which force is being generated (or, in the physical sense, the amount of work being done per unit time, where work is force exerted to cause displacement of an object).

In most athletic events, time to generate force is limited and, therefore, power is of crucial importance for performance (Suchomel, Nimphius, and Stone 2016). The way in which force, power, and velocity interact was first described by Hill (1938) and has gained prominence as the *force-velocity relationship*. This relationship is at the heart of any examination of muscular strength and power. It states that as the velocity of a concentric muscle action increases, less force can be generated during that contraction. The product of force and velocity is power. While the relationship between force and

velocity is strictly negative, there exists an optimal combination of force and velocity, yielding maximum power output. Therefore, maximal muscular power is determined by the parameters of the force-velocity relationship: maximal isometric force (F_{max}), maximal velocity of shortening (V_{max}), as well as the degree of curvature, defined by a/F_{max} or b/V_{max} (Cormie, McGuigan, and Newton 2011).

The second important law of strength and power performance is the *length-tension relationship*, which signifies the relationship between the force being generated by a muscle and the sarcomere length of that muscle (Gordon, Huxley, and Julian 1966). Following this relationship, the greatest potential for force production exists at a certain sarcomere length of a muscle, which permits an optimal overlap between the actin and myosin filaments, which are responsible for the cross-bridge interaction creating muscle contraction. Suppose a muscle is stretched beyond or shortened below the optimal sarcomere length (which is specific to the muscle under examination). In that case, the tension developed by the muscle will not be at its maximum.

Another important discovery with respect to the workings of skeletal muscle was the stretch-shortening cycle (Bosco and Komi 1979). It was found that when a muscle fibre contracts immediately after being stretched, more force and power can be generated than through a concentric-only contraction (Komi and Bosco 1978). This finding had important implications for understanding dynamic athletic performance.

Building upon these seminal works on muscular contraction and the laws of force generation, researchers have continued to investigate determinants of muscular strength and power from a physiological point of view. In this endeavor, different types of muscle fibre have been identified (Komi and Karlsson 1978), and it was shown that the fibre composition of a muscle has a drastic impact on the capacity of force generation (A. Thorstensson, Grimby, and Karlsson 1976; Tihanyi, Apor, and Fekete 1982). Specifically, for a given cross-sectional area, so-called type IIa and IIx fibres exhibit 5- to 10-fold greater peak power compared to the so-called type I fibres (Widrick et al. 2002).

Furthermore, the force being generated by a muscle fibre is directly proportional to its cross-sectional area, irrespective of the fibre type (Bodine et al. 1982), and the maximal shortening velocity (V_{max}) is proportional to its fascicle length (Bodine et al. 1982). Finally, the pennation angle, which characterizes the angle between the muscle's fascicles and the line of action, determines the force being produced by a muscle (Spector et al. 1980).

Besides these morphological factors, research has identified neural factors driving the ability of the nervous system to activate the muscles involved in contraction. Firstly, the number of motor units recruited for a contraction is a determinant of force

production, where the order of recruitment is driven by the size principle (Henneman, Somjen, and Carpenter 1965). Additionally, the frequency at which these motor units send neural impulses to the muscle fibre is of importance (Enoka 1995) as much as synchronization of motor units (Komi 1986) and inter-muscular coordination (i.e., increased synergist activity and decreased antagonist activity, Sale 1988).

After identifying the laws of muscular strength and power in basic research, a more applied perspective has begun to shed light on the relationship between these capabilities and athletic performance in various types of sports. In this sense, strength and power have been demonstrated repeatedly to be strongly related to sport-specific athletic performance (Suchomel, Nimphius, and Stone 2016).

For example, upper body strength was positively associated with throwing velocity in handball players (Gorostiaga et al. 2005). Stone and colleagues (1980) demonstrated a strong correlation between maximal strength and power output in Olympic weightlifting. In football players, strong correlations were found between lower-body strength and sprinting performance as well as vertical jump height (U. Wisløff et al. 2004). Muscular power, both of the legs and the arms, has been demonstrated to predict performance in a swimming sprint (Hawley et al. 1992).

2.1.3 Physical Performance Analysis in Football

As a team sport involving a variety of movement types and modes (running, sprinting, jumping at different intensities, as well as interactions with the ball and tackling), football has a specific profile of physical requirements. Research into this profile has been a crucial task for performance analysis in order to determine the strength and endurance capabilities that football players should be trained and tested in (Hoff and Helgerud 2004).

While certain tests and measures from the general exercise science domain can be directly used in the team sports setting, essential differences in how team sports are played require more specificity in some respects. For example, compared to long-distance running events, the sport-specific endurance performance in team sports exhibits slightly different properties with respect to underlying physiological determinants. During a football match, players spend more than 50% of match time standing or walking and roughly 10% of playing time at high-intensity running or sprinting (Mohr, Krstrup, and Bangsbo 2002). Furthermore, football matches are characterized by frequently occurring, short bouts of sprints (J. Bangsbo, Nørregaard, and Thorsø 1991) or changes of direction (Withers 1982).

Due to this intermittent nature of team sports, football in particular, measures like velocity at the lactate threshold, which is a key predictor of performance in prolonged

endurance exercise (Coyle et al. 1988), are not as predictive of in-game physical performance (Stølen et al. 2005).

While the capacity for a high on-field physical performance is best approximated by means of field tests accounting for the intermittent nature of football (see below), VO_{2max} seems to be an essential determinant of the ability to perform such intermittent exercise. Specifically, a significant positive association between players' VO_{2max} and running performance during football matches could be shown (Krustrup et al. 2003) (Jens Bangsbo, Mohr, and Krustrup 2005). Given the high importance of VO_{2max} for aerobic performance, these findings suggest the measure to be a useful target for training interventions and performance assessment.

Building upon these findings, more pragmatic and ecologically valid assessments of football-specific endurance performance have been sought. To this end, several field tests have been developed and investigated with respect to their validity for football-specific endurance performance. One of the most frequently investigated field tests, the *Yo-Yo Intermittent Recovery Test (YYIR)*, has been shown in numerous studies to be associated both with VO_{2max} and match running performance (Jens Bangsbo, Iaia, and Krustrup 2008; Krustrup et al. 2003; Modric, Versic, and Sekulic 2021; Paul Simon Bradley, Carling, Diaz, et al. 2013).

These findings regarding the importance of VO_{2max} for on-field physical performance in football have informed training strategies specifically targeted at increasing VO_{2max} in footballers. It has been stated that training for football-specific endurance performance should aim at generating intensities at 90-95% of an individual's maximum heart rate (Garber et al. 2011). Based on this requirement, high-intensity interval training has generally been found to be an effective training mode for football players, with intervals at a length of 3-8 minutes, aiming at an intensity range as stated above (Hoff and Helgerud 2004). In an intervention study adopting such a training regimen, a significant improvement in VO_{2max} was accompanied by an enhanced on-field performance, signified by an increase in total distance covered, number of sprints, and involvement with the ball (Helgerud et al. 2001).

As stated above, the physical demands for football players are not exclusively determined by endurance requirements. Up to 11% of the distance covered during a match is covered sprinting (J. Bangsbo, Nørregaard, and Thorsø 1991), with a sprint occurring every 90 seconds (Reilly and Thomas 1976). Similarly, professional football players perform 50 changes of direction per match (Withers 1982). These observations emphasize the importance of strength and power for football players.

Research into testing such strength and power capabilities has revealed that maximal lower-body strength, expressed as the One-Repetition Maximum (1RM) in the back squat, is highly correlated with both jumping and sprinting performance in football

players (U. Wisløff et al. 2004). Interestingly, maximum strength in leg extension (machine-based) did not show a similar association with sprinting performance (Cometti et al. 2001), suggesting the need for specificity in strength testing. This is corroborated by the findings by Thorstensson and colleagues (1976) who demonstrated that training-induced strength gains of 70% in squat 1RM translated to only 20% improvement in knee extension strength. These findings stress the principle of specificity of training (Rutherford et al. 1986).

The research into the physical performance profile of football, and especially the physiological foundations of football-specific physical performance, has generated important insights for the design of training and testing in football. While weak associations between physical performance and football success have been established for the primary purpose of identifying meaningful parameters of physical performance (e.g., Mohr, Krstrup, and Bangsbo 2002; Apor 1988), the prediction of success has not been the main objective of this strand of research. The modern attempts of *Performance Analysis* to identify *Key Performance Indicators* were, therefore, more likely inspired by the use of deterministic models in the field of biomechanics, which will be discussed below.

2.1.4 Deterministic Models in Performance Analysis

Another approach to *Performance Analysis*, which has heavily inspired the later developments in team sports, is the use of deterministic models in biomechanics (Chow and Knudson 2011). A deterministic model is a modeling paradigm seeking to identify the biomechanical factors that determine the outcome of a movement. A closer look at deterministic models is important, since they might have been the basis of the concept of *Performance Indicators* in *Match Analysis Research* in team sports (Hughes and Bartlett 2002).

An example of such a deterministic model is the one proposed by Hay and colleagues, outlining the factors influencing the outcome of a long jump (Hay, Miller, and Canterna 1986). Deterministic models seek relationships between certain biomechanical aspects of the movement (e.g., the horizontal velocity at takeoff) and the outcome of interest (in this case, long jump distance). These relationships are established in a hierarchical structure with multiple levels, where each level consists of factors that determine the variables at the next highest level. The top level represents the outcome of interest. In deterministic models in biomechanics, the factors included must be mechanical quantities (e.g., velocity, force, angular momentum) or appropriate combinations of them. A key characteristic of these models is that they are constructed using knowledge of mechanical properties and physical laws, and are,

therefore, not inherently statistical models. While these deterministic models might be validated through statistical modeling of empirical data, their initial development is based on causal relationships that have been established or are assumed based on previous research. In this way, these models do not rely on exploratory statistical methods (Chow and Knudson 2011).

In the domain of *Performance Analysis* in team sports, the idea of a deterministic model has been taken up to model the impact of certain *Performance Indicators* on successful outcomes (Hughes and Bartlett 2002). Under this approach, an association between certain *Performance Indicators* and corresponding *Success Indicators* is proposed based on domain knowledge about the respective sport. However, in contrast to the approach in biomechanics, these associations are not based on established causal relationships or mechanical properties. Still, they are rather proposed on the grounds of assumptions about the nature of the respective game and then investigated through exploratory statistics.

This difference is mainly due to two reasons. First, performance or success in a team sport is generally not as easily defined as in sports such as gymnastics or track and field, where deterministic models were initially applied. Long-jump distance, for example, as demonstrated by Hay and colleagues (1986), is a clearly measurable variable and therefore differs fundamentally from a multi-component construct like soccer performance (Atkinson 2002). While the former can be approximated by a known physical law (projectile motion), the latter has to be well-defined before even attempting to link it to predictor variables.

Second, even the most agreed-upon indicators of successful outcomes in team sports (like match outcome) are prone to a high degree of randomness (see Section 2.2.3 for further detail) and can, therefore, not be modeled in a deterministic way.

Despite these structural differences between the original applications of deterministic models and the case of *Performance Analysis* in team sports, the seminal paper by Hughes and Bartlett proposing the systematic study of *Performance Indicators* utilizing an adaptation of deterministic models (Hughes and Bartlett 2002) has been succeeded by two decades of *Performance Analysis* research trying to identify significant indicators of success in football (Lepschy, Wäsche, and Woll 2018). However, the above-mentioned characteristics of team sports (and football in particular) should be kept in mind as a caveat when reviewing this *Performance Analysis* literature.

2.2 Match Analysis Research

When *Performance Analysis* in team sports is applied during competition in an attempt to investigate the performances of athletes or teams, this endeavor is

more specifically termed *Match Analysis Research* (Memmert 2021). The following discussions will, therefore, use this terminology.

As described in detail in the previous section, the idea of viewing *Match Analysis Research* (*MAR*) in team sports through the lens of a deterministic model (with *Performance Indicators* predicting outcomes) has likely driven much of the developments in the field. This movement has also been strongly motivated through the idea of *Moneyball* – identifying indicators of player or team performance that will give an edge over the competition – popularized through the massive success of the data-driven approach taken by the Oakland Athletics baseball team under General Manager Billy Beane (Michael Lewis 2004). As a response to the impressive results that Oakland achieved for several years with a comparably low budget, researchers as well as practitioners from other sports started to utilize data analysis to experience a similar “Moneyball moment” (Szymanski 2020, 58).

Specifically, statistical designs exploring the associations between select *Performance Indicators* and *Success Indicators* in an attempt to identify *Key Performance Indicators* have gained popularity and coined the concept of *Quantitative Match Analysis Research* (Rein and Memmert 2016). Innovations with respect to data collection techniques in football, as well as advances in the field of Machine Learning, have further attracted researchers and data analysts to this area. As will be outlined in the following, this large area of research has generated several important insights into the dynamics of the game, but has at the same time produced controversial results that pose a threat to the trustworthiness of research findings. In Section 2.4.1, it will be discussed how the observational nature of *MAR* and potential methodological inconsistencies might be responsible for these contradictory results and which steps *MAR* might have to take in order to avoid biased results. In order to understand the current state of the field, the history of *MAR* is reviewed next.

The first attempts to analyze football matches can be traced back to the early works of two separate camps of researchers studying different aspects of the game.

The first camp consisted of Charles Reep, Richard Pollard, and Bernard Benjamin (Reep and Benjamin 1968; Reep, Pollard, and Benjamin 1971) who studied the mathematical properties of team sports and identified various sports phenomena to follow mathematical probability distributions. Specifically, in the 1968 work, Reep and Benjamin examined passing sequences of varying length. They demonstrated that the length of a passing sequence can be viewed as a random variable which can be approximated very well by a Negative-Binomial probability distribution (Reep and Benjamin 1968). Furthermore, they showed that the conversion rate of shots to goals was variable across games but nearly constant when taking into account a larger sample of matches. In a 1971 paper, Reep, Pollard, and Benjamin identified

multiple phenomena from different sports to follow Negative-Binomial distributions, including innings at cricket, goals in ice hockey, or the length of rallies in tennis (Reep, Pollard, and Benjamin 1971). These works have laid an important foundation for understanding the coexistence of a stochastic element and some regularity within the reasonable limits of probabilistic uncertainty for sports phenomena.

In the other camp, Thomas Reilly and Vaughan Thomas (Reilly and Thomas 1976) were the first ones to describe movement profiles of football players, highlighting, for example, the frequency and duration of sprints within a match. Follow-up studies revealed insights, for instance, into the physiological energy cost of dribbling a soccer ball compared to running without a ball (Reilly and Ball 1984). These early studies of movement profiles have built the basis for later strands of research quantifying the load and movement patterns in training (Guerrero-Calderón et al. 2020) or matches (Guerrero-Calderón et al. 2021).

Since no scientific journal specifically targeting performance analysis in team sports was around at the time, these articles were published in statistics and movement science journals. This was addressed in 1983, when the first issue of the *Journal of Sports Sciences* was published. This event, which roughly coincided with the advent of personal computers, played a considerable role in the development of the study of *Performance Analysis* in sports, specifically of the field of *Match Analysis Research*.

The journal saw the first contributions in this area, which was at the time dominated by the emerging methodology of notational analysis. This methodology, which had implicitly been used already by Reep and colleagues (Reep and Benjamin 1968), relied on writing down the occurrence and certain features of events within a sports game following a predefined schema. According to Alan Nevill, long-standing Editor-In-Chief of the *Journal of Sports Sciences*, “notational analysis is an objective way of recording performance so that key elements of that performance can be quantified in a valid and consistent manner” (Nevill, Atkinson, and Hughes 2008, 418).

While in the early days, notational data was generated by human observers witnessing the performance live and writing down the events (Reep, Pollard, and Benjamin 1971), first video recordings and later Personal Computers eased the process for annotators (Hughes 1984).

The majority of research utilizing notational analysis was conducted at the end of the 1970s and the start of the 1980s (Hughes 1985; Sanderson and May 1977; M. Lewis and Hughes 1988). In this endeavour, the advent of Personal Computers transformed the way that data could be collected. For example, Lewis and Hughes (1988) used an Acorn BBC computer to gather and analyze data on attacking plays in the 1986 World Cup of football. Other applications than the study of successful tactics were the examination of the effects of rule changes, such as the “back pass” (Hughes and

Sykes 1994).

Later, systematic coding systems were developed and used throughout multiple studies in order to provide reproducible designs. For example, Tenga and colleagues developed a reliable annotation scheme for analyzing offensive match play (A. Tenga et al. 2017) and applied it in order to evaluate the effect of certain tactics on offensive success (Albin Tenga et al. 2009b, 2009a). Another system was developed to classify the playing style of teams (Fernandez-Navarro et al. 2016) and was subsequently used to study the relationship between contextual factors and playing style (Fernandez-Navarro et al. 2018) as well as the effectiveness of different playing styles for match-play success (Fernandez-Navarro et al. 2019).

With the advent of new data collection techniques, facilitating a broader availability of data (see Section 2.2.4) and sophisticated statistical modeling techniques (see Section 2.3), a new era of *MAR* began (Rein and Memmert 2016), characterized by the use of large datasets and advanced statistical or machine learning techniques to model these data. Another feature of most *MAR* studies around this time was the desire to predict match-play success and thereby identify *Key Performance Indicators* (Lames and McGarry 2017; Lepschy, Wäsche, and Woll 2018). Examples of different approaches to modeling this predictive problem are presented in Section 2.3.2.

The first two decades of the 21st century have seen a vast number of studies examining potential predictors of footballing success or the relationships between aspects of team performance and behavior. Two important review articles have summarized this movement and pointed out major trends as well as shortcomings of the literature.

The systematic review by Sarmiento and colleagues (2014) provides an overview of 53 studies in the field of *MAR*. The authors categorize the articles into *descriptive*, *comparative*, and *predictive* studies, depending on the main aim of the respective papers. They point out that, although many studies attempted to reveal links between certain behaviors and successful outcomes, only a minority of included articles actually performed predictive analysis in the statistical sense. A multitude of studies aimed to analyze effects between *Performance Indicators* and success, but modeled the *Success Indicator* (e.g., ranking in the final table) as an Independent Variable within their statistical design (which the authors then categorize as a *comparative* instead of a *predictive* study). The authors therefore recommend increasing efforts to publish actual predictive studies while also including contextual variables in the analyses (Sarmiento et al. 2014).

Around the same time, MacKenzie and Cushion (2012) reviewed the literature on match analysis research and stressed the lack of contextual information within *MAR* study designs. They state that out of 59 match analysis studies, only eleven

integrated information on the interaction with the opponent, and only ten accounted for the match location (which might be important in terms of home advantage). This is especially concerning given the enormous impact that *Contextual Factors* have on both team behavior and performance, and team success. The following section provides an overview of the relationship between *Contextual Factors* and *Performance Indicators* as well as *Success Indicators*. The third publication in this thesis (Section 3.3) contains an even more extensive discussion of these relationships, as well as their potential to introduce biases into conclusions drawn from studies if careful consideration is not given to the potentially moderating effect of context on performance and success.

2.2.1 Contextual Factors in MAR

The *Contextual Factors* commonly described within *MAR* (sometimes also referred to as situational variables) can be categorized with respect to different taxonomies. One distinction that is particularly relevant to the study of *MAR* and deeply linked to study design is the one between *Contextual Factors* on the match level and *Contextual Factors* beyond the match level. While the former remain constant over the course of one match (e.g., game location or the strength of teams assessed pre-match), the latter vary according to the evolution of the match or even drive the evolution of the match (e.g., scoreline, time, or tactical context). It is obvious that the inclusion of *Contextual Factors* beyond match level into research studies is dependent on a parameter of study design, the *level of observation*, which is also discussed in detail in the third publication (Section 3.3) as well as in Section 2.2.2 of this thesis.

2.2.1.1 Context on Match Level

Contextual Factors on the match level consist of those contextual variables that are not changing during a match and whose manifestation is mostly known pre-match. While this overview is most likely not exhaustive, the most important of such factors are game location and team strength.

Game Location

Game location is one of the most robust factors influencing team performance. Home teams benefit from a well-established *home advantage* – winning more often (Jamieson 2010; Allen and Jones 2014; Pollard and Gómez Ruano 2014; Riedl et al. 2022) and scoring first more frequently than away teams (Lago-Peñas et al. 2016). Home advantage is present in almost all leagues worldwide to a different extent (Pollard

and Gómez Ruano 2014) with an average worldwide home advantage of 61.9% (Riedl et al. 2022).

Besides being more successful, home teams dominate in terms of ball possession (Lago-Peñas and Martín 2013; Paul Simon Bradley et al. 2014) and offensive actions (e.g., more crosses, dribbles, shots) while away teams tend to show an increased number of defensive actions (Taylor et al. 2008; Lago-Peñas and Lago-Ballesteros 2011). Moreover, home teams are positioned higher up the pitch and adopt more attacking formations, whereas away teams recover the ball closer to their own goal and defend deeper (Bialkowski et al. 2014; Almeida, Ferreira, and Volossovitch 2014; Santos, Lago-Peñas, and García-García 2017; Lago-Peñas 2009). Studies on physical performance report mixed findings, either indicating a higher work rate of away teams (Augusto et al. 2021; García-Unanue et al. 2018) or for home teams (Castellano, Blanco-Villaseñor, and Álvarez 2011; Lago-Peñas et al. 2010).

In general, it might be concluded that home teams appear more dominant, which is supported by a more dominant and aggressive body language in home teams (Furley, Schweizer, and Memmert 2018).

Team Strength

Team strength, reflecting the overall quality of a team, is a key contextual factor but is challenging to quantify. Common proxies include end-of-season or external rankings (Paul Simon Bradley et al. 2014; Lago-Peñas and Martín 2013; Almeida, Ferreira, and Volossovitch 2014; Collet 2012), player market values (Herm, Callsen-Bracker, and Kreis 2014; Franck and Nüesch 2010; Peeters 2018; He, Cachucho, and Knobbe 2015), and pre-match betting odds (Wunderlich and Memmert 2018, 2020; Spann and Skiera 2009; Hvattum and Arntzen 2010). Each operationalization has its limitations – for instance, rankings may introduce temporal bias (Kaufman, Rosset, and Perlich 2011) and circularity (Aldrich 1995), while market values and betting odds incorporate different sources of information.

Despite these challenges, team strength has been found to be an important predictor of team behavior and success. Unsurprisingly, it was demonstrated that a team’s cumulated market value (sometimes logarithmic) was positively related to successful match outcomes (Andreas Heuer and Rubner 2014; Lepschy, Wäsche, and Woll 2020). Player market value is associated with a higher probability of scoring from a shot (Mead, O’Hare, and McMenemy 2023), a higher probability of scoring a penalty (Brinkschulte et al. 2023), and a lower probability of conceding a penalty goal as a goalkeeper (Brinkschulte et al. 2023). Better teams with respect to betting odds

were more likely to win penalty shootouts (Wunderlich et al. 2020) and to score individual penalties (Brinkschulte et al. 2023).

With respect to tactical performance, stronger teams consistently show higher ball possession rates (Lago-Peñas 2009; Lago-Peñas and Dellal 2010; Paul Simon Bradley et al. 2014), more passes (Paul Simon Bradley et al. 2015), more shots, crosses, a higher ball distance, and more space control (Klemp, Wunderlich, and Memmert 2021) than weaker teams. Better teams tend to regain the ball in more advanced zones (Almeida, Ferreira, and Volossovitch 2014) while weaker teams tend to recover balls closer to their own goal, positioning their defensive lines closer to their own goals (Santos, Lago-Peñas, and García-García 2017).

Again, the relationship with physical performance remains mixed, with better teams covering less total distance, but more distance in possession, even when normalized for possession time (Klemp, Memmert, and Rein 2021).

2.2.1.2 Context beyond Match Level

Contextual Factors beyond match level are those variables that might change their value throughout the match. While the fluctuations within these variables might reflect the ordinary rhythm of a match (such as the tactical context, also referred to as match phases), other variables might drastically alter the dynamics of the game when they change (such as scoreline, which changes when a goal is scored and is directly related to match outcome). Depending on the *level of observation* (Section 2.2.2), certain *Contextual Factors* might be included or averaged out in *MAR* studies.

Tactical Context

Beyond static match-level factors, the evolving *tactical context* within a match plays a crucial role. Matches can be segmented into distinct phases – such as established attack, transitions, and set pieces (Hewitt, Greenham, and Norton 2016; Eusebio, Prieto-González, and Marcelino 2024; Gonçalves et al. 2024; Gollan, Bellenger, and Norton 2020) – which markedly influence player behavior and physical demands. For instance, defensive phases often require a higher running distance (Llana et al. 2022; Lorenzo-Martinez et al. 2020), more (Gonçalves et al. 2024) and stronger accelerations (Gregory et al. 2022) as well as more curved running paths (Gregory et al. 2022) compared to offensive phases.

Both offensive and defensive transitions are associated with higher physical demands compared to established attack or defense phases (Gonçalves et al. 2024). Most

sprints in a match were performed during established defense, offensive transition, or defensive transition (Caldbeck and Dos'Santos 2022).

A more granular breakdown of offensive phases into build-up, progression, and finishing further revealed that teams move with higher speed during *progression* and *finishing* phases and in more curved paths during *opposition progression* (Gregory et al. 2022).

Besides match phases, an important aspect of *tactical context* is the tactical function associated with any behavior that players or teams exhibit on the pitch. In their seminal paper, Bradley and Ade (2018) have pointed out that a lot of the ongoing research on physical performance in football is missing the point of asking why players are performing certain runs. They developed a notational system categorizing high-intensity runs with respect to their tactical function and describe how this approach can be used to take physical performance analysis in football to the next level. Recently, the reliability and validity of this contextualization approach have been demonstrated (Ju, Hawkins, et al. 2023; Ju, Doran, et al. 2023).

Scoreline

The current scoreline is arguably the most influential in shaping team motivation and tactics, as it directly reflects the anticipated match outcome (Lago-Ballesteros, Lago-Peñas, and Rey 2012). Researchers have operationalized scoreline in several ways. A common approach is to use a three-level categorical variable (winning, drawing, losing).

Under this scheme, studies have shown that losing teams tend to achieve a higher frequency of score-box possessions (Lago-Ballesteros, Lago-Peñas, and Rey 2012) but to display a reduced probability of scoring from a given shot (Gonzalez-Rodenas et al. 2020). Losing teams further exhibit increased ball possession rates (e.g. 10.97% higher than winning teams, Lago-Peñas 2009), perform more offensive actions such as passes, dribbles, and crosses while making fewer clearances or interceptions (Taylor et al. 2008), and recover the ball further up the pitch (Santos, Lago-Peñas, and García-García 2017; Almeida, Ferreira, and Volossovitch 2014) while taking less time to regain the ball (Vogelbein, Nopp, and Hökelmann 2014), indicating a more intense pressing strategy.

With respect to tactics, losing teams, especially when trailing by two or more, use more buildup play, sustained threat, crossing, and even an increase in fast tempo (Fernandez-Navarro et al. 2018). Conversely, winning teams often adopt a more defensive and risk-averse approach, with evidence also suggesting that work rate is highest when matches are drawn (Redwood-Brown et al. 2012).

When the scoreline is quantified as a goal difference (a continuous measure), findings indicate an inverted U-shaped relationship with physical work rate: players cover the most distance when the score is close (i.e., the goal difference is between -1 and 1), with work rate decreasing as the margin widens (Redwood-Brown et al. 2018). In contrast, passing accuracy shows a U-shaped pattern, with the lowest values observed at small leads (Redwood-Brown et al. 2019).

Overall, these results underscore that scoreline profoundly influences team behavior: losing teams intensify their offensive efforts to equalize, while winning teams tend to play more conservatively. This dynamic impact on both tactical and physical performance highlights the critical need to account for scoreline as a contextual factor in match analysis.

2.2.1.3 Implications for Study Design

As discussed above, *Contextual Factors* influence both team behavior and performance (*Performance Indicators*) and outcomes of events, sequences, and matches (*Success Indicators*). This effect exists because *Contextual Factors* directly reflect player capabilities and have the potential to influence players' aims and motivation. As is described in Section 2.4.3, a covariate (sometimes referred to as a background variable) affecting two variables whose association with each other is to be inferred can be a potential confounder and thus introduce a bias into the association under investigation.

This effect is discussed in more detail in Section 2.4.3; however, at this point, an important implication for study design needs to be mentioned. In order to avoid potential confounding effects of *Contextual Factors*, it is mandatory to include these factors in a statistical model estimating the relationship between *Performance Indicators* and *Success Indicators* (see Section 2.3.1). As pointed out above, the ability of a researcher to include certain covariates, especially those beyond the match level, depends on the study design, specifically the *level of observation*. When observations are aggregated at the match level, variables like game location can be easily included, whereas dynamic factors such as scoreline require a more granular segmentation.

While the *level of observation* is a crucial parameter of study design, considerable heterogeneity exists in the literature with respect to this parameter. More importantly, most studies do not justify their choice of the observation level, likely making it purely a consequence of data availability or convenience. In the following, different decisions with regard to the level of observation made in the literature are outlined, and the consequence of this heterogeneity is discussed afterwards.

2.2.2 Level of Observation in *MAR*

When working with complex data types like the ones prevalent in the modern era of *MAR* (Section 2.2.4), the researcher is usually presented with data in the form of time series with potentially further degrees of freedom regarding the different players and different spatial dimensions. Consequently, one major challenge in designing *MAR* studies using this data is the issue of spatial and temporal aggregation (Floris R. Goes et al. 2020). The choice that a researcher makes with respect to this spatio-temporal aggregation manifests in a parameter of study design, the *level of observation*.

The most prominent examples of observation levels include the season level, the match level, the sequence level, and the action level. The distinction between these categories and examples for each of them is outlined below.

2.2.2.1 Season Level

Probably the least granular level of observation means aggregation on the *season level*. Working with data on the season level typically means that performance data for teams is aggregated over the course of a whole season, thereby averaging or summing over multiple matches. Under this methodology, *MAR* studies have been conducted examining the relationship between *Performance Indicators* and the number of points won by a team in a season (Phatak, Rein, and Memmert 2021; Hoppe et al. 2014; Andreas Heuer and Rubner 2014; Collet 2012), the degree of home advantage present in leagues (Pollard and Gómez Ruano 2014; Riedl et al. 2022), or the number of *Expected Goals* and *Expected Goals Against* (Phatak et al. 2022).

2.2.2.2 Match Level

Going one step more granular, data is aggregated on the *match level*. In these studies, mostly the outcome of matches is modeled as a function of outcome predictors. These predictors can consist of either pre-game information, such as team strength (Wunderlich et al. 2020), home advantage (Lago-Peñas et al. 2016), or opponent quality (Paul Simon Bradley et al. 2014), or performances or behaviors observed over the course of one match, like ball possession (Collet 2012), passing performance (Rein, Raabe, and Memmert 2017), or a combination of several available *Performance Indicators* (Lepschy, Wäsche, and Woll 2020).

2.2.2.3 Sequence Level

A less common level of observation is the sequence level. Under this approach, a game is broken down into segments that can be differentiated from each other according to some logical segmentation. This could be, on the highest level, match halves (Klemp, Wunderlich, and Memmert 2021); however, mostly, more fine-grained segmentations are used. For example, matches have been segmented according to the scoreline, i.e., a new segment starts whenever a goal is scored (Klemp, Memmert, and Rein 2021; Klemp, Rein, and Memmert 2023; Taylor et al. 2008; Lago-Peñas 2009). A very popular segmentation approach is the segmentation of team ball possessions, where each segment starts when one team gains control over the ball (for a definition, see, for example, Reep and Benjamin 1968). Due to the fact that team possessions occur frequently and success in such a phase can be defined differently than by goal scoring, this level of observation is popular in *MAR* research because it provides a partial remedy to the “low-scoring” problem of football and the rarity of goals as a prediction target. Using team possessions, researchers have linked different tactics to successful outcomes, where the *Success Indicators* were either goal scoring (Albin Tenga et al. 2009b), score-box possessions (Lago-Ballesteros, Lago-Peñas, and Rey 2012; Albin Tenga et al. 2009a), or a more sophisticated measure combining different outcome measures (Pollard and Reep 1997). Lastly, an emerging field in the analysis of possession phases is *Expected Possession Value* (EPV) models, which model the probability of scoring in the near future based on features of the current possession (Fernández, Bornn, and Cervone 2019; Llana et al. 2022).

2.2.2.4 Event Level

The most granular approach to the level of observation is clearly the analysis of distinct events in a match. These kinds of studies have experienced promotion through the emergence of event data, providing detailed contextualization for each on-ball action (and potentially even more). However, the decision made by data providers about which actions are recorded and which are not obviously impacts analyses (see Section 2.2.4 for a discussion on the limitations of event data).

Clearly, the most prominent example of event-level analysis is the case of Expected Goals (xG) models, where shots are considered as a subset of all events and their success probability is predicted based on the context of the shot. A variety of works aiming to build xG models can be found in the literature (Pollard, Ensum, and Taylor 2004; Lucey et al. 2015; Mead, O’Hare, and McMenemy 2023; Eggels, Van Elk, and Pechenizkiy 2016; Cavus and Biecek 2022; Anzer and Bauer 2021; Ruiz et al. 2015; Gonzalez-Rodenas et al. 2020).

An extension of the xG methodology that is strongly related to the above-mentioned EPV approach is the idea of action value modeling. These models aim to value individual actions (beyond shots) based on their contribution to the probability of scoring in the near future. A prominent approach to this problem has been the *VAEP* (Valuing Actions by Estimating Probabilities) model (Van Haaren 2021; Decroos and Davis 2020; Decroos et al. 2019).

Other popular targets of event-level analysis include penalty kicks (Brinkschulte et al. 2021, 2023) or ball regains (Almeida, Ferreira, and Volossovitch 2014; Santos, Lago-Peñas, and García-García 2017; Raabe, Nabben, and Memmert 2022).

As can be seen from this overview, a multitude of different manifestations of the *level of observation* can be found in the literature. While the observation level is barely explicitly stated as a study parameter, no justification for its choice can be found in any of the studies. On the other hand, it is obvious from the discussions in Section 2.2.1 that the choice of the observation level has a crucial effect on the potential to include *Contextual Factors* in the statistical analysis. For example, if the scoreline shall be included as a *Contextual Factor*, studies have to be segmented further into sequences of respectively constant scoreline states (Lago-Ballesteros, Lago-Peñas, and Rey 2012; Gregory et al. 2022; Klemp, Memmert, and Rein 2021; Bassek et al. 2023). This is an important consideration that needs to be acknowledged in *MAR* and will be discussed in more detail in the Synopsis of this thesis (Chapter 3).

Another important aspect that deserves special consideration in the context of predictive *MAR* studies is the stochastic nature of any behaviors or outcomes that can be observed in team sports. It has been described in Section 2.1.4 that much of the *MAR* literature has been implicitly or explicitly (Hughes and Bartlett 2002) motivated by the use of deterministic models. However, deterministic modeling underlies the assumption that the future behavior of a system can be definitively predicted based on knowledge about the current and past state of the system. As stated by Franks and Goodman (1986), this assumption clearly does not hold for the nature of team sports, and rather, stochastic models have to be used, allowing for random contributions to any observable of team performance or success. The nature and degree of randomness inherent in the game of football, as well as its implications for *MAR*, are reviewed in the following section.

2.2.3 Randomness and Uncertainty in Football

Part of the great popularity of football stems from the random aspect that is inherent in any football match. In practical terms, the random aspect manifests in the fact

that the better team does not always win a match. In technical terms, this means that the outcomes of single matches are less predictable compared to outcomes in other sports.

A general notion in this regard is the phenomenon that outcomes (or manifestations of specific variables) are prone to a considerable influence of chance in the short term (e.g., one match) while they tend to be more robust and predictable in the long term (e.g., one season).

One of the first demonstrations of this effect was made by Reep and Benjamin (1968) who showed that the number of shots needed to score one goal (the inverse of goal conversion rate) was variable across single matches, but when averaging over seasons, the number tended to converge to about 10 shots (equivalent to a 10% conversion rate). This initial observation was later extended to the observation that the empirically observed number of goals in football matches can be well approximated by probability distributions.

While Reep and Benjamin (1968) and later Pollard (2017) proposed the Negative-Binomial probability distribution, the majority of research seems to have agreed upon the Poisson distribution as the best approximation (A. Heuer, Müller, and Rubner 2010; Chu 2003; Karlis and Ntzoufras 2003)². While the Poisson provides a good fit for the goal sum observed in a match, modeling the actual score line (i.e., home and away goals) using two independent Poisson distributions has been more difficult. Adjustments using correction terms (Mark J. Dixon and Coles 1997; Mark J. Dixon and Pope 2004) or bivariate Poisson distributions (Karlis and Ntzoufras 2003) have been shown to be more accurate, but a further discussion of this issue is beyond the scope of this thesis.

The relevant conclusion from these findings related to the present work is the observation that outcomes of football matches follow known statistical properties when a large enough sample is considered, but that the prediction of individual match outcomes is extremely difficult. An implicit, theoretical derivation from these results, which is, for example, exploited in sports betting, is the assumption that there exists an underlying probability for each possible outcome in a football match and that the observed outcome is purely one realization of all these possible outcomes.

²The main reason for Reep and Benjamin as well as for Pollard to choose the Negative-Binomial over the Poisson distribution was the observed over-dispersion (i.e. the observed variance was higher than the theoretical variance in a Poisson distribution would have allowed). They attributed this observation to the effect of different teams having different scoring rates. In later applications of the Poisson distribution, it has become a common approach to allow varying scoring rates for individual teams (in statistical terms, estimating the parameters of the Poisson, the expected value and the variance, separately for each team). This approach has yielded a better fit of the empirical data to the Poisson distribution, potentially because the over-dispersion caused by the variation of teams has been ruled out of the distribution.

Several attempts have been made to quantify the degree of randomness inherent in football scores. As stated by Wunderlich and colleagues (2021), there are two main approaches to deal with randomness in football: either by means of probabilistic statistical modeling or by analyzing observable signs of randomness within matches.

Following the first approach, Heuer and Rubner (A. Heuer and Rubner 2009) utilized the variance of goal differences among teams in a league on different match days. By modeling this variance as a function of the match day, a strong linear relationship could be observed, delineating a decreasing variance with an increasing number of matches played. In practical terms, with every new match being played, the estimate of a team's strength, represented as their goal difference, became less variable and more reliable. By extrapolating this relationship to a theoretically infinite number of matches played, where the effect of a limited sample size would have averaged out, the contribution of pure statistical variation to the variance in goal differences can be estimated. Then, for every match day, it can be estimated how much this statistical effect of unavoidable uncertainty contributes to the variation in goal differences. Following this approach, the authors conclude that at the end of a season (i.e., at match day 34 in the German Bundesliga), the statistical contribution to the total variance is still as large as 30% (A. Heuer and Rubner 2009).

Under the second approach, researchers normally focus directly on events during a match. A natural choice for an event to study randomness on is the event of goal scoring, since the effect of randomness in football is normally signified by the number of goals scored and generally explained by the low-scoring nature of football (see also below for differences to other sports).

Therefore, Lames (2018) aimed to estimate the proportion of goals scored in football that displayed a considerable effect of chance involved in their occurrence. Specifically, the video footage of 875 goals scored in the German Bundesliga and 1056 goals scored in the English Premier League was examined for the presence of at least one out of six indicators of chance (e.g., a deflection or an own goal). The results showed that 47% of all goals were affected by chance. Extending this study with a larger sample and considering the temporal evolution of the effect, Wunderlich and colleagues (2021) replicated the effect in a sample of 7263 goals scored across seven seasons in the English Premier League. Specifically, they found that the proportion of chance-related goals decreased from 50% to 45% over the seven seasons between 2012/2013 and 2018/2019. Despite this trend, the total contribution of randomness to the results observed in professional football can still be considered considerably high.

Given the enormous presence of random contributions in football, explanations for this phenomenon have been sought throughout the literature. A common statement

to be found in this context is the “low-scoring nature” of football (Robberechts, Van Haaren, and Davis 2019; Van Haaren 2021; Merhej et al. 2021; Anzer and Bauer 2021; Floris R. Goes, Kempe, and Lemmink 2019; Macdonald 2012; Bunker and Sunsjak 2019; Mead, O’Hare, and McMenemy 2023), highlighting that goals occur rarely and, therefore, are hard to predict while at the same time having a substantial impact.

A more nuanced investigation into the reasons behind this crucial feature of football has been conducted by Lamas and colleagues (2020). In this work, different sports were compared to each other with respect to the structure of possession phases, scoring attempts, and finally, scoring.

In general, the authors stated that scoring points in invasion games is a two-step process, consisting of (1) creating a scoring opportunity (commonly referred to as a shot) from a possession and (2) converting a shot into one or more points. In most sports, a successful shot is rewarded with one point, except in basketball, where a shot can yield 1, 2, or 3 points depending on the context and shot location.

The authors then proceed to quantify the frequency of occurrence as well as the success rates for both steps, i.e., the number of possessions per minute, the probability of generating a shot from a possession, and the number of points generated from a shot. It is demonstrated that football displays a unique combination of very low values for all of these figures. In particular, in football, 1.5 possessions take place per minute, .09 shots are generated per possession, and a shot yields .11 points (i.e., goals). Combining the latter two numbers yields a points-per-possession value of .01. This is paralleled only by field hockey (also .01), which, however, has twice as many possessions per minute as football.

So in effect, football has by far the lowest density of points over time, with field hockey and ice hockey coming close.

This investigation is potentially the first one to uncover the details of why football is a “low-scoring” game: a low density of scoring opportunities per unit time, combined with a low success rate for shots. The latter is also pointed out by Lames as a reason for the high proportion of random goals (Lames 2018).

To summarize, the game of football is characterized by a high degree of randomness or uncertainty accompanying the observed outcomes of matches. This random contribution can be quantified to amount to between 30 and 50%, depending on the approach to the problem, and can be explained by the structure of ball possessions and the difficulty of taking a shot. While the extent to which randomness plays a role in football might seem surprisingly high to the reader, it is not at all surprising from a statistical point of view, as the pioneers Reep and Benjamin have already stated in 1968:

“The observation that there is a stochastic element in the number of goals arising from a particular number of shots in one match (as well as near-constant proportion over a larger series of matches) is easy for a statistician to accept; indeed he would be surprised if it were otherwise. It indicates, of course, that an excess of shots by one team does not mean that, by chance, the other side will not get more goals and thus win the match.” (Reep and Benjamin 1968, 585)

The high degree of randomness inherent in football match outcomes is another argument for researching *Performance Indicators* that are less prone to random variability. The effects and results discussed above explain why a valid *Performance Indicator* should predict success in the long term and for larger sets of matches and events. The modeling approach utilized in *MAR* is outlined below, first introducing the data types commonly used and then formulating the statistical approach to the problem.

2.2.4 Types of Data Utilized in *MAR*

2.2.4.1 Notational Data

The earliest approaches to *MAR* mentioned above utilized notational analysis. Specifically, this meant counting certain events during a match, potentially differentiated by certain features of the events such as the location on the field (Pollard and Reep 1997). Franks and Goodman (1986) were probably the first ones to formally and systematically describe the generation of notational data. After stating the limitations of subjective, qualitative observations, they suggest that “one method of aiding this process of observation with a view to improving its objectivity and accuracy is to record the occurrence of behavioural events in some coded form” (Franks and Goodman 1986, 50). In *MAR* studies until around the early 2000s, this notational data was the most prevalent choice to represent team or player behavior (Hughes and Franks 2004).

2.2.4.2 Event Data

The development of *event data* (Garnica-Caparrós 2024) can be viewed as an evolution of notational data, increasing the depth of detail associated with each event and the granularity of data collection (notational data typically resulted in aggregation of counts at some point, whereas event data retains every event as a single observation). This evolution was largely driven by the broad accessibility of Personal Computers

as well as advances in data collection techniques, allowing semi-automated collection of certain features of events (Garnica-Caparrós 2021). Nowadays, event data is used to describe the inherent sequentiality of soccer matches (Garnica Caparrós 2024) or to analyze performances on the event level. The most prominent development in this regard can surely be stated to be the Expected Goals methodology (Mead, O’Hare, and McMenemy 2023), while approaches to modeling action or possession value (Decroos and Davis 2020) are clearly on the verge of becoming the most important application of event data analysis.

Although providing a huge advancement to the field, the boost in the use and sophistication of event data has brought its own challenges. Data providers began to steadily collect more information on each event, trying to enrich the representation of a player action as much as possible.

The first, obvious consequence of this movement is a considerable heterogeneity across the range of data representations created by different vendors, which makes it hard to consider event data a universal data type or format (Decroos et al. 2019). Another challenge is created by the necessity to attach some kind of semantics to any event that is being annotated by a human observer. While tracking the location of players or the ball on the pitch at a given time is quite an objective task whose accuracy can be determined given a ground-truth label (see below), annotating an event during a soccer match is prone to varying interpretations by annotators. For example, suppose a player releases the ball in the direction of the goal. In that case, it can sometimes not be definitely stated whether the player was attempting a pass, a cross, or a shot at the goal, since the player’s intention is unknown to the observer. This makes the classification of an event hard in some cases. Biermann and colleagues have pointed out this issue and have proposed a hierarchical taxonomy of events that allows for some uncertainty with respect to such details (Biermann et al. 2021). However, the current state of research as well as technology relies on data formats as they are provided by vendors, therefore bearing this inherent limitation.

2.2.4.3 Position Data

The second type of data that has emerged through the digital revolution in sports is position data, also referred to as tracking data. This type of data contains the precise locations of players, referees, and the ball throughout a match. These positions are measured as x- and y-coordinates within a two-dimensional Cartesian system (Low et al. 2019). This data is typically supplemented with contextual information, such as the play status and the team in possession. The status of play differentiates between active gameplay and interruptions (e.g., when the ball is out of play or the referee

stops the game). Ball possession is defined as the period from when a player contacts the ball until an opponent makes the next touch.

Position data acquisition systems are primarily divided into sensor-based solutions—like global (Scott, Scott, and Kelly 2016) or local (Blauberger, Marzilger, and Lames 2021) positioning systems—and optical tracking systems. GPS-based systems are mainly utilized for physical performance analysis since they encounter limitations in tactical analysis. Challenges include converting GPS data to a pitch-centered coordinate system and reduced positional accuracy due to obstacles (such as stadium roofs) that interfere with the signal (Pons et al. 2019). In addition, most GPS systems do not provide data on the ball location, and since sensors are normally worn only by the team collecting the data, opponent information is missing. Furthermore, these systems are vulnerable to technical failures during a match (Hennessy and Jeffreys 2018; Buchheit et al. 2014).

The second approach to generating position data employs computer vision algorithms to derive player coordinates from video footage (Thomas et al. 2019). This video footage is generated using multiple cameras typically installed within professional football stadiums, which record the game from different angles (Taberner et al. 2019). A combination of computer vision algorithms and triangulation is then applied to generate position data for the players of both teams and the ball (Manaffard, Ebadi, and Moghaddam 2017). The method has been shown to accurately represent the correct locations of the tracked objects (Linke, Link, and Lames 2020, 2018). While this method automatically detects and tracks moving objects on the field, assigning the identity of players to these objects requires operator intervention, thereby rendering the process semi-automatic.

Both event and position data have become available to practitioners and researchers only at the beginning of the 21st century and have since improved in quality and thereby usability. These innovations with respect to data collection techniques have paved the way for the digital revolution in football (Rein and Memmert 2016) and have caused a considerable increase in studies making use of these types of data to investigate the game (Memmert 2024). A major challenge in this regard is posed by the necessity to process the raw data produced by each match (Klemp 2024), providing a meaningful way of data aggregation (Floris R. Goes et al. 2020). A multitude of indicators of team behavior exist (Low et al. 2019). Still, it is mandatory to derive the choice of performance or behavioral variables from theoretical knowledge instead of using a multitude of variables that do not necessarily represent any meaningful construct (Rein, Perl, and Memmert 2017). Even before the advent of these new data types, this caveat has been formulated by Tim McGarry (McGarry 2017), who discussed the finding that even intra-individual behavior varied largely from game to

game. He concluded that either there was too much variability in sports performance to derive any signature behavior or the data gathered were not able to identify these latent features. Consequently, any operationalization of player or team performance should always be questioned with respect to stability and validity, also keeping in mind the vital distinction between clearly measurable entities and hard-to-measure constructs (Atkinson 2002).

2.3 Statistical Learning

Having established the need to model sporting performance to identify important indicators and introduced the main types of data commonly utilized, the next step is to present the statistical approach behind this endeavor. As will be demonstrated in the following, this approach entails a broad range from simple linear models to sophisticated machine learning algorithms. As will also be shown, the challenges that threaten the validity of research findings are less due to a lack of sophistication in the modeling approach and more due to misconceptions with respect to steps taken before the actual modeling.

As has been worked out in section Section 2.2, the predominant aim of *MAR* is the identification of significant *Performance Indicators* that explain game success. Following the tradition of deterministic models, the approach entails the selection of one or more candidate *Performance Indicators* that are thought to be related to success. Contrary to the biomechanics approach to deterministic models, the relationships between *Performance Indicators* and success are not based upon known physical laws or mechanical properties, but are established based on statistical relationships between the *Performance Indicators* and some *Success Indicator* that are established empirically from data. According to McGarry (2017), a requirement for *Performance Indicators* used in *MAR* from both a theoretical and an applied perspective is that they should account for match outcomes. Consequently, identifying *Key Performance Indicators* in football is a problem of modeling the statistical association between *Performance Indicators* and a *Success Indicator* (which is commonly defined as the match outcome).

Modeling and interrogating the statistical relationship between an input and an output in this way is referred to as *statistical learning* (V. N. Vapnik 1999). The statistical learning problem is considered “as a problem of finding a desired dependence using a *limited* number of observations” (Vladimir N. Vapnik 2000, 17). Statistical learning theory has emerged from efforts with respect to convergence in probability theory, aimed to understand how learning machines generalize from observed to unseen

data (V. N. Vapnik and Chervonenkis 1971). It represents one of the more modern developments in the field of statistics, which has its roots in the early 19th century.

The first meaningful contributions to the modern statistical theory stem from Carl Friedrich Gauss and Adrien-Marie Legendre, who (potentially independently) invented the method of least squares [Stigler (1981); a discussion on which of them was actually the first to use it is beyond the scope of this work]. Towards the end of the 19th century, Francis Galton formalized the method of regression (Galton 1889), and around the same time, correlation was introduced by Karl Pearson (1900). In the 20th century, Ronald Aylmer Fisher revolutionized statistical methodology by introducing the methods that form the basis of today's statistics in scientific studies, maximum likelihood estimation, and hypothesis testing (Ronald Aymeric Fisher 1922). Fisher provided a comprehensive mathematical grounding for statistical inference, thereby establishing the frequentist paradigm that dominated statistical thinking throughout much of the 20th century.

Statistical learning emerged in the 1960s and 1970s through the pioneering work of Vladimir Vapnik and Alexey Chervonenkis. Their landmark paper, "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities" (V. N. Vapnik and Chervonenkis 1971), laid the theoretical foundations for understanding how learning machines generalize from observed data to unseen observations. Central to their contribution was the introduction of the Vapnik-Chervonenkis (VC) dimension, a measure of the complexity or flexibility of learning models. This measure directly links a model's capacity to its ability to generalize well, thus formally capturing the critical trade-off between model complexity and predictive accuracy.

Vapnik's further work formalized the statistical learning problem and caused a paradigm shift from Fisher's paradigm to a new one. Compared to the previous approach to statistical modeling, which posed restrictive constraints on the a priori knowledge needed to perform analyses, the new approach allowed the dependencies between inputs and outputs to be unknown and put inductive principles in place that helped in learning the associations as well as the form of the functions to be used from the data (Vladimir N. Vapnik 2000). Statistical learning theory was a crucial step in laying the theoretical foundations and formulating the main principles needed for the development of machine learning and later artificial intelligence.

For the present work, statistical learning theory provides a formalization of the general problem that the *MAR* researcher is facing. It can therefore be utilized to provide a statistical framework for the general *MAR* problem without relying on specific statistical models and individual approaches.

As detailed by Vapnik (1999), the problem of statistical learning is that of choosing

from a given set of functions, $f(x, \alpha)$, $\alpha \in \Lambda$, the one that predicts the actual outcome in the best possible way. In Vapnik’s terminology, the actual outcome is referred to as the “supervisor’s response,” as a learning machine has to be supervised, and the response of the supervisor is the output that the machine is learning to predict (see components of the learning model below).

According to Vapnik, the learning model is described as consisting of three components. The first component is a generator G of random vectors, x , which are drawn independently from a fixed but unknown distribution $P(x)$. In modern machine learning jargon, these vectors x would be referred to as predictors or feature vectors. The second component is a supervisor S returning an output vector y for every input vector x , according to a (fixed but unknown) conditional distribution function $P(y|x)$. This is equivalent to “ground-truth labels” or a target variable. The third component is a learning machine LM capable of implementing a set of functions, $f(x, \alpha)$, $\alpha \in \Lambda$, where Λ is a set of parameters. It is important to note that the elements $\alpha \in \Lambda$ are not necessarily vectors; they can be any abstract parameters. Consequently, any type of model can be chosen. For example, if the selected function were a linear regression, the α in $f(x, \alpha)$ would indicate the coefficients of the linear regression. In a neural network, α would, for example, indicate weights and biases (and other parameters of the architecture for the more sophisticated modeling approaches that are implemented nowadays).

The end of the 20th and the early 21st century have seen incredible advances in the field of *machine learning*, which have paved the way for *artificial intelligence*. Within this development, the notion has spread that statistics and machine learning are separate approaches to the modeling problem. Although many of the differences referred to in the public discourse are likely more superficial and related to terminology than conceptual (van Iterson, Haagen, and Goeman 2012), obviously, there exist generic differences between the main assumptions and paradigms underlying both approaches (Breiman 2001). These more serious differences seem to be between Fisher’s paradigm of inference and the learning paradigm as described above. While the methods introduced as early as 1808 (Legendre/Gauss) and 1889 (Bacon), specifically linear regression, are still used in statistical learning applications today (James et al. 2013), the conceptual approach might be different in the sense that instead of requiring assumptions about data and their dependencies, these dependencies are estimated from the data in an inductive way (V. N. Vapnik 1999).

An often-referred-to consequence of this inductive approach is the so-called performance-interpretability trade-off (Johansson et al. 2011; Assis, Dantas, and Andrade 2024). As a result of the capability of learning machines to derive dependencies and model parameters from the data without imposing assumptions,

the predictive models tend to become very complex and less interpretable to a human trying to understand how the model makes inferences. The consequence is that the researcher implementing a model of the association between inputs and outputs must balance accuracy (or performance) and the interpretability of the model, depending on the research objectives.

Both high accuracy and high interpretability are equally valuable to the *MAR* problem. While there might be clear cases for very powerful (and thus less interpretable) machine learning approaches as well as clear cases for very interpretable statistical models, often a sweet spot is sought, providing an efficient way to handle complex data and generating accurate predictions while also being able to derive insights about the form of the association between input variables and model predictions. While beyond the scope of this thesis, the efforts to develop Explainable ML and AI should be mentioned here as a valuable development in this context (Das and Rad 2020). In the sports analytics domain, for example, *Explainable Boosting Machines* (Nori et al. 2019) have been demonstrated for action-value models to be competitive while providing transparent insights into how predictor variables affect the predicted outcome (Decroos and Davis 2020).

2.3.1 Formalization of the Statistical Learning Problem in *MAR*

Translating the statistical learning problem to the *MAR* case, the three components would manifest in the following way. First, the generator G would be the process that generates the behavior or performance of football teams (which, in turn, is complex and might have driving factors that are, for now, irrelevant for the problem formulation). Consequently, the input vector x would contain one or more *Performance Indicators* as well as potential covariates. Second, the supervisor S would equal the observed *Success Indicator* (e.g., match outcome), generating output labels y . Lastly, the learning machine LM represents the modeling approach taken to learn the dependency between x and y . This machine is characterized by the model class utilized and has the task of finding the function $f(x, \alpha)$.

In the *MAR* literature, a variety of approaches have been utilized to model the associations between Performance Indicators and Success Indicators, while the choice of the exact modeling design often depends on slightly varying aims (such as explaining match outcomes vs estimating game state values). Generally speaking, and deriving from the above terminology, the statistical learning problem associated with the *MAR* problem consists of finding the function

$$\hat{Y} = f(x, \alpha), \tag{1}$$

where $\hat{Y}$ is the model prediction for the *Success Indicator*, x is the vector of *Performance Indicators* and covariates, f is the function that needs to be found, and α indicates the parameters of the modeling approach. The function f is to be found so that it best approximates the “supervisor’s response”, which, in this case, is the measured *Success Indicator* Y . Now, as it has been pointed out in Section 2.2 and is described in detail in the third publication of this thesis (Section 3.3), an important set of covariates in this learning problem of *MAR* is given by the set of *Contextual Factors*, i.e. situational variables of the *MAR* setting that potentially affect team behavior and team success, such as the game location or team strength. It is therefore considered to be appropriate to mention these covariates explicitly in the formalization of the statistical approach to *MAR*. Consequently, the function to be found can be formulated as

$$\hat{Y} = f(P, C, \alpha), \quad (2)$$

where P is the vector of one or more *Performance Indicators* under investigation and C is the set of *Contextual Factors* to be considered. α is the set of parameters of the approach, where the nature of the parameters depends on the functions taken into consideration. So, for example, a familiar representation in the form of linear regression for the case where the *Success Indicator* would be assumed to be approximately normally distributed would be

$$\hat{Y} = \beta_0 + \beta_1 P + \gamma C, \quad (3)$$

where β_0 corresponds to the intercept, β_1 corresponds to the vector of coefficients associated with the vector of *Performance Indicators*, P and γ corresponds to the vector of coefficients associated with the vector of *Contextual Factors* C . With respect to the formulation in (1) and (2), α would in this case consist of the coefficients in β_0, β_1 , and γ , since for linear regression, the parameterization of the model is given by the coefficients.

In the case of a non-normal *Success Indicator*, one would aim to find the Generalized Linear Model (McCullagh and Nelder 1989) of the form

$$g(E[Y]) = \beta_0 + \beta_1 P + \gamma C, \quad (4)$$

where $g(\bullet)$ denotes the link function to transform the output of the linear predictor on the right-hand side of the equation to the desired distribution (such as Poisson or Logit).

As demonstrated below in the review of statistical approaches to the *MAR* problem, a multitude of studies have utilized categorical *Success Indicators*, thereby requiring

models capable of modeling categorical outcomes. In this regard, the majority of studies have utilized *logistic regression* models of the general form

$$\text{logit}(E[Y]) = \beta_0 + \beta_1 P + \gamma C, \quad (5)$$

where $\text{logit}(p) = \log(\frac{p}{1-p})$ represents the logit link.

For any approach utilizing machine learning models, as stated above, there are no assumptions about the form of f being imposed on the problem, so the general formulation in (1) can be used.

Building upon the above formalization of the *MAR* problem in statistical terms, different approaches taken in the previous literature are discussed in the following. First, different statistical modeling approaches are discussed, and second, an important design choice for *MAR* studies, which shows a strong interdependency with the architecture of the statistical model, the *level of observation*, is highlighted.

2.3.2 Statistical Modeling Approaches in *MAR*

Different approaches to statistically modeling the *MAR* problem are outlined in the following. While several studies have used linear regression models, the majority of studies have utilized some type of classification model, either logistic regression or some kind of machine learning algorithm. Some rare examples of other Generalized Linear Models exist, including ordered logit models, Negative-Binomial models, or mixed effects models.

2.3.2.1 Linear Models

Since the type of model utilized is mainly dependent on the nature of the target variable, i.e., the *Success Indicator*, classification approaches dominate due to the categorical nature of match, shot, or event outcomes. However, certain study designs allow the use of “ordinary” linear regression models.

For example, the points won by a team over a season or one match represent a continuous variable that can be predicted using a linear model. Points won have been targeted in linear regression as a function of fouling behavior (Phatak, Rein, and Memmert 2021), running performance (Hoppe et al. 2014), market value (Andreas Heuer and Rubner 2014), ball possession rates (Collet 2012), or the home advantage (Lago-Peñas et al. 2016). Other *Success Indicators* that have been modeled with linear regression are *Expected Goals* (Phatak et al. 2022) or the magnitude of home advantage (Pollard and Gómez Ruano 2014; Riedl et al. 2022).

Besides these applications of linear regression to actually predict success, some studies have incorporated linear regression models in order to model some kind of *Performance Indicator* as a function of *Contextual Factors*. For example, the effect of contextual variables on ball possession rates has been a popular research topic (Paul Simon Bradley et al. 2014; Lago-Peñas 2009; Lago-Peñas and Martín 2013; Lago-Peñas and Dellal 2010). Similarly, the effect of context on running distances has been analyzed using linear regression (Lago-Peñas et al. 2010). Lastly, one study investigated the impact of contextual variables on the location of ball recoveries and the height of the defensive line (Santos, Lago-Peñas, and García-García 2017).

Other linear models have been used frequently to model the relationship between *Contextual Factors* and certain *Performance Indicators*. A very frequent method has been *analysis of variances* (ANOVA), applied mostly to study the effects of various factors on physical performance in football. In this way, running distance has been modeled as a function of game location and team quality (Paul Simon Bradley et al. 2015), playing time (Paul Simon Bradley et al. 2008, 2010), playing position (Paul Simon Bradley and Noakes 2013), or playing standards (Paul Simon Bradley, Carling, Diaz, et al. 2013). Also, ANOVA has been used to investigate the effect of playing formation on ball possession rates (Paul Simon Bradley, Carling, Archer, et al. 2013).

2.3.2.2 Logistic or Binomial Regression

As stated before, the categorical nature of most *Success Indicators* dictates the use of Generalized Linear Models or machine learning approaches suitable for classification. In this regard, logistic regression models have been very popular, probably due to their prominence in multiple research fields and established methods for their analysis (Mood 2009). A related procedure to logistic regression is binomial regression, which operates under the same assumptions about the data type but examines data on a different aggregation level (Faraway 2016). Consequently, multiple studies have utilized binary outcome variables and modeled effects using logistic regression. Since the outcome of a football match can take on one of three possible values (home win, draw, away win), logistic regression is not suited to predict match outcome in the standard case. However, other aspects of the game display binary outcome variables.

For example, logistic regression has been used to model the outcome of penalty shootouts as a function of team strength and home advantage (Wunderlich et al. 2020) or the outcome of individual penalty kicks as a function of the latent variables skill and pressure (Brinkschulte et al. 2023). In a case where only aggregated

information on the number of saved penalties was available, binomial regression was applied to investigate the effect of goalkeeper nationality on penalty kick outcome (Brinkschulte et al. 2021).

One common application of logistic regression for predictive *MAR* studies has been the prediction of success for ball possession phases. Several authors have considered varying sets of *Contextual Factors* as well as playing tactics as predictors for the success of ball possession phases in either achieving score box possessions (Lago-Ballesteros, Lago-Peñas, and Rey 2012; Albin Tenga et al. 2009a), goal scoring (Albin Tenga et al. 2009b; Gonzalez-Rodenas et al. 2020), or a more sophisticated measure of success, taking into account several possible outcomes (Pollard and Reep 1997).

One specific application has used logistic regression to predict the type and location of ball regains (Almeida, Ferreira, and Volossovitch 2014).

A suitable venue for the application of logistic regression models is the *Expected Goals* (xG) problem, which is very prominent in the *MAR* literature and aims to model a binary outcome. Since, for xG models, machine learning algorithms are a very popular choice (see below), there is only a limited number of works utilizing logistic regression. Among these articles, the one by Richard Pollard and colleagues (2004) stands out as probably the first scientific publication to fit an xG model, although this terminology is not used in the paper. While several papers presenting xG models have utilized logistic regression merely as a benchmark model to be compared to machine learning approaches (Eggels, Van Elk, and Pechenizkiy 2016; Robberechts et al. 2020; Anzer and Bauer 2021), the paper by Lucey and colleagues (2015) uses logistic regression as the main modeling approach.

2.3.2.3 Other Generalized Linear Models

Alternative modeling approaches have made use of Generalized Linear Models (GLM, Nelder and Wedderburn 1972) for target variables other than binary ones, where GLM technically allow modeling of a variety of target distributions.

As stated above, the outcome of a match is a categorical variable with three possible levels, while there is also a certain order present: from the perspective of one team, a loss is worth less than a draw, which is worth less than a win (which is also represented in the number of points awarded for the different outcomes). For this reason, ordered logit models have been a popular choice to model the outcome of a match as a function of different *Performance Indicators* (Lepschy, Wäsche, and Woll 2020; Collet 2012). One study used such an ordered logit model while accounting for

the nested structure in the data by means of a cumulative link mixed model (Rein, Raabe, and Memmert 2017).

The study by Taylor and colleagues (2008) aimed to predict the occurrence of certain events based on contextual information and utilized a log-linear model due to the arrangement of the raw data in a contingency table.

Besides these methods from the rather traditional area of statistical modeling, as stated before, machine learning has gained great popularity in *MAR* in recent years. Therefore, a considerable number of studies can be found modeling the *MAR* problem with machine learning methods, with a selection outlined below.

2.3.2.4 Machine Learning

The existing applications of machine learning in *Performance Analysis* in team sports are both manifold and versatile. Machine learning has been applied to forecast injuries (for a review see Eetvelde et al. 2021), performance analysis (for a review see Herold et al. 2019), and outcome prediction (for a review see Horvat and Job 2020; Bunker and Sunsjak 2019). In the following, while likely not exhaustive, approaches to link *Performance Indicators* to *Success Indicators* in match play are listed together with their aims and the specific algorithms that have been used.

As stated above, the *Expected Goals* problem is a research subject that invites the application of classification machines. Logically, various authors have utilized machine learning algorithms to model the *Expected Goals* value of shots (i.e., the probability of a shot resulting in a goal) as a function of a variety of *Contextual Factors*. As is the standard approach in the computer science domain, most articles include more than one modeling approach and compare the different models with each other. In effect, besides logistic regression, the xG problem has seen the use of decision trees (Eggels, Van Elk, and Pechenizkiy 2016), random forests (Eggels, Van Elk, and Pechenizkiy 2016; Cavus and Biecek 2022), AdaBoost (Eggels, Van Elk, and Pechenizkiy 2016; Mead, O'Hare, and McMenemy 2023), Catboost and LightGBM (Cavus and Biecek 2022), XGBoost (Cavus and Biecek 2022; Mead, O'Hare, and McMenemy 2023; Robberechts et al. 2020) as well as Multilayer Perceptrons (Ruiz et al. 2015; Mead, O'Hare, and McMenemy 2023).

Action-value models represent an extension to the xG problem, aiming to assign a value not only to shots, but to any action taken by a player in a match. A very prominent approach to valuing actions is the *VAEP* (Valuing Actions by Estimating Probabilities) model that was originally developed using the Catboost algorithm (Decroos et al. 2019), but has more recently been replicated by means of the more

interpretable *Explainable Boosting Machine* (Decroos and Davis 2020; Van Haaren 2021).

Further extending the idea of action-value models, the broad family of *Expected Possession Value* (EPV) models aims to predict the success probability of a current ball possession for any given moment in the game. To this end, EPV models have been proposed using Graph-Neural-Networks (Dick, Tavakol, and Brefeld 2021) and have been extended using multi-layer Deep Learning architectures (Fernández, Bornn, and Cervone 2019).

Other applications of machine learning in *MAR* have predicted the match outcome using decision trees (Lago-Peñas et al. 2016) or possession regains through Graph-Neural-Networks (Raabe, Nabben, and Memmert 2022) (while in the latter article, the focus was rather on data representation for position data in machine learning applications than on actual prediction).

To summarize, a broad variety of statistical and machine learning methods have been used to model the relationship between *Performance Indicators* and *Success Indicators* in *MAR*. The choice of modeling approach is clearly dependent on the types of variables included in the analysis, while machine learning approaches place less demands on the distributions of variables or knowledge about dependencies. While machine learning models are flexible and applicable to a broad variety of problems, they are not immune to biases that might be introduced through flawed study design (Christodoulou et al. 2019). Therefore, *MAR* has to be discussed in the light of *causal inference* if meaningful insights are to be derived from its studies.

2.4 Causal Inference

2.4.1 *MAR* as a Case of Observational Research

The type of research discussed so far in this thesis concerning *MAR* can almost exclusively be classified as so-called *observational research*. The term *observational* means that the data analyzed in the study do not stem from randomized, controlled experiments but from real-world observations of the subjects under examination. It is commonly agreed upon within science that the highest degree of conviction regarding the causal effect of an intervention on certain outcomes can be achieved only through experimental manipulation while ruling out as many other possible explanations as possible by means of randomization (Pearson 1911).

While this approach to science is clearly the gold standard, there are scenarios in which randomized controlled trials are not an option.

First, ethical considerations deem certain intervention studies impossible. While the effects of certain, potentially harmful behaviors like smoking or alcohol consumption on health are obviously of great interest, it would be highly unethical to deliberately expose people to these behaviors in order to study their effects. Consequently, no controlled experimental data will ever be available for these kinds of questions.

Second, studies conducted in laboratory settings suffer from limited ecological validity because they fail to replicate the complex and interactive nature of the phenomena that they are supposed to model. Observational research, on the other hand, benefits from high ecological validity because data are collected in naturalistic settings, reflecting real-world human behavior.

As a result, certain research questions will likely never be answered within an experimental design. At the same time, such questions might be of great interest and importance to the public, so alternative solutions to answering these questions have to be pursued. The general approach to this is observational research, i.e., observing people in their “natural habitat” and gathering data about behavior and outcomes of interest. For example, the study of public health tries to collect data on as many individuals as possible through surveys and to infer associations between aspects of behavior and long-term health (Rothman and Greenland 2005). Although sometimes researchers are cautious about deriving causal conclusions from observational data, it is being argued that, given a diligent approach to science and use of the correct methods, important insights might be generated from such studies (Hernán 2018), especially when controlled experiments are not possible or not able to mirror the specifics of the real-life situation under investigation.

In the following sections, the *causal inference* framework is introduced, which provides a strong theoretical foundation for the endeavor of deriving causal conclusions from empirical data. Especially for the case of observational research, causal inference is an invaluable piece of theory that allows to formalize the research problem in rather general terms and to interrogate the research design for potential sources of bias.

Before proceeding, a few terms have to be introduced and related to the concepts discussed so far in this thesis. Since causal inference was mainly developed and made the greatest contributions in the field of public health, the causal inference terminology is tightly related to the terms commonly used in this field. Consequently, most study designs generally incorporate an investigation of the effects of certain *treatments* or *exposure* (in practice, often medical treatments or exposure to risk factors in everyday life) on outcomes (in these studies, outcomes often reflect health-related outcomes). The relationship between treatment or exposure and outcomes is referred to as the *association* between these two, while an important distinction has to be made between a possible statistical association and a causal effect between the

two. Potential circumstantial information on the sample under investigation might be referred to as a *background* or simply as *covariates*.

If the causal inference problem were approached through statistical learning and the formalism introduced in Equation 1, the presence and potential form of a treatment or exposure as well as the background variables would be reflected in the input vector x and the outcome would form the supervisor's response y . The aim of the study would be to find the association between inputs and outputs, i.e. f and its set of parameters α . In a more traditional statistical sense, treatment or exposure would be termed the *independent variable*, the outcome would be termed the *dependent variable*, and any potential background variable would be termed the *covariate* or *control variable*. An analogy between causal inference terminology and the actual *MAR* problem is provided in a later subsection.

In order to introduce causal inference, first, the underlying model of causation and an operational mathematical definition of it are reviewed. It is then explained how confounding bias arises, and again, a mathematical definition of confounding is provided. Lastly, the *MAR* problem is viewed in the light of causal inference and confounding, discussing potential sources of bias in *MAR* studies and giving an example. The section concludes with a short consideration of experimental studies in *MAR*, including the arguments for and against using such a paradigm.

2.4.2 The Counterfactual Model of Causation

In order to perform causal inference on any data, whether they are observational or experimental, requires a solid definition of the concept of a *cause*. While several attempts to this definition in the public domain suffer from some kind of circularity (Greenland, Pearl, and Robins 1999), the definition that has probably gained most acceptance in science is the *counterfactual definition of causation* that goes back to David Hume (1748):

“We may define a cause to be an object, followed by another, . . . where, if the first object had not been, the second had never existed” (Hume 1748, 115)

This approach to the term *causation* is termed *counterfactual* because it considers alternative realities in which the “object” that is termed a cause is not present. Only if, in this alternative reality where the cause is not present, the effect would also not be present, can the relationship between the two be considered causal (Greenland, Pearl, and Robins 1999). Of course, such an investigation will always be impossible.

The hypothetical scenario is rather drawn in order to understand the true meaning of *cause and effect*.

For the study of causal inference, the *counterfactual model of causation* is particularly relevant because one is frequently concerned with estimating an effect that cannot be observed. In experimental studies, if possible, cross-over designs are used, ensuring that each participant receives every treatment that is under investigation to rule out potential individual participant effects. While this is not even possible for any kind of experimental study, it is definitely impossible for observational studies. In these, researchers normally try to infer the effects that a treatment or exposure would have had on those units that did not actually receive the treatment, thereby making inferences about *counterfactuals* (also referred to as potential outcomes).

This formulation of the objective of counterfactual models provides a good introduction to the formalization of the counterfactual model of causation that has been proposed by Neyman (1923) and can be summarized as follows:

Consider an experiment involving N units, which could be individuals, populations, or objects. The study design will assign one of $K + 1$ treatments, denoted as $x_0, x_1, \dots, x_k$, to each unit. The notation of $K + 1$ is chosen because normally one of the treatment conditions refers to a control condition (which is here specified as x_0). For each unit i , the outcome of interest is the value of the response variable Y_i . If unit i receives treatment x_k , the corresponding observed outcome is denoted by y_{ik} .

The reference treatment x_0 typically serves as a benchmark for comparison with the other treatments. This reference condition is usually a placebo, no treatment, or a standard intervention.

The causal effect of treatment x_k (for $k \geq 1$) on the outcome Y_i , relative to the reference treatment x_0 is defined as the difference $y_{ik} - y_{i0}$. If the response variable is strictly positive, an alternative definition of the causal effect can be given as the ratio y_{ik}/y_{i0} or, equivalently, the log difference $\log y_{ik} - \log y_{i0}$.

In essence, a causal effect represents a *counterfactual contrast*, comparing the potential outcomes of the same unit under different treatment conditions. This is why this model is also referred to as the potential-outcomes model of causation.

It is important to note, however, that because only one of the potential outcomes y_{ik} can be observed in any one unit, the effect of a treatment on an individual unit $y_{ik} - y_{i0}$ can never be observed.

This formalism of the counterfactual model of causation can be logically extended to discussions on biases, specifically confounding, since confounding arises due to the fact that it is not possible to observe the difference between the effects of two treatments

on the same individual. Comparison of treatment effects must always include a comparison of individuals, and if important differences exist in the characteristics of these individuals, the effect might be biased. The following section deals with the emergence and detection of confounding.

2.4.3 Confounding in Causal Inference

It is natural that observational studies are not immune to biases since the systems that are being studied (e.g., humans in their daily life) are highly complex and a multitude of unobserved influences exist. If these unobserved influences affect the variables of the main study, biases can arise. One prominent bias is *confounding*.

Confounding arises when the association between a treatment and an outcome is distorted by the presence of one or more extraneous variables that influence both. In causal inference, a confounder is a variable that is associated with both the treatment assignment and the outcome, creating a biased estimate of causal effects (Greenland, Pearl, and Robins 1999; Hernan 2023). Failure to account for confounding can result in misleading conclusions, as observed associations may not reflect true causal relationships. Addressing confounding is crucial in both experimental and observational studies, typically through study design strategies such as randomization (Ronald A. Fisher 1935), or statistical adjustments such as stratification, regression modeling, or propensity score methods (Rosenbaum and Rubin 1983).

One of the earliest systematic discussions of “confounded effects” can be found in John Stuart Mill’s book “A System of Logic” (1843), in chapter 10 “Of Plurality of Causes, and the Intermixture of Effects”. Earlier in the book, Mill acknowledges Francis Bacon as having dealt with the issues of confounding even earlier.

Mill formulates a requirement for experiments aiming to derive causal effects:

“ . . . none of the circumstances [of the experiment] that we do know shall have effects susceptible of being confounded with those of the agents whose properties we wish to study.” (Mill 1843, 490)

It should be noted that in Mill’s terms, an “experiment” rather refers to an observation whose circumstances can be at least partly controlled by the experimenter and not to the controlled trial that the word experiment implies nowadays.

In other words, Mill’s definition of confounding states that if certain circumstances of the observation affect the outcomes under study, it might happen that these effects caused by the circumstances are *confounded* with the effects of the primary study agents (mostly treatments or exposure). In simple terms, this means that the

association that is observed in the study can not necessarily be attributed to the study agents (which would be desirable), but the option has to be taken into account that the association is due to any of the circumstances of the observation.

Most of modern literature follows this informal conceptualization of confounding, although terminology has become more precise, with “*treatment*” used to refer to an agent administered by the investigator and “*exposure*” often used to denote an unmanipulated agent (Greenland and Morgenstern 2001). The “circumstances of the experiment” are mostly referred to as background variables or covariates.

Now it is not too problematic if a covariate has an effect on the outcomes independently of the effect of the treatment or exposure. In this case, the respective associations could be estimated using statistical techniques and would most likely not be confounded. Problems arise especially when the covariate affects both the outcomes and the primary study agents (for example, the assignment of treatment or the probability of exposure). This scenario is expressed in the classical causal diagram for a confounded relationship among three variables, shown in Figure 2.1, adapted from Hernán and colleagues (2011), where C reflects a confounder, A reflects treatment, and B reflects the outcome. One of the most prominent pictures of confounder bias can be found in the classical example of *Simpson’s paradox* (Simpson 1951).

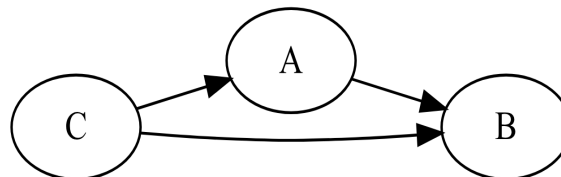


Figure 2.1: Causal diagram of a confounder C (Hernán, Clayton, and Keiding 2011).

Simpson’s paradox is generally known as the phenomenon that a certain association between variables observed in multiple groups of data disappears or is reversed when the groups are aggregated. In the original example, Simpson presents (hypothetical) data of men and women being administered to some medical treatment. Within both sexes, the Odds Ratio of the effect of the medication on death suggests a protective effect of medication. However, when considering the whole sample and ignoring sex, this protective effect disappears (Simpson 1951).

Simpson’s paradox has been discussed in the light of varying explanations, and it has been pointed out that the causal structure of the problem is crucial to decide which of the effect measures to choose (Hernán, Clayton, and Keiding 2011). However, a very general mathematical explanation for the occurrence of this paradox is the fact that, in the data presented by Simpson, men were less likely to receive treatment

than women, while, at the same time, men were less likely to die overall. In other words, sex had an effect on both the probability of treatment and on the outcome variable. As was described above, this effect is a necessary condition for confounding. The causal diagram in Figure 2.1 applies to this problem if one considers A to be the medical treatment, B to be the outcome (death/survival), and C to be sex. Graph theory shows that in such a constellation, a common cause like C will create an association between its effects A and B that is independent of the actual (causal) relationship between the two variables (Pearl 1995). So while the actual association between A and B is negative (implying a protective effect of medication on the probability of death), a positive association is added through the confounder C . This type of confounding, where the effect of the confounder is in the opposite direction compared to the actual effect, can lead not only to a biased estimate of the effect, but actually to a reversed association (as is the case in Simpson’s example).

Having established the symptoms of confounding as well as the general mechanism of how a confounder acts upon a relationship between two variables, it is discussed how this can apply to the *MAR* setting and which common indicators of performance or success are particularly prone to confounder bias.

2.4.4 Confounding in *MAR*

As outlined above, *MAR* in its common form relies on observational, correlational studies to gain insights into the dynamics of the game and associations between *Performance Indicators* and *Success Indicators*. Building upon the *causal inference* framework, it is generally possible to derive causal effects from purely observational data. However, *MAR* is no less vulnerable to the same potential for biases as any research field in the realm of causal inference.

The problem of confounding has been outlined in detail above. In summary, confounder bias might be introduced to an association between two variables A and B if there exists a background variable C that affects both A and B (see Figure 2.1) and is not accounted for in the statistical analysis. If we again consider the terminology from Section 2.3, these statements translate to the following in *MAR*-specific terms: if a *Contextual Factor* C affects both the *Performance Indicator* X and the *Success Indicator* Y , these associations might confound the marginal association between X and Y , which is the main target of the statistical analysis. It is important to note that in this case, the marginal association should not be used and instead the conditional associations between X and Y , conditional on levels of C , should be considered. Consequently, confounder bias can be avoided when C is accounted for statistically.

Coming back to the common statistical model deployed in *MAR* that is depicted in Equation 2, it is therefore in general interesting to have a look at the interrelationships between *Contextual Factors* C , *Performance Indicators* X , and *Success Indicators* Y . As is demonstrated in detail in Section 2.2.1, strong dependencies between the game context and team behavior, as well as success, exist. It is therefore more than reasonable to assume that C does, in most cases, affect both X and Y . The *MAR* setting seems, therefore, to be an endangered endeavor with respect to confounder bias.

In fact, for the example of *Contextual Factor* scoreline (C), *Performance Indicator* running performance (P) and *Success Indicator* next goal (Y), it was demonstrated that C affected both P and Y and consequently, the marginal association between P and Y was considerably different from the conditional associations for all levels of C , effectively demonstrating a confounding effect of scoreline on the marginal association between running performance and goal scoring (Klemp, Rein, and Memmert 2023). Under the impression of the evidence on effects of *Contextual Factors* outlined above, this seems to be a regular rather than an exceptional phenomenon. It should therefore be common practice to take these interdependencies into account and investigate whether potential confounding due to context could be present.

2.4.4.1 Controversial Findings in *MAR*

The above sections have detailed the mechanism by which interrelationships between context and team behavior or success in football might be responsible for confounding in the *MAR* setting. Also, one study was mentioned that explicitly investigated the potential presence of confounder bias. However, the *MAR* literature has seen a concerning number of controversial findings with studies on the same research objective demonstrating different evidence and conclusions. While this might be an issue of replicability, it quickly becomes obvious that inconsistencies in the adopted methodologies exist, potentially providing a causal explanation for the variation in results. One prominent example of such controversy is discussed below, making a case for more rigorous inclusion of *Contextual Factors* into statistical design.

One very clear example of the effect of context on the associations analyzed in *MAR* can be found in the work by Phatak and colleagues (2022). Here, it is demonstrated that the overall effect of defensive *Performance Indicators* (such as number of tackles or interceptions) on success is found to be negative (i.e., making more tackles and interceptions is harmful for success). However, the authors argue that the numbers of offensive and defensive events, respectively, are linked to ball possession rates (since offensive events are only possible during ball possession and vice versa). After

controlling for ball possession rates within matches, the direction of the effects of the defensive indicators changes, rendering them beneficial for success. Elaborating on the above discussion on confounding, it is obvious that here possession (C) might have had an effect on both defensive performance (P) and success (Y). It could have caused confounding of the actual association between P and Y with the association introduced by the effect of C on both P and Y (compare Figure 2.1).

This example highlights how, in a common study design within *MAR*, the inclusion of contextual information drastically changes the inferences drawn from the statistical results and emphasizes the need for rigorous evaluation of whether such control is needed. As the question might be asked, how should the decision be made about whether to include a Contextual Factor in statistical analysis or not? A detailed account of this problem can be found in the third publication of this thesis (Section 3.3). There, it is laid out specifically how the presence of confounding in the statistical approach to the study of *MAR* can be investigated using the concept of *collapsibility* and how tests of collapsibility can be performed in order to decide whether or not a given *Contextual Factor* should be considered for inclusion.

2.4.5 The Experimental Match Analysis Paradigm

Having acknowledged the inherent limitations of observational *MAR* research, it is necessary to mention one parallel development that has the potential to overcome these limitations: the *experimental match analysis paradigm*.

As stated before, the controlled experiment is considered the gold standard for answering scientific questions empirically. This is equally true for the case of *MAR* studies (McGarry 2017). For most of the history of *MAR*, the closest approximation of an experiment has been attempts to study individual or group-level behaviors in Small-Sided-Games studies (for an overview see Halouani et al. 2014). However, it is questionable whether research findings from these studies can be extended to the formal football game due to differences in field size and player numbers (Silva et al. 2014).

A major limitation of conventional *MAR* methodologies, whether observational or SSG-based, is the difficulty in balancing ecological validity with sufficient control over confounding variables. The dynamic nature of football, including transitions between match phases, game contexts such as competition level, psychological pressure, and scoreline effects, poses challenges for designing experiments that accurately capture real-game behaviors. Furthermore, conducting experimental studies in full-scale matches is highly resource-demanding and often constrained by logistical and ethical considerations, such as access to elite players and ensuring competitive fairness.

An innovative approach to addressing these challenges has been introduced by Memmert and colleagues, who proposed an experimental, trial-based match analysis paradigm designed to investigate causal effects in football (Memmert et al. 2019). This approach follows the fundamental scientific progression from description to explanation, followed by hypothesis-driven testing through empirical data collection (McGarry 2017). Unlike traditional *MAR* methodologies, this paradigm allows for the systematic manipulation of specific game conditions while maintaining a controlled experimental framework. In this line of research, effects of playing formations (Memmert et al. 2019; Low, Rein, Schwab, et al. 2021), pressing strategies (Low, Rein, Raabe, et al. 2021), or cognitive capacities (Memmert et al. 2023) have been investigated utilizing experimental manipulation and controlled observation using *Performance Indicators*.

However, its implementation requires overcoming substantial obstacles, such as ensuring adequate sample sizes to mitigate the risk of pseudo-replication (Hurlbert 1984). The sample sizes that can be realistically generated under this approach are way lower than the sample sizes normally available in *MAR* studies utilizing existing, observational data (Klemp, Memmert, and Rein 2021) and, therefore, the potential insights are limited in this regard. Related to the sample size, it has to be noted that the samples generally available to researchers in the context of the *experimental match analysis paradigm* will rather consist of amateur players whose performance and behavior provide potentially little external validity with respect to elite players who are, after all, the main subject of research in *MAR*.

It can therefore be concluded that the experimental match analysis paradigm presents a valuable approach to a problem for which no perfect solution exists. However, due to the above-mentioned limitations, it will not be capable of replacing the observational *MAR* paradigm in the near future. On the contrary, it is to be expected that both approaches will develop further and hopefully draw from each other with respect to insights generated into their methodology.

3 Synopsis

3.1 Study 1: In-play forecasting in football using event and positional data

As has been illustrated in the Literature Review of this thesis, the common *MAR* paradigm generally seeks to identify relationships between *Performance Indicators* and *Success Indicators* by means of statistical modeling (see Section 2.2). However, it has been pointed out that the multitude of works published in this realm, although implicitly pursuing a similar aim, have been conducted using highly heterogeneous methodologies. In particular, varying statistical approaches (Section 2.3.2) and observation levels (Section 2.2.2) have been utilized, leading to variations with respect to the inclusion of *Contextual Factors* (Section 2.2.1).

This heterogeneity makes it difficult to compare the results generated by different studies, since it is mostly not possible to disentangle the effects of the actual primary study agents from potential artifacts due to the respective study design choices.

Therefore, the first publication of this thesis aimed to investigate the relationship between various *Performance Indicators* and *Success Indicators*. *Performance Indicators* were chosen from all known categories commonly investigated in the literature (i.e., physical, technical, and tactical) in order to provide an overview and enable comparison of their respective explanatory power for match success. In order to better understand the relationship between the *Performance Indicators* and immediate as well as future success, performance was observed in the first half of each match, and then relationships with success in both halves were estimated separately. Furthermore, *team strength* was incorporated as the main *Contextual Factor*, and the pairwise associations between team strength, performance, and success were analyzed. The purpose of this procedure was to understand the different facets of *Performance Indicators*, i.e., (1) their dependency on pre-game knowledge of team strength, (2) their explanatory power for explaining immediate success, and (3) their prognostic value of predicting future success (in the second half). Importantly, game location and interaction with the opponent were accounted for by study design.

The results of the first publication generated important insights with respect to the relative contribution of various aspects of team performance to match success. To the best of the author's knowledge, no previous study in *MAR* has provided

a comparable overview of the effects of so many *Performance Indicators* under a unified methodological framework. Furthermore, the results revealed for the first time dependencies between team strength, performance, and success in a systematic way, aiding in understanding which aspects of performance characterize team strength and which are important for game-play success.

However, the insights generated in this study are still limited in an important way. Since performance and success were both aggregated on the level of match halves, it cannot be stated with certainty whether the observed performances preceded and thereby caused successful outcomes (i.e., goals) or whether they occurred as a consequence of success. Since the goal difference per half is utilized as *Success Indicator*, but no information on the timing of goals was included in the analysis, a more fine-grained analysis was needed to clarify the effect of performance on success. This conclusion informed the second publication of this thesis, adopting a more granular approach to the *MAR* problem.

In-play forecasting in football using event and positional data

Klemp, M., Wunderlich, F., & Memmert, D. (2021). In-play forecasting in football using event and positional data. *Scientific Reports*, 11(1).

<https://doi.org/10.1038/s41598-021-03157-3>

Abstract: Two highly relevant aspects of football, namely forecasting of results and performance analysis by means of performance indicators, are combined in the present study by analysing the value of in-play information in terms of event and positional data in forecasting the further course of football matches. Event and positional data from 50 matches, including more than 300 million datapoints were used to extract a total of 18 performance indicators. Moreover, goals from more than 30,000 additional matches have been analysed. Results suggest that surprisingly goals do not possess any relevant informative value on the further course of a match, if controlling for pre-game market expectation by means of betting odds. Performance indicators based on event and positional data have been shown to possess more informative value than goals, but still are not sufficient to reveal significant predictive value in-play. The present results are relevant to match analysts and bookmakers who should not overestimate the value of in-play information when explaining match performance or compiling in-play betting odds. Moreover, the framework presented in the present study has methodological implications for performance analysis in football, as it suggests that researchers should increasingly segment matches by scoreline and control carefully for general team strength.

3.2 Study 2: The influence of running performance on scoring the first goal in a soccer match

As was concluded following the results of the first study of this thesis, study designs aggregating performance and success variables over longer time intervals are limited in their capability to derive any predictive effects of performance on success. While goals are the most important *Success Indicator* in football, they also directly affect the *scoreline*, one of the most crucial *Contextual Factors* driving team behavior (see Section 2.2.1). This circularity is a unique challenge in the *MAR* setting and suggests that contextualization on scoreline should be an important premise for any *MAR* study that aims to derive associations between performance and success. Furthermore, it became clear from the first study that the prediction of success from performance should respect the temporal order of performance and success and should, therefore, measure success only after a performance has occurred.

Therefore, the aim of the second study of this thesis was to perform a more granular analysis of the association between performance and success while specifically considering the scoreline as a contextual factor as well as the temporal relationship between performance and success. The *Performance Indicator* chosen for investigation in this study was the physical performance of football teams, and the *Success Indicator* was chosen to be goal scoring. In order to fulfil the above-mentioned requirements, the study was designed so that individual goals could be predicted based on the physical performance that was observed in the time interval *before* the goal was scored. In order to control for potential scoreline effects, only the first goal within each match in the sample was considered, respectively, ensuring a standardized scoreline of 0:0.

The most important result of this study was that teams' running performance was positively related to goal-scoring, irrespective of game location, team strength, and the specific parameter of physical performance (e.g., total distance, high-intensity running distance, in-possession running distance). This is particularly relevant because previous studies on the association between physical performance and success have yielded mixed results. As is described in detail in the article, these controversial results might have been due to heterogeneous study designs. For example, none of these studies considered the current scoreline as a *Contextual Factor*, which might limit the validity of study results, as scoreline is known to affect physical performance and could therefore be a potential confounder (see Section 2.4.3) of this association. Consequently, the results from this study, where only standardized scorelines were considered, provide an important contribution to this discussion.

A secondary finding, which is an implication of the above discussion, is that the segmentation of matches into sequences of standardized scorelines might reverse the

association observed between a *Performance Indicator* and a *Success Indicator* as compared to previous results from the literature. This is noteworthy especially for the broader context of *MAR* study design and the potential problems of confounding. In general, the results highlight the importance of contextualization and, if necessary, of segmentation in order to properly account for context.

In general, the results from the study suggest that segmentation and contextualization might alter effects observed in *MAR* studies. This might be true for other *Contextual Factors* than scoreline and other *Performance Indicators* than running performance. In the present study, only a segmented approach was utilized, and the comparison is only made with previous results from the literature. It is, however, important to clarify the effects of segmentation within the same sample. Additionally, if two approaches to analyzing the same data produce conflicting results, one needs to provide arguments for why one of the two approaches is more appropriate than the other. This argument can be made using the causal inference framework introduced in Section 2.4.

Consequently, the third study of this thesis approached the *MAR* problem through the lens of causal inference and aimed to further investigate the effects of segmentation on *MAR* study results.

The influence of running performance on scoring the first goal in a soccer match

Klemp, M., Memmert, D., & Rein, R. (2021). The influence of running performance on scoring the first goal in a soccer match. *International Journal of Sports Science & Coaching*, 17 (3), 558–567.

<https://doi.org/10.1177/17479541211035382>

Abstract: Previous studies investigating running distance in high performance soccer have led to contradictory evidence, potentially due to ignoring contextual information during match phases. The present study therefore examined the relationship between running performance and goal scoring in a football match for a standardised score line. In a sample of 302 matches from the first German Bundesliga, the first goal was modelled as a function of the teams' running performance and team strength using logistic regression. Goodness of fit was assessed by the prediction accuracy of the model utilising cross-validation. The best model showed a mean accuracy of 77%, reflecting a strong relationship between running performance and the probability of scoring the first goal. This relationship was strongest for total running distance compared to high-speed, sprint or in-possession running distance. We propose two different potential mechanisms to explain the relationship between running performance and goal-scoring found in the present study. These are (1) better ability to reach tactical aims or (2) accumulation of fatigue in the opponent. Future studies should build on these results by further examining the relationship between running performance and success using a more granular segmentation of matches.

3.3 Study 3: The Role of Contextual Factors in Match Analysis Research and Their Potential for Confounding Effects: A Case for a Segmented Match Analysis Paradigm

The first two studies contained in this dissertation revealed important insights into the importance of *Contextual Factors* within the realm of *MAR* as well as the interrelationship between study design and study results. Specifically, it was demonstrated that an approach taking into account the contextual effects of the scoreline in a match produced different results than approaches ignoring this context. Furthermore, it was shown that certain aspects of contextualization require segmentation of matches, thereby modifying the *level of observation*, an important parameter of study design, compared to most previous works.

As outlined above, the contrasts of these effects were observed between multiple studies. Thereby, variations could be due to varying samples or other differences with respect to study design. Therefore, the aim of the third study was to investigate the effects of segmentation as well as the importance of *Contextual Factors* for the study of *MAR*.

In addition, the study aimed to provide a justification for the procedure of contextualization and the accompanying need for segmentation. If two approaches (with and without segmentation/contextualization) produce contrasting results, it is not straightforward which approach should be considered the appropriate one. Consequently, a theoretical foundation needs to be laid upon which this decision can be made. As is outlined in detail in this article, this theoretical foundation can be found in the framework of causal inference. Following the causal structure of a problem, it can be examined whether conditioning on a certain parameter is valid or not (see Section 2.4.3). More specifically, viewing the *MAR* problem through the lens of causal inference leads to the notion of *collapsibility*, a statistical phenomenon closely related to the problem of confounding. *Collapsibility* describes whether the inclusion and exclusion of a covariate (and segmentation if necessary) has an effect on the statistical association that is of main interest to the study.

In the third study, first, the *MAR* literature is reviewed to establish an overview of the effect of *Contextual Factors* on team performance and success. It is demonstrated how various aspects of game context (see Section 2.2.1 for an overview) influence how teams behave and perform, as well as their expectation of success. Next, the study establishes a framework to analyze collapsibility in *MAR* studies to determine the need for contextualization. This framework is applied to the *Contextual Factor* scoreline, which has already been utilized in previous studies, and it is shown how the scoreline can cause noncollapsibility, possibly causing confounding.

The results of the third study provide important insights both in a practical and a theoretical sense. First, the strong effects of scoreline as a *Contextual Factor* on team behavior, as well as the associations between performance and success, could be demonstrated. Second, these results can be viewed in a more general and theoretical sense, providing researchers with a method to repeat this analysis for other *Contextual Factors* and other *Performance Indicators*. It was worked out in the study that conditioning on a *Contextual Factor* is mandatory if (1) the causal structure of the problem generally allows that *Contextual Factor* to affect team behavior and success, and (2) the *Contextual Factor* causes noncollapsibility with respect to the association between the *Performance Indicator* and the *Success Indicator* of interest.

4 Discussion and Conclusion

The field of *Match Analysis Research (MAR)* has its roots in several decades ago and therefore looks back to a considerable history and evolution. The advent of big data and machine learning in the field has transformed the potential of *MAR* while also posing greater challenges for researchers regarding appropriate data analysis and modeling. The present thesis has worked out methodological challenges and shortcomings in the previous literature of *MAR* and revealed potential sources of systematic bias due to study design choices.

4.1 Summary

This thesis began with a comprehensive review of the historical evolution of *Match Analysis Research*, emerging from the broader area of Performance Analysis within as well as outside of team sports. Here, it was demonstrated how Performance Analysis arose, on the one hand, from exercise science and the attempt to uncover the physiological basis of human movement and, on the other hand, from biomechanics and the aspiration to build deterministic models of team sport performance.

From the latter, it was hoped to uncover underlying laws governing performance in sports. While the approach has its origins in biomechanics, where deterministic models have successfully described mechanical phenomena, the complexity and inherent uncertainty present in football data complicate the direct application of this paradigm.

With the advent of big data and machine learning (ML), *MAR* has experienced a significant methodological shift. The rapid technological innovations in data collection have dramatically increased both the volume and complexity of data available, fostering the need for methodological approaches capable of dealing with such volume and complexity.

Advanced data processing procedures, as well as statistical learning and, more specifically, machine learning techniques have proven useful in modeling the millions of data points being produced per football match. Insightful metrics have been proposed, which would never have seen the light of day without machine learning. However, although these methods offer flexibility and predictive power, they have inadvertently nourished the misleading perception of robustness against causal biases,

which were thought to be largely related to limited sample sizes and rigid modeling assumptions. These biases, however, can find their way into causal conclusions being drawn from statistical analyses at various levels of sample size or modeling flexibility.

This thesis has uncovered the sources of such biases in the literature review on the causal inference framework, illustrating that causal bias is not a consequence of a lack of data or modeling power, but of the structure of the causal problem underlying the study design. Specifically, confounder bias was identified as the most prevalent issue in observational studies, originating from a confounding covariate (or background variable) that is related to both the predictors (or independent variables) and the target variable (or dependent variable) in a statistical analysis.

In an attempt to understand the specific causes of confounder bias in the *MAR* setting, the central pillar of this thesis was the exploration of *Contextual Factors*. It was demonstrated that *Contextual Factors* have an enormous impact on the dynamics of the game through influencing the aims and behaviors of players and teams. At the same time, context impacts success rates on different levels of observations. Considering the structure of the classic *MAR* problem, these features of *Contextual Factors* form the basis for introducing systematic biases into analyses, specifically confounder bias.

By employing the causal inference framework, this dissertation provided a theoretical foundation to explain precisely how confounder bias arises within *MAR*.

The synopsis, comprising the core studies of this dissertation, empirically demonstrated the substantial impact of Contextual Factors on performance indicators and their relationship with success outcomes. Specifically, the studies illustrated that varying the level of observation and adopting a segmented match analysis approach profoundly influence the results and the interpretation of the effects of *Performance Indicators*.

These differences in empirically observed effects were explained in the language of causal inference and traced back to the potential of confounder bias in *MAR* studies, which might be the central and most crucial finding of this dissertation.

4.2 Main Findings

As has been shown throughout the literature review in Chapter 2 as well as the research articles in the synopsis in Chapter 3, the classical *MAR* paradigm, trying to link *Performance Indicators* to *Success Indicators* using statistical learning, is vulnerable to biases, most prominently confounder bias. One of the most prevalent mechanisms by which this bias might be introduced to *MAR* studies is the insufficient

consideration of *Contextual Factors* in study designs. Context is crucial to the study of *MAR* for several reasons: First, a game's context carries a strong potential to affect the behavior, performance, and success of football teams or players. Second, accounting for *Contextual Factors* in *MAR* studies might alter the study design due to changes in the *level of observation* and data aggregation strategies.

As has been discussed in Section 2.4 and demonstrated in study 3 (Section 3.3), a *Contextual Factor* influencing both the *Performance Indicator* and the *Success Indicator* in an *MAR* study has the potential to introduce a confounding effect into the association between *Performance Indicator* and *Success Indicator*, thereby undermining the robustness of study results.

Controversial results in previous *MAR* studies can be viewed in the light of these mechanisms, potentially explaining the controversy as being the result of study design choices.

Consequently, it is mandatory to investigate these interrelationships and to assess the risk of confounder bias in *MAR* studies. In the present thesis, the framework of causal inference has been introduced as a valid theoretical background for conducting these assessments, providing a terminology as well as statistical methods to account for the causal structures inherent in *MAR* problems. Specifically, for the common statistical model of *MAR* (see Equation 2), *collapsibility* analysis has been introduced in study three (Section 3.3) as a way to examine the effect of *Contextual Factors* and the presence of confounding.

These main findings and contributions pose several important implications for both the practice of match analysis and research on the associations between performance and success. An exploration of these implications is presented in the following.

4.3 Practical Implications

The implications arising from the main findings in this thesis affect the scientific as well as the applied approach to match analysis in football. While the former has to acknowledge the causal structure of an *MAR* problem and take it into account in terms of study design, consequences for the latter mainly come in the form of how *Performance Indicators* are generally reported and evaluated in practice.

4.3.1 Implications for Research

In the context of research, the most important findings presented in this thesis are (1) the strong effect that *Contextual Factors* have on both performance and success,

(2) the potential for confounder bias that arises from this effect of *Contextual Factors*, and (3) the observation that different data aggregation strategies may produce different results.

An important implication of these findings is the need to adjust study design and data aggregation techniques accordingly. Specifically, when *Contextual Factors* shall be included which are normally not constant over the course of a game (e.g., scoreline), the classical match-level analysis is not the appropriate design choice. Instead, a *segmented match analysis paradigm* is needed, ensuring that the respective contextual variable is constant within each observation of the sample and can, therefore, be included in the analysis.

While the thesis has presented collapsibility analysis as a method that can be used to investigate the need for segmentation (Section 3.3), it is valid to suggest that a more granular analysis will mostly be the better choice for study design, due to the inherent differences in the aims and behaviors of teams when leading versus trailing, or when attacking versus defending. Robust aggregation levels might be segments of constant scoreline (Klemp, Memmert, and Rein 2021; Lago-Peñas 2009; Taylor et al. 2008) or possession phases (Bassek et al. 2023; Pollard and Reep 1997; Fernández, Bornn, and Cervone 2019). Studies on the event or action level provide the highest granularity and contextualization (Brinkschulte et al. 2021; Decroos et al. 2019), but might be limited to certain types of actions, such as penalties (Brinkschulte et al. 2023) or shots (Mead, O’Hare, and McMenemy 2023), and might, therefore, follow a different approach than the classical *MAR* paradigm, trying to capture the tactical dynamics of team behavior and their relation to success.

This need for segmented match analysis is an important consequence of the findings of this thesis and can be regarded as one of the main practical implications that can be derived. Another implication is the need to view any *MAR* problem in the light of its causal structure, even if this structure cannot be directly derived from the data. It has been demonstrated that the same data can be attributed to different causal semantics, resulting in varying needs for covariate control (Hernán, Clayton, and Keiding 2011). Utilizing a priori domain knowledge to recognize the causal structure of an *MAR* problem is essential and can prevent mistakes of circularity, trying to predict the past from the future. Moreover, *MAR* researchers should transparently communicate their assumptions about the causal structure of their analyses, promoting greater methodological transparency and enabling more accurate assessments of study validity.

A last important implication is the finding that machine learning techniques, despite showing strong predictive performance and posing few constraints on the data in terms of statistical assumptions, are not per se immune to biases. Ignoring the causal

structure of the problem or the effect of crucial contextual variables can cause the same biases in machine learning approaches as in “traditional” statistical modeling (Davis and Robberechts 2024).

4.3.2 Implications for Sports Practice

For the case of applied match analysis in the realm of sports practice, the insights provided by this thesis offer practical guidance for interpreting performance data. Awareness of *Contextual Factors* and their potential biases allows practitioners to critically assess performance indicators in training and match contexts. Specifically, this means that any *Performance Indicator* that is used to evaluate a player or a team should be viewed in its context and potentially be adjusted.

An important example of this necessity is the evaluation of defensive actions. It is an almost trivial insight that the number of defensive actions a team performs is negatively correlated to the team’s ball possession rate, since a team can only perform defensive actions when not in possession (Phatak, Rein, and Memmert 2021). However, this might lead to a spurious negative correlation between the number of defensive actions and success because teams that have high ball possession have more opportunities to create shots and goals while having fewer opportunities to perform defensive actions. The severity of this phenomenon has been demonstrated by Phatak and colleagues (2022), illustrating that the number of defensive actions is negatively correlated to success when not taking into account ball possession rates, but positively correlated to success when adjusting for ball possession. Consequently, practitioners should refrain from looking at aggregate statistics like the number of tackles in order to compare players or teams, without taking into consideration which contextual aspects of the match could have contributed to these numbers. High ball possession rates might —as outlined in detail in Section 2.2.1 — be a result of team strength or scoreline. Therefore, a high number of tackles, when not adjusting for ball possession, might be a second-order consequence of one of these two *Contextual Factors*. Practitioners seeking to evaluate the defensive performance of their team or players should aim to get their hands on metrics of contextualized defensive performance¹.

¹The claim that the pure count of defensive actions is a valid indicator of defensive performance is highly debatable and most likely not true. Good defending based on closing up spaces and covering opponents can not be reflected in an action-based assessment of performance and any tackle or interception might be viewed as being rather a reaction than an action. However, the discussion of approaches to quantify defensive performance is beyond the scope of this thesis. For the purpose of the present discussion, the number of defensive actions serves as a vivid example and is appropriate because the effects discussed in the main text have been demonstrated empirically. Still, caution should be exercised when evaluating the number of defensive actions as a *Performance Indicator*.

The example of defensive actions is backed by empirical evidence and constitutes a vivid illustration of the problem. However, the same mechanism applies for any *Performance Indicator* that is subject to contextual influences, be it the number of offensive actions, running performance, or ball possession itself.

In summary, the findings of this thesis imply the need for a rethink in how data-driven match analysis is approached in both academic research and sports practice. In general, the importance of a match's context cannot be overstated and should, in one way or another, be incorporated into scientific studies as well as practical evaluations. A methodology that might help in accounting for these contextual effects is the *segmented match analysis paradigm*, which advocates breaking down a football match into smaller functional segments based on context. Acknowledging the effects of context and segmentation is key to securing the validity of findings as well as the theoretical foundation of match analysis.

4.4 Contributions

This dissertation makes several contributions to the field of *Match Analysis Research (MAR)*, both from a theoretical and practical perspective. These contributions are outlined below, referencing the objectives of the thesis formulated in Section 1.2.

First, by formally describing the *MAR* problem in conceptual as well as statistical terms, a common ground was created to discuss issues and controversies in the field, enabling the search for their origins.

The conceptual description of the *MAR* problem was conducted by first describing the history and origins of MAR, specifically highlighting how the use of Performance Indicators was inspired by the use of deterministic models and motivated by the strong randomness inherent in the game. Further, the main objective of *MAR* studies was defined, resulting in the conceptual formulation of the “standard” *MAR* study design. This was then translated into statistical terms, deriving the formalization of the statistical approach to the *MAR* problem. A review of the literature was performed to outline the various approaches to the problem from the perspectives of study design and statistical modeling.

This overview of the literature was secondly utilized to demonstrate the role of contextual variables in *MAR*. The thesis systematically illustrates how contextual factors influence both *Performance Indicators* and *Success Indicators*. While previous studies have acknowledged the significance of these variables, this dissertation provides a comprehensive overview of their effects and highlights the risks of drawing misleading conclusions when they are neglected.

Third, by framing *MAR* as an observational research challenge with inherent causal complexities, a potential explanation for previous contradictory results in the literature is proposed. This is achieved by utilizing the framework of causal inference and introducing the notion of confounding to explain how bias may arise due to the misrepresentation of context in *MAR* studies. The mechanism by which a confounding variable might introduce a bias in the main association of interest is described against the background of causal inference and known phenomena such as Simpson's paradox.

Fourth, this explanation is further developed to understand the interdependencies between contextual variables, study design, and the risk of bias arising through differences in data aggregation and the failure to control for confounding contextual factors. Specifically, the level of observation is identified as a crucial parameter of study design that needs attention to avoid confounder bias.

Lastly, these insights regarding the emergence of bias in *MAR* studies are extended to propose a set of methods that enable the assessment of the risk of confounding. Furthermore, methods to investigate whether a certain *Contextual Factor* should be accounted for are provided. These methods include collapsibility analysis in the well-known regression setting normally present in *MAR*, as well as data aggregation techniques that help to avoid noncollapsibility.

In this regard, on a practical level, this dissertation introduces the *segmented match analysis paradigm*, a methodological shift that advocates for the segmentation of match data into contextually meaningful units rather than treating matches as homogeneous entities. This is different from the standard approach to *MAR*, where the match level is commonly utilized as the level of observation.

This segmentation approach enables a more precise analysis of player and team behavior under different match conditions. Empirical evidence presented in this thesis underscores the necessity of segmentation, showing that aggregated analyses failing to consider contextual segmentation may obscure critical patterns and lead to erroneous conclusions.

By addressing these theoretical and practical challenges, this dissertation significantly advances the methodological rigor of *MAR* and contributes to the ongoing evolution of football analytics. The insights provided herein not only clarify existing limitations but also offer concrete solutions that can be applied in future research and practice.

4.5 Limitations and Scope

While this dissertation makes significant contributions to the field of *Match Analysis Research (MAR)*, it is important to acknowledge its limitations. As with any research endeavor, certain methodological constraints and conceptual boundaries define the scope of the findings presented.

One primary limitation concerns the granularity of the analyses conducted. Although the *segmented match analysis paradigm* introduced in this dissertation improves upon traditional methods by incorporating more refined levels of observation, further granularity could always be pursued. More detailed segmentation strategies, potentially incorporating real-time tactical adjustments or individual player-level contextual factors, could provide even more precise insights into the mechanisms underlying team performance and behavior. However, increasing granularity comes with trade-offs, particularly in terms of computational feasibility, data availability, and interpretability.

Another limitation relates to the mechanisms driving the observed effects of contextual information. This research highlights the influence of contextual variables, such as scoreline, on various performance and success indicators. However, the present thesis deliberately ignores the underlying mechanisms driving these effects for reasons of scope and feasibility. While the reasons for the effects of context on behavior are highly relevant, the focus of the present dissertation is on the mathematical nature of these relationships and their statistical implications. An in-depth analysis of these associations is not only beyond the scope of this thesis but also likely not feasible given the type of data utilized in the research forming the core of this work.

Future research could complement the present findings by integrating experimental methods to explore these underlying mechanisms in greater depth.

Additionally, while this dissertation emphasizes the importance of controlling for confounding in MAR, there are inherent limitations in fully accounting for all possible confounders. Despite methodological advances in causal inference and statistical modeling, observational research remains subject to the risk of unmeasured confounding.

This is discussed in the causal inference literature as the problem of *sufficient control* (Greenland and Morgenstern 2001). The absence of randomization in MAR means that causal interpretations should always be made with caution, acknowledging the potential for hidden biases that may still influence the observed relationships.

Finally, this dissertation focuses primarily on methodological advancements within MAR rather than on practical implementation in applied football settings. While

the insights gained from this research have direct implications for analysts, coaches, and decision-makers in football, the translation of these findings into real-world applications remains an area for further exploration. Future work should seek to bridge the gap between research and practice by developing tools and frameworks that facilitate the integration of advanced MAR methodologies into professional football environments.

Despite these limitations, the contributions of this dissertation represent a substantial step forward in enhancing the methodological rigor of MAR. By identifying key challenges and proposing innovative solutions, this research lays the groundwork for future studies that can build upon its findings, further refining the theoretical and practical understanding of football analytics.

4.6 Future Directions

The present thesis has generated insights into the interplay between context, performance, and success in *Match Analysis Research* in football. Furthermore, it has revealed methodological shortcomings in the current approach to the problem and proposed potential remedies for some of them. While these insights and proposals might help to further develop the research field, future advancements with respect to the research are possible and should be the aim of research projects to follow.

While the necessity of adopting a *segmented match analysis paradigm* has been worked out in detail in this thesis, it has to be stated that applying this kind of segmentation to the data commonly collected in match analysis brings its own unique challenges. The *Contextual Factors* that have been shown to be so crucial to the study of MAR are not always provided in the standard data formats that researchers are required to work with. While the contextual information on scoreline is comparably easy to obtain through knowledge of the timing of goals, ball possession phases, as well as their respective tactical context, must normally be annotated by human analysts. This approach requires an enormous amount of labor and, therefore, drastically reduces the available sample sizes. Automated contextualization or segmentation of matches would be a ground-breaking innovation to boost the feasibility of such studies. Recently, AI-based approaches utilizing graph representations of position data (Raabe, Nabben, and Memmert 2022) have successfully been used to automatically segment and label ball possession phases in handball (Bassek et al. 2023). An extension of this approach to the use case of football, where the detection of ball possession and tactical context might be slightly more difficult, would greatly develop the field.

Another logical extension of the findings presented in this thesis is a more formal description of the MAR problem in terms of the causal inference framework. So

far, the conceptual analogy between the *MAR* paradigm and the general problem of causal inference in observational studies has been drawn. Furthermore, the statistical methodology of collapsibility analysis has been introduced from causal inference to statistically model the potential for confounding through covariates. However, the biggest strength of the causal inference framework for establishing robust associations between variables from non-experimental studies comes through the careful building of causal diagrams, describing the interplay between a multitude of factors. In the traditional statistical modeling paradigm, described in Section 2.3, it is only possible to model the associations between a set of inputs and a set of outputs (and mostly, even, only one output). While it is possible to distinguish the contributions of the various input or predictor variables, hierarchical or circular relationships among the predictor variables are generally ignored for the sake of modeling assumptions. In causal inference, one normally establishes a causal diagram, which is essentially a graph, allowing bivariate relations between all variables in the graph (Pearl 1995). These relations can then also be modeled, respectively, creating essentially a structural equation model (Pearl 2021).

Adopting this modeling approach for the *MAR* problem would provide even greater insights into the interplay between *Performance Indicators*, *Contextual Factors*, and *Success Indicators* and may shed further light on various kinds of bias that might be present in the research field (also moving beyond confounder bias). It could, for example, become possible to disentangle the seemingly complex interrelations between team strength, team performance, and team success (Klemp, Wunderlich, and Memmert 2021). This application of causal diagrams seems to be a promising venue for further research in *MAR*, further developing the research field’s capabilities to learn about the dynamics of the game.

References

- Aldrich, John. 1995. "Correlations Genuine and Spurious in Pearson and Yule." *Statistical Science* 10 (4): 364–76. <https://doi.org/10.1214/ss/1177009870>.
- Allen, Mark S., and Marc V. Jones. 2014. "The "Home Advantage" in Athletic Competitions." *Current Directions in Psychological Science* 23 (1): 48–53. <https://doi.org/10.1177/0963721413513267>.
- Almeida, Carlos Humberto, António Paulo Ferreira, and Anna Volossovitch. 2014. "Effects of Match Location, Match Status and Quality of Opposition on Regaining Possession in UEFA Champions League." *Journal of Human Kinetics* 41 (1): 203–14. <https://doi.org/10.2478/hukin-2014-0048>.
- Anzer, Gabriel, and Pascal Bauer. 2021. "A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer)." Edited by Arno Knobbe José Luis Felipe. *Frontiers in Sports and Active Living* 3 (March). <https://doi.org/10.3389/fspor.2021.624475>.
- . 2022. "Expected Passes." *Data Mining and Knowledge Discovery* 36 (1): 295–317. <https://doi.org/10.1007/s10618-021-00810-3>.
- Apor, Peter. 1988. "Successful Formulae for Fitness Training." In *Science and Football*, edited by Thomas Reilly, 95–107. Routledge.
- Assis, André, Jamilson Dantas, and Ermeson Andrade. 2024. "The Performance-Interpretability Trade-Off: A Comparative Study of Machine Learning Models." *Journal of Reliable Intelligent Environments* 11 (1). <https://doi.org/10.1007/s40860-024-00240-0>.
- Åstrand, Per-Olof, and Bengt Saltin. 1961a. "Maximal Oxygen Uptake and Heart Rate in Various Types of Muscular Activity." *Journal of Applied Physiology* 16 (6): 977–81. <https://doi.org/10.1152/jappl.1961.16.6.977>.
- . 1961b. "Oxygen Uptake During the First Minutes of Heavy Muscular Exercise." *Journal of Applied Physiology* 16 (6): 971–76. <https://doi.org/10.1152/jappl.1961.16.6.971>.
- Atkinson, Greg. 2002. "Sport Performance: Variable or Construct?" *Journal of Sports Sciences* 20 (4): 291–92. <https://doi.org/10.1080/026404102753576053>.
- Augusto, Diêgo, João Brito, Rodrigo Aquino, Pedro Figueiredo, Fabio Eiras, Márcio Tannure, Bruno Veiga, and Fabrício Vasconcellos. 2021. "Contextual Variables Affect Running Performance in Professional Soccer Players: A Brief Report." Edited by Clare Minahan Alliance Kubayi. *Frontiers in Sports and Active Living*

- 3 (December). <https://doi.org/10.3389/fspor.2021.778813>.
- Baker, Julien S., Marie Clare McCormick, and Robert A. Robergs. 2010. "Interaction Among Skeletal Muscle Metabolic Energy Systems During Intense Exercise." *Journal of Nutrition and Metabolism* 2010: 1–13. <https://doi.org/10.1155/2010/905612>.
- Bangsbo, Jens, F Marcello Iaia, and Peter Krstrup. 2008. "The Yo-Yo Intermittent Recovery Test." *Sports Medicine* 38 (1): 37–51. <https://doi.org/10.2165/00007256-200838010-00004>.
- Bangsbo, Jens, Magni Mohr, and Peter Krstrup. 2005. "Physical and Metabolic Demands of Training and Match-Play in the Elite Football Player." *Journal of Sports Sciences* 24 (7): 665–74. <https://doi.org/10.1080/02640410500482529>.
- Bangsbo, J, L Nørregaard, and F Thorsø. 1991. "Activity Profile of Competition Soccer." *Canadian Journal of Sport Sciences* 16 (2): 110—116. <http://europepmc.org/abstract/MED/1647856>.
- Bassek, Manuel, Dominik Raabe, Alexander Banning, Daniel Memmert, and Robert Rein. 2023. "Analysis of Contextualized Intensity in Men's Elite Handball Using Graph-Based Deep Learning." *Journal of Sports Sciences*, October, 1–10. <https://doi.org/10.1080/02640414.2023.2268366>.
- Bassett, David R. 2002. "Scientific Contributions of A. V. Hill: Exercise Physiology Pioneer." *Journal of Applied Physiology* 93 (5): 1567–82. <https://doi.org/10.1152/japplphysiol.01246.2001>.
- Bassett, David R., and E. T. Howley. 2000. "Limiting Factors for Maximum Oxygen Uptake and Determinants of Endurance Performance." *Medicine & Science in Sports & Exercise*, January, 70. <https://doi.org/https://doi.org/10.1097/00005768-200001000-00012>.
- Bergh, U., B. Sjödén, A. Forsberg, and J. Svedenhag. 1991. "The Relationship Between Body Mass and Oxygen Uptake During Running in Humans." *Medicine & Science in Sport & Exercise* 23 (2): 205–11.
- Bialkowski, Alina, P. Lucey, Peter Carr, Yisong Yue, and I. Matthews. 2014. "'Win at Home and Draw Away': Automatic Formation Analysis Highlighting the Differences in Home and Away Team Behaviors." In *Proceedings of the 8th MIT Sloan Sports Analytics Conference*, 1–7. Citeseer.
- Biermann, Henrik, Jonas Theiner, Manuel Bassek, Dominik Raabe, Daniel Memmert, and Ralph Ewerth. 2021. "A Unified Taxonomy and Multimodal Dataset for Events in Invasion Games." In *Proceedings of the 4th International Workshop on Multimedia Content Analysis in Sports*, 1–10. MM '21. ACM. <https://doi.org/10.1145/3475722.3482792>.
- Biermann, Henrik, Franz-Georg Wieland, Jens Timmer, Daniel Memmert, and Ashwin A. Phatak. 2023. "Towards Expected Counter - Using Comprehensible

- Features to Predict Counterattacks.” In *Machine Learning and Data Mining for Sports Analytics*, 3–13. Springer Nature Switzerland.
https://doi.org/10.1007/978-3-031-27527-2_1.
- Blauberger, Patrick, Robert Marzilger, and Martin Lames. 2021. “Validation of Player and Ball Tracking with a Local Positioning System.” Edited by Elena Simona Lohan. *Sensors* 21 (4): 1465. <https://doi.org/10.3390/s21041465>.
- Bodine, S C, R R Roy, D A Meadows, R F Zernicke, R D Sacks, M Fournier, and V R Edgerton. 1982. “Architectural, Histochemical, and Contractile Characteristics of a Unique Biarticular Muscle: The Cat Semitendinosus.” *Journal of Neurophysiology* 48 (1): 192–201.
<https://doi.org/10.1152/jn.1982.48.1.192>.
- Bosco, Carmelo, and Paavo V. Komi. 1979. “Potentiation of the Mechanical Behavior of the Human Skeletal Muscle Through Prestretching.” *Acta Physiologica Scandinavica* 106 (4): 467–72.
<https://doi.org/10.1111/j.1748-1716.1979.tb06427.x>.
- Bradley, Paul S. 2020. *Football Decoded: Using Match Analysis & Context to Interpret the Demands*. [Erscheinungsort nicht ermittelbar]: Paul Bradley.
- Bradley, Paul Simon, and Jack D. Ade. 2018. “Are Current Physical Match Performance Metrics in Elite Soccer Fit for Purpose or Is the Adoption of an Integrated Approach Needed?” *International Journal of Sports Physiology and Performance* 13 (5): 656–64. <https://doi.org/10.1123/ijsp.2017-0433>.
- Bradley, Paul Simon, David T. Archer, Bob Hogg, Gabor Schuth, Michael Bush, Chris Carling, and Chris Barnes. 2015. “Tier-Specific Evolution of Match Performance Characteristics in the English Premier League: It’s Getting Tougher at the Top.” *Journal of Sports Sciences* 34 (10): 980–87.
<https://doi.org/10.1080/02640414.2015.1082614>.
- Bradley, Paul Simon, Chris Carling, Dave Archer, Jenny Roberts, Andrew Dodds, Michele Di Mascio, Darren J. Paul, Antonio Gomez Diaz, Dan Peart, and Peter Krstrup. 2013. “The Effect of Playing Formation on High-Intensity Running and Technical Profiles in English FA Premier League Soccer Matches.” *Journal of Sports Sciences* 29 (8): 821–30. <https://doi.org/10.1080/02640414.2011.561868>.
- Bradley, Paul Simon, Christopher Carling, Antonio Gomez Diaz, Peter Hood, Chris Barnes, Jack Ade, Mark Boddy, Peter Krstrup, and Magni Mohr. 2013. “Match Performance and Physical Capacity of Players in the Top Three Competitive Standards of English Professional Soccer.” *Human Movement Science* 32 (4): 808–21. <https://doi.org/10.1016/j.humov.2013.06.002>.
- Bradley, Paul Simon, Carlos Lago-Peñas, Ezequiel Rey, and Jaime Sampaio. 2014. “The Influence of Situational Variables on Ball Possession in the English Premier League.” *Journal of Sports Sciences* 32 (20): 1867–73.

- <https://doi.org/10.1080/02640414.2014.887850>.
- Bradley, Paul Simon, Michele Di Mascio, Dan Peart, Peter Olsen, and Bill Sheldon. 2010. “High-Intensity Activity Profiles of Elite Soccer Players at Different Performance Levels.” *Journal of Strength and Conditioning Research* 24 (9): 2343–51. <https://doi.org/10.1519/jsc.0b013e3181aeb1b3>.
- Bradley, Paul Simon, and Timothy D. Noakes. 2013. “Match Running Performance Fluctuations in Elite Soccer: Indicative of Fatigue, Pacing or Situational Influences?” *Journal of Sports Sciences* 31 (15): 1627–38. <https://doi.org/10.1080/02640414.2013.796062>.
- Bradley, Paul Simon, William Sheldon, Blake Wooster, Peter Olsen, Paul Boanas, and Peter Krstrup. 2008. “High-Intensity Running in English FA Premier League Soccer Matches.” *Journal of Sports Sciences* 27 (2): 159–68. <https://doi.org/10.1080/02640410802512775>.
- Breiman, Leo. 2001. “Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author).” *Statistical Science* 16 (3). <https://doi.org/10.1214/ss/1009213726>.
- Brinkschulte, Michel, Philip Furley, Maximilian Klemp, and Daniel Memmert. 2021. “English Goalkeepers Are Not Responsible for England’s Poor Performance in Penalty Shootouts in the Past.” *Scientific Reports* 11 (1). <https://doi.org/10.1038/s41598-021-04118-6>.
- Brinkschulte, Michel, Fabian Wunderlich, Philip Furley, and Daniel Memmert. 2023. “The Obligation to Succeed When It Matters the Most—the Influence of Skill and Pressure on the Success in Football Penalty Kicks.” *Psychology of Sport and Exercise* 65 (March): 102369. <https://doi.org/10.1016/j.psychsport.2022.102369>.
- Brooks, George A. 1985. “Anaerobic Threshold: Review of the Concept and Directions for Future Research.” *Medicine & Science in Sports & Exercise* 17 (1): 22–34. <https://doi.org/10.1249/00005768-198502000-00005>.
- Buchheit, Martin, Adam Allen, Tsz Kit Poon, Mattia Modonutti, Warren Gregson, and Valter Di Salvo. 2014. “Integrating Different Tracking Systems in Football: Multiple Camera Semi-Automatic System, Local Position Measurement and GPS Technologies.” *Journal of Sports Sciences* 32 (20): 1844–57. <https://doi.org/10.1080/02640414.2014.942687>.
- Bunker, Rory P., and Teo Sunsjak. 2019. “The Application of Machine Learning Techniques for Predicting Results in Team Sport: A Review.” *Journal of Artificial Intelligence Research* 73 (December): 1285–1322. <https://doi.org/10.31236/osf.io/wq542>.
- Bunker, Rory P., Calvin Yeung, and Keisuke Fujii. 2023. “Machine Learning for Soccer Match Result Prediction.”
- Caldbeck, Paul, and Thomas Dos’Santos. 2022. “How Do Soccer Players Sprint

- from a Tactical Context? Observations of an English Premier League Soccer Team.” *Journal of Sports Sciences* 40 (23): 2669–80.
<https://doi.org/10.1080/02640414.2023.2183605>.
- Castellano, Julen, A. Blanco-Villaseñor, and D. Álvarez. 2011. “Contextual Variables and Time-Motion Analysis in Soccer.” *International Journal of Sports Medicine* 32 (06): 415–21. <https://doi.org/10.1055/s-0031-1271771>.
- Cavus, Mustafa, and Przemyslaw Biecek. 2022. “Explainable Expected Goal Models for Performance Analysis in Football Analytics.” In *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*. przemyslaw.biecek@pw.edu.pl. IEEE.
<https://doi.org/10.1109/dsaa54385.2022.10032440>.
- Chow, John W., and Duane V. Knudson. 2011. “Use of Deterministic Models in Sports and Exercise Biomechanics Research.” *Sports Biomechanics* 10 (3): 219–33. <https://doi.org/10.1080/14763141.2011.592212>.
- Christodoulou, Evangelia, Jie Ma, Gary S. Collins, Ewout W. Steyerberg, Jan Y. Verbakel, and Ben Van Calster. 2019. “A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models.” *Journal of Clinical Epidemiology* 110 (June): 12–22.
<https://doi.org/10.1016/j.jclinepi.2019.02.004>.
- Chu, Singfat. 2003. “Using Soccer Goals to Motivate the Poisson Process.” *INFORMS Transactions on Education* 3 (2): 64–70.
<https://doi.org/10.1287/ited.3.2.64>.
- Collet, Christian. 2012. “The Possession Game? A Comparative Analysis of Ball Retention and Team Success in European and International Football, 2007–2010.” *Journal of Sports Sciences* 31 (2): 123–36.
<https://doi.org/10.1080/02640414.2012.727455>.
- Cometti, G., N. A. Maffiuletti, M. Pousson, J.-C. Chatard, and N. Maffulli. 2001. “Isokinetic Strength and Anaerobic Power of Elite, Subelite and Amateur French Soccer Players.” *International Journal of Sports Medicine* 22 (1): 45–51.
<https://doi.org/10.1055/s-2001-11331>.
- Conley, DL, and GS Krahenbuhl. 1980. “Running Economy and Distance Running Performance of Highly Trained Athletes.” *Medicine and Science in Sports and Exercise* 12 (5): 357–360. <http://europepmc.org/abstract/MED/7453514>.
- Cormie, Prue, Michael R. McGuigan, and Robert U. Newton. 2011. “Developing Maximal Neuromuscular Power: Part 1 – Biological Basis of Maximal Power Production.” *Sports Medicine* 41 (1): 17–38.
<https://doi.org/10.2165/11537690-000000000-00000>.
- Coyle, E. F., A. R. Coggan, M. K. Hopper, and T. J. Walters. 1988. “Determinants of Endurance in Well-Trained Cyclists.” *Journal of Applied Physiology* 64 (6):

- 2622–30. <https://doi.org/10.1152/jappl.1988.64.6.2622>.
- Das, Arun, and Paul Rad. 2020. “Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey.” *arXiv*.
<https://doi.org/10.48550/ARXIV.2006.11371>.
- Davis, Jesse, and Pieter Robberechts. 2024. “Biases in Expected Goals Models Confound Finishing Ability.” *arXiv*.
<https://doi.org/10.48550/ARXIV.2401.09940>.
- Decroos, Tom, Lotte Bransen, Jan Van Haaren, and Jesse Davis. 2019. “Actions Speak Louder Than Goals: Valuing Player Actions in Soccer.” In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD ’19. ACM. <https://doi.org/10.1145/3292500.3330758>.
- Decroos, Tom, and Jesse Davis. 2020. “Interpretable Prediction of Goals in Soccer.” In *Proceedings of the AAAI-20 Workshop on Artificial Intelligence in Team Sports*.
- Di Salvo, Valter, Warren Gregson, Greg Atkinson, Phil Tordoff, and Barry Drust. 2009. “Analysis of High Intensity Activity in Premier League Soccer.” *International Journal of Sports Medicine* 30 (03): 205–12.
<https://doi.org/10.1055/s-0028-1105950>.
- Dick, Uwe, Maryam Tavakol, and Ulf Brefeld. 2021. “Rating Player Actions in Soccer.” *Frontiers in Sports and Active Living* 3 (July).
<https://doi.org/10.3389/fspor.2021.682986>.
- Dixon, Mark J, and Stuart G Coles. 1997. “Modelling Association Football Scores and Inefficiencies in the Football Betting Market.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 46 (2): 265–80.
<https://doi.org/10.1111/1467-9876.00065>.
- Dixon, Mark J., and Peter F. Pope. 2004. “The Value of Statistical Forecasts in the UK Association Football Betting Market.” *International Journal of Forecasting* 20 (4): 697–711. <https://doi.org/10.1016/j.ijforecast.2003.12.007>.
- Eetvelde, Hans Van, Luciana D. Mendonça, Christophe Ley, Romain Seil, and Thomas Tischer. 2021. “Machine Learning Methods in Sport Injury Prediction and Prevention: A Systematic Review.” *Journal of Experimental Orthopaedics* 8 (1). <https://doi.org/10.1186/s40634-021-00346-x>.
- Eggels, Harm, Ruud Van Elk, and Mykola Pechenizkiy. 2016. “Explaining Soccer Match Outcomes with Goal Scoring Opportunities Predictive Analytics.” In *Mlsa@ Pkdd/Ecml*. <https://www.semanticscholar.org/paper/549b3dc2b9601dbde0807c35d629b6c3ddb40489>.
- Enoka, Roger M. 1995. “Morphological Features and Activation Patterns of Motor Units.” *Journal of Clinical Neurophysiology* 12 (6): 538–59.
<https://doi.org/10.1097/00004691-199511000-00002>.

- Eusebio, Pedro, Pablo Prieto-González, and Rui Marcelino. 2024. “Decoding the Complexities of Transitions in Football: A Comprehensive Narrative Review.” *German Journal of Exercise and Sport Research*, April.
<https://doi.org/10.1007/s12662-024-00951-9>.
- Faraway, Julian J. 2016. *Extending the Linear Model with r: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman; Hall/CRC.
- Farrell, Peter A., Jack H. Wilmore, Edward F. Coyle, John E. Billing, and David L. Costill. 1993. “Plasma Lactate Accumulation and Distance Running Performance.” *Medicine & Science in Sports & Exercise* 25 (10): 1091–97.
<https://doi.org/10.1249/00005768-199310000-00002>.
- Fernandez, Javier, and Luke Bornn. 2018. “Wide Open Spaces: A Statistical Technique for Measuring Space Creation in Professional Soccer.” In *Sloan Sports Analytics Conference*. Vol. 2018. <https://www.semanticscholar.org/paper/a9f6975e93f8c86dad9249750712305511d5b69>.
- Fernández, Javier, Luke Bornn, and Dan Cervone. 2019. “Decomposing the Immeasurable Sport: A Deep Learning Expected Possession Value Framework for Soccer.” In *Proceedings of the 13th MIT Sloan Sports Analytics Conference*. <https://www.semanticscholar.org/paper/fc78b144a531a8ffdf3216a677f3a65e70dad3c7>.
- Fernandez-Navarro, Javier, Luis Fradua, Asier Zubillaga, Paul R. Ford, and Allistair P. McRobert. 2016. “Attacking and Defensive Styles of Play in Soccer: Analysis of Spanish and English Elite Teams.” *Journal of Sports Sciences* 34 (24): 2195–2204. <https://doi.org/10.1080/02640414.2016.1169309>.
- Fernandez-Navarro, Javier, Luis Fradua, Asier Zubillaga, and Allistair P. McRobert. 2019. “Evaluating the Effectiveness of Styles of Play in Elite Soccer.” *International Journal of Sports Science & Coaching* 14 (4): 514–27.
<https://doi.org/10.1177/1747954119855361>.
- Fernandez-Navarro, Javier, Luis Fradua, Asier Zubillaga, and Allistair P. McRobert. 2018. “Influence of Contextual Variables on Styles of Play in Soccer.” *International Journal of Performance Analysis in Sport* 18 (3): 423–36.
<https://doi.org/10.1080/24748668.2018.1479925>.
- Fisher, Ronald A. 1935. *The Design of Experiments*. Macmillan Pub Co.
- Fisher, Ronald Aymeric. 1922. “On the Mathematical Foundations of Theoretical Statistics.” *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222 (594–604): 309–68. <https://doi.org/10.1098/rsta.1922.0009>.
- Flint, Austin. 1878. *On the Sources of Muscular Power, Arguments and Conclusions Drawn from Observations Upon the Human Subject, Under Conditions of Rest and of Muscular Exercise*. New York: D. Appleton; Company.

- Foster, Carl. 1983. "VO2 Max and Training Indices as Determinants of Competitive Running Performance." *Journal of Sports Sciences* 1 (1): 13–22.
<https://doi.org/10.1080/02640418308729657>.
- Franck, Egon, and Stephan Nüesch. 2010. "Talent and/or Popularity: What Does It Take to Be a Superstar?" *Economic Inquiry* 50 (1): 202–16.
<https://doi.org/10.1111/j.1465-7295.2010.00360.x>.
- Franks, Ian M., and David Goodman. 1986. "A Systematic Approach to Analysing Sports Performance." *Journal of Sports Sciences* 4 (1): 49–59.
<https://doi.org/10.1080/02640418608732098>.
- Furley, Philip, Geoffrey Schweizer, and Daniel Memmert. 2018. "Thin Slices of Athletes' Nonverbal Behavior Give Away Game Location: Testing the Territoriality Hypothesis of the Home Game Advantage." *Evolutionary Psychology* 16 (2): 147470491877645.
<https://doi.org/10.1177/1474704918776456>.
- Galton, Francis. 1889. *Natural Inheritance*. Macmillan.
- Garber, Carol Ewing, Bryan Blissmer, Michael R. Deschenes, Barry A. Franklin, Michael J. Lamonte, I-Min Lee, David C. Nieman, and David P. Swain. 2011. "Quantity and Quality of Exercise for Developing and Maintaining Cardiorespiratory, Musculoskeletal, and Neuromotor Fitness in Apparently Healthy Adults: Guidance for Prescribing Exercise." *Medicine & Science in Sports & Exercise* 43 (7): 1334–59.
<https://doi.org/10.1249/mss.0b013e318213febf>.
- García-Unanue, Jorge, Jorge Pérez-Gómez, Jesús-Vicente Giménez, José Luis Felipe, Santiago Gómez-Pomares, Leonor Gallardo, and Javier Sánchez-Sánchez. 2018. "Influence of Contextual Variables and the Pressure to Keep Category on Physical Match Performance in Soccer Players." Edited by Slavko Rogan. *PLOS ONE* 13 (9): e0204256. <https://doi.org/10.1371/journal.pone.0204256>.
- Garnica Caparrós, Marc. 2024. "Process Mining." In *Computer Science in Sport*, 149–55. Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-662-68313-2_18.
- Garnica-Caparrós, Marc. 2021. "KPIs on the Basis of Match Events Data." In *Match Analysis*, 159–67. Routledge.
- . 2024. "Event Data." In *Computer Science in Sport*, 35–41. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-68313-2_5.
- Glass, Alana. 2019. "FIFA Women's World Cup Breaks Viewership Records." <https://www.forbes.com/sites/alanaglass/2019/10/21/fifa-womens-world-cup-breaks-viewership-records/>.
<https://www.forbes.com/sites/alanaglass/2019/10/21/fifa-womens-world-cup-breaks-viewership-records/>.

- Goes, Floris R, Matthias Kempe, and Koen APM Lemmink. 2019. "Predicting Match Outcome in Professional Dutch Football Using Tactical Performance Metrics Computed from Position Tracking Data." *EasyChair Preprints*, May, 993. <https://doi.org/10.29007/4jjb>.
- Goes, Floris R., L. A. Meerhoff, M. J. O. Bueno, D. M. Rodrigues, F. A. Moura, M. S. Brink, M. T. Elferink-Gemser, et al. 2020. "Unlocking the Potential of Big Data to Support Tactical Performance Analysis in Professional Soccer: A Systematic Review." *European Journal of Sport Science* 21 (4): 481–96. <https://doi.org/10.1080/17461391.2020.1747552>.
- Gollan, Stuart, Clint Bellenger, and Kevin Norton. 2020. "Contextual Factors Impact Styles of Play in the English Premier League." *Journal of Sports Science & Medicine* 19 (1): 78.
- Gonçalves, Luiz Guilherme, Ana Filipa Silva, Diego Augusto, Bruno Pasquarelli, Alejandro Pastor, Felipe de Okato Plato, Bruno L. S. Bedo, Fabrício Vasconcellos, and Rodrigo Aquino. 2024. "Attack, Defense, and Transitions in Soccer: Analyzing the Running Performance of Match-Play." *Sport Sciences for Health*, April. <https://doi.org/10.1007/s11332-024-01210-y>.
- Gonzalez-Rodenas, Joaquin, Michalis Mitrotasios, Rafael Aranda, and Vasilis Armatas. 2020. "Combined Effects of Tactical, Technical and Contextual Factors on Shooting Effectiveness in European Professional Soccer." *International Journal of Performance Analysis in Sport* 20 (2): 280–93. <https://doi.org/10.1080/24748668.2020.1743163>.
- Gordon, A. M., A. F. Huxley, and F. J. Julian. 1966. "The Variation in Isometric Tension with Sarcomere Length in Vertebrate Muscle Fibres." *The Journal of Physiology* 184 (1): 170–92. <https://doi.org/10.1113/jphysiol.1966.sp007909>.
- Gorostiaga, E. M., C. Granados, J. Ibáñez, and M. Izquierdo. 2005. "Differences in Physical Fitness and Throwing Velocity Among Elite and Amateur Male Handball Players." *International Journal of Sports Medicine* 26 (3): 225–32. <https://doi.org/10.1055/s-2004-820974>.
- Grant, S., I. Craig, J. Wilson, and T. Aitchison. 1997. "The Relationship Between 3 Km Running Performance and Selected Physiological Variables." *Journal of Sports Sciences* 15 (4): 403–10. <https://doi.org/10.1080/026404197367191>.
- Greenland, Sander, and Hal Morgenstern. 2001. "Confounding in Health Research." *Annual Review of Public Health* 22 (1): 189–212. <https://doi.org/10.1146/annurev.publhealth.22.1.189>.
- Greenland, Sander, Judea Pearl, and James M. Robins. 1999. "Confounding and Collapsibility in Causal Inference." *Statistical Science* 14 (1). <https://doi.org/10.1214/ss/1009211805>.
- Gregory, Sam, Sam Robertson, Robert Aughey, and Grant Duthie. 2022. "The

- Influence of Tactical and Match Context on Player Movement in Football.” *Journal of Sports Sciences* 40 (9): 1063–77.
<https://doi.org/10.1080/02640414.2022.2046938>.
- Gregory, Sam, Sam Robertson, Robert Aughey, Bartholomew Spencer, and Jeremy Alexander. 2024. “Assigning Goal-Probability Value to High Intensity Runs in Football.” Edited by Bruno Gonçalves. *PLOS ONE* 19 (9): e0308749.
<https://doi.org/10.1371/journal.pone.0308749>.
- Guerrero-Calderón, Berni, Maximilian Klemp, Alfonso Castillo-Rodriguez, José Alfonso Morcillo, and Daniel Memmert. 2020. “A New Approach for Training-Load Quantification in Elite-Level Soccer: Contextual Factors.” *International Journal of Sports Medicine* 42 (08): 716–23.
<https://doi.org/10.1055/a-1289-9059>.
- Guerrero-Calderón, Berni, Maximilian Klemp, José Alfonso Morcillo, and Daniel Memmert. 2021. “How Does the Workload Applied During the Training Week and the Contextual Factors Affect the Physical Responses of Professional Soccer Players in the Match?” *International Journal of Sports Science & Coaching* 16 (4): 994–1003. <https://doi.org/10.1177/1747954121995610>.
- Halouani, Jamel, Hamdi Chtourou, Tim Gabbett, Anis Chaouachi, and Karim Chamari. 2014. “Small-Sided Games in Team Sports Training: A Brief Review.” *The Journal of Strength & Conditioning Research* 28 (12): 3594–3618.
<https://doi.org/10.1519/jsc.0000000000000564>.
- Hawley, J A, M M Williams, M M Vickovic, and P J Handcock. 1992. “Muscle Power Predicts Freestyle Swimming Performance.” *British Journal of Sports Medicine* 26 (3): 151–55. <https://doi.org/10.1136/bjism.26.3.151>.
- Hay, James G., John A. Miller, and Ron W. Canterna. 1986. “The Techniques of Elite Male Long Jumpers.” *Journal of Biomechanics* 19 (10): 855–66.
[https://doi.org/10.1016/0021-9290\(86\)90136-3](https://doi.org/10.1016/0021-9290(86)90136-3).
- He, Miao, Ricardo Cachucho, and Arno J Knobbe. 2015. “Football Player’s Performance and Market Value.” In *Mlsa@ Pkdd/Ecml*, 87–95.
- Helgerud, Jan. 1994. “Maximal Oxygen Uptake, Anaerobic Threshold and Running Economy in Women and Men with Similar Performances Level in Marathons.” *European Journal of Applied Physiology and Occupational Physiology* 68 (2): 155–61. <https://doi.org/10.1007/bf00244029>.
- Helgerud, Jan, Lars Christian Engen, Ulrik Wisløff, and Jan Hoff. 2001. “Aerobic Endurance Training Improves Soccer Performance.” *Medicine and Science in Sports and Exercise* 33 (11): 1925–31.
<https://doi.org/10.1097/00005768-200111000-00019>.
- Henneman, Elwood, George Somjen, and David O. Carpenter. 1965. “Functional Significance of Cell Size in Spinal Motoneurons.” *Journal of Neurophysiology* 28

- (3): 560–80. <https://doi.org/10.1152/jn.1965.28.3.560>.
- Hennessy, Liam, and Ian Jeffreys. 2018. “The Current Use of GPS, Its Potential, and Limitations in Soccer.” *Strength & Conditioning Journal* 40 (3): 83–94. <https://doi.org/10.1519/ssc.0000000000000386>.
- Herm, Steffen, Hans-Markus Callsen-Bracker, and Henning Kreis. 2014. “When the Crowd Evaluates Soccer Players’ Market Values: Accuracy and Evaluation Attributes of an Online Community.” *Sport Management Review* 17 (4): 484–92. <https://doi.org/10.1016/j.smr.2013.12.006>.
- Hernan, Miguel A. 2023. *Causal Inference: What If*. Edited by James M Robins. First edition. Boca Raton: Taylor; Francis.
- Hernán, Miguel A. 2018. “The c-Word: Scientific Euphemisms Do Not Improve Causal Inference from Observational Data.” *American Journal of Public Health* 108 (5): 616–19. <https://doi.org/10.2105/ajph.2018.304337>.
- Hernán, Miguel A., David Clayton, and Niels Keiding. 2011. “The Simpson’s Paradox Unraveled.” *International Journal of Epidemiology* 40 (3): 780–85. <https://doi.org/10.1093/ije/dyr041>.
- Hernán, Miguel A., John Hsu, and Brian Healy. 2019. “A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks.” *CHANCE* 32 (1): 42–49. <https://doi.org/10.1080/09332480.2019.1579578>.
- Herold, Mat, Floris Goes, Stephan Nopp, Pascal Bauer, Chris Thompson, and Tim Meyer. 2019. “Machine Learning in Men’s Professional Football: Current Applications and Future Directions for Improving Attacking Play.” Edited by Ros Frederic. *International Journal of Sports Science & Coaching* 14 (6): 798–817. <https://doi.org/10.1177/1747954119879350>.
- Heuer, A., C. Müller, and O. Rubner. 2010. “Soccer: Is Scoring Goals a Predictable Poissonian Process?” *EPL (Europhysics Letters)* 89 (3): 38007. <https://doi.org/10.1209/0295-5075/89/38007>.
- Heuer, Andreas, and Oliver Rubner. 2014. “Optimizing the Prediction Process: From Statistical Concepts to the Case Study of Soccer.” Edited by Dominik Wodarz. *PLoS ONE* 9 (9): e104647. <https://doi.org/10.1371/journal.pone.0104647.g001>.
- Heuer, A., and O. Rubner. 2009. “Fitness, Chance, and Myths: An Objective View on Soccer Results.” *The European Physical Journal B* 67 (3): 445–58. <https://doi.org/10.1140/epjb/e2009-00024-8>.
- Hewitt, Adam, Grace Greenham, and Kevin Norton. 2016. “Game Style in Soccer: What Is It and Can We Quantify It?” *International Journal of Performance Analysis in Sport* 16 (1): 355–72. <https://doi.org/10.1080/24748668.2016.11868892>.
- Hill, A. V. 1922. “The Mechanism of Muscular Contraction.” *Physiological Reviews*

- 2 (2): 310–41. <https://doi.org/10.1152/physrev.1922.2.2.310>.
- Hill, A. V., and H. Lupton. 1923. “Muscular Exercise, Lactic Acid, and the Supply and Utilization of Oxygen.” *Quarterly Journal of Medicine* os-16 (62): 135–71. <https://doi.org/10.1093/qjmed/os-16.62.135>.
- Hill, Archibald Vivian. 1938. “The Heat of Shortening and the Dynamic Constants of Muscle.” *Proceedings of the Royal Society of London. Series B - Biological Sciences* 126 (843): 136–95. <https://doi.org/10.1098/rspb.1938.0050>.
- Hoff, Jan, and Jan Helgerud. 2004. “Endurance and Strength Training for Soccer Players: Physiological Considerations.” *Sports Medicine* 34 (3): 165–80. <https://doi.org/10.2165/00007256-200434030-00003>.
- Hoppe, Matthias W., M. Slomka, C. Baumgart, H. Weber, and J. Freiwald. 2014. “Match Running Performance and Success Across a Season in German Bundesliga Soccer Teams.” *International Journal of Sports Medicine* 36 (07): 563–66. <https://doi.org/10.1055/s-0034-1398578>.
- Horvat, Tomislav, and Josip Job. 2020. “The Use of Machine Learning in Sport Outcome Prediction: A Review.” *WIREs Data Mining and Knowledge Discovery* 10 (5). <https://doi.org/10.1002/widm.1380>.
- Hughes, Mike D. 1984. “Using a Microcomputer for Notational Analysis in Squash.” *Journal of Sports Sciences* 4: 189–90.
- . 1985. “A Comparison of the Patterns of Play of Squash.” *International Ergonomics* 85: 139–41.
- Hughes, Mike D., and Roger M. Bartlett. 2002. “The Use of Performance Indicators in Performance Analysis.” *Journal of Sports Sciences* 20 (10): 739–54. <https://doi.org/10.1080/026404102320675602>.
- Hughes, Mike D., and Ian M. Franks. 2004. “Notational Analysis—a Review of the Literature.” In *Notational Analysis of Sport*, 71–116. Routledge.
- Hughes, Mike D., and I. Sykes. 1994. “Computerised Notational Analysis of the Effects of the Law Changes in Soccer Upon Patterns of Play.” *Journal of Sports Sciences* 12: 180.
- Hume, David. 1748. *An Enquiry Concerning Human Understanding*. LaSalle:Open Court.
- Hurlbert, Stuart H. 1984. “Pseudoreplication and the Design of Ecological Field Experiments.” *Ecological Monographs* 54 (2): 187–211. <https://doi.org/10.2307/1942661>.
- Hvattum, Lars Magnus, and Halvard Arntzen. 2010. “Using ELO Ratings for Match Result Prediction in Association Football.” *International Journal of Forecasting* 26 (3): 460–70. <https://doi.org/10.1016/j.ijforecast.2009.10.002>.
- Ingebrigtsen, Jørgen, Mads Bendiksen, Morten Bredsgaard Randers, Carlo Castagna, Peter Krstrup, and Andreas Holtermann. 2012. “Yo-Yo IR2 Testing of Elite

- and Sub-Elite Soccer Players: Performance, Heart Rate Response and Correlations to Other Interval Tests.” *Journal of Sports Sciences* 30 (13): 1337–45. <https://doi.org/10.1080/02640414.2012.711484>.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning. Springer Texts in Statistics*. Vol. 112. Springer New York. <https://doi.org/10.1007/978-1-4614-7138-7>.
- Jamieson, Jeremy P. 2010. “The Home Field Advantage in Athletics: A Meta-Analysis.” *Journal of Applied Social Psychology* 40 (7): 1819–48. <https://doi.org/10.1111/j.1559-1816.2010.00641.x>.
- Johansson, Ulf, Cecilia Sönströd, Ulf Norinder, and Henrik Boström. 2011. “Trade-Off Between Accuracy and Interpretability for Predictive in Silico Modeling.” *Future Medicinal Chemistry* 3 (6): 647–63. <https://doi.org/10.4155/fmc.11.23>.
- Jones, Rory. 2023. “Qatar 2022: World Cup Final Scores 1.5 Bn Global Viewers.” <https://www.sportspro.com/news/qatar-2022-fifa-world-cup-final-argentina-france-viewers-engagement/>. <https://www.sportspro.com/news/qatar-2022-fifa-world-cup-final-argentina-france-viewers-engagement/>.
- Joyner, M. J. 1991. “Modeling: Optimal Marathon Performance on the Basis of Physiological Factors.” *Journal of Applied Physiology* 70 (2): 683–87. <https://doi.org/https://doi.org/10.1152/jappl.1991.70.2.683>.
- Joyner, Michael J., and Edward F. Coyle. 2008. “Endurance Exercise Performance: The Physiology of Champions.” *The Journal of Physiology* 586 (1): 35–44. <https://doi.org/10.1113/jphysiol.2007.143834>.
- Ju, Wonwoo, Dominic Doran, Richard Hawkins, Mark Evans, Andy Laws, and Paul Bradley. 2023. “Contextualised High-Intensity Running Profiles of Elite Football Players with Reference to General and Specialised Tactical Roles.” *Biology of Sport* 40 (1): 291–301. <https://doi.org/10.5114/biolsport.2023.116003>.
- Ju, Wonwoo, Richard Hawkins, Dominic Doran, Antonio Gómez-Díaz, Andrés Martín-García, Mark Evans, Andy Laws, and Paul Bradley. 2023. “Tier-Specific Contextualised High-Intensity Running Profiles in the English Premier League: More on-Ball Movement at the Top.” *Biology of Sport* 40 (2): 561–73. <https://doi.org/10.5114/biolsport.2023.118020>.
- Karlis, Dimitris, and Ioannis Ntzoufras. 2003. “Analysis of Sports Data by Using Bivariate Poisson Models.” *Journal of the Royal Statistical Society: Series D (The Statistician)* 52 (3): 381–93. <https://doi.org/10.1111/1467-9884.00366>.
- Kaufman, Shachar, Saharon Rosset, and Claudia Perlich. 2011. “Leakage in Data Mining: Formulation, Detection, and Avoidance.” In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’11. ACM. <https://doi.org/10.1145/2020408.2020496>.

- Kempe, Matthias, and Daniel Memmert. 2018. ““Good, Better, Creative”: The Influence of Creativity on Goal Scoring in Elite Soccer.” *Journal of Sports Sciences* 36 (21): 2419–23. <https://doi.org/10.1080/02640414.2018.1459153>.
- Klemp, Maximilian. 2024. “Python.” In *Computer Science in Sport*, 125–31. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-68313-2_15.
- Klemp, Maximilian, Daniel Memmert, and Robert Rein. 2021. “The Influence of Running Performance on Scoring the First Goal in a Soccer Match.” *International Journal of Sports Science & Coaching* 17 (3): 558–67. <https://doi.org/10.1177/17479541211035382>.
- Klemp, Maximilian, Robert Rein, and Daniel Memmert. 2023. “Analyzing Collapsibility in Performance Analysis in Sports.” In *13th World Congress of Performance Analysis of Sport and 13th International Symposium on Computer Science in Sport*, 108–11. Springer Nature Switzerland. <http://dx.doi.org/10.1080/02640414.2021.1930685>. https://doi.org/10.1007/978-3-031-31772-9_23.
- Klemp, Maximilian, Fabian Wunderlich, and Daniel Memmert. 2021. “In-Play Forecasting in Football Using Event and Positional Data.” *Scientific Reports* 11 (1). <https://doi.org/10.1038/s41598-021-03157-3>.
- Komi, Paavo V. 1986. “Training of Muscle Strength and Power: Interaction of Neuromotoric, Hypertrophic, and Mechanical Factors.” *International Journal of Sports Medicine* 07 (S 1): S10–15. <https://doi.org/10.1055/s-2008-1025796>.
- Komi, Paavo V., and C. Bosco. 1978. “Utilization of Stored Elastic Energy in Leg Extensor Muscles by Men and Women.” *Medicine and Science in Sports* 10 (4): 261–265. <http://europepmc.org/abstract/MED/750844>.
- Komi, Paavo V., and Karlsson. 1978. “Skeletal Muscle Fibre Types, Enzyme Activities and Physical Performance in Young Males and Females.” *Acta Physiologica Scandinavica* 103 (2): 210–18. <https://doi.org/10.1111/j.1748-1716.1978.tb06208.x>.
- Krustrup, Peter, Magni Mohr, Tommas Amstrup, Torben Rysgaard, Johnny Johansen, Adam Steensberg, Preben K. Pedersen, and Jens Bangsbo. 2003. “The Yo-Yo Intermittent Recovery Test: Physiological Response, Reliability, and Validity.” *Medicine & Science in Sports & Exercise* 35 (4): 697–705. <https://doi.org/10.1249/01.mss.0000058441.94520.32>.
- Lago-Ballesteros, Joaquin, Carlos Lago-Peñas, and Ezequiel Rey. 2012. “The Effect of Playing Tactics and Situational Variables on Achieving Score-Box Possessions in a Professional Soccer Team.” *Journal of Sports Sciences* 30 (14): 1455–61. <https://doi.org/10.1080/02640414.2012.712715>.
- Lago-Peñas, Carlos. 2009. “The Influence of Match Location, Quality of Opposition, and Match Status on Possession Strategies in Professional Association Football.”

- Journal of Sports Sciences* 27 (13): 1463–69.
<https://doi.org/10.1080/02640410903131681>.
- Lago-Peñas, Carlos, Luis Casais, Eduardo Dominguez, and Jaime Sampaio. 2010. “The Effects of Situational Variables on Distance Covered at Various Speeds in Elite Soccer.” *European Journal of Sport Science* 10 (2): 103–9.
<https://doi.org/10.1080/17461390903273994>.
- Lago-Peñas, Carlos, and Alexandre Dellal. 2010. “Ball Possession Strategies in Elite Soccer According to the Evolution of the Match-Score: The Influence of Situational Variables.” *Journal of Human Kinetics* 25 (2010): 93–100.
<https://doi.org/10.2478/v10078-010-0036-z>.
- Lago-Peñas, Carlos, Miguel Ángel Gómez Ruano, Diego Megías-Navarro, and Richard Pollard. 2016. “Home Advantage in Football: Examining the Effect of Scoring First on Match Outcome in the Five Major European Leagues.” *International Journal of Performance Analysis in Sport* 16 (2): 411–21.
<https://doi.org/10.1080/24748668.2016.11868897>.
- Lago-Peñas, Carlos, and Joaquin Lago-Ballesteros. 2011. “Game Location and Team Quality Effects on Performance Profiles in Professional Soccer.” *Journal of Sports Science & Medicine* 10 (3): 465.
- Lago-Peñas, Carlos, and Rafael Martín. 2013. “Determinants of Possession of the Ball in Soccer.” *Journal of Sports Sciences* 25 (9): 969–74.
<https://doi.org/10.1080/02640410600944626>.
- Lamas, Leonardo, José Vitor Senatore, and Gilbert Fellingham. 2020. “Two Steps for Scoring a Point: Creating and Converting Opportunities in Invasion Team Sports.” Edited by Miguel Ángel Gómez Ruano. *PLOS ONE* 15 (10): e0240419.
<https://doi.org/10.1371/journal.pone.0240419>.
- Lames, Martin. 2018. “Chance Involvement in Goal Scoring in Football—an Empirical Approach.” *German Journal of Exercise and Sport Research* 48 (2): 278–86. <https://doi.org/10.1007/s12662-018-0518-z>.
- Lames, Martin, and Tim McGarry. 2017. “On the Search for Reliable Performance Indicators in Game Sports.” *International Journal of Performance Analysis in Sport* 7 (1): 62–79. <https://doi.org/10.1080/24748668.2007.11868388>.
- Lepschy, Hannes, Hagen Wäsche, and Alexander Woll. 2018. “How to Be Successful in Football: A Systematic Review.” *The Open Sports Sciences Journal* 11 (1): 3–23. <https://doi.org/10.2174/1875399x01811010003>.
- . 2020. “Success Factors in Football: An Analysis of the German Bundesliga.” *International Journal of Performance Analysis in Sport* 20 (2): 150–64.
<https://doi.org/10.1080/24748668.2020.1726157>.
- Lewis, M, and Mike D. Hughes. 1988. “Attacking Play in the 1986 World Cup of Association Football.” *Journal of Sport Science* 6: 169.

- Lewis, Michael. 2004. *Moneyball: The Art of Winning an Unfair Game*. WW Norton & Company.
- Linke, Daniel, Daniel Link, and Martin Lames. 2018. “Validation of Electronic Performance and Tracking Systems EPTS Under Field Conditions.” Edited by Luca Paolo Ardigò. *PLOS ONE* 13 (7): e0199519. <https://doi.org/10.1371/journal.pone.0199519>.
- . 2020. “Football-Specific Validity of TRACAB’s Optical Video Tracking Systems.” Edited by Hugo A. Kerhervé. *PLOS ONE* 15 (3): e0230179. <https://doi.org/10.1371/journal.pone.0230179>.
- Llana, Sergio, Borja Burriel, Pau Madrero, and Javier Fernández. 2022. “Is It Worth the Effort? Understanding and Contextualizing Physical Metrics in Soccer.” *arXiv Preprint arXiv:2204.02313*.
- Lorenzo-Martinez, Miguel, Anton Kalén, Ezequiel Rey, Roberto López-Del Campo, Ricardo Resta, and Carlos Lago-Peñas. 2020. “Do Elite Soccer Players Cover Less Distance When Their Team Spent More Time in Possession of the Ball?” *Science and Medicine in Football* 5 (4): 310–16. <https://doi.org/10.1080/24733938.2020.1853211>.
- Low, Benedict, Diogo Coutinho, Bruno Gonçalves, Robert Rein, Daniel Memmert, and Jaime Sampaio. 2019. “A Systematic Review of Collective Tactical Behaviours in Football Using Positional Data.” *Sports Medicine* 50 (2): 343–85. <https://doi.org/10.1007/s40279-019-01194-7>.
- Low, Benedict, Robert Rein, Dominik Raabe, Sebastian Schwab, and Daniel Memmert. 2021. “The Porous High-Press? An Experimental Approach Investigating Tactical Behaviours from Two Pressing Strategies in Football.” *Journal of Sports Sciences* 39 (19): 2199–2210. <https://doi.org/10.1080/02640414.2021.1925424>.
- Low, Benedict, Robert Rein, Sebastian Schwab, and Daniel Memmert. 2021. “Defending in 4-4-2 or 5-3-2 Formation? Small Differences in Footballers’ Collective Tactical Behaviours.” *Journal of Sports Sciences* 40 (3): 351–63. <https://doi.org/10.1080/02640414.2021.1993655>.
- Lucey, Patrick, Alina Bialkowski, Mathew Monfort, Peter Carr, and Iain Matthews. 2015. ““Quality Vs Quantity”: Improved Shot Prediction in Soccer Using Strategic Features from Spatiotemporal Data.” In *Proceedings of 9th MIT Sloan Sports Analytics Conference*.
- Macdonald, Brian. 2012. “An Expected Goals Model for Evaluating NHL Teams and Players.” In *Proceedings of the 2012 MIT Sloan Sports Analytics Conference*.
- Mackenzie, Rob, and Chris Cushion. 2012. “Performance Analysis in Football: A Critical Review and Implications for Future Research.” *Journal of Sports Sciences* 31 (6): 639–76. <https://doi.org/10.1080/02640414.2012.746720>.

- Manafifard, M., H. Ebadi, and H. Abrishami Moghaddam. 2017. "A Survey on Player Tracking in Soccer Videos." *Computer Vision and Image Understanding* 159 (June): 19–46. <https://doi.org/10.1016/j.cviu.2017.02.002>.
- McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. Routledge. <https://doi.org/10.1201/9780203753736>.
- McGarry, Tim. 2017. "Applied and Theoretical Perspectives of Performance Analysis in Sport: Scientific Issues and Challenges." *International Journal of Performance Analysis in Sport* 9 (1): 128–40. <https://doi.org/10.1080/24748668.2009.11868469>.
- Mead, James, Anthony O'Hare, and Paul McMenemy. 2023. "Expected Goals in Football: Improving Model Performance and Demonstrating Value." Edited by Rabi Muazu Musa. *PLOS ONE* 18 (4): e0282295. <https://doi.org/10.1371/journal.pone.0282295>.
- Mehta, Saumya, Philip Furley, Dominik Raabe, and Daniel Memmert. 2023. "Examining How Data Becomes Information for an Upcoming Opponent in Football." *International Journal of Sports Science & Coaching* 19 (3): 978–87. <https://doi.org/10.1177/17479541231187871>.
- Memmert, Daniel. 2021. *Match Analysis: How to Use Data in Professional Sport*. Edited by Daniel Memmert. London: Routledge,.
- . 2024. *Computer Science in Sport: Modeling, Simulation, Data Analysis and Visualization of Sports-Related Data*. Springer Nature.
- Memmert, Daniel, Maximilian Klemp, Sebastian Schwab, and Benedict Low. 2023. "Individual Attention Capacity Enhances in-Field Group Performances in Soccer." *International Journal of Sport and Exercise Psychology*, April, 1–18. <https://doi.org/10.1080/1612197x.2023.2204364>.
- Memmert, Daniel, Koen A. P. M. Lemmink, and Jaime Sampaio. 2016. "Current Approaches to Tactical Performance Analyses in Soccer Using Position Data." *Sports Medicine* 47 (1): 1–10. <https://doi.org/10.1007/s40279-016-0562-5>.
- Memmert, Daniel, Dominik Raabe, Sebastian Schwab, and Robert Rein. 2019. "A Tactical Comparison of the 4-2-3-1 and 3-5-2 Formation in Soccer: A Theory-Oriented, Experimental Approach Based on Positional Data in an 11 Vs. 11 Game Set-up." Edited by Alessandro Moura Zagatto. *PLOS ONE* 14 (1): e0210191. <https://doi.org/10.1371/journal.pone.0210191>.
- Merhej, Charbel, Ryan Beal, Sarvapali Ramchurn, and Tim Matthews. 2021. "What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data." <https://doi.org/10.48550/ARXIV.2106.01786>.
- Midgley, Adrian W, Lars R McNaughton, and Michael Wilkinson. 2006. "Is There an Optimal Training Intensity for Enhancing the Maximal Oxygen Uptake of Distance Runners?: Empirical Research Findings, Current Opinions,

- Physiological Rationale and Practical Recommendations.” *Sports Medicine* 36 (2): 117–32. <https://doi.org/10.2165/00007256-200636020-00003>.
- Mill, John Stuart. 1843. *A System of Logic: Ratiocinative and Inductive : Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. Repr. Honolulu, Hawaii: Univ. Press of the Pacific.
- Modric, Toni, Sime Versic, and Damir Sekulic. 2021. “Relationship Between Yo-Yo Intermittent Endurance Test-Level 1 and Match Running Performance in Soccer: Still on the Right Path?” *Polish Journal of Sport and Tourism* 28 (4): 16–20. <https://doi.org/10.2478/pjst-2021-0021>.
- Modric, Toni, Sime Versic, Damir Sekulic, and Silvester Liposek. 2019. “Analysis of the Association Between Running Performance and Game Performance Indicators in Professional Soccer Players.” *International Journal of Environmental Research and Public Health* 16 (20): 4032. <https://doi.org/10.3390/ijerph16204032>.
- Mohr, Magni, Peter Krstrup, and Jens Bangsbo. 2002. “Match Performance of High-Standard Soccer Players with Special Reference to Development of Fatigue.” *Journal of Sports Sciences* 21 (7): 519–28. <https://doi.org/10.1080/0264041031000071182>.
- Mood, C. 2009. “Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do about It.” *European Sociological Review* 26 (1): 67–82. <https://doi.org/10.1093/esr/jcp006>.
- Nelder, J. A., and R. W. M. Wedderburn. 1972. “Generalized Linear Models.” *Journal of the Royal Statistical Society. Series A (General)* 135 (3): 370. <https://doi.org/10.2307/2344614>.
- Nevill, Alan M., Greg Atkinson, and Mike D. Hughes. 2008. “Twenty-Five Years of Sport Performance Research in the Journal of Sports Sciences.” *Journal of Sports Sciences* 26 (4): 413–26. <https://doi.org/10.1080/02640410701714589>.
- Neyman, Jerzy. 1923. “Sur Les Applications de La Theorie Des Probabilites Aux Experiences Agricoles: Essay Des Principes. Excerpts Reprinted (1990) in English.” *Statistical Science* 5 (463–472): 4.
- Nori, Harsha, Samuel Jenkins, Paul Koch, and Rich Caruana. 2019. “InterpretML: A Unified Framework for Machine Learning Interpretability.” arXiv. <https://doi.org/10.48550/ARXIV.1909.09223>.
- O’Donoghue, Peter. 2009. *Research Methods for Sports Performance Analysis*. Routledge. <https://doi.org/10.4324/9780203878309>.
- Pate, Russell R., and Andrea Kriska. 1984. “Physiological Basis of the Sex Difference in Cardiorespiratory Endurance.” *Sports Medicine* 1 (2): 87–98. <https://doi.org/10.2165/00007256-198401020-00001>.
- Pearl, Judea. 1995. “Causal Diagrams for Empirical Research.” *Biometrika* 82 (4):

- 702–10. <https://doi.org/10.1093/biomet/82.4.702>.
- . 2021. *Causal Inference in Statistics: A Primer*. Edited by Madelyn Glymour and Nicholas P. Jewell. Reprinted with revisions. Chichester: Wiley.
- Pearson, Karl. 1900. “X. On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling.” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50 (302): 157–75. <https://doi.org/10.1080/14786440009463897>.
- . 1911. *The Grammar of Science*. Adam; Charles Black.
- Peeters, Thomas. 2018. “Testing the Wisdom of Crowds in the Field: Transfermarkt Valuations and International Soccer Results.” *International Journal of Forecasting* 34 (1): 17–29. <https://doi.org/10.1016/j.ijforecast.2017.08.002>.
- Phatak, Ashwin A., Saumya Mehta, Franz-Georg Wieland, Mikael Jamil, Mark Connor, Manuel Bassek, and Daniel Memmert. 2022. “Context Is Key: Normalization as a Novel Approach to Sport Specific Preprocessing of KPI’s for Match Analysis in Soccer.” *Scientific Reports* 12 (1). <https://doi.org/10.1038/s41598-022-05089-y>.
- Phatak, Ashwin A., Robert Rein, and Daniel Memmert. 2021. “The Dirty League: English Premier League Provides Higher Incentives for Fouling as Compared to Other European Soccer Leagues.” *Journal of Human Kinetics* 80 (October): 263–76. <https://doi.org/10.2478/hukin-2021-0095>.
- Pollard, Richard. 2017. “Goal-Scoring and the Negative Binomial Distribution.” *The Mathematical Gazette* 69 (447): 45–47. <https://doi.org/10.2307/3616453>.
- Pollard, Richard, Jake Ensum, and Samuel Taylor. 2004. “Estimating the Probability of a Shot Resulting in a Goal: The Effects of Distance, Angle and Space.” *International Journal of Soccer and Science* 2 (1): 50–55.
- Pollard, Richard, and Miguel Ángel Gómez Ruano. 2014. “Components of Home Advantage in 157 National Soccer Leagues Worldwide.” *International Journal of Sport and Exercise Psychology* 12 (3): 218–33. <https://doi.org/10.1080/1612197x.2014.888245>.
- Pollard, Richard, and Charles Reep. 1997. “Measuring the Effectiveness of Playing Strategies at Soccer.” *Journal of the Royal Statistical Society: Series D (The Statistician)* 46 (4): 541–50. <https://doi.org/10.1111/1467-9884.00108>.
- Pons, Eduard, Tomás García-Calvo, Ricardo Resta, Hugo Blanco, Roberto López del Campo, Jesús Díaz García, and Juan José Pulido. 2019. “A Comparison of a GPS Device and a Multi-Camera Video Technology During Official Soccer Matches: Agreement Between Systems.” Edited by Caroline Sunderland. *PLOS ONE* 14 (8): e0220729. <https://doi.org/10.1371/journal.pone.0220729>.
- Prampero, P. E. di, G. Atchou, J. -C. Brückner, and C. Moia. 1986. “The

- Energetics of Endurance Running.” *European Journal of Applied Physiology and Occupational Physiology* 55 (3): 259–66. <https://doi.org/10.1007/bf02343797>.
- Prampero, Pietro Enrico di. 2003. “Factors Limiting Maximal Performance in Humans.” *European Journal of Applied Physiology* 90 (3–4): 420–29. <https://doi.org/10.1007/s00421-003-0926-z>.
- Raabe, Dominik, Reinhard Nabben, and Daniel Memmert. 2022. “Graph Representations for the Analysis of Multi-Agent Spatiotemporal Sports Data.” *Applied Intelligence* 53 (4): 3783–803. <https://doi.org/10.1007/s10489-022-03631-z>.
- Redwood-Brown, Athalie, Peter O’Donoghue, Alan M. Nevill, Chris Saward, and Caroline Sunderland. 2019. “Effects of Playing Position, Pitch Location, Opposition Ability and Team Ability on the Technical Performance of Elite Soccer Players in Different Score Line States.” Edited by Anthony C. Constantinou. *PLOS ONE* 14 (2): e0211707. <https://doi.org/10.1371/journal.pone.0211707>.
- Redwood-Brown, Athalie, Peter ODonoghue, Alan M. Nevill, Chris Saward, Nicholas Dyer, and Caroline Sunderland. 2018. “Effects of Situational Variables on the Physical Activity Profiles of Elite Soccer Players in Different Score Line States.” *Scandinavian Journal of Medicine & Science in Sports* 28 (12): 2515–26. <https://doi.org/10.1111/sms.13271>.
- Redwood-Brown, Athalie, Peter, Gemma Robinson, and Paul Neilson. 2012. “The Effect of Score-Line on Work-Rate in English FA Premier League Soccer.” *International Journal of Performance Analysis in Sport* 12 (2): 258–71. <https://doi.org/10.1080/24748668.2012.11868598>.
- Reep, Charles, and Bernard Benjamin. 1968. “Skill and Chance in Association Football.” *Journal of the Royal Statistical Society. Series A (General)* 131 (4): 581–85. <https://doi.org/10.2307/2343726>.
- Reep, Charles, Richard Pollard, and Bernard Benjamin. 1971. “Skill and Chance in Ball Games.” *Journal of the Royal Statistical Society. Series A (General)* 134 (4): 623. <https://doi.org/10.2307/2343657>.
- Reilly, Thomas, and David Ball. 1984. “The Net Physiological Cost of Dribbling a Soccer Ball.” *Research Quarterly for Exercise and Sport* 55 (3): 267–71.
- Reilly, Thomas, and Vaughan Thomas. 1976. “A Motion Analysis of Work-Rate in Different Positional Roles in Professional Football Match-Play.” *Journal of Human Movement Studies* 2: 87–97.
- Rein, Robert, and Daniel Memmert. 2016. “Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science.” *SpringerPlus* 5 (1). <https://doi.org/10.1186/s40064-016-3108-2>.
- Rein, Robert, Jürgen Perl, and Daniel Memmert. 2017. “Maybe a Tad Early for a

- Grand Unified Theory: Commentary on ‘Towards a Grand Unified Theory of Sports Performance.’” *Human Movement Science* 56 (December): 173–75.
<https://doi.org/10.1016/j.humov.2017.04.011>.
- Rein, Robert, Dominik Raabe, and Daniel Memmert. 2017. “‘Which Pass Is Better?’ Novel Approaches to Assess Passing Effectiveness in Elite Soccer.” *Human Movement Science* 55 (August): 172–81.
<https://doi.org/10.1016/j.humov.2017.07.010>.
- Riedl, Dennis, Kathrin Staufienbiel, Bernd Strauss, and Andreas Heuer. 2022. “The Global Home Advantage in Soccer: Status, Developments and Causes,” February.
<https://doi.org/10.31234/osf.io/zaujy>.
- Robberechts, Pieter, Jesse Davis, K Leuven, and Belgium. 2020. “How Data Availability Affects the Ability to Learn Good xG Models.” In *Machine Learning and Data Mining for Sports Analytics: 7th International Workshop, MLSA 2020, Co-Located with ECML/PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings* 7, 17–27. Springer. https://doi.org/10.1007/978-3-030-64912-8_2.
- Robberechts, Pieter, Jan Van Haaren, and Jesse Davis. 2019. “Who Will Win It? An in-Game Win Probability Model for Football.” In *ArXiv*. Vol. abs/1906.05029. <https://api.semanticscholar.org/CorpusID:186206589>.
- Robergs, Robert A. 2003. *Fundamentals of Exercise Physiology: For Fitness, Performance, and Health*. McGraw-Hill.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70 (1): 41–55. <https://doi.org/10.1093/biomet/70.1.41>.
- Rothman, Kenneth J., and Sander Greenland. 2005. “Causation and Causal Inference in Epidemiology.” *American Journal of Public Health* 95 (S1): S144–50.
<https://doi.org/10.2105/ajph.2004.059204>.
- Ruiz, Héctor, Paulo Lisboa, Paul Neilson, and Warren Gregson. 2015. “Measuring Scoring Efficiency Through Goal Expectancy Estimation.” In *ESANN*. <https://www.semanticscholar.org/paper/e135ee90bc78512923c02eb5c9f44b6996f95387>.
- Rutherford, O. M., C. A. Greig, A. J. Sargeant, and D. A. Jones. 1986. “Strength Training and Power Output: Transference Effects in the Human Quadriceps Muscle.” *Journal of Sports Sciences* 4 (2): 101–7.
<https://doi.org/10.1080/02640418608732105>.
- Sale, Digby G. 1988. “Neural Adaptation to Resistance Training.” *Medicine & Science in Sports & Exercise* 20 (Sup 1): S135–45.
<https://doi.org/10.1249/00005768-198810001-00009>.
- Saltin, Bengt, and Per-Olof Åstrand. 1967. “Maximal Oxygen Uptake in Athletes.” *Journal of Applied Physiology* 23 (3): 353–58.
<https://doi.org/10.1152/jappl.1967.23.3.353>.

- Sanderson, F. H., and K. I. May. 1977. "The Development of Objective Methods of Game Analysis in Squash Rackets [Proceedings]." *British Journal of Sports Medicine* 11 (4): 188–88. <https://doi.org/10.1136/bjism.11.4.188>.
- Santos, Pedro, Carlos Lago-Peñas, and Oscar García-García. 2017. "The Influence of Situational Variables on Defensive Positioning in Professional Soccer." *International Journal of Performance Analysis in Sport* 17 (3): 212–19. <https://doi.org/10.1080/24748668.2017.1331571>.
- Sarmento, Hugo, Rui Marcelino, M. Teresa Anguera, Jorge Campaniço, Nuno Matos, and José Carlos Leitão. 2014. "Match Analysis in Football: A Systematic Review." *Journal of Sports Sciences* 32 (20): 1831–43. <https://doi.org/10.1080/02640414.2014.898852>.
- Schauberger, Gunther, Andreas Groll, and Gerhard Tutz. 2017. "Analysis of the Importance of on-Field Covariates in the German Bundesliga." *Journal of Applied Statistics* 45 (9): 1561–78. <https://doi.org/10.1080/02664763.2017.1383370>.
- Scott, Macfarlane T. U., Tannath J. Scott, and Vincent G. Kelly. 2016. "The Validity and Reliability of Global Positioning Systems in Team Sport." *Journal of Strength and Conditioning Research* 30 (5): 1470–90. <https://doi.org/10.1519/jsc.0000000000001221>.
- Shah, Nilay D., Ewout W. Steyerberg, and David M. Kent. 2018. "Big Data and Predictive Analytics: Recalibrating Expectations." *JAMA* 320 (1): 27. <https://doi.org/10.1001/jama.2018.5602>.
- Siff, M. C. 2000. "Biomechanical Foundations of Strength and Power Training." In *Biomechanics in Sport*, edited by Vladimir M. Zatsiorsky, 103–39. Wiley. <https://doi.org/10.1002/9780470693797.ch6>.
- Silva, Pedro, Ricardo Duarte, Jaime Sampaio, Paulo Aguiar, Keith Davids, Duarte Araújo, and Júlio Garganta. 2014. "Field Dimension and Skill Level Constrain Team Tactical Behaviours in Small-Sided and Conditioned Games in Football." *Journal of Sports Sciences* 32 (20): 1888–96. <https://doi.org/10.1080/02640414.2014.961950>.
- Simpson, E. H. 1951. "The Interpretation of Interaction in Contingency Tables." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 13 (2): 238–41. <https://doi.org/10.1111/j.2517-6161.1951.tb00088.x>.
- Spann, Martin, and Bernd Skiera. 2009. "Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters." *Journal of Forecasting* 28 (1): 55–72. <https://doi.org/10.1002/for.1091>.
- Spector, S. A., P. F. Gardiner, R. F. Zernicke, R. R. Roy, and V. R. Edgerton. 1980. "Muscle Architecture and Force-Velocity Characteristics of Cat Soleus and Medial Gastrocnemius: Implications for Motor Control." *Journal of*

- Neurophysiology* 44 (5): 951–60. <https://doi.org/10.1152/jn.1980.44.5.951>.
- Stigler, Stephen M. 1981. “Gauss and the Invention of Least Squares.” *The Annals of Statistics*, 465–74.
- Stølen, Tomas, Karim Chamari, Carlo Castagna, and Ulrik Wisløff. 2005. “Physiology of Soccer: An Update.” *Sports Medicine* 35 (6): 501–36. <https://doi.org/10.2165/00007256-200535060-00004>.
- Stone, Michael H, Ronald Byrd, John Tew, and Michael Wood. 1980. “Relationship Between Anaerobic Power and Olympic Weightlifting Performance.”
- Suchomel, Timothy J., Sophia Nimphius, and Michael H. Stone. 2016. “The Importance of Muscular Strength in Athletic Performances.” *Sports Medicine* 46 (10): 1419–49. <https://doi.org/10.1007/s40279-016-0486-0>.
- Szymanski, Stefan. 2020. “Sport Analytics: Science or Alchemy?” *Kinesiology Review* 9 (1): 57–63. <https://doi.org/10.1123/kr.2019-0066>.
- Taberner, Matt, Jason O’Keefe, David Flower, Jack Phillips, Graeme Close, Daniel Dylan Cohen, Chris Richter, and Christopher Carling. 2019. “Interchangeability of Position Tracking Technologies; Can We Merge the Data?” *Science and Medicine in Football* 4 (1): 76–81. <https://doi.org/10.1080/24733938.2019.1634279>.
- Taylor, Joseph B., Stephen D. Mellalieu, Nic James, and David A. Shearer. 2008. “The Influence of Match Location, Quality of Opposition, and Match Status on Technical Performance in Professional Association Football.” *Journal of Sports Sciences* 26 (9): 885–95. <https://doi.org/10.1080/02640410701836887>.
- Tenga, A., D. Kanstad, L. T. Ronglan, and R. Bahr. 2017. “Developing a New Method for Team Match Performance Analysis in Professional Soccer and Testing Its Reliability.” *International Journal of Performance Analysis in Sport* 9 (1): 8–25. <https://doi.org/10.1080/24748668.2009.11868461>.
- Tenga, Albin, Ingar Holme, Lars Tore Ronglan, and Roald Bahr. 2009a. “Effect of Playing Tactics on Achieving Score-Box Possessions in a Random Series of Team Possessions from Norwegian Professional Soccer Matches.” *Journal of Sports Sciences* 28 (3): 245–55. <https://doi.org/10.1080/02640410903502766>.
- . 2009b. “Effect of Playing Tactics on Goal Scoring in Norwegian Professional Soccer.” *Journal of Sports Sciences* 28 (3): 237–44. <https://doi.org/10.1080/02640410903502774>.
- Thomas, Graham, Rikke Gade, Thomas B. Moeslund, Peter Carr, and Adrian Hilton. 2019. “Computer Vision for Sports: Current Applications and Research Topics.” *Computer Vision and Image Understanding* 159 (September): 3–18. <https://doi.org/10.1016/j.cviu.2017.04.011>.
- Thorstensson, A., G. Grimby, and J. Karlsson. 1976. “Force-Velocity Relations and Fiber Composition in Human Knee Extensor Muscles.” *Journal of Applied*

- Physiology* 40 (1): 12–16. <https://doi.org/10.1152/jappl.1976.40.1.12>.
- Thorstensson, Alf, Bodil Hultén, Wilhelm von Döbeln, and Jan Karlsson. 1976. “Effect of Strength Training on Enzyme Activities and Fibre Characteristics in Human Skeletal Muscle.” *Acta Physiologica Scandinavica* 96 (3): 392–98. <https://doi.org/10.1111/j.1748-1716.1976.tb10207.x>.
- Tihanyi, J., P. Apor, and Gy. Fekete. 1982. “Force-Velocity-Power Characteristics and Fiber Composition in Human Knee Extensor Muscles.” *European Journal of Applied Physiology and Occupational Physiology* 48 (3): 331–43. <https://doi.org/10.1007/bf00430223>.
- Van Haaren, Jan. 2021. “”Why Would i Trust Your Numbers?” on the Explainability of Expected Values in Soccer.” *arXiv Preprint arXiv:2105.13778*. <https://www.semanticscholar.org/paper/09ff80f6990632e75ac6cd5f5f481a1c43f6d06b>.
- van Iterson, Maarten, Herman H. B. M. van Haagen, and Jelle J. Goeman. 2012. “Resolving Confusion of Tongues in Statistics and Machine Learning: A Primer for Biologists and Bioinformaticians.” *Proteomics* 12 (4–5): 543–49. <https://doi.org/10.1002/pmic.201100395>.
- Van Roy, Maaïke, Pieter Robberechts, Tom Decroos, and Jesse Davis. 2020. “Valuing on-the-Ball Actions in Soccer: A Critical Comparison of xT and VAEP.” In *Proceedings of the AAAI-20 Workshop on Artificial Intelligence in Team Sports*. AI in Team Sports Organising Committee. <https://www.semanticscholar.org/paper/99a1ac4279cdab37f4021f924481ba0517c10c54>.
- Vapnik, V. N. 1999. “An Overview of Statistical Learning Theory.” *IEEE Transactions on Neural Networks* 10 (5): 988–99. <https://doi.org/10.1109/72.788640>.
- Vapnik, V. N., and A. Ya. Chervonenkis. 1971. “On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities.” *Theory of Probability & Its Applications* 16 (2): 264–80. <https://doi.org/10.1137/1116025>.
- Vapnik, Vladimir N. 2000. *The Nature of Statistical Learning Theory*. Springer New York. <https://doi.org/10.1007/978-1-4757-3264-1>.
- Vogelbein, Martin, Stephan Nopp, and Anita Hökelmann. 2014. “Defensive Transition in Soccer – Are Prompt Possession Regains a Measure of Success? A Quantitative Analysis of German Fußball-Bundesliga 2010/2011.” *Journal of Sports Sciences* 32 (11): 1076–83. <https://doi.org/10.1080/02640414.2013.879671>.
- Wahl, Patrick, Wilhelm Bloch, and Joachim Mester. 2009. “Moderne Betrachtungsweisen Des Laktats: Laktat Ein Überschatztes Und Zugleich Unterschätztes Molekul.” *Schweizerische Zeitschrift Fur Sportmedizin Und Sporttraumatologie* 57 (3): 100.

- Wasserman, Karlman, and Malcolm B. McIlroy. 1964. "Detecting the Threshold of Anaerobic Metabolism in Cardiac Patients During Exercise." *The American Journal of Cardiology* 14 (6): 844–52.
[https://doi.org/10.1016/0002-9149\(64\)90012-8](https://doi.org/10.1016/0002-9149(64)90012-8).
- Weil, Joy, E., and Encyclopaedia Britannica. 2024. "Football (Soccer)." *Encyclopaedia Britannica*. <https://www.britannica.com/sports/football-soccer>.
- Widrick, Jeffrey J., Julian E. Stelzer, Todd C. Shoepe, and Dena P. Garner. 2002. "Functional Properties of Human Muscle Fibers After Short-Term Resistance Exercise Training." *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 283 (2): R408–16.
<https://doi.org/10.1152/ajpregu.00120.2002>.
- Wisløff, U, C Castagna, J Helgerud, R Jones, and J Hoff. 2004. "Strong Correlation of Maximal Squat Strength with Sprint Performance and Vertical Jump Height in Elite Soccer Players." *British Journal of Sports Medicine* 38 (3): 285–88.
<https://doi.org/10.1136/bjsm.2002.002071>.
- Wisløff, Ulrik, Jan Helgerud, and Jan Hoff. 1998. "Strength and Endurance of Elite Soccer Players." *Medicine & Science in Sports & Exercise* 30 (3): 462–67.
<https://doi.org/10.1097/00005768-199803000-00019>.
- Withers, R. T. 1982. "Match Analyses of Australian Professional Soccer Players." *Journal of Human Movement Studies* 8: 159–76.
- Wunderlich, Fabian, Felix Berge, Daniel Memmert, and Robert Rein. 2020. "Almost a Lottery: The Influence of Team Strength on Success in Penalty Shootouts." *International Journal of Performance Analysis in Sport* 20 (5): 857–69.
<https://doi.org/10.1080/24748668.2020.1799171>.
- Wunderlich, Fabian, and Daniel Memmert. 2018. "The Betting Odds Rating System: Using Soccer Forecasts to Forecast Soccer." Edited by Anthony C. Constantinou. *PLOS ONE* 13 (6): e0198668. <https://doi.org/10.1371/journal.pone.0198668>.
- . 2020. "Forecasting the Outcomes of Sports Events: A Review." *European Journal of Sport Science* 21 (7): 944–57.
<https://doi.org/10.1080/17461391.2020.1793002>.
- Wunderlich, Fabian, Alessandro Seck, and Daniel Memmert. 2021. "The Influence of Randomness on Goals in Football Decreases over Time. An Empirical Analysis of Randomness Involved in Goal Scoring in the English Premier League." *Journal of Sports Sciences* 39 (20): 2322–37.
<https://doi.org/10.1080/02640414.2021.1930685>.

A Appendix

This appendix contains the articles that are part of the dissertation, but have not been discussed in the main text. All articles have been published in peer-reviewed journals and have been written in English. Impact Factors and Quartiles have been extracted from Scopus (CiteScore).

Study 4**Individual attention capacity enhances in-field group performances in soccer.**

Memmert, D., **Klemp, M. (shared first-authorship)**, Schwab, S., & Low, B. (2023). Individual attention capacity enhances in-field group performances in soccer. *International Journal of Sport and Exercise Psychology*, 1–18. <https://doi.org/10.1080/1612197X.2023.2204364>

[Impact Factor 2023: 7.3, Q1 Social Psychology]

Abstract: Visual attention capacity impacts performance in different laboratory and real-world tasks positively. The present field-study extends for the first time performance gains from individuals to team performance. Our study examined the attention capacity of 88 soccer players to create low vs. high attention capacity teams (LACT vs. HACT) who competed in a controlled field-study. Positional data were gathered using player tracking systems (1 Hz) in a standardised 11 vs. 11 soccer game. Key performance indicators (KPIs), which have been established in match analysis research, were measured to reflect tactical performance of the teams. As predicted from attention frameworks, HACT outperformed LACT in five out of seven KPIs, suggesting that attention capacity might play a crucial role for/in searching and gathering important space on the pitch. These findings provide evidence for attention being a predictor of team success. Practice task designs with an emphasis on attention capacity therefore could be a useful tool to study the emergent behavioural dynamics in a real-world environment.

Study 5**English goalkeepers are not responsible for
england's poor performance in penalty
shootouts in the past**

Brinkschulte, M., Furley, P., **Klemp, M.**, & Memmert, D. (2021). English goalkeepers are not responsible for england's poor performance in penalty shootouts in the past. *Scientific Reports*, 11(1).

<https://doi.org/10.1038/s41598-021-04118-6>

[Impact Factor 2023: 7.5, Q1]

Abstract: Scrutinizing public opinion is one of the central goals of science as the divergence between public opinion and scientific evidence can have negative consequences. The present study aims to further investigate the alleged English ‘penalty curse’ and determine if it can be linked to the prevalent stereotype of the ‘English goalkeeper problem’. We analyzed a large sample of 2379 penalty kicks that 629 different goalkeepers faced in World Cups and European Championships, as well as in the Champions and Europa League by comparing the goalkeeper success rates of different nations by fitting a generalized linear model (binomial regression) to the data. However, the results do not reveal meaningful differences between the success rates (on average 22.23%). Consequently, we conclude that English goalkeepers are not responsible for England’s poor performance in penalties in the past as they perform as well as goalkeepers from other nations and, in turn, provide a counterargument to the widespread stereotype that ‘England has a goalkeeper problem’.

Study 6**A new approach for training-load quantification in elite-level soccer: Contextual factors**

Guerrero-Calderón, B., **Klemp, M.**, Castillo-Rodriguez, A., Morcillo, J. A., & Memmert, D. (2020). A new approach for training-load quantification in elite-level soccer: Contextual factors. *International Journal of Sports Medicine*, 42(08), 716–723.

<https://doi.org/10.1055/a-1289-9059>

[Impact Factor 2023: 4.8, Q1 Physical Therapy, Sports Therapy and Rehabilitation]

Abstract: The aims of this study were to analyse the physical responses of professional soccer players during training considering the contextual factors of match location, season period, and quality of the opposition; and to establish prediction models of physical responses during training sessions. Training data was obtained from 30 professional soccer players from Spanish La Liga using global positioning technology (N = 1365 performances). A decreased workload was showed during training weeks prior to home matches, showing large effects in power events, equivalent distance, total distance, walk distance and low-speed running distance. Also, the quality of the opposition also affected the training workload ($p < 0.05$). All regression-models showed moderate effects, with an adjusted R² of 0.37 for metabolic-work, 0.34 for total distance covered, 0.25 for high-speed running distance (18–21 $km \cdot h^{-1}$), 0.29 for very high-speed running distance (21–24 $km \cdot h^{-1}$), 0.22 for sprint running distance ($> 24 km \cdot h^{-1}$) and 0.34 for equivalent distance. The main finding of this study was the great association of match location, season period and quality of opposition on the workload performed by players in the training week before the match; and the development of workload prediction-models considering these contextual factors, thus proposing a new and innovative approach to quantify the workload in soccer.

Study 7**How does the workload applied during the training week and the contextual factors affect the physical responses of professional soccer players in the match?**

Guerrero-Calderón, B., **Klemp, M.**, Morcillo, J. A., & Memmert, D. (2021). How does the workload applied during the training week and the contextual factors affect the physical responses of professional soccer players in the match? *International Journal of Sports Science & Coaching*, 16(4), 994–1003.

<https://doi.org/10.1177/1747954121995610>

[Impact Factor 2023: 3.5, Q1 Social Sciences (miscellaneous)]

Abstract: The aim of this study was to examine whether match physical output can be predicted from the workload applied in training by professional soccer players. Training and match load records from two professional soccer teams belonging to the Spanish First and Second Division were collected through GPS technology over a season ($N = 1678$ and $N = 2441$ records, respectively). The factors playing position, season period, quality of opposition, category and playing formation were considered into the analysis. The level of significance was set at $p = .05$. The prediction models yielded a conditional R-squared in match of 0.51 in total distance (TD); 0.58 in high-intensity distance (HIRD, from 14 to 24 $km \cdot h^{-1}$); and 0.60 in sprint distance (SPD, $> 24 km \cdot h^{-1}$). The main finding of this study was that the physical output of players in the match was predicted from the training-load performed during the previous training week. The training-TD negatively affected the match physical output while the training-HIRD showed a positive effect. Moreover, the contextual factors – playing position, season period, division and quality of opposition – affected the players’ physical output in the match. Therefore, these results suggest the appropriateness of programming lower training volume but increasing the intensity of the activity throughout the weekly microcycle, and considering contextual factors within the load programming.