KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Beyond the Scoreline: Using Machine Learning Techniques to Understand (Women's) Soccer

Lotte Bransen

Dissertation presented in partial fulfillment
of the requirements for the degree of
Doctor of Engineering Science (PhD):
Computer Science

Supervisor:
Prof. dr. Jesse Davis

November 2025

BEYOND THE SCORELINE: USING MACHINE LEARNING TECHNIQUES TO UNDERSTAND (WOMEN'S) SOCCER

Lotte BRANSEN

Supervisor:
Prof. dr. Jesse Davis

Members of the
Examination Committee:
Em. Prof. dr. ir. Lucie Vandewalle, chair
Prof. dr. Danny Hughes
Dr. Wannes Meert
Dr. ir. Jan Van Haaren
Dr. Sigrid Olthof
(Liverpool John Moores University)
Dr. Matthias Kempe
(Universität Wien)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Computer Science

November 2025

© 2025 Lotte Bransen
Uitgegeven in eigen beheer, Lotte Bransen, Utrecht, the Netherlands (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

Acknowledgements

Ik heb begrepen dat het dankwoord het meest gelezen stuk van een proefschrift is. Dat bracht nogal wat druk op het schrijven van dit stuk. Dan ben je bijna vijf jaar met het schrijven van een proefschrift en zijn de meeste mensen vooral geïnteresseerd in wie je allemaal benoemt in het dankwoord. Ik hoop dan maar dat dit dankwoord jullie bevalt. Het is in ieder geval het meest persoonlijke stuk van dit proefschrift geworden.

Als kind had ik de grote droom om ooit in het mannenelftal van FC Utrecht te mogen voetballen. Er waren nog geen vrouwenvoetbalteams en het leek me ook wel haalbaar om tussen de mannen te voetballen als ik later groot zou zijn. Eerst bracht een verhuizing naar Bemmel bijna roet in het eten: uit boosheid over de verhuizing werd ik een week fan van NEC. Later bleek ook dat mijn snelheid wat achter bleef bij die van Daphne Schippers en er nog wel meer kwaliteiten misten om het eerste mannenelftal van de FC te halen. Naast voetbal was ik ook altijd gefascineerd door getallen. Dat was natuurlijk een beetje nerdy, maar ik ben dat steeds meer gaan omarmen. Dat die twee werelden van voetbal en getallen samen zouden komen in een PhD, daar gingen mijn dromen dan weer niet over.

In januari 2021, piek coronatijd, begon ik de PhD onder supervisie van Jesse. Ik begon de PhD zoals zovelen hem beginnen. Met de ambitie om de wereld te veranderen. In mijn geval om ervoor te zorgen dat vrouwenvoetbal meer aandacht zou krijgen en ik al mijn honderden ideeën kon uitwerken. Dat is uiteraard niet allemaal gelukt.

Het proces tot dit “boekje” was er een van pieken en dalen. Tijdens die tocht waren er veel mensen die erbij waren om de pieken te vieren en me te steunen bij de dalen. En ook gewoon om eens in de zoveel tijd te vragen hoe het toch met die PhD ging. Die vraag vond ik de laatste periode wat lastiger omdat het einde zwaar was, en de motivatie af en toe uit de tenen moest komen om op papier te zetten wat ik de afgelopen vijf jaar heb gedaan. Hulde aan de mensen

die toch nog bleven vragen hoe het ermee ging.

Op dit boekje staat mijn naam maar in mijn eentje was me dit nooit gelukt.

Ten eerste wil ik Jesse bedanken. Je hebt me de kans gegeven een PhD te doen, in een bijzondere vorm die zeker ook van jou veel inspanning heeft gevergd. Bedankt voor de support gedurende de afgelopen vijf jaar. Je hebt me overtuigd überhaupt een PhD te gaan doen, gaf me de vrijheid om met mijn eigen projecten te komen, gaf me sturing op momenten dat ik afgedwaald dreigde te raken en beurde me op bij de zoveelste afwijzing van een paper. Ook bedankt dat je een ally bent voor vrouwen (in het) voetbal, en vrouwen in de academische wereld in het algemeen.

Verder wil ik de jury bedanken: Jan, Wannes, Danny, Sigrid, Matthias en Lucie. Jullie hebben me uitgedaagd om dit proefschrift tot zijn beste versie te krijgen. In het bijzonder wil ik Jan nog bedanken met wie ik verder terugga. Jij gaf me ooit het gevoel dat ik een PhD aan zou kunnen, iets wat ik zelf nooit in mezelf had gezien.

Daarnaast ook de mede-PhDers. In het bijzonder Pieter en Maaike. Ik was natuurlijk een vreemde eend in de bijt: kwam sporadisch naar Leuven en sprak ook net een ander taaltje, maar ik heb me altijd heel erg welkom gevoeld als ik langskwam. Het betekende voor jullie ook een pizzalunch, dus ik snap wel dat jullie blij waren als ik weer eens langs kwam.

Het combineren van een PhD met een fulltime baan was niet altijd even makkelijk en soms ook best eenzaam. Bedankt aan alle (oud-)collega's voor de interesse in mijn PhD en het begrip op momenten dat ik tijd nodig had voor mijn PhD.

In het bijzonder nog dank aan SciSports voor de ondersteuning bij de start van de PhD en het leveren van de data voor een groot deel van de onderzoeken. Bedankt Tim voor de fijne samenwerking. Ik heb veel van je geleerd. Bedankt Mirelle, voor de goede gesprekken over vrouwenvoetbal en de vele waardevolle introducties bij mensen in deze wereld.

De meeste support heb ik toch wel gehad van de mensen om me heen. Te beginnen met mijn voetbalteam. Bedankt dat jullie de verhalen over halfspaces, expected goals, penalty's, en AI (voordat het een hype was) aan wilden horen. En voor mij is de beste remedie voor een vol hoofd altijd een potje voetbal, waar de gedachten uit kunnen en de voetjes doen wat ze willen (1 hakje is geen hakje). Ik vind jullie een prachtig gezelschap en heb altijd plezier met jullie.

Lars (Arie), Elke en Jorieke, die me juist konden laten inzien dat we het niet altijd te serieus moeten nemen.

Anouk, Iris en Zaja, die juist meer met data dan met voetbal hebben, maar wel altijd interesse toonden.

Dan zijn er nog een aantal personen die vanaf het begin van de PhD nauw betrokken waren en de mensen die ook weten dat het leven niet altijd een vega bitterbal is. Jen, Rens, Kate, Marjolein en Irina: bedankt dat jullie er ook zijn op de moeilijkere momenten.

Stijn en Camiel, en Daphne en Fransje. En sinds kort ook Emma, die altijd vrolijk is.

Als laatste mijn ouders. Jullie hebben me altijd laten zijn wie ik wilde zijn. Als enige meisje op de voetbalclub. Als enige meisje op de schaakclub. Toen vond ik dat normaal, later begreep ik dat niet alle ouders dat goed hadden gekeurd. Met andere ouders was ik wellicht op turnen gezet en had ik moeten tekenen en knutselen. Daar ligt mijn talent en plezier niet. Dan was dit proefschrift er waarschijnlijk ook niet geweest.

Ik voel jullie trots en dan eigenlijk niet zozeer op het goed zijn met getallen, dat was er altijd al. Ik voel dat jullie trots zijn dat dat ooit zo ontzettend verlegen meisje het nu leuk vindt om op een podium te staan. “Soms ben je gewoon te bang”. Bedankt dat jullie me altijd mezelf hebben laten zijn en mijn dromen hebben laten volgen. Profvoetbalster bij de mannen van FC Utrecht ben ik dan niet geworden, maar ik ben een hele gelukkige vrouwmeid met allemaal leuke, lieve en grappige mensen om me heen.

Bedankt dus aan iedereen die me waardeert zoals ik ben. Ik zou willen dat iedereen overal ter wereld elkaar zo laat zijn zoals jullie mij laten zijn wie ik ben. Dan zou de wereld voor veel mensen een beter plekje zijn.

Popularized Abstract

In recent years, soccer has entered a data revolution. From passes and dribbles, to tackles and player positioning, nearly every detail of a match can now be recorded and analyzed. As a result, clubs are increasingly using data in their decision-making process. Data can especially play a big role in player recruitment and match preparation.

Player recruitment encompasses finding the right player to acquire. However, the global player pool is too large to monitor manually and player performance in one league may not transfer directly to the other due to differences in playing style and level. Additionally, successful transfers depend not only on individual performance but also on tactical and team fit. Match preparation is equally demanding: coaching staff must analyze opponents and prepare set pieces with only a few days between matches. Although video analysis supports this process, time constraints limit how thoroughly staff can assess footage. The increasing availability of complex data from matches provides opportunities to address these challenges using Artificial Intelligence (AI) and Machine Learning (ML) techniques.

Quantifying player performance is one of those challenges that has been addressed using AI techniques. Traditionally, player performance was measured using simple counting statistics such as the number of shots on target or the percentage of successful passes. The introduction of more sophisticated data enabled the development of metrics that more accurately measure player performance. One of the first metrics was Expected Goals (xG), which measures the likelihood of a shot yielding a goal. Since shots only constitute a small proportion of the actions in a match, other metrics have been introduced that value all on-the-ball actions such as passes, dribbles and interceptions. However, these metrics often ignore the broader context in which they were executed. That is, they ignore context such as the player's teammates and the mental aspect. This dissertation addresses this gap by proposing novel AI-based frameworks to 1) quantify player chemistry and 2) player performance under mental pressure.

Furthermore, this dissertation takes a closer look at women's soccer and addresses whether AI models developed for the men's game can simply be transferred to the women's game. Furthermore, it provides the first large-scale data-driven analysis of the technical and tactical developments in the women's game and introduces a novel metric to measure a player's pace of play.

Finally, data-driven analysis of penalty kicks is limited. Moreover, existing work assumes that goalkeepers and kickers act independently, oversimplifying the complexity of penalty kicks. This dissertation introduces a simulation framework using AI techniques to optimize goalkeeper policies taking into account player action capacities.

In summary, this dissertation has three main contributions to the field of soccer analytics: 1) understanding women's soccer using data, 2) capturing more context in data-driven recruitment, and 3) optimizing goalkeeper policies for penalty kicks.

Gepopulariseerde Samenvatting

In de afgelopen jaren is het voetbal een datarevolutie ingegaan. Van passes en dribbels tot tackles en spelerposities — vrijwel elk detail van een wedstrijd kan nu worden vastgelegd en geanalyseerd. Als gevolg daarvan maken clubs in toenemende mate gebruik van data in hun besluitvormingsprocessen. Data kan vooral een grote rol spelen in het scoutingsproces en de wedstrijdvoorbereiding.

Het scoutingsproces omvat het vinden van de juiste speler om aan te trekken voor jouw team. De wereldwijde spelerspool is echter te groot om handmatig te monitoren, en prestaties van spelers in de ene competitie vertalen zich niet per se naar een andere competitie vanwege verschillen in speelstijl en niveau. Daarnaast hangt het succes van een transfer niet alleen af van individuele prestaties, maar ook van tactische en teammatige fit. De wedstrijdvoorbereiding is minstens zo veeleisend: de technische staf moet tegenstanders analyseren en standaardsituaties voorbereiden met slechts een paar dagen tussen de wedstrijden. Hoewel videoanalyse dit proces ondersteunt, zorgen tijdsdruk en deadlines ervoor dat stafleden niet alle relevante beelden grondig kunnen bekijken. De toenemende beschikbaarheid van complexe data uit wedstrijden biedt kansen om deze uitdagingen aan te pakken met behulp van technieken uit Kunstmatige Intelligentie (AI) en Machine Learning (ML).

Het kwantificeren van spelersprestaties is één van die uitdagingen die inmiddels met AI-technieken is aangepakt. Traditioneel werd de prestatie van een speler gemeten aan de hand van eenvoudige statistieken zoals het aantal schoten op doel of het percentage geslaagde passes. De introductie van meer geavanceerde data heeft het mogelijk gemaakt om metrieken te ontwikkelen die spelersprestaties beter in kaart brengen. Een van de eerste metrieken was Expected Goals (xG). Deze metriek meet de kans dat een schot in een doelpunt resulteert. Aangezien schoten maar een klein deel van de acties in een wedstrijd uitmaken, werden er

metrieken geïntroduceerd om alle aan-de-bal acties te waarderen, zoals passes, dribbels en intercepties. Echter, deze metrieken negeren vaak de bredere context waarin deze acties plaatsvinden. Ze houden bijvoorbeeld geen rekening met de medespelers van de speler of met mentale factoren. Dit proefschrift pakt die tekortkoming aan door nieuwe AI-gebaseerde raamwerken voor te stellen om 1) spelerschemie te kwantificeren en 2) prestaties onder mentale druk te meten.

Daarnaast richt dit proefschrift zich op het vrouwenvoetbal en onderzoekt het of AI-modellen die ontwikkeld zijn voor het mannenvoetbal zomaar overgezet kunnen worden naar het vrouwenvoetbal. Verder biedt het de eerste grootschalige, data-gedreven analyse van de technische en tactische ontwikkelingen in het vrouwenvoetbal en introduceert het een nieuwe metriek om de handelingssnelheid van een speler te meten.

Ten slotte is data-gedreven analyse van penalty's beperkt. Bestaande studies gaan er bovendien van uit dat keepers en nemers onafhankelijk van elkaar handelen, wat de complexiteit van penalty's oversimplificeert. Dit proefschrift introduceert een simulatiekader dat met behulp van AI-technieken keeperstrategieën optimaliseert, waarbij rekening wordt gehouden met de capaciteiten van spelers.

Samenvattend levert dit proefschrift drie belangrijke bijdragen aan het veld van voetbalanalyse: 1) het beter begrijpen van vrouwenvoetbal met behulp van data, 2) het vastleggen van meer context in data-gedreven scouting, en 3) het optimaliseren van keeperstrategieën bij penalty's.

Abstract

The rapid growth in the amount of data collected from soccer matches has transformed how the game is analyzed. Soccer clubs increasingly rely on data in their decision-making processes. Player recruitment and match preparation are two critical tasks in professional soccer that could benefit from the use of data.

Player recruitment encompasses finding the right player to acquire for your team. However, the global player pool is too large to monitor manually and player performance in one league may not transfer directly to the other due to differences in playing style and level. Additionally, successful transfers depend not only on individual performance but also on tactical and team fit. Match preparation is equally demanding: coaching staff must analyze opponents and prepare set pieces with only a few days between matches. Although video analysis supports this process, time constraints limit how thoroughly staff can assess footage. The increasing availability of complex data from matches provides opportunities to address these challenges using Artificial Intelligence (AI) and Machine Learning (ML) techniques.

Quantifying player performance is one of those challenges that has been addressed using AI techniques. Traditionally, player performance was measured using simple counting statistics such as the number of shots on target or the percentage of successful passes. The introduction of more sophisticated data enabled building metrics that better measure player performance. One of the first metrics was Expected Goals (xG), which measures the likelihood of a shot yielding a goal. Since shots only constitute a small proportion of the actions in a match, other metrics have been introduced that value all on-the-ball actions such as passes, dribbles and interceptions. However, these metrics often ignore the broader context in which they were executed. That is, they ignore context such as the player's teammates and the mental aspect. This dissertation addresses this gap by proposing novel AI-based frameworks to 1) quantify player chemistry and 2) player performance under mental pressure.

Furthermore, the field of soccer analytics has largely focused on men's soccer, with limited attention to the women's game. Most existing performance models were trained on men's data. Given the machine learning principle that models should be trained on representative data, it raises the question whether such metrics could simply be transferred to the women's game. This dissertation addresses this question. Furthermore, it provides the first large-scale data-driven analysis of the technical and tactical developments in the women's game. It also introduces a novel metric to measure a player's pace of play.

Finally, data-driven analysis of penalty kicks is limited and often models penalty kicks as a two-player zero-sum game, assuming that goalkeepers and kickers choose their actions independently. However, this oversimplifies reality as kickers and goalkeepers may adapt their decisions based on the opponent's movements. Moreover, not all players can adapt their actions in this way, and small sample sizes make it hard to assess individual player capacities. This dissertation introduces a simulation framework using AI techniques to optimize goalkeeper policies taking into account these action capacities.

In summary, this dissertation has three main contributions to the field of soccer analytics: 1) understanding women's soccer using data, 2) capturing more context in data-driven recruitment, and 3) optimizing goalkeeper policies for penalty kicks.

Beknopte samenvatting

De snelle groei in de hoeveelheid data die wordt verzameld bij voetbalwedstrijden heeft de manier waarop het spel geanalyseerd wordt veranderd. Voetbalclubs vertrouwen steeds meer op data in hun besluitvormingsprocessen. Het scoutingsproces en de wedstrijdvoorbereiding zijn twee cruciale taken in het professionele voetbal die kunnen profiteren van het gebruik van data.

Het scoutingsproces omvat het vinden van de juiste speler om aan te trekken voor jouw team. De wereldwijde spelerspool is echter te groot om handmatig te monitoren, en prestaties van spelers in de ene competitie vertalen zich niet per se naar een andere competitie vanwege verschillen in speelstijl en niveau. Daarnaast hangt het succes van een transfer niet alleen af van individuele prestaties, maar ook van tactische en teammatige fit. Wedstrijdvoorbereiding is minstens zo veeleisend: de technische staf moet tegenstanders analyseren en standaardsituaties voorbereiden met slechts een paar dagen tussen de wedstrijden. Hoewel videoanalyse dit proces ondersteunt, zorgen tijdsdruk en deadlines ervoor dat stafleden niet alle relevante beelden grondig kunnen bekijken. De toenemende beschikbaarheid van complexe data uit wedstrijden biedt kansen om deze uitdagingen aan te pakken met behulp van technieken uit Kunstmatige Intelligentie (AI) en Machine Learning (ML).

Het kwantificeren van spelersprestaties is één van die uitdagingen die inmiddels met AI-technieken is aangepakt. Traditioneel werd de prestatie van een speler gemeten aan de hand van eenvoudige statistieken zoals het aantal schoten op doel of het percentage geslaagde passes. De introductie van meer geavanceerde data heeft het mogelijk gemaakt om metrieken te ontwikkelen die spelersprestaties beter in kaart brengen. Een van de eerste metrieken was Expected Goals (xG). Deze metriek meet de kans dat een schot in een doelpunt resulteert. Aangezien schoten maar een klein deel van de acties in een wedstrijd uitmaken, werden er metrieken geïntroduceerd om alle aan-de-bal acties te waarderen, zoals passes, dribbels en intercepties. Echter, deze metrieken negeren vaak de bredere context waarin deze acties plaatsvinden. Ze houden bijvoorbeeld geen rekening met de

medespelers van de speler of met mentale factoren. Dit proefschrift pakt die tekortkoming aan door nieuwe AI-gebaseerde raamwerken voor te stellen om 1) spelerschemie te kwantificeren en 2) prestaties onder mentale druk te meten.

Daarnaast heeft het veld van voetbalanalyses zich grotendeels gericht op het mannenvoetbal, met beperkte aandacht voor het vrouwenvoetbal. De meeste bestaande prestatiemodellen zijn getraind op data uit het mannenvoetbal. Gezien het principe in Machine Learning dat modellen getraind moeten worden op representatieve data, roept dit de vraag op of dergelijke metrieken zomaar overgedragen kunnen worden naar het vrouwenvoetbal. Dit proefschrift onderzoekt deze vraag. Daarnaast levert het de eerste grootschalige data-gedreven analyse van de technische en tactische ontwikkelingen in het vrouwenvoetbal. Ook introduceert het een nieuwe metriek om de handelingssnelheid van spelers te meten.

Ten slotte is data-gedreven analyse van penalty's beperkt en wordt deze vaak gemodelleerd als een zero-som-spel tussen twee spelers, waarbij wordt aangenomen dat keepers en nemers hun acties onafhankelijk van elkaar kiezen. Dit vereenvoudigt echter de werkelijkheid, aangezien keepers en nemers hun beslissingen kunnen aanpassen op basis van de bewegingen van de tegenstander. Bovendien is niet elke speler in staat om zijn of haar acties op deze manier aan te passen, en maken het kleine aantal penalty's per speler het moeilijk om individuele capaciteiten te beoordelen. Dit proefschrift introduceert een simulatiekader dat met behulp van AI-technieken keeperstrategieën optimaliseert, waarbij rekening wordt gehouden met de capaciteiten van spelers.

Samenvattend levert dit proefschrift drie belangrijke bijdragen aan het veld van voetbalanalyse: 1) het beter begrijpen van vrouwenvoetbal met behulp van data, 2) het vastleggen van meer context in data-gedreven scouting, en 3) het optimaliseren van keeperstrategieën bij penalty's.

List of Abbreviations

AI Artificial Intelligence.

ML Machine Learning.

PoP Pace of Play.

SPADL Soccer Player Action Description Language.

VAEP Valuing Actions by Estimating Probabilities.

xG Expected Goals.

xT Expected Threat.

Contents

Popularized Abstract v

Gepopulariseerde Samenvatting vii

Abstract ix

Beknopte samenvatting xi

List of Abbreviations xv

List of Symbols xvii

Contents xvii

List of Figures xxiii

List of Tables xxix

1 Introduction 1
- 1.1 Gaps in the Literature . . . 5
- 1.2 Dissertation Statement . . . 6
- 1.3 Contributions and Structure of the Dissertation . . . 7
  - 1.3.1 Background . . . 7
  - 1.3.2 Contribution 1: Understanding Women's Soccer Using Data . . . 7
  - 1.3.3 Contribution 2: Measuring Player Performance in Broader Context . . . 9
  - 1.3.4 Contribution 3: Optimizing Goalkeeper Policies for Penalty Kicks . . . 10
  - 1.3.5 Conclusion . . . 10

2 Background 11

2.1 Soccer . . . 11
2.2 Data . . . 12
2.2.1 Matchsheet Data . . . 12
2.2.2 Match Event Stream Data . . . 13
2.2.3 Tracking Data . . . 16
2.3 Measuring Player Performance in Soccer . . . 17
2.3.1 Shots . . . 17
2.3.2 Passes and Other Actions . . . 19
2.3.3 xT . . . 20
2.3.4 VAEP . . . 22

3 Women's Football Analyzed: Interpretable Expected Goals Models for Women 25
3.1 Introduction . . . 25
3.2 Data . . . 27
3.3 Observations . . . 28
3.3.1 Matches and Competitions . . . 28
3.3.2 Shot Behaviour . . . 30
3.4 xG Models . . . 31
3.4.1 Challenges . . . 32
3.4.2 Data Set Sampling . . . 33
3.4.3 xG Models . . . 33
3.5 Results . . . 34
3.5.1 Model Inspection . . . 34
3.5.2 Model Evaluation . . . 37
3.5.3 Individual Shots with Big xG Differences . . . 37
3.5.4 xG Values Across Leagues and Gender . . . 39
3.6 Conclusion . . . 40

4 A Data-driven Analysis of the Technical and Tactical Evolution of Elite Women's Football 41
4.1 Introduction . . . 42
4.2 Data . . . 44
4.3 Approach . . . 46
4.3.1 Tactical Behaviour . . . 46
4.3.2 Technical Skills . . . 48
4.3.3 Sensitivity Analysis . . . 49
4.4 Results . . . 50
4.4.1 Evolution in Tactical Behavior . . . 50
4.4.2 Evolution in Technical Skill . . . 50
4.5 Discussion . . . 53

5 Measuring Player Pace of Play in Elite Female Association Football 57

5.1 Introduction . . . 57
5.2 Methods . . . 59
5.2.1 Data . . . 60
5.2.2 Relationship Between PoP and League/Team Strength . 60
5.2.3 Relationship Between PoP and Player Position . . . 62
5.2.4 Correlation Between PoP and Playing Time in a New League . . . 62
5.2.5 Identifying Interesting Players . . . 63
5.3 Results . . . 63
5.3.1 Relationship Between PoP and League/Team Strength . 63
5.3.2 Relationship Between PoP and Player Position . . . 65
5.3.3 Correlation Between PoP and Playing Time in a New League . . . 65
5.3.4 Identifying Interesting Players . . . 65
5.3.5 Discussion . . . 65

6 Measuring Player Performance under Mental Pressure 69
6.1 Introduction . . . 69
6.2 Measuring Mental Pressure . . . 71
6.3 Measuring Player Performance . . . 73
6.3.1 Contribution Rating . . . 75
6.3.2 Decision Rating . . . 75
6.3.3 Execution Rating . . . 76
6.4 Measuring Player Performance Under Mental Pressure . . . 77
6.5 Use Cases . . . 78
6.5.1 Player Acquisition: Replacing Riyad Mahrez at Leicester City . . . 79
6.5.2 Training: Dom Dwyer's Fouls . . . 79
6.5.3 Tactical Planning . . . 80
6.5.4 Lineups and Substitutions . . . 83
6.6 Observations . . . 85
6.6.1 Neymar's Performances Decline Under Mental Pressure 85
6.6.2 MLS Stars of the Future . . . 86
6.6.3 Liverpool Recruitment . . . 88
6.6.4 Fatigue . . . 88
6.7 Related Work . . . 88
6.8 Conclusions . . . 91

7 Player Chemistry: Striving for a Perfectly Balanced Soccer Team 93
7.1 Introduction . . . 93
7.2 Dataset . . . 94
7.3 Measuring the Mutual Chemistry for a Pair of Players . . . 95
7.3.1 Joint Offensive Impact (JOI) . . . 96

7.3.2 Joint Defensive Impact (JDI) 97
7.4 Predicting the Mutual Chemistry for a Pair of Players 101
7.4.1 Representing Players as Feature Vectors 101
7.4.2 Predicting the Joint Impact for a Pair of Players 102
7.4.3 Analyzing the Machine-Learned Models 102
7.5 Automatically Assembling a Maximum-Chemistry Team 103
7.6 Observations 104
7.6.1 Highest-Chemistry Player Pairs 104
7.6.2 Mutual Chemistry for the 2018/2019 Liverpool Side . . 107
7.6.3 Mesut Özil's Chemistry at Arsenal Over Time 108
7.7 Use Cases 109
7.7.1 Which Center Back Should Manchester City Acquire? . 111
7.7.2 Which Right Winger Should Real Madrid Field? 111
7.7.3 Which Club Should Hakim Ziyech Move To? 113
7.8 Conclusions 117

8 Stop Guessing: Optimizing Goalkeeper Policies for Soccer Penalty Kicks 119
8.1 Introduction 119
8.2 Preliminaries 121
8.2.1 Penalties from a Takers' Perspective 122
8.2.2 Penalties from a Goalkeepers' Perspective 122
8.2.3 Data 123
8.3 Penalty Game Theory with Player Tactics 124
8.4 Modeling Goalkeeper Policies for Penalty Kicks 125
8.4.1 Specifying Educated Policies with Learned Models . . . 126
8.4.2 Computing $p(c|k, \pi, gk)$ 127
8.4.3 Computing $p(s|k, c, \pi, gk)$ 128
8.5 Evaluation 129
8.5.1 Evaluating Goalkeeper Policies 129
8.5.2 Initial Goalkeeper Position 130
8.5.3 Evaluating the Learned Models 131
8.5.4 Illustrative Use Cases 132
8.6 Related Work 134
8.7 Conclusion 135

9 Conclusion 137
9.1 Summary 137
9.1.1 Understanding Women's Soccer Using Data 137
9.1.2 Capturing More Context in Data-Driven Recruitment . 138
9.1.3 Optimizing Goalkeeper Policies for Penalty Kicks 139
9.2 Critical Reflections 140
9.2.1 Limitations 141

9.3 Future Work 142
9.3.1 Player Recruitment 142
9.3.2 Match Preparation 143

A Stop Guessing: Optimizing Goalkeeper Policies for Soccer Penalty Kicks 147
A.1 Data 147
A.1.1 World Cup Penalty Dataset 147
A.1.2 Event Stream Dataset 148
A.1.3 Merging the Data 148
A.1.4 Order of Takers 149
A.2 Features for the Penalty Distance and Direction Models 150
A.3 Learning the Uncertainty Values for Equation 2 151
A.4 Model Learning 153
A.4.1 Hyperparameter Tuning 153
A.4.2 Calibration Plots for the Penalty Direction Model . . . 153
A.4.3 Insights in Feature Combinations for Learned Models . 154

Bibliography 157

Curriculum Vitae 179

List of Figures

2.1 Zones across the width of the pitch. Direction of attack is from left to right. 13
2.2 Matchsheet data for the 2019 Women's World Cup semi-final between the Netherlands and Sweden. Available from: https://women.soccerway.com/matches/2019/07/03/world/womens-world-cup/netherlands/sweden/2972754/ [Accessed 7 April 2025] 14
2.3 The sequence leading up to Jackie Groenen's winning goal in the 2019 FIFA Women's World Cup semi-final between the Netherlands and Sweden. 16

3.1 Distribution of goals for women's (orange) and men's games per (a) each quarter of the season, and, (b) each 15 minute interval of a match. 29
3.2 Difference in proportion of shots with the (a) foot and (b) head from each location. Orange (blue) areas indicate more shots in women's (men's) matches. 31
3.3 We split our data sets into a train and a test set using a temporal split and train our xG models on three different train sets; a women only, a men only and a combined set. 32
3.4 Partial dependence plots for the angle to the goal and the distance to the goal for the WomenGAM (orange lines) and the MenGAM (blue lines). The histograms on the bottom show the features' distributions of values in the combined test set. 34
3.5 Partial dependence plots for features describing the body part used for the shot and action preceding the shot for the WomenGAM (orange lines) and the MenGAM (blue lines). The histograms on the bottom show the features' distributions of values in the combined test set. 36

3.6 Partial dependence plots for the features describing the game state for the WomenGAM (orange lines) and the MenGAM (blue lines). The histograms on the bottom show the distribution of values for the feature in the combined test set. 37
3.7 xG calibration plots 38
3.8 Marcus Rashford's 62nd minute goal from the "Robben" location in the 2-1 win against Brighton on April 4th, 2021. 39
3.9 Tip-in by Jordan Nobbs following a Beth Mead cross in the 4-0 Arsenal win against Everton on December 20th, 2020. 39
3.10 Late Victoria Williams miss in the 95th minute following a freekick in the 3-1 loss against Tottenham on December 6th, 2020. 40

4.1 The sequence leading to Aitana Bonmati's goal in the 2021 Champions League final. For each action, the data records its type, the player and team performing the action, the x and y coordinate of the action's start and end location, and the time in the match the action was executed. 44
4.2 The accuracy on deep passes for each position line as a function of time. The accuracy of these passes has increased for all position groups over time. 52
4.3 The success probability for passes that would be in the top 5% of xT value if they were completed. The accuracy of these passes has increased over time. 52
4.4 The difference in end locations for shots from inside the box plotted on the goal. Green (orange) indicates a higher proportion of shots end in the location in the NEW (OLD) dataset end in a location. The values in the colorbar correspond to the proportion difference between the two datasets in percentage points. We observe a higher proportion of shots aimed at the far corner in the NEW dataset. 53

5.1 Event data for Aitana Bonmati's header goal in the 41st minute in the match against Finland during the 2021 European Championship. For all actions the time of execution, the player and team involved, the type of action, the result and the bodypart used are annotated. 61
5.2 The distribution of average PoP by position group. Goalkeepers and Center backs have on average the slowest pace, whereas midfielders (both center and attacking) show the fastest pace. . 65

5.3 The relationship between the percentage of matches played after a transfer and the z-score comparing the player's PoP to the new league's PoP when making an interleague transfer. We observe that a lower z-score, which means that the player had a lower average PoP than the new league's average PoP, corresponds with appearing in a higher percentage of matches on average at their new club. 66

6.1 During the Tottenham - Manchester City match on October 29th, 2018, David Silva chose to pass to his teammate Raheem Sterling instead of to shoot when he was several meters in front of an open goal. This is the lowest-rated decision for the 2018/2019 Premier League season up until November 11th, 2018. 74
6.2 A comparison between Riyad Mahrez' and his suggested replacements at Leicester City: Aouar, Ghezzal and Shaqiri. Graph A compares the contribution ratings per action type for these players. Graphs B, C and D compare their performance under pressure. 80
6.3 A comparison between the fouls committed by Dom Dwyer in low-pressure and high- pressure situations. While under pressure, Dwyer commits significantly more needless fouls, handing ball possession to the opponent. 81
6.4 Manchester United typically attempts to use dribbles to force a breakthrough in high- pressure situations. Yet, the contribution from these dribbles decreases, while the contribution from crosses increases under pressure. This observation suggests that, in tense games, crossing the ball might be a more valuable strategy than dribbling. 82
6.5 Average decision ratings per 90 minutes of Leicester City's throw-ins, compared to the league average (left) and probability maps of where these throw-ins end up (right). Data is modeled as a two-dimensional Gaussian distribution with red zones showing where Leicester is more likely to throw under pressure (compared to when the pressure is low) and blue zones where they are less likely to throw. 84
6.6 A comparison of the average contribution per 90 minutes between Juventus' defenders, PSG's offensive midfielders and Chelsea's strikers. These graphs show that it might be worthwhile to select different players in high-pressure situations. For Giroud, we include data from his period at Arsenal due to lack of enough high-pressure moments at Chelsea. 85

6.7 Under pressure, Neymar evolves from a player with a very high average contribution per 90 minutes to an average player. This decrease in performance can mainly be attributed to poor decision making in high-pressure situations. 86
6.8 The percentile rank contribution ratings for MLS talents Alphonso Davies, Tyler Adams and Ezequiel Barco under different pressure levels. 87
6.9 Average contribution per 90 minutes of Liverpool's recent signings in low- and high-pressure situations at their former clubs. For comparison, we include the average contributions of an average player (by position) irrespective of the mental pressure situation. Since the 2017/2018 season, Liverpool has almost consistently signed players who perform better in high-pressure situations. Naby Keïta is the only exception. 89
6.10 A comparison between each player's performance in high-pressure situations and performance under fatigue (i.e., in the last quarter of a game, excluding the high-pressure situations). Performance is measured as the average contribution per 90 minutes. The low Pearson correlation of 0.136 indicates that our pressure metric does not measure fatigue. 90

7.1 The mutual offensive chemistry between the Liverpool players in the 2018/2019 English Premier League season. Green lines reflect high offensive chemistry links, whereas dark red lines reflect low offensive chemistry links. 108
7.2 The mutual defensive chemistry between the Liverpool players in the 2018/2019 English Premier League season. Green lines reflect high defensive chemistry links, whereas dark red lines reflect low defensive chemistry links. 109
7.3 The evolution of the average Joint Offensive Impact per 90 minutes for Mesut Özil, Arsenal and the average Premier League team over the course of the past five seasons. Both Özil's and Arsenal's average have rapidly declined since Alexis Sánchez' departure in 2018. 110
7.4 The top-ten-ranked Arsenal player pairs in terms of Joint Offensive Impact per 90 minutes since the start of the 2015/2016 season. Mesut Özil features in six of these ten player pairs, including the top-four-ranked pairs. Player pairs involving Mesut Özil are shown in orange. 110

7.5 Predicted defensive chemistry between potential targets Koulibaly, Škriniar and Alderweireld on one hand and Fernandinho, Otamendi, Stones and Rodrigo on the other hand. In terms of defensive chemistry with the defensive compartment, Koulibaly would be the best fit with all players but Otamendi. 112
7.6 Predicted offensive chemistry between potential targets Koulibaly, Škriniar and Alderweireld on one hand, and Fernandinho, Otamendi, Stones and Rodrigo on the other hand. In terms of offensive chemistry with the defensive compartment, Koulibaly would be the best fit. 113
7.7 Predicted offensive chemistry between potential targets Koulibaly, Škriniar and Alderweireld on one hand, and Jesus, Mahrez, Agüero, De Bruyne, D. Silva and B. Silva on the other hand. In terms of offensive chemistry with the offensive compartment, Alderweireld would be the best fit. 114
7.8 The average predicted offensive chemistry between Bale, Rodrygo and Vinícius Júnior on one hand, and all other Real Madrid players on the other hand. While Bale had the highest offensive chemistry at the start of the 2019/2020 season, he is expected to be quickly overtaken by both Rodrygo and Vinícius Júnior as they collect more matches. 115
7.9 The mutual offensive chemistry between each of Bale, Rodrygo and Vinícius Júnior on one hand, and Benzema, Kroos, Hazard and Isco on the other hand. Bale has the highest mutual offensive chemistry with all players but Hazard. 116
7.10 Predicted offensive chemistry between Hakim Ziyech and the other ten players in the maximum-chemistry teams of Bayern Munich, Chelsea, Roma, Internazionale, and Arsenal. 116
7.11 The mutual offensive chemistry between the Bayern Munich players for the maximum-chemistry team that includes Hakim Ziyech. 117

8.1 Each plot shows how the expected save percentage varies as function of the early dive range for a fixed late dive range. . . . 130
8.2 Expected save percentage for on-target penalties as function of moving x meters toward the kicker's natural corner for all goalkeeper policies. 131
8.3 Accuracy per range threshold for our Penalty Distance Model and baselines. Most shots are placed within 2.5 and 3 meters from the center, and this is also the area in which our model slightly outperforms the baselines. 132

A.1 The distribution of penalties (both shootout and non-shootout) taken by players prior to a shootout, per order in the shootout. We observe that the further we progress into the shootout, the less experienced players get. 150
A.2 The influence of within-reach save probability ρ and uncertainty factor μ on the save probability. 152
A.3 Calibration plot for predicted probabilities of saving a penalty given that you are in the correct corner using $r(\pi_e, gk) = 3.1$, $r(\pi_l, gk) = 2.8$, $\mu = 0.7$, $\rho = 0.7$. 153
A.4 Calibration plots for the predictions by the Penalty Direction Model for each of our classes (non-natural, center, natural). . . 155

List of Tables

3.1 Number of shots in our data set per league per season. Some of the 19/20 seasons are incomplete due to the Covid-19 pandemic. The 2020 NWSL season only contains 25 shots because it was cancelled due to the pandemic. 28
3.2 Brier scores for all six models on each of the three test sets. . . 37
3.3 The average xG value per shot and the percentage of low xG shots per league. We observe lower average xG values and a higher percentage of low xG shots in the men leagues. 40

4.1 The number of matches per season for each of the leagues in our dataset. 45
4.2 The average number of shots per match, the proportion of all shots and the average xG value per shot for all shots, shots taken with a foot, shots from outside the box, and shots from inside the box. 51
4.3 Results of our sensitivity analysis for our 4 datasets. For all findings, except for the shot conversion rate, we observe significant differences between the OLD versus NEW_{ALT} datasets, OLD_{ALT} versus NEW datasets, and OLD_{ALT} versus NEW_{ALT} datasets. For the NEW_{ALT} dataset we even observe some bigger differences in shot placement values and pass accuracy values. 54

5.1 The average and standard deviation of the PoP for each league. The average is computed over all matches in each league. The number of matches per league is also reported. The * indicates a significant difference according to a two sample t-test with Bonferonni correction between the WCL knockout stages PoP and the league's PoP. The WCL knockout stages, Primera División and WSL show the lowest PoP values, indicating that the ball is moved quickly by players in these matches. 64
5.2 The team with the lowest average PoP in each league for the 2021(/2022) season. We also report the team's end of season Elo rating and for the European teams, their absolute ranking out of the 66 teams (i.e., excluding the 10 NWSL teams). Chicago Red Stars is ranked third in terms of Elo rating in the NWSL. We also report the team's final position in their league table. Olympique Lyon and FC Barcelona have a much higher Elo rating than the other teams. For comparison, the third team in terms of Elo is Vfl Wolfsburg with an Elo rating of 1897. 64
5.3 The top 5 players in terms of their average PoP value in the 2021/2022 (or 2021 for the NWSL) season, and the team they played for in that season and their transfer value according to Soccerdonna (retrieved from soccerdonna.de) in January 2025. The last column shows the rank of the player's transfer value across all transfer values on Soccerdonna. 67

6.1 Some illustrative examples about what constitutes low, average, and high ratings according to each of our metrics: Contribution, Decision and Execution. 74

7.1 We obtain the responsibility share for each combination of player pair and opponent by computing the Euclidean distance between the positions of the players using this 5-by-5 grid. 100
7.2 Root Mean Square Error (RMSE) on the test set for both our models and the baselines. In both settings, our models outperform the baseline models. 103
7.3 Player pairs with the highest Joint Offensive Impact per 90 minutes across all seasons in our dataset with at least 900 minutes played together. 105
7.4 Player pairs with the highest Joint Offensive Impact per 90 minutes in the 2018/2019 season with at least 900 minutes played together or 450 minutes for the Champions League. 106
7.5 Player pairs with the highest Joint Defensive Impact per 90 minutes in the 2018/2019 season with at least 900 minutes played together or 450 minutes for the Champions League. 107

8.1 The empirical payoff matrix for the Penalty Kick on our dataset. 124
8.2 The expected save percentage for U1 (left) and UC2 (right) for goalkeepers GK1, GK2 and GK3 following the **early educated**, **late** and **mixed** strategies. 133

A.1 Optimal hyperparameters per fold for the Penalty Direction Model154
A.2 Optimal hyperparameters per fold for the Penalty Distance Model154

Chapter 1

Introduction

Soccer is one of the most popular sports in the world [149]. With an estimated 250 million soccer players around the globe, around 3% of the world's population plays soccer [80]. The sport also attracts a vast global audience, with 1.4 billion people watching the 2022 Men's World Cup final and 222 million watching the 2023 Women's World Cup final [83, 81]. Within the professional space, the main objective of soccer clubs is to win matches.

Among the many aspects that affect winning matches, two crucial factors to consider are squad composition and match preparation. First, winning requires recruiting players who form a strong squad [56]. This means selecting players whose strengths complement one another [24]. Second, every match demands a strong tactical plan that addresses the opportunities and threats posed by the opposing team [189, 98]. Moreover, it requires that a team's players are both physically and mentally fit. However, these two tasks have their challenges.

Player recruitment is hard for several reasons. First, the pool of players is large: approximately 19 thousand (semi-)professional female players [85] and 128 thousand professional male players [79] play the game globally. This makes it impossible to monitor all players closely, which raises the question of how to select the players that should be more critically examined. Second, these players are spread across multiple leagues. The quality of play (e.g., competitiveness, level of skill) can vary significantly among leagues. Moreover, the style of play varies across leagues. Consequently, it is difficult to assess how a player's performance will translate to a different league. Third, soccer is a team sport. Therefore, the success of a transfer also depends on how well a player fits the team's tactical game plan and the playing styles of their new teammates.

Match preparation brings its own set of challenges. The technical staff must analyze the upcoming opponent to identify their strengths and weaknesses in order to make a strong game plan. This includes selecting the starting line-up and determining a tactical plan, taking into account the players' individual qualities, playing styles, fitness levels, and their mutual chemistry [19]. In addition, set-pieces, such as free kicks, corner kicks and penalty kicks, require dedicated preparation [168]. Due to the high frequency of matches, especially for top teams competing both domestically and internationally, there is limited time to prepare for each game. Video analysis plays a crucial role in match preparation, supporting opponent scouting and set-piece analysis [139]. However, time constraints make it impossible to watch all relevant footage in depth.

Consequently, soccer practitioners must consider a wide array questions when thinking about how to optimize on-pitch performance. These questions are inherently difficult to answer, but data can play a role. In recent years, an increasing amount of data has been collected about soccer [58]. While, initially, such data was very coarse-grained consisting primarily of basic match information such as the lineups, substitutions, yellow and red cards, scoreline and goalscorers. Nonetheless, such data was widely collected. However, over the past 25 years much more granular data has been collected about soccer. For example, new types of data include:

Event stream data contains more detailed information about the game as it describes all on-ball actions in a game such as passes, dribbles and shots. This type of data is available for hundreds of professional men's leagues and tens of professional women's leagues. Event stream data only contains the position of the players involved in the action.

Tracking data contains the positions of all players and the ball at 10-50 frames per second. This type of data is very rich in information at the expense of lower availability. In women's soccer, up until recently no tracking data was collected. Moreover, for the women's game there is no publicly available tracking data. However, several clubs do track their own players.

Data can play a key role in addressing these challenges in player recruitment and match preparation. The vast amounts of data now collected from matches worldwide can be used to support decision-making. For example, in player recruitment, data could help to automatically identify players who excel at lower-tier clubs. In match preparation, it could be used to find player pairs that gel well together when selecting the starting line-up. However, the volume and complexity of this data make it difficult to analyze using traditional methods. Finding such patterns therefore requires advanced techniques from Artificial Intelligence (AI) and Machine Learning (ML). This thesis explores

how developing novel paradigms based on AI and ML applied to data from soccer matches can support clubs in making more informed decisions.

Quantifying player performance Player performance in soccer consists of many elements, which can be quantified in physical and technical performance metrics [157]. This dissertation specifically focuses on the technical performance of soccer players which reflects the skill-related abilities of soccer players to execute actions [186]. Before the introduction of more sophisticated data, player performance was measured using simple counting statistics such as the number of shots on target or the percentage of successful passes. Such counting statistics ignore important context in which actions occur. For example, one might think a team performed well in a match because it generated a large number of shots on target. However, when watching closely it turns out they attempted many long-distance shots which were easily saved by the opposing goalkeeper. Similarly, a center back might have a very high percentage of successful passes over the course of a season and therefore attract a scout's attention. However, when looking more closely, it turns out this player consistently opts for the safe option to pass the ball to their fellow center back. This yields a high percentage of successful passes, but does not tell us much about the player's passing qualities. Therefore, specific metrics have been developed to take such aspects into account.

An example of such a metric is the expected goals (xG) metric [96, 44]. xG quantifies the likelihood that a given shot yields a goal. This task can be formulated as a binary classification problem and most xG models are based on machine learning techniques [133, 175, 179, 134, 8]. This requires access to a large set of historical shots where a domain expert would typically hand-craft features about context under which the shot was taken. This may include the distance and angle to the goal, the type of assist (e.g., dribble, open play cross, corner kick), the body part used (e.g., foot, head) and the time in the match. Given this data, machine learning algorithms attempt to find patterns in how these features affect a shot's chance of resulting in a goal. For instance, an xG model might find that shots by foot within the penalty box are more likely to yield a goal than long-distance shots. This metric can assist recruitment by identifying players who convert more shots than expected based on the xG values of their shots [61]. Additionally, in match preparation, xG is useful for analyzing an opponent's chance creation.

However, xG solely focuses on shots, which account for less than 2% of the on-the-ball actions [62]. Consequently, other metrics have been developed to quantify the impact of other on-the-ball actions such as passes, dribbles and tackles [183, 144, 129, 36, 202, 62, 132, 165]. Two examples of such metrics

are the expected Threat (xT) metric [202], and the VAEP metric [62] of which I am one of the co-developers. Both metrics value actions by evaluating the value of the game state prior to the action and after the action. The game state is the current situation in the match, which may include the locations of the players and the ball, the scoreline, and time in the match. xT uses a discrete Markov Decision Process to assess the impact of passes and solely uses the start and end location of the pass. VAEP uses two machine learning models that estimate the probabilities to score and to concede a goal in the near future for each game state.

These models will thus assign more value to a pass from the center back to a teammate in the final third than to a lateral pass to the other center back. They can automatically identify players with a high expected impact with their on-the-ball actions. This holds true even if these players play in a lower-ranked team and have few assists, due to their teammates failing to convert their strong passes into goals. Action-value models can thus be of great value in the recruitment process.

Providing advice about game play Data also has the potential to inform the decisions that teams make during game play. Two areas that have received significant attention are set pieces such as corner kicks and penalty kicks.

Power et al. [168] showed in a data-driven manner that teams are more likely to score from in-swinging corners than from out-swinging or driven corners. They also demonstrated that teams defending corners with players on both posts concede most goals, whereas teams with a player on the back post concede the least. Other studies have explored the optimal placement of corner kicks using large-scale datasets [198, 209]. Wang et al. [218] used graph neural networks to predict the receiver of a corner kick and its outcome. These predictions could then be used by an AI assistant to suggest adjustments for corner kick tactics.

While corner kicks involve dynamic positioning of multiple players, penalty kicks are a static game between just two players: the goalkeeper and the kicker. This structured nature makes them well-suited for data-driven analysis using AI techniques. Yet, penalty kicks have been studied primarily from a psychological and social science perspective [90, 120, 11], while receiving less attention in data science.

Penalty kicks have been modeled as a 2-player zero-sum game between Goalkeeper and Kicker [12, 160, 210]. Here, the actions are shooting (diving) to the natural or non-natural side for the kicker. Few studies have analyzed goalkeeper and shooting strategies in a data-driven manner. Hunter et al. [111] proposed a model to predict the penalty success by incorporating goalkeeper

leave-times. Using this model, they analyze different shooting strategies and conclude that when facing a late goalkeeper, aiming on the ground close to the posts is most effective. Tuyls et al. [210] showed that players with different styles followed different strategies and had different Nash-derived recommendations. Busca et al. [42] trained a regression model on 412 penalties to predict the probability of a penalty kick being directed to the left or right side of the goal.

1.1 Gaps in the Literature

This thesis focuses on addressing the following three gaps in the literature.

1. A lack of data-driven analysis for women's soccer A common characteristic of all data-driven work mentioned in this introduction is that it is based solely on data from men's soccer. Although it has been noted that especially in women's soccer there is a lot to gain from data-driven recruitment [104], data-driven research on women's soccer is limited. Especially with regard to player performance, research is limited to traditional counting statistics [188, 161, 47, 93] or physical performance [27, 163, 46]. Moreover, performance metrics, such as xG, xT and VAEP, were all first developed using data from men's soccer. However, as is well-known in machine learning, models should be trained on data that is representative of the data in which they are applied on. This raises the question whether such models can be directly applied to data about the women's game.

Moreover, the women's game has evolved rapidly in recent years in terms of money involved, attention, and the number of players. These increased investments and professionalism might indicate evolutions in player performance and team tactical behavior. But such developments have not been investigated yet in a data-driven manner as most research in women's soccer focuses mostly on the physical aspects of the game.

2. Broader context in which actions are performed is ignored A player's performance can be affected by a multitude of factors, including mental and physical state, teammates, opponent, league level, weather conditions, and tactical fit. Performance metrics such as xG, xT and VAEP ignore such context. They are also often used without taking the broader context in consideration such as what league they play in, who their teammates are, and the mental pressure they feel during a match. For instance, player performance in lower-level leagues may not translate to higher level-leagues, as slower play in lower-level leagues allows players more time on the ball. To illustrate the importance

of team fit, consider a tall center forward that has excellent heading skills. However, this is apparent or exploitable if the team's tactics do not encourage crossing or if the team lacks players who excel at this skill. Finally, the mental pressure a player feels might affect their performance. However, this varies depending on a number of contextual factors. On the one hand, certain games are more "important" than others (e.g., a tournament final vs. friendly match, or a derby against a strong team vs. playing the worse side in the league). On the other hand, mental pressure also depends on the in-game context (e.g., there is less pressure when winning 3-0 than during a tied game with 10 minutes left). However, research on quantifying these contextual factors and their impact on player performance remains limited.

3. Penalty kick research ignores player capacities There is little research on penalty kicks from a data science perspective. Moreover, data-driven works model penalty kicks as a two-player zero-sum game [160, 210]. These works suffer from a significant limitation: they make the unrealistic simplifying assumption that goalkeepers and takers select their action – where to dive and where to the place the kick – independently of each other. In reality, this is not the case [74, 150, 151]. Some takers watch the movements of the goalkeeper during their run up and base the kick's direction on their analysis. Similarly, some goalkeepers wait and analyze a kicker's movements before deciding where to dive. This adds substantial complexity to the problem because not all players have the same action capacities, that is, only some players are capable of basing their decisions on their opponent's movements. Moreover, a player's action capacities may not be known to the opposing team due to the relatively small number of penalties that a player is involved in.

1.2 Dissertation Statement

This dissertation aims to address the gaps in the current literature using data science and machine learning techniques. Using event stream data the following questions will be answered:

1. Can we use a data-driven approach to understand (the developments in) women's soccer?
2. Can machine learning techniques be used to contextualize player performance?
3. Can goalkeeper penalty kick policies be optimized by taking into account action capacities?

1.3 Contributions and Structure of the Dissertation

This section describes the structure of this dissertation and summarizes its three contributions.

1.3.1 Background

Chapter 2 describes the background information to understand the content of the rest of this dissertation. It describes the game of soccer and the three data sources in more detail. Furthermore, it contains an overview of the relevant related work in the field of soccer analytics. The content of this chapter is based on the following publications where I am a (co)author:

L. Bransen et al. "How does context affect player performance in football?" In: *Football Analytics 2021: The Role of Context in Transferring Analytics to the Pitch* (2020)

T. Decroos et al. "Actions speak louder than goals: Valuing player actions in soccer". In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 1851–1861

T. Decroos et al. "VAEP: An objective approach to valuing on-the-ball actions in soccer". In: *International Joint Conference on Artificial Intelligence*. 2020, pp. 4696–4700

1.3.2 Contribution 1: Understanding Women's Soccer Using Data

This part of the dissertation describes the work performed on understanding women's soccer in a data-driven way.

Chapter 3 explores building xG models from data from the women's game and data from the men's game. Most soccer metrics were first developed for men's soccer. Typically, these models have been trained on data arising from men's matches and then they are also applied to data from women's matches. This chapter explores whether or not this approach is viable for xG models. To

address this question, event data from 9,076 matches across the five strongest domestic leagues in women's and men's soccer from 2017 to 2021 was used. Three subsets were created: **Women only**, **Men only** and **Combined** (all shots from both). xG models were trained separately on each subset using a standard train-test split. Next, these trained models were applied to all three test sets to assess how well the models generalized. The resulting xG values were further analyzed to explore differences between the models. This work shows that for xG models this is not optimal as there are differences in which goals are scored between men's and women's soccer. Therefore, models trained specifically for women's soccer should be used.

The content of this chapter is based on the following publication [32].

L. Bransen and J. Davis. "Women's football analyzed: Interpretable expected goals models for women". In: *Proceedings of the AI for Sports Analytics (AISA) Workshop at the International Joint Conference on Artificial Intelligence*. 2021

Chapter 4 performs the first large-scale, longitudinal analysis of women's soccer by looking at event data collected from 2013 to 2022. It uses analytics techniques to assess changes in technical skills and tactical behaviours in the women's game. Specifically, the dataset was split chronologically into an OLD and a NEW dataset. Several comparative analyses were conducted on the two datasets. Distinct xG and xT models were trained seperately on each dataset to capture the tactical and technical differences. Tactically, we observe longer possession sequences and fewer long-distance shots. Technically, we observe that players are completing more difficult and valuable passes at a higher rate, and putting a higher percentage of shots on target, with a particular emphasis on aiming for the corners of the goal.

The content of this chapter is based on the following publication [29].

L. Bransen and J. Davis. "A data-driven analysis of the technical and tactical evolution of elite women's football". In: *International Journal of Sports Science & Coaching* 19.6 (2024), pp. 2449–2459

Chapter 5 introduces a novel simple yet effective metric for the recruitment process in women's soccer: the Pace of Play (PoP) metric. It simply measures the amount of time a player takes to play a pass after gaining control of the ball either via receiving a pass or intercepting the ball. We assess the utility of

our metric by analyzing five seasons of data about six top women's leagues. We find that PoP is correlated with (1) team strength as measured by Elo ratings: stronger teams play at a faster pace, and (2) the amount of playing time after an interleague transfer: players that exhibit a faster pace appear in a greater percentage of games after transferring to a new league.

The content of this chapter is based on the following publication [30].

L. Bransen and J. Davis. "Measuring player pace of play in elite female association football". In: *International Journal of Sports Science & Coaching* (2025), p. 17479541251334022

1.3.3 Contribution 2: Measuring Player Performance in Broader Context

Chapter 6 introduces a framework to measure player performance under mental pressure. This framework consists of two main components: 1) measuring mental pressure and 2) measuring player performance. The mental pressure component combines two models: a pre-game mental pressure and an in-game mental pressure model. The pre-game model estimates the pre-game pressure based on factors such as team ambition (e.g., win league, avoid relegation), current league standing, rivalrousness and recent coaching changes. The in-game model utilizes an in-game win probability model to capture how pressure evolves during a match. The player performance component evaluates each on-the-ball action on three aspects: the total contribution (using a VAEP model), the quality of the decision and the quality of the execution. This framework was applied to data from 6,858 matches from seven men's leagues. It was shown that the framework has valuable applications in the recruitment process, match preparation and training.

The content of this chapter is based on the following publication [35].

L. Bransen et al. "Choke or shine? Quantifying soccer players' abilities to perform under mental pressure". In: *Proceedings of the 13th MIT Sloan Sports Analytics Conference*. 2019

Chapter 7 introduces a framework to measure and predict the chemistry between a pair of players. Using the VAEP metric it measures offensive and

defensive chemistry for player pairs. A machine-learned model is then trained to predict player-pair chemistry using features such as their playing style, age and key qualities. Next, the chapter also introduces the Team Builder, an approach to optimize line-up composition for maximum chemistry.

The content of this chapter is based on the following publication [38].

L. Bransen and J. Van Haaren. "Player Chemistry: Striving for a Perfectly Balanced Soccer Team". In: *Proceedings of the 14th MIT Sloan Sports Analytics Conference*. 2020

1.3.4 Contribution 3: Optimizing Goalkeeper Policies for Penalty Kicks

Chapter 8 introduces a player-agnostic simulation framework that can evaluate the efficacy of different goalkeeper strategies. It considers a rich set of choices and incorporates information about a goalkeeper's skills. Our work is grounded in a large dataset of penalties that were annotated by penalty experts and include aspects of both kicker and goalkeeper strategies. We show how our framework can be used to optimize goalkeeper policies in real-world situations.

The content of this chapter is based on the following paper [33].

L. Bransen et al. "Stop Guessing: Optimizing Goalkeeper Policies for Soccer Penalty Kicks". In: *arXiv preprint arXiv:2505.24629* (2025)

1.3.5 Conclusion

Chapter 9 concludes this dissertation. It summarizes the main findings and provides some directions for future work.

Chapter 2

Background

The content of this chapter is based on the following publications where I was a coauthor:

L. Bransen et al. "How does context affect player performance in football?" In: *Football Analytics 2021: The Role of Context in Transferring Analytics to the Pitch* (2020)

T. Decroos et al. "Actions speak louder than goals: Valuing player actions in soccer". In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 1851–1861

T. Decroos et al. "VAEP: An objective approach to valuing on-the-ball actions in soccer". In: *International Joint Conference on Artificial Intelligence*. 2020, pp. 4696–4700

2.1 Soccer

Soccer is an invasion sport in which two teams compete against each other. The game is played on a grass pitch with standard dimensions of 105 by 68 meters. There is a goal (7.32 meters by 2.44 meters) at each end of the pitch; see Figure

2.1. The objective is to score more goals than your opponent over the course of a 90-minute game (which is divided into two 45-minute halves). A goal is awarded when the ball fully crosses the plane of the goal mouth. Players can move the ball using any body part except for their arms and hands. There is one exception, each team's goalkeeper may touch the ball with their arms and hands within their team's penalty box.

Both teams start the game with eleven players, which we call the starting lineup. Typically, the players position themselves according to a formation like 4-3-3 or 4-4-2. These numbers refer to the number of outfield players across the position lines. For example, a 4-3-3 formation uses a goalkeeper, four defenders, three midfielders and three attackers. In modern soccer, formations are dynamic and often vary between the offensive, defensive and transition phases of team play [89]. During the game, the coach can substitute a maximum of five (since 2020) players.

In club soccer, a player can only play for one club. Twice a year there is a transfer window in which players can switch clubs. A player can make a permanent move (buy) or a temporary move (loan).

From an analysis perspective, it is common to talk about certain regions of the pitch. As visualized in Figure 2.1, it is common to divide the pitch heightwise into three equal-sized regions called the **own third** (i.e., first 35 metres of the pitch for the attacking team), **middle third** (i.e., middle 35 metres of the pitch) and **final third** (i.e., the last 35 metres of the pitch for the attacking team).

2.2 Data

While there are many types of data that can be collected about soccer matches [7], we will now briefly describe three types of data that feature prominently in the AI and machine learning literature: match sheet data, match event data and tracking data. These types of data differ in their availability and granularity. We explain each data type in more detail.

2.2.1 Matchsheet Data

Matchsheet data contains high-level information about the match such as the lineups, substitutions, yellow and red card bookings, the scoreline, goalscorers and assist-givers. It thus provides a high-level summary of what happened in the match. While this type of data is not very detailed, its advantage is that it is widely available. For men's soccer it is available for all professional matches

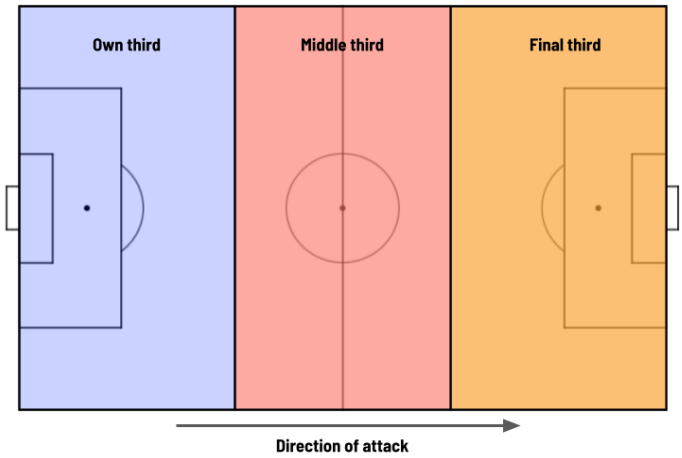


Figure 2.1: Zones across the width of the pitch. Direction of attack is from left to right.

and also for higher-level amateur matches across the world. In women's soccer, match sheet data is mostly available for the highest league in every country with a (semi)-professional league. Figure 2.2 shows an example of match sheet data from Soccerway for the match between the Netherlands and Sweden in the 2019 FIFA Women's World Cup.

2.2.2 Match Event Stream Data

Event stream data annotates the times and locations of specific events (e.g., passes, shots, and cards) that occur in a game. Vendors such as Opta, WyScout, and StatsBomb collect this type of data by having humans annotate match videos.

The nature of event stream data complicates analysis. Specifically, event data poses three key challenges from a data science perspective:

Challenge 1: Event stream data serves multiple different objectives (e.g., reporting information to broadcasters, analysts at soccer clubs). Hence, the data is not necessarily designed to facilitate data analysis. Some important information is missing (e.g., WyScout does not record exact

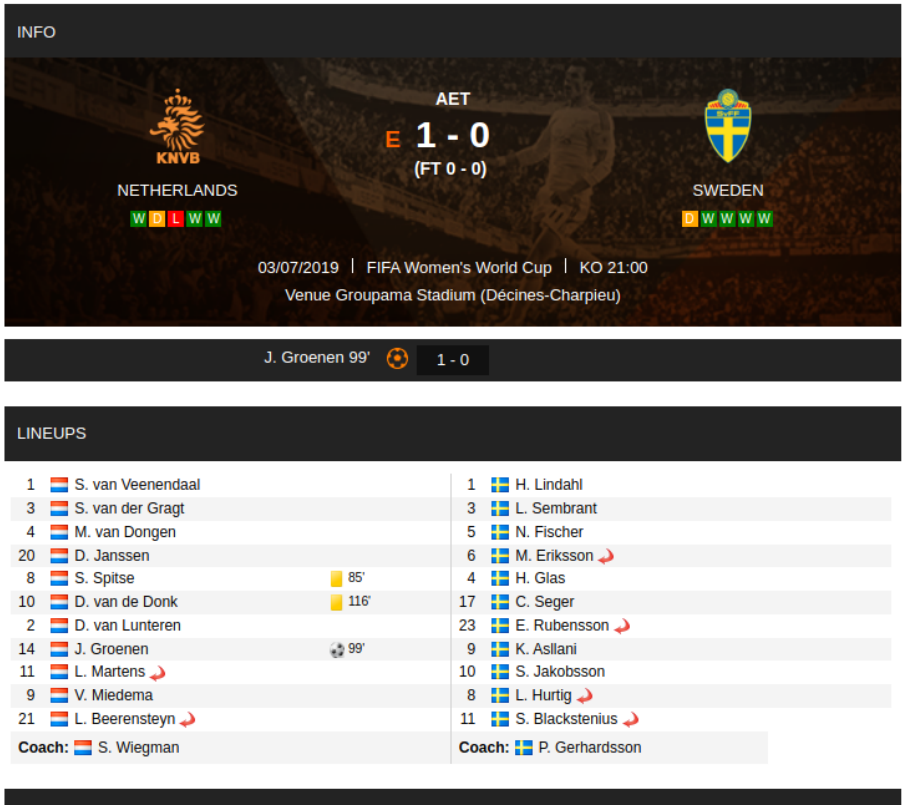


Figure 2.2: Matchsheet data for the 2019 Women's World Cup semi-final between the Netherlands and Sweden. Available from: https://women.soccerway.com/matches/2019/07/03/world/womens-world-cup/netherlands/sweden/2972754/ [Accessed 7 April 2025]

end locations for shots) or irrelevant information is included (e.g., Opta records weather changes).

Challenge 2: Each vendor of event stream data uses their own unique terminology and definitions to describe the events that occur during a game. Hence, software written for event stream data has to be tailored to a specific vendor and cannot be used without modifications to analyze data from another vendor.

Challenge 3: Most vendors offer optional information snippets per type of event. For example, Opta often specifies details about the exact type of foul that was committed. While sometimes useful, this information makes it extremely hard to apply automatic analysis tools.

SPADL The need for a uniform representation language is a typical AI problem. To this end, we [63] proposed SPADL, a language to describe soccer actions that unifies event stream formats from different vendors (e.g., WyScout, StatsBomb, Opta) into a common vocabulary that enables subsequent data analysis. SPADL describes player *actions*, as opposed to the formats by commercial vendors that describe *events*. The distinction is that actions are a subset of events that require a player to perform the action. For example, a passing event is an action, whereas an event signifying the end of the game is not an action.

Unlike other event stream formats, SPADL always stores the same eight attributes for each action: time, start location, end location, involved player, team of the involved player, the action type (e.g., *pass*, *shot*, *dribble*), the body part used to perform the action (e.g., *foot*, *head*), and the result of the action (e.g., *success* or *fail*). Storing the same eight attributes for each action addresses the third challenge, as excluding optional information snippets enables us to more easily apply automatic analysis tools.

SPADL distinguishes between 21 possible types of actions including, among others, *passes*, *crossed corners*, *dribbles*, *throw-ins*, *tackles*, *shots*, *penalty shots*, *clearances*, and *keeper saves*. These action types were, in collaboration with domain experts, designed to be interpretable and specific enough to accurately describe what happens on the pitch yet general enough such that similar actions have the same type.

Figure 2.3 shows the possession sequence leading to Jackie Groenen's winning goal in the Women's 2019 World Cup semi-final between the Netherlands and Sweden. The table on top of the figure shows how this possession sequence is represented in the SPADL format. The sequence started with a successful headed pass by Stephanie van der Gragt to Dominique Janssen, who received the ball and passed to Danielle van de Donk in midfield. Van de Donk then

	time	player	team	actiontype	result	bodypart
1	98m45s	Stephanie van der Gragt	Netherlands	pass	success	head
2	98m48s	Dominique Janssen	Netherlands	reception	success	foot
3	98m50s	Dominique Janssen	Netherlands	pass	success	foot
4	98m51s	Danielle van de Donk	Netherlands	reception	success	foot
5	98m53s	Danielle van de Donk	Netherlands	pass	success	foot
6	98m53s	Vivianne Miedema	Netherlands	challenge	success	foot
7	98m54s	Jackie Groenen	Netherlands	reception	success	foot
8	98m54s	Jackie Groenen	Netherlands	shot	success	foot

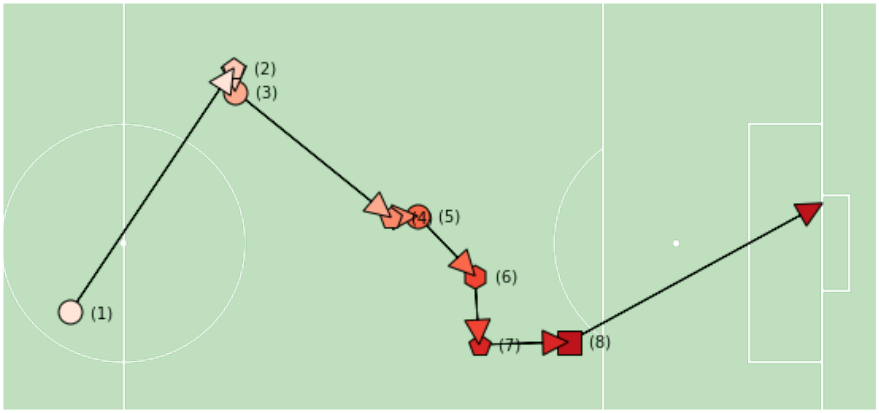


Figure 2.3: The sequence leading up to Jackie Groenen's winning goal in the 2019 FIFA Women's World Cup semi-final between the Netherlands and Sweden.

managed to get the ball to Vivianne Miedema who won a challenge which led to Jackie Groenen receiving the ball just outside the penalty box. From this position, Groenen shot and scored.

2.2.3 Tracking Data

Tracking data contains all player and ball positions at a high frequency (typically between 10 and 50 frames per seconds). This type of data enables for a more fine-grained analysis as all positions are available at all times. In contrast, traditional event stream data only contains the locations of the players involved in the action such as the passer and receiver for a pass event. Tracking data is either collected by a GPS tracker (e.g. Catapult Sports), in-stadium cameras (e.g. TRACAB, Second Spectrum) or broadcast videos (e.g. Stats Perform, SkillCorner). Providers that collect tracking data from broadcast videos estimate the positions of the players not visible in the broadcast videos. Tracking data

is thus very granular. However, it is more expensive and less widely available. For example, in-stadium tracking data is typically only available to a team in its own league or tournament [123].

2.3 Measuring Player Performance in Soccer

Football professionals and fans nowadays have access to insights from data that enable them to make better-informed decisions. Technological advances have enabled capturing more fine-grained information about the context in which football players perform their actions, which has spurred the development of novel metrics for quantifying player performance. Attention has gradually shifted from simply counting how often players perform certain actions to also accounting for how well players perform those actions. While quantity-based metrics usually treat all actions of a particular type equally (e.g., number of shots on target), quality-based metrics consider additional contextual information in order to differentiate among actions of the same type (e.g., expected-goals or xG values for shots). Furthermore, attention has gradually shifted from evaluating mostly offensive actions such as shots and passes, where the impact is relatively easy to quantify, to other types of actions such as interceptions and tackles where the impact is much harder to quantify.

Due to the low-scoring nature of soccer, the impact of the many actions that a player performs in a match is hard to quantify in isolation. As a result, the focus in the soccer analytics community has gradually shifted from quantifying each action's impact on the goals that were scored (e.g., counting assists or pre-assists) to quantifying each action's impact on increasing or decreasing the likelihood of scoring a goal from a particular game situation (e.g., computing xG values for shots). By doing so, the performance of a player can be assessed from a much larger, more representative set of actions.

2.3.1 Shots

Shot-based performance metrics have quickly evolved from context-free metrics such as the shot conversion rate (i.e., the fraction of a player's shots that result in a goal) to more context-aware metrics such as expected goals (xG). The objective of xG is to value the quality of a scoring opportunity. This is done by estimating the probability that a shot will result in a goal by relating characteristics of the shot (e.g., where the shot was taken from, scoreline, etc.) to its outcome (i.e., goal vs. no goal). xG models are trained using machine learning by framing this as a binary classification problem: each shot either

results in a goal (positive label) or not (negative label). The models are then trained on historical databases of shots.

xG models have evolved over time from considering little contextual information such as the location of the shot [96] to more sophisticated models that account for the locations of opponents and the height of the ball at the time of the shot [121]. The early xG models mostly include characteristics of the shot such as the distance to the goal, the shot angle and whether the shot was headed or not [96, 167]. These models distinguish between short- and long-distance shots and recognize that shots with the foot are easier to finish than shots with the head. However, they do ignore important contextual information such as the build-up to the shot. For instance, scoring from a fast counter attack is often easier than scoring from a slow build-up when the defense has more time to organize themselves. Therefore, most early xG models employed a set of models where each model addressed a particular situation such as footed open-play shots, headed open-play shots, and direct freekicks [45, 114]. The introduction of more granular match event data and spatio-temporal tracking data has eventually led to models that can identify, for instance, shots taken under pressure of defenders, shots where the goalkeeper is out of position and shots that require difficult body movements from the goalkeeper to reach the ball [121, 133, 8, 49, 224]. This also led to new work that contextualizes xG values based on the player taking the shot, the team, the league level and gender [176, 208, 147, 106, 193]. More recently, the introduction of skeleton data has led to a new generation of xG models [225, 223].

Beyond looking at the quality of the scoring opportunities, xG can also be used to evaluate the impact of a shot. Typically, this can be done by looking at a player's goals over expected by taking the difference between a shot's actual outcome (i.e., goal or no goal) and its xG value [16, 60]. With this metric, a player is positively rewarded for scoring a difficult long-range shot and negatively rewarded for missing a sitter. However, few players are able to consistently outperform their cumulative xG [166].

Although xG models are primarily used to evaluate the performances of attackers, they are also increasingly used to evaluate the performances of goalkeepers. While xG values provide insights into how many goals an attacker should have scored, expected-saves values (xS values) provide insights into how many on-target shots a goalkeeper should have saved. Attackers are typically positively rewarded for scoring more goals than expected, whereas goalkeepers are typically positively rewarded for conceding fewer goals than expected. However, two important differences exist between evaluating attackers and goalkeepers. First, unlike attackers, goalkeepers are only evaluated based on the shots that were on target and hence could possibly be saved. Second, while xS models are similar in spirit to xG models, they usually also take relevant post-shot information

into account such as the direction of the shot [128, 184].

2.3.2 Passes and Other Actions

Since shots constitute less than 2% of the on-the-ball actions in a match and primarily involve attacking players and goalkeepers, assessing the performances of defenders and midfielders remains hard using the aforementioned xG and xS models. Other types of on-the-ball actions happen much more often with passes being the most frequent type of action [62]. As a result, quantifying the impact of other types of on-the-ball actions, including passes, dribbles, interceptions and tackles among others, has gained more attention in recent years. Widespread count-based metrics are limited as they ignore important contextual information. For example, metrics such as passing accuracy and number of completed passes ignore the circumstances in which passes are executed. For example, a lateral pass from one central defender to the other is different from a perfect through ball from a midfielder that puts the striker in a one-on-one with the goalkeeper. While both passes would be counted as accurate, the latter pass has a considerably higher positive impact on the team's scoring chances than the former pass.

Analogous to shots, the general approach to quantifying the quality of an arbitrary on-the-ball action is to measure to what extent the action increases or decreases the likelihood of scoring a goal in the near future [183, 62, 202]. The commonality among most of these approaches is that they value the game states right before and after the action. Intuitively, the game state can be thought of as the current situation in the match, which may include the locations of the players and the ball, the scoreline, and the time remaining in the match. Since an action transitions a match from one game state to another, its value corresponds to the difference in value between the game states before and after the action. Hence, an action that transitions the match to a more favorable game state is positively rewarded, whereas an action that transitions the match to a less favorable game state is negatively rewarded.

Different approaches have been proposed to value game states with the key difference being the amount of contextual information that each approach takes into account. Some approaches solely consider the sequence that an action belongs to [66, 129], whereas other approaches also consider the location of the ball [202, 102]. However, these approaches cannot distinguish between, for example, a forward pass on the own half after regaining possession that initiates a counter-attack on one hand and a dribble in the build-up phase without any other players around the ball on the other hand. Therefore, several approaches that leverage more contextual information to determine the game state value

have been proposed such as xT [183, 202], VAEP [62], Goals Added [6] and PV [165]. These approaches can differentiate between different match situations (e.g., penalty, corner, free kick; [183]), type of action [207] and include the trajectory and speed of play to describe the game state [144, 6, 28, 36, 63, 132, 165]. These approaches can better differentiate between the aforementioned actions, but still have difficulties to capture contextual information such as pressure on the ball and players possibly blocking the pass lane.

To account for this contextual information, Power et al. [169] introduce an approach to measure the impact of single passes using spatio-temporal tracking data. They account for factors such as the pressure on the player in possession of the ball, and whether any players are between the passing player and the intended recipient. Moreover, they consider tactical concepts such as whether the pass is part of a counter-attack. Similarly, Goes et al. [94] use spatio-temporal tracking data to value passes. They introduce a defensive disruptiveness score to value the effectiveness of passes in disrupting the opponent's defensive organization. Fernandez et al.'s EPV metric [75] and Link et al.'s Action Value [131] use spatio-temporal tracking data to value actions besides passes. In their work, Fernandez et al. state that "a critical aspect to properly evaluate soccer situations is to have a clear understanding of the ongoing context". Both approaches account for the locations of all players to value on-the-ball actions.

More recently, work has been introduced to quantify availability, deep runs, pressing, and defensive actions [142, 205, 67, 13, 20, 97, 87, 158]. Moreover, there has been an increase of work on player decision-making [216, 95, 172, 181, 173, 218]. Transformers have also found their way to soccer analytics and have been used to model soccer data to better contextualize player performance in various ways [201, 140, 141, 182].

For this thesis, xT and VAEP are particularly important. Therefore, the following subsections discuss them in greater detail.

2.3.3 xT

Formally, xT is a discrete Markov Decision Process which is characterized by the following components:

- A finite set of states S. A state is either (a) transient, meaning that it can be exited, or (b) absorbing, meaning that once entered it cannot be exited and the episodic task ends. Typically, one considers 768 transient field states by partitioning the pitch into non-overlapping tiles of 3 meters by 3 meters. xT considers three absorbing states: (1) goal, which means that the team scores a goal, (2) no goal, which corresponds to shooting

and missing, and (3) loss of possession, which occurs when the other team acquires possession (e.g., ball goes out of bounds, pass is intercepted).

- A finite set of actions A. xT considers all actions that move the ball from one state to another. This is done by either dribbling or passing the ball from one of the transient states to another.
- A transition function S x A x $S \rightarrow [0, 1]$. This function represents the probability of moving from one state to another. In our case, a player possessing the ball in state s can either decide to shoot ($z(s)$) or decide to move the ball by passing or dribbling ($m(s)$). We denote the probability that a shot will yield a goal from state s as $xG(s)$. Given that the player decides to move the ball, the probability of moving from s to s' is represented by $T_{s \rightarrow s'}$. The probability of transitioning from s to s' is thus given by $m(z) * T_{s \rightarrow s'}$.
- A reward function R. This associates a reward with each possible transition. A reward of 1 is associated with the transition from each of the 768 field states to the absorbing goal state. All other transitions have a reward of 0.
- A discount factor γ. The discount factor is set to 1.

The transition function is learned from data by empirically counting for each state the total number of move actions s_m, and shoot actions s_z. The probability of moving $m(s) = s_m/(s_m + s_z)$ and the probability of shooting $z(s) = s_z/(s_m + s_z)$. $T_{s \rightarrow s'}$ is the empirical probability of moving from state s to s'. Next, for each state s the probability of scoring $xG(s)$ is simply the shot conversion rate from that location.

The value function $V(s)$ assigns a value to each transient state $s \in S$. This can be assigned by applying the Value Iteration algorithm [21] which employs an iterative algorithm to solve this equation for states:

$$V(s) = z(s) \cdot xG(s) + m(s) \cdot \sum_{s' \in S} T_{s \rightarrow s'} \cdot V(s')$$

Intuitively, this value represents the probability that a team will score before losing possession of the ball if they possess the ball in a specific location (i.e., state). Consequently, xT takes a longer-term perspective when assessing value as it considers what is likely to happen after multiple subsequent actions. In contrast, xG only considers the chance of directly scoring from the shot. For example, Dominique Janssen's pass to Van de Donk in Figure 2.3 has an xT value of 0.005 indicating that her pass increased her team's chance of scoring by 0.5 percentage points.

2.3.4 VAEP

Like xT, VAEP values game states, but it does so in a different way. It is based on the intuition that a helpful action is one that benefits a player's team, either by increasing the chance that their team scores or decreasing the chance that the opposing team scores. These actions should be positively valued. In contrast, actions that do the opposite hurt the player's team and should be negatively valued. This observation suggests that one way to value a game state is:

$$V(s) = P^k_{scores}(s) - P^k_{concedes}(s)$$

where $P^k_{scores}(s)$ ($P^k_{concedes}(s)$) is the probability that the team possessing the ball in state s will score (concede) within the next k actions. Typically, k is a small number like five or ten as the effects of most actions are temporally limited. Practically, the task then becomes estimating these two probabilities for each state, which can be solved by training a machine learning model that predicts a probability.

Estimating Scoring and Conceding Probabilities VAEP requires estimating the scoring and conceding probabilities. This task can be formally described as follows:

Given: A game state s;

Do: Estimate $P^k_{scores}(s)$ and $P^k_{concedes}(s)$.

For both binary classification problems VAEP trains a probabilistic classifier to estimate the probabilities. In principle, any machine learning model (e.g., logistic regression, random forest, neural network) that predicts a probability could be used to address these tasks. Empirically, it was concluded that gradient boosted decision tree ensembles such as XGBoost [52] and CatBoost [170] work well in practice.

Applying a standard machine learning algorithm requires converting the sequence of actions $[a_1, a_2, \ldots, a_m]$ describing an entire game into examples in a feature-vector format. Thus, one training example is constructed for each game state s. VAEP computes 151 features that impact the probability of a goal being scored in the near future. These features are based on the last three actions of a game state and can be divided into three categories.

First, the *SPADL features* are categorical and real-valued features based on information explicitly included in the SPADL representation. VAEP considers categorical features for an action's type, result, and the body part used by the

player performing the action. Similarly, VAEP considers real-valued features for the (x, y)-coordinates of the action's start and end locations, and the time elapsed since the start of the game.

Second, the *complex features* combine information within and across consecutive actions. Within each action, these features include (1) the distance and angle to the goal for both the action's start and end locations, and (2) the distance covered during the action in both the x and y directions. Between two consecutive actions, VAEP computes the distance and elapsed time between them and whether the ball changed possession. These features provide some intuition about the current speed of play.

Third, the *game context features* include the scoreline and goal difference in state s. This information is relevant because scoring probabilities depend on these factors [180].

Comparison with xT xT and VAEP are thus similar approaches because they both value actions by comparing the pre-action and post-action game states. However, there are three key differences between the two approaches [214]. First, xT only values actions that move the ball (e.g., passes, dribbles, crosses), whereas VAEP values all on-the-ball actions (e.g., also tackles, interceptions). Second, VAEP accounts for both the offensive and defensive contribution of an action, while xT only considers the offensive contribution. As a result, VAEP also captures the potential risks of an action. Third, the xT framework is generally more interpretable than VAEP as the value assigned to a game state is solely based on the location of the ball. In contrast, VAEP uses a broad set of features and a learned model to value game states.

Chapter 3

Women's Football Analyzed: Interpretable Expected Goals Models for Women

Most soccer metrics are initially developed for the men's game. This chapter investigates whether such models can simply be transferred to the women's game. More specifically, we focus on shots and investigate whether xG models can be transferred from one gender to the other.

The content of this chapter is based on the following publication [32].

L. Bransen and J. Davis. "Women's football analyzed: Interpretable expected goals models for women". In: *Proceedings of the AI for Sports Analytics (AISA) Workshop at the International Joint Conference on Artificial Intelligence*. 2021

3.1 Introduction

The women's football game has made huge advances in recent years. When the first official FIFA Women's World Cup took place in 1991, it featured matches that lasted only 80 minutes and the final was not even shown on TV [105]. Just 28 years later, more than 1 billion viewers watched the 2019 World Cup final between the Netherlands and the USA [82]. Concurrently, financial investment

in the women's club realm has increased the number of players who are able to play professionally. Recently, UEFA announced a new Women's Champions League format with more prize money and more teams [211]. The net effect is an increased level of competitiveness.

Despite its increased popularity, there has been less analysis of data related to the women's game. Most existing studies focus on physical aspects of the game. [163] state that the differences in the style of play between women and men are mostly due to the physical differences like endurance, kicking velocity, height, speed and foot length. [27] discovered that male players covered more distance at a higher speed and that these differences were bigger in the second half. [46] found the largest differences in sprints, jumps and intermittent endurance between female and male football players. In contrast, analyzing the technical aspects of the game has received less attention. [222] compared the shots taken in the English Women's Super League and the Premier League and [188] compared the number of goals in national team matches and concluded that in most cups women score more goals per match. [161] performed a more extensive comparison of women's and men's football in World Cup matches.

They concluded that men shoot from further away and have higher pass accuracies, while women regain possession quicker and prefer short passes over long balls. Furthermore, they use several indicators to train a model that predicts whether these represent a women's or men's match. Similarly, [47] saw more accurate passes in Spanish men matches than in women matches. [93] analyzed event data from the men's 2016 and women's 2017 European Championships. They trained models to predict a player's gender based on features such as the number of times a player (un)successfully performed a certain type of action.

This research takes a first in-depth analytical look using machine learning at the technical data that is now being collected from professional women's matches. Specifically, we focus on shots as the fundamental objective of football is to score more goals than your opponent, and as Johan Cruijff famously said: "you can't score if you don't shoot." A natural way to analyse shots and shot behaviour is through the lens of the well-known expected goals (xG) metric, which gives the probability that a shot will yield a goal. This metric helps cope with the fact that evaluating players and teams based on the number of goals scored can be undesirable because goals are relatively rare and subject to random fluctuations and luck (e.g., deflections). In contrast, xG quantifies the quality of the chances created which enables better understanding the performance of a team or player. Expected goals has become a relatively mainstream metric and is discussed on TV shows like BBC's Match of the Day and has even been used in the popular Football Manager computer game [195].

Needless to say, expected goals has been extensively researched [167, 133, 44, 175, 179, 134, 8]. However, as far as we know, no research has focused on the women's game. This could be because until recently, little data about the women's game was collected and even less of it is publicly available, with a notable exception being StatsBomb releasing some data about the English WSL. While the data contains xG values for taken shots, it is too small to enable training an accurate xG model or performing a more in-depth analysis. In this chapter, we fill this gap by analysing multiple seasons of event data from five top professional women's leagues. Moreover, we will contrast the shooting behaviour between women's and men's competitions. Specifically, we will address the following questions:

1. What, if any, are the differences on a match and competition level between women's and men's leagues?
2. What, if any, are the differences in shot selection and conversions between top women's and men's leagues?
3. Are xG models learned on data from women's matches transferable to men's matches and vice versa?

Our analysis produced a number of intriguing findings such as that women's professional matches have less extra time than men's matches, women convert a higher percentage of their shots than men, and women both head closer to the goal and score more often from headers. Our analysis of xG models finds that they are indeed transferable across matches of different genders though some specific shot types are valued differently. These observations greatly enhance our understanding of the way that the women's game is being played.

3.2 Data

Our analysis considers the five strongest women's and men's football leagues. The women's leagues included are: the English WSL, American NWSL, German Frauen Bundesliga, French Feminine Division 1 and Spanish Primera Division Femenina. The considered men's leagues are the English Premier League, Spanish Primera Division, French Ligue 1, German Bundesliga and Italian Serie A. The data set contains event stream data encoded in the SPADL format [63] from matches in the 2017 and 2017/2018 seasons until the 25th of April 2021 and was provided by SciSports.[1] In total, we have information about 9,076 matches of which 2,100 are women's matches and 6,976 are men's matches.

[1] https://www.scisports.com

Table 3.1: Number of shots in our data set per league per season. Some of the 19/20 seasons are incomplete due to the Covid-19 pandemic. The 2020 NWSL season only contains 25 shots because it was cancelled due to the pandemic.

League	M/F	17/18	18/19	19/20	20/21
WSL	F	606	2292	2005	2488
Division 1	F	2147	2737	2048	2224
Bundesliga	F	2859	3298	3356	2301
Primera Div.	F	2041	2382	2917	4685
NWSL	F	2865	2621	2698	25
Premier League	M	8401	8505	8287	6855
Ligue 1	M	8274	8180	6008	6874
Bundesliga	M	6884	7288	7111	5981
Serie A	M	8849	9375	9632	7166
Primera Div.	M	8090	8122	7485	6001

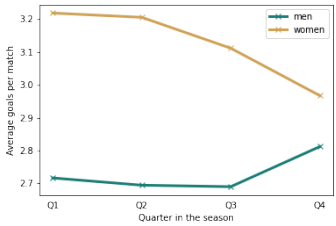
We focus on open play shots and hence omit penalties, direct freekicks and own goals. Table 3.1 shows the number of shots for each league in each considered season. As women's leagues typically have fewer teams, these leagues have fewer matches and shots per season. Furthermore the 2019/20 and 2020 seasons contain less data as the Covid-19 pandemic led to some matches being cancelled in these seasons. Due to the unfinished 2019/20 Primera Division Femenina, two extra teams were added to the league for the ongoing 2020/21 season. Finally, data for some matches in this league is also missing for the 2017/18 and 2018/19 seasons resulting in a lower number of shots for these seasons.

3.3 Observations

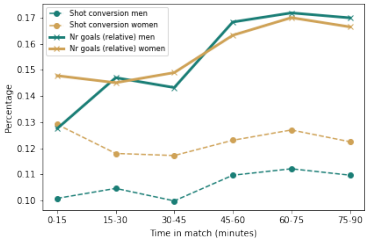
We discuss several interesting insights obtained from our data set in terms of differences between women's and men's football on two levels; (1) goals in matches and competitions, and (2) shot behaviour and outcomes during matches.

3.3.1 Matches and Competitions

First, we explore differences between the women's and the men's game regarding shots across matches and competitions.



(a) The average number of goals per match over the season. The number of goals per women's match (orange) declines over the season whereas men's matches (blue) see an increase near the end of the season.



(b) The average number of goals and the shot conversion rate for each 15 minute interval in a match. Women (orange) tend to score more and have a higher shot conversion rate than men (blue) at the start of the match.

Figure 3.1: Distribution of goals for women's (orange) and men's games per (a) each quarter of the season, and, (b) each 15 minute interval of a match.

Women score more goals at the start of the season We analyze matches from complete seasons (i.e., excluding the Covid-19 seasons (2019/20 and 2020) and the ongoing 2020/21 seasons) to see how the number of goals varies over the course of a season. We divide each season into quarters based on the number of matches in the season and compute the number of goals scored per quarter. Figure 3.1(a) shows that more goals are scored in men's matches near the end of the season. However, this is different in women's leagues, where more goals are scored at the start of the season.

Women's leagues contain more less-competitive matches In general, larger goal differences arise in women's matches, with 18% of the matches having a goal difference of more than 3 goals opposed to 6% in the men's matches. This likely arises because the women's game is still rapidly evolving and an increasing number of players can nowadays make a living from playing football [77, 85].

Draws are rare in the women's game In our data set, 17% of the women's matches end in draws versus 25% of the men's matches. However, women's matches have more away wins and a slightly higher number of home wins.

Women matches have less extra time We observe that women's matches tend to have on average around 37 seconds less time added to the regular 90 minutes of play. This includes time added in both the first and second half.

This is quite a big difference, given that women's matches have about 5 minutes extra time. The two English leagues have the longest extra time with the WSL having 6.8 and the Premier League 6.5 added minutes.

3.3.2 Shot Behaviour

Next, we explore differences in shot behaviour in terms of conversion rates and where shots arise.

Women have a higher shot conversion rate On average 12.2% of the women's shots find the net opposed to 10.7% in men's matches. The exception is the American NWSL where the conversion rate is 9.9%.

Shot conversion rates change throughout the course of the match Research showed that men tend to score more goals near the end of a match [10]. We investigate whether this claim also holds in the women's game. Figure 3.1(b) shows the average number of goals and the shot conversion rates for both games for every fifteen minutes in a match. The number of goals in both games shows a similar pattern throughout the match, with more goals near the end of the match. The shot conversion rates seem to be stable throughout the match.

Fewer "Robben" shots in the women's game Figure 3.2(a) shows the difference in proportion of shots with the foot from each location with orange (blue) areas indicating more shots in women's (men's) matches. Women tend to shoot from locations with a smaller angle to the goal and in general from closer distances to the goal. Interestingly, men shoot more often than women from the so-called "Robben" location[2], inside the box near the corner of the box where an inverted winger can shoot and curl the ball into the far corner of the goal.

Women head closer to the goal and score more often from headers Figure 3.2(b) shows the difference in proportion of headed shots from each location with orange (blue) areas indicating more shots in women's (men's) matches. Generally, women head closer to the goal than men do with the main difference arising inside the 5 meter box. Women score more often from headers with 15.7% of their headers yielding a goal opposed to 12.7% in the men's game.

[1] `https://github.com/TomDecroos/matplotsoccer`

[2] Arjen Robben was one of the first prominent inverted wingers and he frequently shot from this area of the pitch.



(a) Shots with the foot. Women shoot more often directly in front of goal inside the box whereas men shoot more from "Robben" locations and just outside the box.

(b) Headed shots. We observe that women tend to head closer to the goal than men.

Figure 3.2: Difference in proportion of shots with the (a) foot and (b) head from each location. Orange (blue) areas indicate more shots in women's (men's) matches.

3.4 xG Models

We address the question: are xG models learned on data from women's matches transferable to men's matches and vice versa? To answer this question we consider three datasets:

Women only contains only shots from women's matches.

Men only contains only shots from men's matches.

Combined contains all shots.

We split each data set into a train and test set using a temporal split: the train set has matches from the 2017/18, 2018/19 and 2019/20 seasons and the test set has matches from the 2020/21 season. Figure 3.3 visualizes the breakdown of our data sets. This split results in a total number of 157,346 shots for training (36,868 women, 120,478 men) and 44,589 shots in the test sets (11,719 women, 32,870 men).

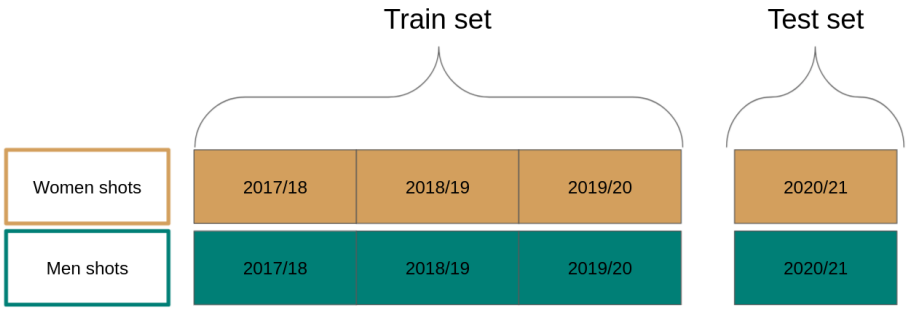


Figure 3.3: We split our data sets into a train and a test set using a temporal split and train our xG models on three different train sets; a women only, a men only and a combined set.

3.4.1 Challenges

Comparing models trained on different data sets is not straightforward for several reasons:

Rapid evolution women's game Due to the rapid evolution of the women's game, historical shot data might not be representative for the game as it is played today. Therefore, we do not go further back in time than the 17/18 season and the 2017 season for the NWSL.

Covid-19 Due to the Covid-19 pandemic, our data set contains many incomplete seasons of data. Our analysis indicates that the number of goals scored varies over the course of a season. Therefore, we need to correct for the fact that our data set contains fewer shots from the last quarter of the seasons due to the pandemic.

Competition set up Women leagues have fewer teams, and thus matches and shots. We do not want the combined model to overfit on the men's shots. Therefore need to correct for the difference in the number of women and men shots such that it contains the same number of shots from both games while still preserving the ratio of shots by the home team to the away team.

3.4.2 Data Set Sampling

To overcome the aforementioned challenges we use sample weighting to ensure that some shots in our data set will be more important during training than others. We assign weights to matches, not to individual shots, to ensure that all shots within a match are weighted similarly. We weight samples as follows:

1. **Women shots** Given that we have more shots by men than women, we weight the shots by women in our train set over three times more than our shots by men. Thus we have the same number of women and men matches in our train set.
2. **Matches per quarter** We weight the matches in such a way that we have an equal number of matches from each quarter of the season in our data set.
3. **Home and away matches** Once we have corrected for the matches per quarter, some teams may have a big difference in the number of home and away matches. We correct for this by weighting shots per team per home/away match based on its home/away ratio in our train set.

3.4.3 xG Models

On each data set, we train two xG models:

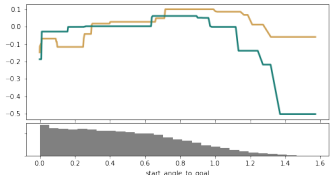
Generalized Additive Models (GAMs) are interpretable models that have been successfully applied to football data [65]. We use the `interpret-ml`[3] package to train the GAMs using their Explainable Boosting Classifiers [153] that use a boosting procedure to learn the feature functions of the GAM and to select pairwise interaction terms.

XGBoost [52] which learns an ensemble of trees, which have yielded excellent performance on xG and other football tasks [8, 134, 63]. We use `xgboost`'s[4] XGBClassifier to train the model.

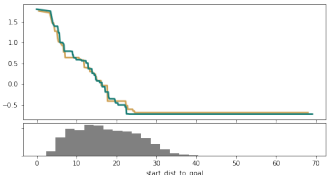
We refer to the six models based on the considered data and type of model: TotalGAM, WomenGAM, MenGAM, TotalXGB, WomenXGB and MenXGB.

We describe each shot using 12 features such as the shot's location, the type of assist (e.g. from dribble, from cross, possession change), the body part

[3] `https://github.com/interpretml/interpret`
[4] `https://xgboost.readthedocs.io/`



(a) Angle between the shot location and center of the goal



(b) Distance to the goal

Figure 3.4: Partial dependence plots for the angle to the goal and the distance to the goal for the WomenGAM (orange lines) and the MenGAM (blue lines). The histograms on the bottom show the features' distributions of values in the combined test set.

used (e.g. foot, head), the scoreline, and time in the match as calculated by the `soccer-xg`[5] package. We predict whether the shot yields a goal and are mostly interested in the probabilities that the models produce which define the expected goals values. We optimize the models' hyperparameters using a gridsearch with 5-fold cross-validation.

3.5 Results

This section describes the results of our experiments. Firstly, we inspect our GAMs to gain insight into whether a feature's importance varies between the women's and men's games. Secondly, we analyze the performance of our six models on a per data set basis to help us answer the question whether we can transfer models trained on one gender to the other. Thirdly, we analyze the predictions made by the models on different types of shots and show some examples where the models clearly do not agree. Finally, we analyze the average xG value of a shot and the percentage of shots with a low xG value per league in the 2020/21 season.

3.5.1 Model Inspection

The additive nature of GAMs permits analyzing the impact of each feature on the predictions by using partial dependency plots. We compare the partial

[5] `https://github.com/ML-KULeuven/soccer_xg`

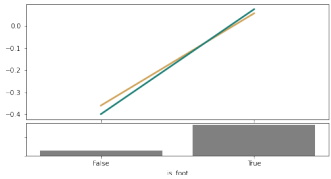
dependence plots for our WomenGAM model to our MenGAM model to analyze the similarities and differences.

Shot location Figure 3.4 shows the partial dependence plots for two important features regarding the shot's location: the angle and distance to the goal. Interestingly, the angle to goal has a differing impact in both models with larger angles negatively impacting the probability of a goal in the men's model while having a negligible impact in the women's model. In contrast, the relationship between the distance to the goal is similar in both models.

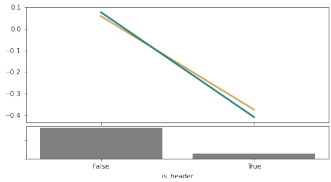
Assist type and body part Figure 3.5 shows the partial dependence plots for a number of features describing the shot's assist and the body part used to execute the shot. We can summarize the insights from these plots as follows:

1. Shooting with the foot has a slightly smaller positive impact in the WomenGAM whereas a header has a slightly less negative impact in the WomenGAM.
2. Shots following a dribble get a higher xG value in the WomenGAM, while shots following a teammate's pass get a higher xG value in the MenGAM.
3. Both models assign lower xG values to shots following a set piece, with the WomenGAM producing a larger negative impact.
4. The WomenGAM assigns lower xG values to shots directly from a cross, whereas this has a slight positive effect in the MenGAM.

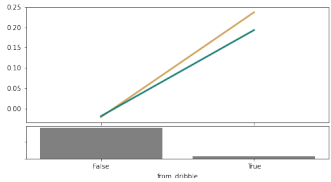
Game state Figure 3.6 shows the partial dependence plots for two features describing the game state when the shot was taken: goal difference and time in the current half. In both models, higher positive goal differences result in higher xG values. This finding agrees with past work [44], which was explained by the fact that less defensive pressure will be applied by the opponent when leading. This feature could also be a proxy for team strength as better teams tend to score more. Slightly larger goal differences have a bigger impact in the MenGAM than the WomenGAM. The WomenGAM's xG values tend to decrease over time, apart from the first minutes when fewer shots are taken in general. However, a different pattern arises in the MenGAM where xG values drop near the end of the half before increasing again in extra time.



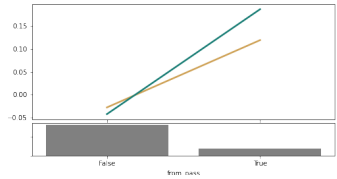
(a) Foot shot



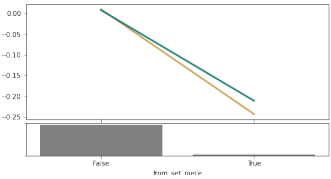
(b) Header



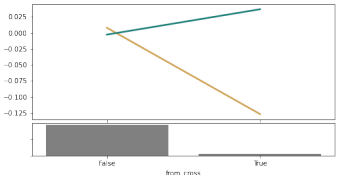
(c) Shot following a dribble



(d) Shot directly from a teammate's pass

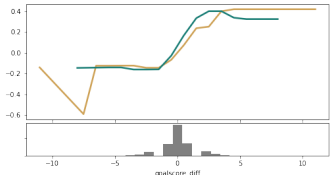


(e) Shot directly from a teammate's set piece

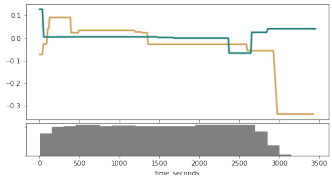


(f) Shot directly from a teammate's cross

Figure 3.5: Partial dependence plots for features describing the body part used for the shot and action preceding the shot for the WomenGAM (orange lines) and the MenGAM (blue lines). The histograms on the bottom show the features' distributions of values in the combined test set.



(a) Goal difference



(b) Number of seconds in the current half

Figure 3.6: Partial dependence plots for the features describing the game state for the WomenGAM (orange lines) and the MenGAM (blue lines). The histograms on the bottom show the distribution of values for the feature in the combined test set.

Table 3.2: Brier scores for all six models on each of the three test sets.

Model/Test set	Women only	Men only	Combined
WomenXGB	0.09705	0.08677	0.08947
WomenGAM	0.09714	0.08680	0.08952
MenXGB	0.09675	0.08636	0.08909
MenGAM	0.09722	0.08674	0.08950
TotalXGB	0.09675	0.08635	0.08908
TotalGAM	0.09690	0.08654	0.08926

3.5.2 Model Evaluation

We estimate the xG values for the shots in our three test sets using the three GAM models and three XGBoost models. Table 3.2 shows the Brier Score each model achieves on each data set. We use this metric as the most important criterion is having a well-calibrated model. All models achieve similar performance. Figure 3.7 shows the calibration plots for the six models on the combined test set. All models are well-calibrated for the more commonly occurring low xG shots and, predictably, struggle for the rarer big chances that yield higher xG values.

3.5.3 Individual Shots with Big xG Differences

We investigate the predictions made by the different models on individual shots with a particular emphasis on identifying shot types where the models produce

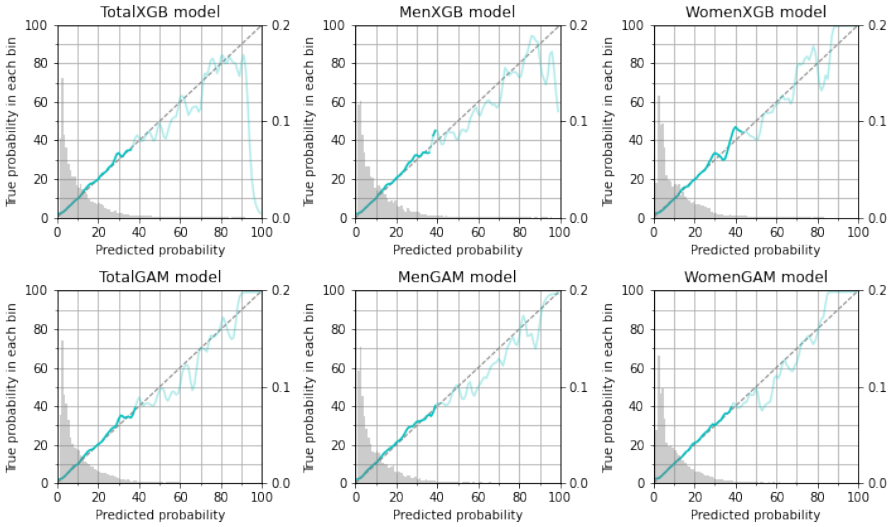


Figure 3.7: The calibration plots for all six models on the combined test set. In general, all models are well-calibrated for the probabilities on the most commonly occurring shots.

differing xG values. The women-based models and men-based models disagree most for the so-called "Robben" shots, tip-ins from short distance following a cross, and headers following a set piece.

"Robben" shots The models trained on men's shots assign higher values to "Robben" shots than the models trained on women's shots. Figure 3.8 shows an example of such a shot by Marcus Rashford. This shot received xG values of 0.069 (MenGAM) and 0.087 (MenXGB). However, the models trained on women's data produced lower values: 0.037 (WomenGAM) and 0.066 (WomenXGB).

Tip-ins following a cross The models trained on women's data assign higher values to tip-ins close to the goal than the models trained on men's data do. Figure 3.9 shows an example of such a goal by Jordan Nobbs which received xG-values of around 0.60 from the women's models (0.586 from WomenGAM and 0.608 from WomenXGB). However, the men's models assigned a much lower xG value of around 0.48 (0.478 from MenGAM and 0.476 from MenXGB).

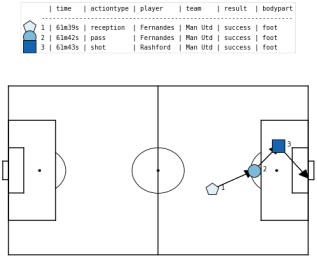


Figure 3.8: Marcus Rashford's 62nd minute goal from the "Robben" location in the 2-1 win against Brighton on April 4th, 2021.

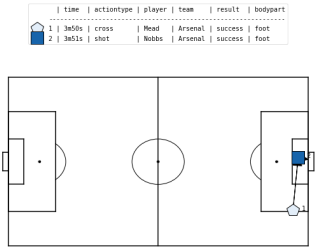


Figure 3.9: Tip-in by Jordan Nobbs following a Beth Mead cross in the 4-0 Arsenal win against Everton on December 20th, 2020.

Headers following a set piece The models trained on men's data assign higher xG values to headers following a set piece than the women's models do. Figure 3.10 shows an example of such a header by Brighton & Hove Albion's Victoria Williams. This opportunity was valued by the women's models at 0.196 (WomenGAM) and 0.359 (WomenXGB), but received substantially higher xG values of 0.654 (MenGAM) and 0.614 (MenXGB) from the men's models.

3.5.4 xG Values Across Leagues and Gender

Table 3.3 shows the average xG value per shot (based on the XGBoost model per gender) in our test set for each league, where we exclude the NWSL due to its small sample size. Women tend to take shots with higher chances of yielding a goal and men take more speculative shots, that is, those having a low chance of yielding a goal.

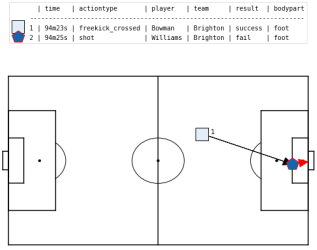


Figure 3.10: Late Victoria Williams miss in the 95th minute following a freekick in the 3-1 loss against Tottenham on December 6th, 2020.

Table 3.3: The average xG value per shot and the percentage of low xG shots per league. We observe lower average xG values and a higher percentage of low xG shots in the men leagues.

League	M/F	avg xG	% low xG (<0.02)
WSL	F	0.128	3.5%
Division 1	F	0.134	3.6%
Bundesliga	F	0.127	3.1%
Primera Div.	F	0.126	4.3%
Premier League	M	0.111	10.3%
Ligue 1	M	0.105	14.0%
Bundesliga	M	0.110	11.2%
Serie A	M	0.112	10.3%
Primera Div.	M	0.110	13.3%

3.6 Conclusion

This chapter showed an extensive analysis of women's football shots. We identified interesting observations such as the fact that women tend to shoot from different locations than men, have a higher shot conversion rate and their goals are differently distributed across the season. We trained six different xG models on different data sets with different machine learning algorithms and found that, in general, models from one gender are applicable to shots from the other. However, when inspecting the models and shots, some interesting differences arose in terms of what features are important and how the models value certain types of shots.

Chapter 4

A Data-driven Analysis of the Technical and Tactical Evolution of Elite Women's Football

Whereas the previous chapter compared shots in women's soccer and men's soccer, it ignored the developments of the game over time. Given the recent increase in investment in women's soccer, it is interesting to investigate the developments of the women's game. This chapter presents a large-scale analysis of the technical and tactical developments of the women's game.

The content of this chapter is based on the following publication [29].

L. Bransen and J. Davis. "A data-driven analysis of the technical and tactical evolution of elite women's football". In: *International Journal of Sports Science & Coaching* 19.6 (2024), pp. 2449–2459

4.1 Introduction

Between the 1920s-1970s many European football federations forbid women from playing [135]. Even once federations began sanctioning matches, opportunities remained limited, particularly with the women's game largely consisting of amateurs. However, the establishment of the FIFA Women's World Cup has coincided with an increased interest and professionalisation of the women's game: many countries have established professional leagues and UEFA started a women's version of the Champions League [220].

Money is slowly beginning to flow into the women's game [77]. The increased opportunities and investment suggest that there may be evolutions in player performance in women's football. While research on the women's game has grown, it still receives less attention than the men's game [115]. Currently, most research studies the physical aspects of women's football [156, 103].

In addition to physical skills, tactical and technical skills are also essential for football players [190]. Tactical behaviour refers to the short-term actions performed by players when adapting to the changing game situation [178]. Tactics are often summarised by defining metrics that capture the (relative) frequencies that specific behaviours arise in matches. For example, some works focus on possession such as its length (e.g., number of actions, duration) [109] and zones of the pitch where a team possesses the ball [127]. More recently, ball possessions specifically in the women's game have been analysed, focusing on goalkeeper distribution (i.e., where goalkeepers move the ball to) [48], possessions that lead to goal scoring opportunities [145], and complete ball possessions [115, 70].

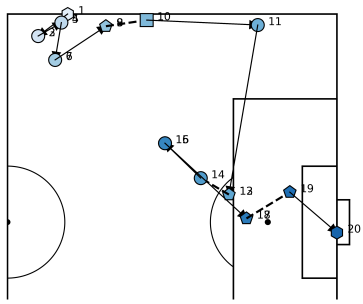
Other works perform locational analysis such as whether players are positioned higher or lower up the pitch [22] or how the prevalence and directionality (e.g., backwards, sidelines, forward) of passes depends on the location [76]. More recently, De Jong et al. [118] performed a network analysis on data from 694 matches from the women's game. They concluded that successful teams are highly connected and centralise the distribution of ball possession. They identified the lack of dependency on key players for the total ball flow as a unique characteristic of women's football.

Similarly, shooting behaviour such as where shots are taken and how their conversion rate depends on contextual factors (e.g., length of preceding possession, shot characteristics) has been extensively studied [32, 216, 133]. However, such research on extensive datasets has largely focused on men's football. Two more recent studies analyse goal scoring opportunities in the women's game. Scanlan et al. [192] analysed 52 matches from the 2015 FIFA Women's World Cup and found that the average time needed to create a goal

scoring opportunity was 12 seconds. Mesquita et al. [143] analysed 174 goals and concluded that goal scoring opportunities in women's football are more often created through the offensive organisation than in the men's game.

Technical skills such as ball control, passing, and shooting are crucial in football performance [4]. Several studies have shown that more successful teams show better technical performance [25, 174]. Accurate passing [206] and accurate shooting [4] are key football skills. Such technical skills are often evaluated by analysing the accuracy of execution [177, 185] such as pass accuracies, although it is also important to consider the difficulty of executing the pass [206]. Similarly, shooting skills can be assessed by looking at the ability to generate shots and convert them to goals [133, 138]. Finally, other research uses statistics and machine learning techniques to analyse various technical and tactical aspects of football [171, 173, 63, 175]. Typically such techniques are applied to large datasets of detailed event data that describe all on-the-ball events. The few papers that study the women's game using machine learning primarily focus on contrasting it to the men's game by looking at metrics such as pass accuracy, possession length, and shot conversions [26, 47]. Pappalardo et al. [161] used machine learning techniques (logistic regression, decision trees, tree ensembles) to predict whether a match involved a men's or women's team based on technical indicators such as the number of shots and passing accuracy. Garnica-Caparrós and Memmert [93] used logistic regression, decision trees, and neural networks to find performance factors (e.g., ground duels won, shots on target, successful passes) that help differentiate between female and male football players. Both studies only considered a small set of matches (108 matches and 82 matches, respectively).

This chapter represents the first large-scale analysis of women's event data which providers have only recently started collecting. Uniquely, we will perform a data-driven analysis of data from matches played in the six top domestic leagues and the Champions League between 2013 and 2022. We hypothesise that there has been an evolution and improvement in tactical and technical skills of players. These can be developed during training sessions [197], which may be more accessible and of higher quality with increased professionalisation. Tactically, we will analyse possession lengths and shooting strategies. We expect the game to have evolved tactically in a similar manner to the men's game with a stronger emphasis on retaining possession [43], more high-quality opportunities and fewer long-distance shots [216]. Technically, we focus on the passing and shooting skills. For passing, we will evaluate whether players are more likely to complete difficult but valuable passes such as ones that increase the chance of scoring. To evaluate shooting skill, we consider shot conversion and shot placement.



	time	player_name	type_name	x	y	dx	dy
1	18m38s	Leila Ouahabi	throw_in	62.1	68.0	-4.7	-3.6
2	18m38s	Alexia Putellas	receival	57.4	64.4	0.0	0.0
3	18m39s	Alexia Putellas	pass	57.4	64.4	3.7	2.2
4	18m39s	Leila Ouahabi	receival	61.1	66.6	0.0	0.0
5	18m40s	Leila Ouahabi	pass	61.1	66.6	-1.0	-6.1
6	18m40s	Alexia Putellas	receival	60.1	60.5	0.0	0.0
7	18m40s	Alexia Putellas	pass	60.1	60.5	8.1	5.5
8	18m41s	Lieke Martens	receival	68.2	66.0	0.0	0.0
9	18m42s	Lieke Martens	dribble	68.2	66.0	6.5	1.0
10	18m43s	Lieke Martens	take_on	74.7	67.0	17.7	-0.8
11	18m47s	Lieke Martens	pass	92.4	66.2	-4.6	-27.7
12	18m48s	Jennifer Hermoso	receival	87.8	38.5	0.0	0.0
13	18m49s	Jennifer Hermoso	dribble	87.8	38.5	-4.5	2.7
14	18m49s	Jennifer Hermoso	pass	83.3	41.2	-5.7	5.7
15	18m50s	Alexia Putellas	receival	77.6	46.9	0.0	0.0
16	18m50s	Alexia Putellas	pass	77.6	46.9	13.0	-12.3
17	18m51s	Aitana Bonmati	receival	90.6	34.6	0.0	0.0
18	18m52s	Aitana Bonmati	dribble	90.6	34.6	6.9	4.3
19	18m52s	Aitana Bonmati	shot	97.5	38.9	7.5	-6.6
20	18m52s	Aitana Bonmati	goal	105.0	32.3	0.0	0.0

Figure 4.1: The sequence leading to Aitana Bonmati's goal in the 2021 Champions League final. For each action, the data records its type, the player and team performing the action, the x and y coordinate of the action's start and end location, and the time in the match the action was executed.

4.2 Data

This work analyses an extensive match event dataset containing 8.5 million actions from 3,510 matches in the top leagues and UEFA Women's Champions League (Table 4.1). The data was provided by SciSports in the SPADL format [63]. The data is typically recorded by human annotators watching video footage and describes all on-the-ball actions that took place in a match by recording the action's start and end location, which players were involved, the action type, and a time stamp (Figure 4.1). Our goal is to evaluate general differences over time. Moreover, we need a sufficient amount of data to train the machine learning models. Therefore, instead of considering a season-by-season analysis, we chronologically partition the data into two similarly sized datasets: an OLD and NEW dataset.

However, splitting the data is challenging. We do not want matches played in the same time window (year/season) to appear in both datasets. Moreover, because we want to test differences over time and not differences across leagues, we want a similar number of matches per league in both datasets. This is complicated because the number of matches per league varies per season for two reasons: 1) due to Covid-19 some matches were cancelled (e.g., NWSL 2020 season) or rescheduled, and 2) some league structures changed resulting in more teams participating (e.g., Champions League, Spanish Primera División). Given these desiderata, the NEW dataset contains all matches (N=1,594) played in the 2020, 2020/2021, 2021 and 2021/2022 (until 23th March, 2022) seasons,

Table 4.1: The number of matches per season for each of the leagues in our dataset.

	'13 (/14)	**'14 (/15)**	**'15 (/16)**	**'16 (/17)**	**'17 (/18)**	**'18 (/19)**	**'19 (/20)**	**'20 (/21)**	**'21 (/22)**
Swedish Damallsvenskan	0	6	3	7	8	6	132	132	0
German Frauen Bundesliga	0	1	9	12	124	131	131	129	101
French Division Féminine 1	0	3	6	12	101	129	93	131	102
American NWSL	0	6	2	35	115	108	111	1	118
Spanish Primera división	0	0	2	3	86	99	116	280	175
English Women's Super League	0	0	0	1	34	97	87	131	98
Champions League	2	2	2	1	5	17	71	79	117
Total	**2**	**18**	**24**	**71**	**473**	**587**	**741**	**883**	**711**

whereas the OLD dataset contains all matches (N=1,916) played before these seasons.

4.3 Approach

We present the methods and metrics used to assess tactical and technical evolutions.

4.3.1 Tactical Behaviour

We analyse tactical behaviour by studying ball possession and shooting behaviour.

1. Ball Possession

A possession sequence is a sequence of consecutive actions performed by one team without an interruption by the opposing team [117]. Longer possession sequences are an indicator of success [109, 117]. Possession sequences are extracted by identifying moments of ball loss (e.g., pass intercepted by opponent, ball out event, foul) which signify the end of the ongoing sequence. We only consider sequences where the team has executed at least two consecutive actions to avoid having the results unduly influenced by short possessions (e.g., being immediately dispossessed after getting the ball). We report the duration and distance travelled for possession sequences. The duration is the time in seconds between the first and last action [117]. Distance travelled is the metres travelled by the ball throughout the sequence (i.e., the sum of the metres travelled between each pair of consecutive actions in the sequence).

2. Goal Attempts

We report the number of shots in a match as well as the number of shots from in and outside the penalty box. We use an expected goals model (xG) [32, 133, 8, 167] to measure the quality for a specific shot. xG models use machine learning to find patterns in historical shots that relate the characteristics of the shot (e.g., where the shot was taken from, scoreline, etc.) to its outcome (goal vs. no goal). Specifically, given a shot the xG model computes the probability that it will yield a goal. We train xG models to compare the xG distributions

for shots in- and outside the box to get an understanding of the difference in the quality of shots taken in both datasets.

Measuring Chance Quality To train our xG models, we first collect all shots from open play in our datasets. Hence shots directly from set pieces such as penalty kicks and free kicks are excluded. Shots directly following a set piece such as a corner or throw in are included. We also remove all own goals from our dataset. This results in a total of 45,606 shots in the OLD dataset and a total of 36,256 shots in the NEW dataset. The average number of shots per match in the OLD and NEW dataset are 23.8±6.1 and 22.8±6.4, respectively.

Training an xG model can be framed as a binary classification problem: each shot either results in a goal (positive label) or not (negative label). Our dataset is imbalanced due to the fact that around 1 out of 10 open play shots result in a goal. We want the probabilities that the model assigns to each shot yielding a goal to be as accurate as possible and will therefore focus on optimising the Brier loss [40] of our xG models. This is an evaluation metric that looks at how well calibrated probabilistic predictions are. A well-calibrated model means that predicted probabilities match the definition of a probability. For example, if all shots with an xG value of 0.13 are considered, in a well-calibrated model 13% of those shots will yield a goal.

We use the publicly available XGBoost Python package[1] to train XGBoost classifiers [52] on different sets of our datasets. As shown by Robberechts and Davis [179], XGBoost classifiers show good performance for xG models. We will use the same feature set as they used which can be constructed using the open source xg-soccer Python package[2].

Typically when training and evaluating a machine learning model, one splits that data set in three sets: a training set, a validation set and a test set (hold-out set). The model learns the patterns using the data in the training set, the hyperparameters are tuned using the validation set and the analysis is performed on the test set. As we want to compare the xG distributions over time, we will not use such a static split with one held-out set. Therefore, we train six models on the OLD dataset and six models on the NEW dataset. One for each of the domestic leagues. We use the shots from the Champions League as a comparison set to analyse the performance of our models on. Each model is trained on five leagues and the model is used to make predictions for the shots in the held-out league. In that way we can fairly use these values for our analysis on the differences over time. For each model, we use the same set of features and use 5-fold cross validation on the training set to optimise the

[1] https://xgboost.readthedocs.io/en/
[2] https://github.com/ML-KULeuven/soccer_xg

number of trees and the tree depth as those are important hyperparameters to avoid overfitting [52].

4.3.2 Technical Skills

We investigate passing and shooting skills because they are important indicators of improved technical performance [217]. As we expect technical skills to have improved, we use one-sided t-tests to test whether we observe positive improvements in passing and shooting skills amongst players.

Passing Skills. While pass completion percentage is a common metric, it can be misleading because it can be increased by favouring safer passes (e.g., backpasses) over more risky, yet valuable, ones (e.g., through balls) [206]. Therefore, we investigate the success rates on two other types of passes. First, we consider deep completions which are defined by SciSports as forward passes (passing angle of 45 degrees) that travel at least 15m and end in the final third of the pitch. Such passes are harder to complete as the final third is typically more crowded and forward passes are harder to complete than backward passes. Furthermore, passes towards crowded areas have been related to technical skills and team success [3].

Second, we use a pass' xT value [215, 202] to assess the impact of the pass. Simply put, a pass' xT value measures the difference in the chances of scoring a goal in the near future of locations where the pass a) ended and b) started. A key difference between xT and xG is that xT values ball progression actions (passes, carries) whereas xG only considers shots which is beneficial since there are many more ball progression actions than shots in a match.

Formally, xT is a discrete Markov Decision Process, in which the game of football is modelled as a system with transitions between states. In each state, the system has a certain probability of transitioning to another state. xT represents states by overlaying a 32x24 grid on the pitch. The current state is given by the location of the player possessing the ball. For each state, the model specifies the probability that a player will (a) shoot or (b) move the ball to every other zone on the pitch. Moreover, it specifies the probability that the action is successful (e.g., shot results in a goal). Second, a value is assigned to each state. Intuitively, the value represents the probability that a team will score before losing possession of the ball if they possess the ball in a specific location (i.e., state). Consequently, xT takes a longer-term perspective when assessing value as it considers what is likely to happen after multiple subsequent

actions. In contrast, xG only considers the chance of directly scoring from the shot.

We use the publicly available socceraction Python package[3] to learn the xT models separately on the OLD and NEW dataset. For example, Putellas' pass to Bonmati in Figure 1 has an xT value of 0.075 indicating that her pass increased her team's chance of scoring by 7.5 percentage points. Valuable passes with the highest expected reward are often also the riskiest ones. Hence, we report the accuracy on the passes that would be in the top 5% of xT value if they were completed. Such passes increase the scoring chances by at least 2 percentage points.

Shooting Skills. Shot conversion rate and percentage of shots on target are common metrics. However, neither are necessarily proxies for shooting skill as their values are affected by the type of shots taken. Distance to goal is the most predictive feature for both metrics [133]. Therefore, we look at shooting skill in two ways. First, we report conversion rate and percent of shots on target inside and outside the penalty box. Conversion rates and the number of shots on target from different locations have been related to player abilities [4, 138]. Second, we look at the placement of the shots in terms of percentage of shots that are within a certain distance of the posts.

4.3.3 Sensitivity Analysis

We perform a sensitivity analysis to check the robustness of our findings. For this, we repeat our analysis on an alternative partition of the data. As the amount of data across the leagues is not equal for the older seasons, we introduce an alternative OLD dataset OLD_{ALT} in which we exclude all matches played in the '13(/14), '14(/15), '16(/17) and '17(/18) seasons. Similarly, we introduce NEW_{ALT} in which we exclude the matches played in the 20/21 season and the single match played in the 2020 NWSL season. Due to the Covid pandemic, the NWSL 2020 season was cancelled, and the matches in the 20/21 season were played under different circumstances. Note that some matches in the 19/20 season were played behind closed doors. However, we do not know which matches were affected so we only excluded the 20/21 season matches. We investigate whether excluding these matches has an impact on our main findings. The OLD_{ALT} dataset contains data from 1,670 matches and the NEW_{ALT} dataset contains data from 711 matches.

[3] https://github.com/ML-KULeuven/socceraction

4.4 Results

4.4.1 Evolution in Tactical Behavior

Ball Possession Possession sequences have become significantly longer, both in time and distance travelled. The duration of possession sequences has increased from 13.0±11.7s in the OLD dataset to 13.5±13.0s ($p<0.001$) in the NEW dataset, whereas the distance travelled has increased from 84.1±65.5m to 85.6±69.2m ($p<0.001$). The same holds for sequences that lead to a shot on target (17.9±14.3s to 19.3±16.0s ($p<0.001$) and 121.8±79.3m to 123.8±86.9m ($p<0.001$)).

Goal Attempts The number of shots from outside the penalty box and the proportion of shots taken with the foot have decreased (Table 4.2). Furthermore, the average xG per shot outside the box has increased, whereas the average xG for shots inside the box has decreased (Table 4.2).

4.4.2 Evolution in Technical Skill

Evolution in Passing Skill While the number of passes per match did not significantly change, success percentages have risen significantly from 77.1±0.034% in the OLD dataset to 78.2±0.037% ($p<0.001$) in the NEW dataset for all passes. The number of deep passes has dropped from 50.3±10.5 to 48.4±11.3 ($p<0.001$) deep passes per match, whereas the success percentage has increased from 48.5±0.16% to 51.5±0.18% ($p<0.001$). Also, the success percentage of passes with high (top 5%) xT values has risen from 39.8±0.18% to 42.5±0.20% ($p<0.001$). Moreover, a year-on-year analysis shows a steady increase in the success percentage for deep passes (Figure 4.2) and passes with a high xT value (Figure 4.3).

Evolution in Shooting Skill Overall shot conversion rates have remained relatively stable (11.9±0.15% in the OLD dataset and 12.2±0.17% in the NEW dataset, $p=0.188$), whereas the proportion of shots on target has increased significantly from 39.1±0.23% in the OLD dataset to 41.2±0.26% in the NEW dataset ($p<0.001$). This is driven by an increase in shots on target from outside of the penalty box, which have risen from 30.1±0.34% in the OLD dataset to 32.8±0.41% in the NEW dataset ($p<0.001$).

There has been large significant increases in the percent of shots placed near the posts (Table 4.3 and Figure 4.4). We observe these increases across all leagues,

Table 4.2: The average number of shots per match, the proportion of all shots and the average xG value per shot for all shots, shots taken with a foot, shots from outside the box, and shots from inside the box.

		OLD	NEW	p value
All shots	**Number per match**	23.90±6.07	22.85±6.43	<0.001
	xG per shot	0.117±0.138	0.120±0.142	0.0015
Shots by foot	**Number per match**	20.55±5.52	19.51±5.77	<0.001
	Proportion of shots	86.0±0.16%	85.4±0.19%	0.014
	xG per shot	0.111±0.136	0.114±0.139	0.0026
Outside box	**Number per match**	9.57±3.67	8.34±3.45	<0.001
	Proportion of shots	40.1±0.23%	36.5±0.25%	<0.001
	xG per shot	0.035±0.033	0.039±0.040	<0.001
Inside box	**Number per match**	14.23±4.8	14.41±4.98	0.28
	Proportion of shots	59.9±0.23%	63.5±0.25%	<0.001
	xG per shot	0.173±0.154	0.167±0.158	<0.001

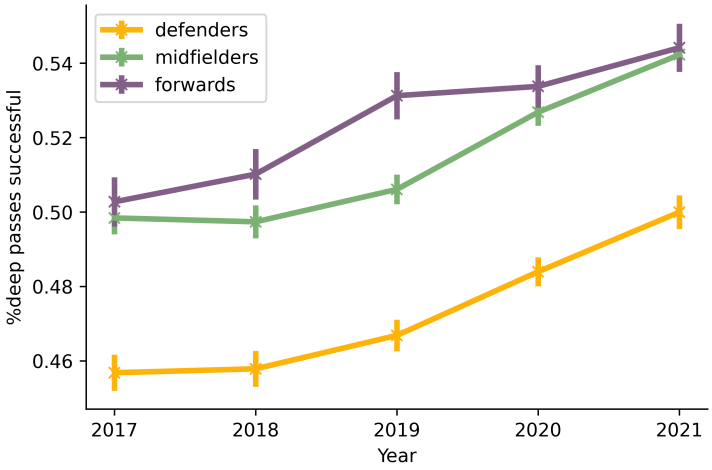


Figure 4.2: The accuracy on deep passes for each position line as a function of time. The accuracy of these passes has increased for all position groups over time.

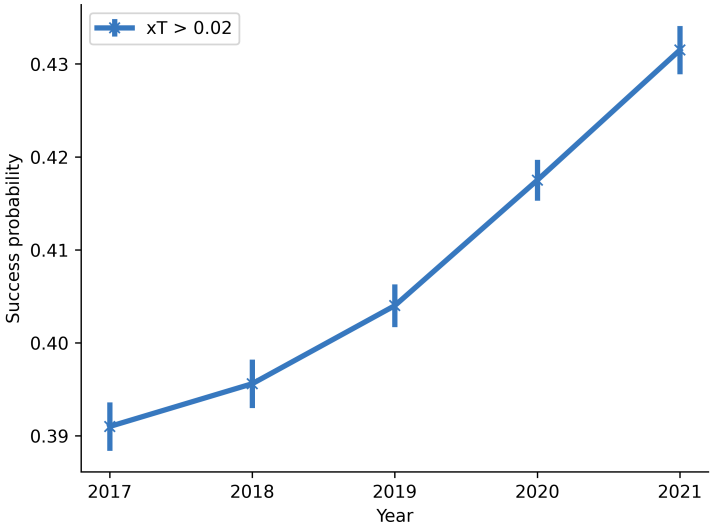


Figure 4.3: The success probability for passes that would be in the top 5% of xT value if they were completed. The accuracy of these passes has increased over time.

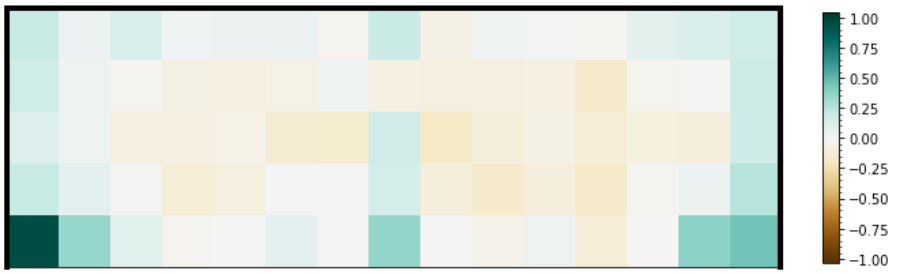


Figure 4.4: The difference in end locations for shots from inside the box plotted on the goal. Green (orange) indicates a higher proportion of shots end in the location in the NEW (OLD) dataset end in a location. The values in the colorbar correspond to the proportion difference between the two datasets in percentage points. We observe a higher proportion of shots aimed at the far corner in the NEW dataset.

with an average increase per league of 57.7±8.4% for the proportion of shots placed within 0.5m of the posts. Furthermore, we observe a clear upward trend for the proportion of shots placed within 0.5m of the posts increasing from 4.9% in 2017, to 5.4% in 2018, 5.3% in 2019, 7.7% in 2020 to 8.6% in 2021.

Sensitivity Analysis Table 4.3 shows the results of our sensitivity analysis for our main findings.

4.5 Discussion

Using standard metrics and machine learning techniques, we have observed changes in tactical behaviour and improvements in proxies for technical skills in an analysis of data from 3,510 matches from top women's leagues plus the UEFA Women's Champions League.

Tactically, it seems that retaining possession has become more important. Possession sequences are longer in both duration and distance, mirroring changes in the men's game [43]. These changes follow the optimal passing characteristics for elite women's football as described by Dipple et al. [70]. Furthermore, players have become more judicious about the shots they take. Mirroring the men's game, there has been a decrease in shots from outside the penalty box [216]. The average xG value of a shot outside the box has increased, as women forgo the lowest quality long shots.

Table 4.3: Results of our sensitivity analysis for our 4 datasets. For all findings, except for the shot conversion rate, we observe significant differences between the OLD versus NEW_{ALT} datasets, OLD_{ALT} versus NEW datasets, and OLD_{ALT} versus NEW_{ALT} datasets. For the NEW_{ALT} dataset we even observe some bigger differences in shot placement values and pass accuracy values.

	OLD	OLD_{ALT}	NEW	NEW_{ALT}	Difference
Possession length	84.1m	83.9m	85.6m	85.1m	Increase
Possession duration	13.0s	13.0s	13.5s	13.5s	Increase
xG per shot	0.117	0.117	0.120	0.122	Increase
xG per shot outside box	0.035	0.034	0.039	0.400	Increase
Pass accuracy	77.1%	77.2%	78.2%	78.3%	Increase
Nr deep passes per match	50.3	50.1	48.4	46.1	Decrease
Success percentage deep passes	48.5%	48.6%	51.5%	52.3%	Increase
Success percentage high xT	39.8%	40.0%	42.5%	42.9%	Increase
Shot conversion rate	11.9%	12.0%	12.2%	12.0%	No change
Shot placement within 0.5m from post (all shots)	5.2%	5.4%	8.1%	8.6%	Increase
Shot placement within 0.5m from post (shots inside box)	5.8%	6.0%	9.2%	9.9%	Increase
Shot placement within 0.5m from post (shots outside box)	4.3%	4.4%	6.2%	6.4%	Increase

Technically, there is evidence that players have improved their passing and shooting skills. Players have become more proficient at completing difficult passes. We observed an increased success percentage for two types of passes that are completed less often than a typical pass. An increased accuracy for such passes has been linked to improved technical skills [3]. However, an alternative hypothesis is that tactical evolutions made these passes somehow easier to complete.

Regardless of whether they shoot from inside or outside the box, players nowadays are placing more shots on target. Similarly, there is a significant increase in the placement of shots near the posts. Such shots are harder to stop and are an indication of improved technical skills [185]. For shots taken inside the box, the proportion and number of on target shots placed within 0.5m of a post have significantly increased. There are also gains for shots outside the box. Regardless of location, the proportion and number of shots ending within 1 and 3.5m of posts have fallen. However, we do not observe an increased conversion rate for on target shots, which might indicate improved shot-stopping by goalkeepers. This requires further investigation.

Our study has some limitations. While carefully selected, our dataset is not perfect. For some seasons, data from a number of matches is missing, particularly for older seasons, meaning the proportion of matches per league varies between our datasets. However, our sensitivity analysis shows that our results are robust and hold for different partitions of the data. However, the differences for shot placement and pass accuracy are even bigger if we exclude matches played during the Covid pandemic. Moreover, the data is annotated by humans, and both the event definition and which events are collected can change over time. Using event data allows us to perform a broader analysis. However, on-the-ball events are mostly offensive in nature. Because we do not know the locations of the defenders, we cannot study defense or other prominent tactics such as pressing.

In the future, if more data becomes available, we would like to perform a more fine-grained analysis on the league level. Furthermore, it would be interesting to investigate developments in other aspects of the game, like goalkeeper and defensive skills. It would also be worthwhile to investigate if these trends also hold in international matches and tournaments. Finally, it would be interesting to compare and contrast these trends to the evolution in the men's game.

Chapter 5

Measuring Player Pace of Play in Elite Female Association Football

Chapter 3 showed that it is better to develop a dedicated xG model specifically for women's soccer, while Chapter 4 showed that the women's game has developed substantially in both technical and tactical aspects. In this chapter, we introduce a new metric to measure a player's pace and show its relevance for data-driven recruitment in the women's game.

The content of this chapter is based on the following publication [30].

L. Bransen and J. Davis. "Measuring player pace of play in elite female association football". In: *International Journal of Sports Science & Coaching* (2025), p. 17479541251334022

5.1 Introduction

Transfers are a key mechanism for roster construction in soccer. From the perspective of a club, this requires assessing players along a variety of different dimensions. Namely, it requires evaluating players' technical, physical, mental, cognitive and tactical skills. Crucially, one also must project how a player

will fit into a new environment (i.e., team and possibly league). In particular, understanding how player performances translate across leagues is important because two-thirds of transfers in men's football involve players changing leagues [155]. The number of international transfers in women's football is increasing rapidly with a 20% increase in 2023 compared to 2022 [78].

Unfortunately, it is challenging to contextualize a player's performance based on their league as leagues differ in both level (i.e., quality) and style of play. One differentiating factor among leagues is the game intensity where higher intensities generally indicate a higher level of play. One line of work measures intensity by looking at the speed in which the ball is moved around and distance it travels on the pitch. Shen, Santo and Akande [199] defined pace of play as the average velocity of the ball during a possession sequence. More recently, Epasinghege and Swartz [71] utilized spatiotemporal tracking data to analyze whether playing with pace is advantageous. They introduced "attacking pace", which is defined as the ball's velocity in the direction of the opposing goal during a possession sequence. Stats Perform [165] introduced a Direct Speed Metric that measures the speed at which the ball moves in the x-direction (from one goal to the other goal) during a possession sequence. These works all focus on team pace and analyzing team strategy.

Another line of work considers the physical intensity of the game, such as the players' running speeds and the number of high-intensity runs has also been investigated. Higher intensity has been related to higher level (professional versus amateur) in women's football [99]. Sahasrabudhe and Bekkers [187] concluded that the speed of the players and ball during a counterattack is related to the success of the attack for both women's and men's football.

Pace and intensity are also relevant concepts in other team sports such as American football, basketball, and ice hockey. For example, in basketball it is common to measure pace in terms of the number of possessions in a game [124]. Similarly, in American Football, the number of plays has been used as a measure of pace [54]. In ice hockey, the movement of the puck has been used to analyze both team and player level pace. Silva, Davis and Swartz [200] evaluated pace using the horizontal distances travelled by the puck. Yu et al. [226] analyzed team-level pace in ice hockey using the speed of the puck in various directions. They proposed an on-off metric to assess an individual player's pace and the player's effect on their team's pace.

This chapter is based on the premise that intensity goes beyond looking at how quickly the ball moves. Specifically, when the intensity is high, players will have less time to execute actions: after receiving a pass, a player in a league with a higher-level of play will have to quickly decide what to do with the ball (e.g., due to being pressured). Moving the ball quickly and accurately requires (a)

the cognitive skills to quickly select the next action, (b) the physical skill to position oneself to perform the selected action, and (c) the technical skill to execute the action. Indeed, the ability to move the ball quickly after receiving it is seen as an indicator of player potential [199]. Players might excel in a lower-level league where they have slightly more time on the ball, but may not have the ability to maintain their performance when confronted with a faster pace where they are afforded less time.

This chapter hypothesizes that it is possible to measure the amount of time using match event data. Concretely, we propose the Pace of Play (PoP) metric that simply calculates the amount of time a player uses to move the ball after receiving it. We focus on evaluating our metric in the context of women's football because assessing the relative level of leagues is especially challenging within the women's game due to its rapid developments, fewer historical transfers to learn from, and fewer matches between teams from different leagues than in men's football [86, 29]. These make it hard to compare leagues in women's football. We hypothesize that our PoP metric can aid in the recruitment process when signing players from other leagues. For this we investigate the following questions:

1. Can we use PoP as a proxy for the level of play?
2. Is there a relationship between a player's PoP and the playing time after a transfer to a new league?
3. Can we identify interesting players using our PoP metric?

5.2 Methods

We want to capture the time used by a player between receiving the ball and playing a footed pass to another teammate. We focus on passes because they are the most common action during a football match (65% of actions in our dataset are passes). Our goal is to focus on situations where a player has to control the ball and then deliberately moves it. Therefore, we will only consider passes that were preceded by either a teammate's pass or the player herself intercepting the ball and subsequently executing a pass. Consequently, we exclude passes preceded by carries or take-ons and clearances. Carries or take-ons require different skills and including them would be unfair as a good carry or take-on followed by a pass can actually indicate quality. Clearances (e.g., removing the ball from your own box to prevent the opponent from scoring) do not reflect moments where a player deliberately decides where to pass towards.

To do so, we use event data which records all on-the-ball actions that occurred during a football match. Each event is annotated with information such as its type (e.g., pass, shot, dribble), location on the pitch, and a time stamp among others.

For a given pass we compute the Pace of Play (PoP) Metric as follows:

$$PoP(a_i) = T_{a_i} - T_{a_{i-1}}$$

where T_{a_i} is the start time of action a_i, which is the moment the ball leaves the foot. This measure serves as a proxy for the time needed to control the ball, to decide where to go next and to execute the next action. PoP is measured in seconds.

5.2.1 Data

We consider match event data for all matches played between 2017 and 2021 in the women's first divisions of England, Spain, Germany, France, and US, and the UEFA Women's Champions League. In total, it consists of data from 3,253 matches and has 4,059,471 actions in total. It was provided in the SPADL format [63]; see Figure 5.1 for a visualization of this format.

Excluding passes that were preceded by clearances or carries yields a total of 1,656,869 passes that will be analyzed. Those passes were executed by 4,404 players playing for 172 teams.

Our analyses require information about interleague transfers. We identify interleague transfers from our match event data as follows. For every player, we determine their team and league for each season (e.g. 2020/2021, etc). An interleague transfer occurs if a player appears in a match in two different leagues in consecutive seasons. To ensure sufficient data to compute the PoP metric, we excluded any player who had < 50 passes that met the criteria for contributing to PoP. Furthermore, we excluded mid-season transfers to ensure sufficient data is available about the player's performance for their new team. This results in a dataset of 202 interleague transfers.

5.2.2 Relationship Between PoP and League/Team Strength

We hypothesize that the PoP metric is related to a league's level. That is, we expect the PoP value to be lower (i.e., indicating a faster pace of play) when the level is higher. As the best teams and players in European club football play in the Champions League knockout stages, we hypothesize that the lowest

	time	player	team	actiontype	result	bodypart
1	40m34s	Patri Guijarro	Spain	pass	success	foot
2	40m36s	Lucia Garcia Cordoba	Spain	pass	success	foot
3	40m38s	Ona Batlle Pascual	Spain	pass	success	foot
4	40m43s	Patri Guijarro	Spain	pass	success	foot
5	40m46s	Maria Pilar Leon Cebrian	Spain	dribble	success	foot
6	40m50s	Maria Pilar Leon Cebrian	Spain	cross	success	foot
7	40m52s	Aitana Bonmati Conca	Spain	shot	success	head

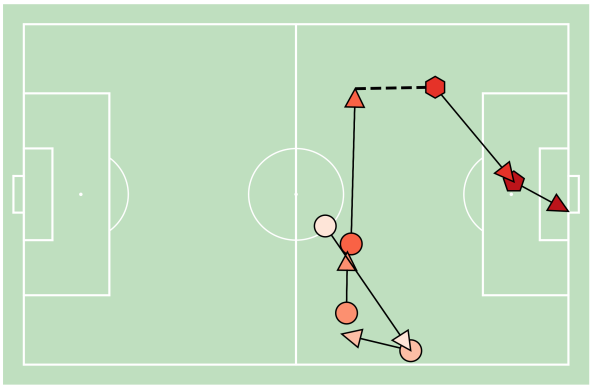


Figure 5.1: Event data for Aitana Bonmati's header goal in the 41st minute in the match against Finland during the 2021 European Championship. For all actions the time of execution, the player and team involved, the type of action, the result and the bodypart used are annotated.

PoP values amongst club teams in Europe will occur in these matches. To investigate this hypothesis, we compute the average PoP per league as follows. We compute the average PoP for each match in the league. We take the mean over these values to compute the league average.

We consider the Champions League group stages and Champions League knockout stages as separate leagues to enable a comparison in pace of play between those two different stages of the tournament. The matches in the knockout stages are between the best teams of Europe, whereas the difference in team level can be quite big in the group stages as the champions of lower-ranked domestic leagues participate in the group stages. We use one-sided t-tests to test whether the PoP in the WCL knockout stages is significantly lower than in the WCL group stages and the domestic leagues.

Similarly, we also hypothesize that higher-ranked teams exhibit a faster pace and thus will have a lower value for the PoP metric. Therefore, we investigate

the correlation between PoP and Elo ratings. Elo ratings are a common measure for team strength [73, 55, 122, 112]. We compute the Elo ratings for all teams in our dataset using the standard Elo algorithm. Since there are no cross-over matches involving NWSL teams and the teams in the European leagues, the Elo ratings for NWSL teams may not be directly comparable to those of teams from Europe. However, within the NWSL the Elo ratings do give an indication of relative team strength. Therefore, we analyze the relationship between PoP and team strength for two disjoint subgroups: one containing only European teams and one containing only the NWSL teams. For each set of teams we compute the Pearson correlation coefficient between a team's PoP in the 2021(/2022) season and its Elo rating at the end of that season.

5.2.3 Relationship Between PoP and Player Position

How quickly a player can move the ball also depends on how quickly they are pressured by an opposing player. However, whether a player is quickly closed down strongly depends on their location on the pitch: opposing teams tend to afford less time in more dangerous areas. Hence, generally speaking, players will have less time as they move further up the pitch towards the opponents goal. Moveover, there is likely less time in central areas than wide ones. This suggests that PoP values may depend on a player's position. Therefore, we investigate the distribution of PoP values for the following position groups: Goalkeepers, center backs, wing backs, center midfielders, attacking midfielders, wingers and center forwards.

5.2.4 Correlation Between PoP and Playing Time in a New League

We focus on interleague transfers and investigate whether the PoP metric can aid in the recruitment process in assessing the player's level compared to the league level. We hypothesize that players with lower average PoP values will be more likely to succeed in a new team. To measure success, we compute the percentage of matches in which the player made an appearance for her new team in the season following her interleague transfer.

A player's chance of success is tied to how her PoP relates to the PoP of their new league. We assume that the player will continue to exhibit the same PoP after transferring to their new league. To account for league differences, we assess how their PoP fits in the new league by computing their relative standing in their new league using a z-score:

$$z = \frac{PoP - \mu}{\sigma}$$

where PoP is the player's average pace of play in her previous league, and μ and σ are the average and standard deviation of her new league PoP. Note that PoP, μ and σ are computed only using data from the season prior to the player's transfer. Intuitively, these z-scores tell us how much faster (negative z-score) or slower (positive z-score) the transferred player moves the ball compared to the average player in her new league.

We compute the Pearson correlation coefficient between the players' z-scores and the percentage of matches she appeared in for their new team in the season following the transfer.

5.2.5 Identifying Interesting Players

As a measure of face validity for our metric [116, 58], we report the five U23 players with the lowest PoP value in the 2021/2022 (or 2021 for the NWSL) season. Because this last season starts in 2021, these correspond to players born after January 1st, 1998. We only include players with at least 100 passes executed in that season.

5.3 Results

5.3.1 Relationship Between PoP and League/Team Strength

Table 5.1 shows the average PoP for each league over the considered time period. We observe a lower average PoP per match in the Women's Champions League (WCL) knockout stages than the domestic leagues in Europe as well as the WCL group stages.

We observe a significantly lower overall PoP in the WCL knockout stages compared to the NWSL according to a two sample t-test with Bonferonni correction (6 tests). We do not find a significant difference with the PoP shown in the other leagues.

In terms of team strength, we observe a negative correlation between a team's Elo rating and its PoP for both sets of teams. For the 10 NWSL teams, we find a Pearson R of -0.7728 (p-value=0.00876) and for the 66 European teams we find a Pearson R of -0.3913 (p-value=0.00116). Table 5.2 shows the best

Table 5.1: The average and standard deviation of the PoP for each league. The average is computed over all matches in each league. The number of matches per league is also reported. The * indicates a significant difference according to a two sample t-test with Bonferonni correction between the WCL knockout stages PoP and the league's PoP. The WCL knockout stages, Primera División and WSL show the lowest PoP values, indicating that the ball is moved quickly by players in these matches.

League name	**Average ± stdev PoP**	**Nr of matches**
WCL knockout stages	2.953±0.259	65
Spanish Primera División	2.974±0.219	756
English Women's Super League	2.982±0.233	447
German Frauen Bundesliga	3.012±0.192	616
French Division Féminine 1	3.024±0.207	556
WCL group stages	3.032±0.218	241
American NWSL	3.035±0.236*	589

Table 5.2: The team with the lowest average PoP in each league for the 2021(/2022) season. We also report the team's end of season Elo rating and for the European teams, their absolute ranking out of the 66 teams (i.e., excluding the 10 NWSL teams). Chicago Red Stars is ranked third in terms of Elo rating in the NWSL. We also report the team's final position in their league table. Olympique Lyon and FC Barcelona have a much higher Elo rating than the other teams. For comparison, the third team in terms of Elo is Vfl Wolfsburg with an Elo rating of 1897.

Team	**League**	**Average PoP**	**Elo (Rank)**	**Final league pos**
Hoffenheim	Frauen Bundesliga	2.6526	1733 (9)	5
Ol. Lyon	Division Féminine 1	2.6987	2025 (1)	1
FC Barcelona	Primera División	2.7287	2016 (2)	1
Chicago Red Stars	NWSL	2.7981	1685 (-)	4
Man. City	Women's Super League	2.8011	1791 (4)	3

team per league in terms of PoP and also reports the Elo rating for these teams. Olympique Lyon and FC Barcelona have the highest Elo rating in our dataset, which is not surprising since all Women's Champions League finals played between 2015 and 2024 were won by either Olympique Lyon or FC Barcelona.

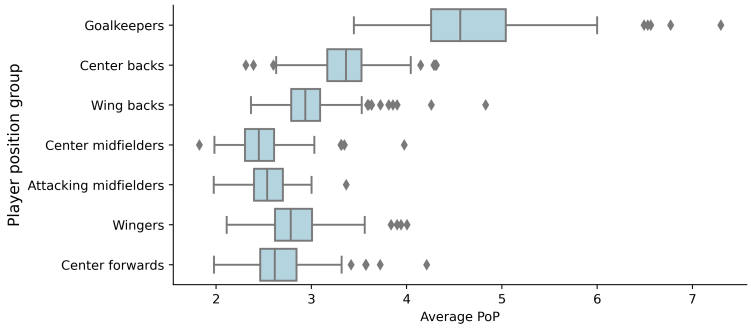


Figure 5.2: The distribution of average PoP by position group. Goalkeepers and Center backs have on average the slowest pace, whereas midfielders (both center and attacking) show the fastest pace.

5.3.2 Relationship Between PoP and Player Position

Figure 5.2 shows a boxplot comparing the distribution of per player average PoP values subdivided by position groups.

5.3.3 Correlation Between PoP and Playing Time in a New League

Figure 5.3 shows a regression plot of the Z-scores of a player's PoP compared to their new league and the percent of matches a player participated in for their new team following an interleague transfer. We observe a negative correlation between the z-scores comparing a player's PoP to their new league's pace and the percentage of matches played at their new team (Pearson's R: -0.1520, p-value=0.0308).

5.3.4 Identifying Interesting Players

Table 5.3 shows the top 5 players in terms of PoP and born after 01-01-1998.

5.3.5 Discussion

We introduce a novel quantitative metric to measure the Pace of Play (PoP) of football players using match event data. This metric can be used to assess

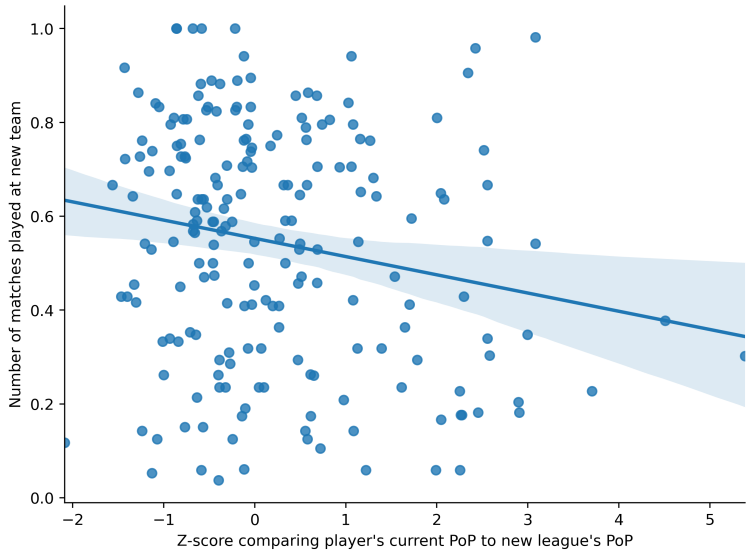


Figure 5.3: The relationship between the percentage of matches played after a transfer and the z-score comparing the player's PoP to the new league's PoP when making an interleague transfer. We observe that a lower z-score, which means that the player had a lower average PoP than the new league's average PoP, corresponds with appearing in a higher percentage of matches on average at their new club.

the speed in which players move the ball and can be used in the recruitment process when assessing a player playing in another league.

We observe that the knockout stages of the Women's Champions League show the fastest pace amongst all leagues. However, we do only find statistical differences between the WCL knockout stages and the NWSL. As the best teams progress to the knockout stages, this can be an indicator that our PoP metric can pick up differences in level of play. This aligns with our findings that our PoP metric is slightly correlated with team Elo ratings, further supporting our hypothesis that PoP is correlated to league and team strength. We found a significant correlation between the teams' average PoP and their Elo rating, both for the European teams and the NWSL teams. However, the correlation between the European teams' PoP values and their Elo ratings is not very strong (R=-0.3913). This suggests that the PoP values might also be related to a team's playing style and improvements to the PoP metric might be needed.

The fact that the NWSL shows a significantly higher PoP than the WCL

Table 5.3: The top 5 players in terms of their average PoP value in the 2021/2022 (or 2021 for the NWSL) season, and the team they played for in that season and their transfer value according to Soccerdonna (retrieved from soccerdonna.de) in January 2025. The last column shows the rank of the player's transfer value across all transfer values on Soccerdonna.

PoP Rank	**Player**	**Team (2021/2022)**	**Average PoP**	**Transfer value Jan 2025**	**Transfer value rank**
1	Claudia Pina	FC Barcelona	2.250	€300,000	30
2	Aitana Bonmati	FC Barcelona	2.296	€800,000	1
3	Patri Guijarro	FC Barcelona	2.325	€475,000	5
4	Catarina Macario	Ol. Lyon	2.387	€175,000	97
5	Ella Toone	Man. United	2.722	€275,000	41

knockout stages could be due to the way the NWSL is set up. The NWSL has rules such as a salary cap, a Transfer Fee Threshold and a yearly expansion draft [154]. These rules make the league more balanced in terms of level and cause the better players to be spread out across the league. Furthermore, American soccer is characterized by a more physical style of play which might cause the teams to show a slower PoP.

The WSL is often considered the deepest league in terms of strength across the teams in Europe given its recent investments in the league and spendings on transfers [72]. The league became fully professional in 2018 and in 2019 it reported the highest average club revenue (996 million USD) across Europe [77]. Also, in 2019, 91% of the players in the WSL have football as their primary source of income [77]. To put this in perspective, this percentage is lower for the French (80%) and German (70%) first-level leagues. These investments are visible in the fact that the WSL teams have the biggest average number of full-time technical staff per club: 9.1 (8.4 in US, 8.2 in Spain, 4.4 in France, and 4.3 in Germany) [78].

The Spanish Primera División also showed a fast pace of play in our analysis. This can be explained by the fact that they contain the team with the lowest PoP: FC Barcelona. Also, Spanish football is often characterized by its tiki-taka style of play [101], which is a style of play with fast, short passes. Our metric might thus not only be related to level of play, but might be biased to certain styles of play.

We also find that a player's PoP is correlated to the proportion of matches played at her new club after transfering to a new league. This observation is

especially interesting for the recruitment process in which the metric could be used to analyze whether a player might be able to match the new league's pace. It must be noted that we took a relatively simple approach to investigate this correlation. For future work, it would be interesting to correct for the strength of the new club and other factors that might influence a player's playing time after a transfer. Such factors include injuries, maternity leave and red card suspensions.

Beyond recruitment, the metric may also provide some insights for practitioners. For example, communicating how quickly players tend to move the ball may be useful as part of the opposition analysis prior to a match. More speculatively, players that exhibit a higher PoP may be more vulnerable to pressure.

Nonetheless, our PoP metric has some limitations. First, it does not account for the context in which the players receive the ball. A player could receive the ball under high pressure which forces her to pass it quickly again. Also, the scoreline of the match, the moment in the match, the positioning of a player's teammates, as well as the tactics of the coach might influence the player in how quickly she passes the ball again. Second, we showed that our metric is related to a player's position. Goalkeepers, defenders and wingers spend more time on the ball than midfielders and center forwards. Therefore, it is more appropriate to compare PoP values for players that play similar positions. Third, the metric ignores the time needed for the ball to move to the receiver of the pass. Therefore, if the ball needs to travel a long distance, the receiver might be wrongly punished for taking more time to play the ball again. This can be solved if the event data contains a timestamp for when the ball was received or by consulting optical tracking data. Finally, all data is imperfect and subject to measurement errors. This is likely partially mitigated by the large sample sizes that event data affords (i.e., errors cancel out). For future work, it might be interesting to investigate the use of tracking data obtained with in-stadium camera's to have very accurate measurements.

In conclusion, we believe that our PoP metric is a simple yet effective metric to assess a relevant aspect of player performance. Our experiments illustrate its potential to be useful in the recruitment process by helping to contextualize performance across different leagues. Moreover, from a face validity perspective it identifies players such as two-time Ballon d'Or winner (2023 and 2023) Aitana Bonmati, Champions League winners Bonmati, Guijarro, Pina (all with FC Barcelona) and Macario (with Olympique Lyon), and European Championship winner Ella Toone (with England). It might not be surprising that three of the top 5 players play for tiki-taka FC Barcelona. Also, these players all play most of their matches as midfielders – the position for which we see in general the lowest PoP values as shown in Section 5.3.2.

Chapter 6

Measuring Player Performance under Mental Pressure

The previous chapters focussed on the women's game. The metrics discussed in these chapters, like xG, xT and PoP, ignore the broader context in which the actions occur. An important factor that may affect player performance is the mental aspect. This chapter explores how mental pressure can influence player performance, both in terms of technical execution and decision-making.

The content of this chapter is based on the following publication [35].

L. Bransen et al. "Choke or shine? Quantifying soccer players' abilities to perform under mental pressure". In: *Proceedings of the 13th MIT Sloan Sports Analytics Conference*. 2019

6.1 Introduction

An iconic recent moment in the English Premier League was the final day of the 2011/2012 season, where Manchester City needed a win to secure the title but trailed Queens Park Rangers by one goal heading into injury time. At that point, the key question was how the players would respond. Would they up their game or wilt under the mental pressure? What happened next has cemented itself into the lore of Manchester City's: Edin Džeko equalized before Sergio Agüero

scored to improbably secure Manchester City's first league title in 44 years. While not all game situations are as pressure packed as this one, soccer players are confronted with numerous situations that impose a high level of mental pressure. While most existing soccer performance metrics focus on a player's technical and physical performances (e.g., [41, 133, 204, 203]), they typically ignore the mental pressure under which these performances were delivered. Yet, mental pressure is a recurrent concept in the analysis of a player's or a team's performance.[1][2] Hence, a metric that quantifies how mental pressure affects the performance of soccer players would have four important use cases for soccer clubs.

1. Player acquisitions: A soccer club would prefer having players who perform well under mental pressure. Being able to more accurately assess this characteristic for transfer targets would yield an additional insight that could guide player acquisition. Example: We identify Houssem Aouar and Xherdan Shaqiri as suitable replacements for Leicester City's star Riyad Mahrez.

2. Training: If a soccer club knows that a player consistently makes poor decisions in certain tense circumstances, this could be addressed. For example, the manager could coach the player on what to do in specific contexts. Furthermore, it may be possible to design training sessions tailored towards addressing these weaknesses. Example: We identify a large number of needless fouls under pressure as one of the fixable weaknesses of Orlando City's striker Dom Dwyer.

3. Tactical decisions: Certain actions are relatively more valuable or likely to succeed in high-pressure situations during soccer matches. Knowing these actions could help a soccer manager in his/her tactical planning for a match. Example: We identify crossing as a strategy with underused potential in high-pressure situations for Manchester United.

4. Lineups and substitutions: Knowing which players on a club perform well under mental pressure could be used to help inform a manager's decision making in terms of which players to line up in anticipation of a crucial game and which players to substitute on or off when a game gets tense. Example: While Juventus' central defenders Benatia, Bonucci, Chiellini and Rugani are of equal strength in normal game situations, we identify Benatia and Bonucci as the best central duo in high-pressure situations.

This chapter takes a first step towards objectively providing insight into the question: How will a soccer player perform and behave during high-mental-

[1] https://www.skysports.com/football/news/11667/11528605/juan-mata-is-still-not-trusted-by-jose-mourinho-in-the-big-games

[2] https://www.theguardian.com/football/blog/2017/feb/02/arsenal-chelsea-watford-arsene-wenger-pressure

pressure game situations? To explore this question, we employ the following approach:

1. For each situation in a soccer game, we develop a machine learned model to estimate how much mental pressure the player possessing the ball experienced using a combination of match context features (e.g., whether the game is a rivalry and the gap between the teams' current and desired league positions) and the current game state (e.g., the score and time left).
2. Since mental pressure could affect a player's performance in different ways such as his decision making (i.e., how he selects an action from several possible choices) or how well he executes a chosen action, we develop machine learned models to evaluate three aspects of each action: the choice of action, the execution of the chosen action, and the action's expected contribution to the scoreline.
3. To assess how a player reacts to mental pressure, we compare his performance metrics across different levels of mental pressure.

Using this methodology, we analyzed event data for 6,858 matches from 7 leagues. Unlike tracking data, event data is widely available for a variety of leagues across the world. As a result, teams can use our metric not only to better evaluate their own players' contributions in the crucial moments of a game, but also for recruiting players and for comparing a team's own players with players in other teams and leagues.

6.2 Measuring Mental Pressure

The models to measure mental pressure were mainly developed by my co-author Pieter Robberechts. Hence, we do not discuss the models in full detail. We refer to the original paper for full details [35].

We hypothesize that high mental pressure arises in situations that may significantly impact a team's ability to achieve its goals. Specifically, two factors affect the pressure level:

1. The context surrounding the match affects the pre-game pressure. For example, a rivalry game or a game directly impacting relegation will be more tense than a typical end-of-season game with little to nothing at stake.

2. The events in the game itself affect the in-game pressure level. Pressure mounts in close games, particularly as time winds down because a goal would increase the chances of a favorable outcome. Conversely, pressure decreases when the margin is big as a goal would only have a small impact on the expected outcome.

Therefore, we develop two novel metrics: one captures the pre-game pressure and the other the in-game pressure. Concretely, for a game g currently at game state x_t, the total pressure is given as a combination of these two metrics:

$$Pressure(g, x_t) = Pressure_{pre-game}(g) + Pressure_{in-game}(x_t) \tag{6.1}$$

Pre-game mental pressure Ideally, each game would have a category denoting its pressure level (e.g., no pressure, low pressure, normal pressure, high pressure). However, no such labeling exists and it may be difficult to define such categories with enough precision to enable manually labeling the data. Therefore, we collect pairwise rankings of games from a number of soccer experts in which they annotate which game has higher stakes. We use these rankings to fit a Gradient Boosted Ranking Trees model. The model uses features that reflect game importance, team ambition (e.g. winning the league, avoid relegation), recent performance, and game context (e.g. location, rivalrousness, match attendance). The model learns to assign scores to pairs of games, such that a higher score is assigned to the game with the higher pre-game pressure. These scores define our pressure metric.

In-game mental pressure During the game, the teams' pressure levels will change over time based on the current game state. We argue that pressure should mount when scoring a goal increases the chance of a favorable match outcome, and subside when a goal would only have a small impact on the expected outcome. One way to estimate the impact of scoring or conceding a goal on the expected match outcome is to measure the difference in win probability between the current game state and the two hypothetical game states where the home or away team has scored an additional goal. A team's in-game pressure level at time t is the sum of the increase in win probability and the increase in tie probability if the team would score at time t.

We estimate these scoring intensities from the current game state, which we describe using the following features: the number of goals scored, the goal difference, the number of yellow and red cards, the difference in Elo-ratings of the teams, the average number of attacking passes in the previous 10 time frames, and the average percentage of duels won in the previous 10 time frames. Since the importance of game state features varies over time in a non-linear

way, we model the scoring intensities using a temporal stochastic process. We build an in-game win probability model using PyMC3's Auto-Differentiation Variational Inference (ADVI) algorithm [125] on data for the 2014/2015 through 2017/2018 seasons of the English Premier League, Spanish LaLiga, and German Bundesliga.

6.3 Measuring Player Performance

Pressure could affect players' performances in different ways. Hence, when a player performs an action in a match, there are arguably three important aspects of the action that should be evaluated:

1. **Total contribution:** How helpful was the result of an action in terms of increasing the team's chance of scoring or preventing the other team from scoring? For example, a successful through ball that puts a teammate 1-on-1 with the goalkeeper will have a high contribution.

2. **Quality of the decision:** Did the player take the best possible action? Even if an action helped his team, it is possible that another, even better action was possible. That is, the player could have made a better decision. For example, in the Tottenham - Manchester City match on October 29th, 2018, David Silva chose to pass instead of shoot when he was several meters in front of an open goal (see Figure 3). This was clearly a poor choice.

3. **Quality of the execution:** How well did the player perform the chosen action? A player may make the correct decision, such as shooting at an open goal, but simply execute the action poorly (e.g., sky the shot over the bar).

In recent years, several performance metrics have been introduced to capture the contributions of actions (e.g., [37, 63, 133]). However, to the best of our knowledge, no such performance metrics have been proposed yet for measuring the quality of a decision or an action's execution. Therefore, we use an existing metric to capture an action's contribution and introduce two novel metrics to evaluate decisions and executions. The following table gives some illustrative examples about what constitutes low, average, and high ratings according to each of our metrics.

Table 6.1: Some illustrative examples about what constitutes low, average, and high ratings according to each of our metrics: Contribution, Decision and Execution.

Example	Contribution	Decision	Execution
A midfielder shoots from 35 meters out and scores, while a simple pass could have put a teammate 1-on-1 with the goalkeeper.	High	Low	High
The left back crosses the ball diagonally over 60 meters and reaches the right winger.	Average	Average	High
A midfielder attempts a through ball to put the striker in front of the opponent's goal. However, the through ball is inaccurate and the goalkeeper picks up the ball.	Low	High	Low

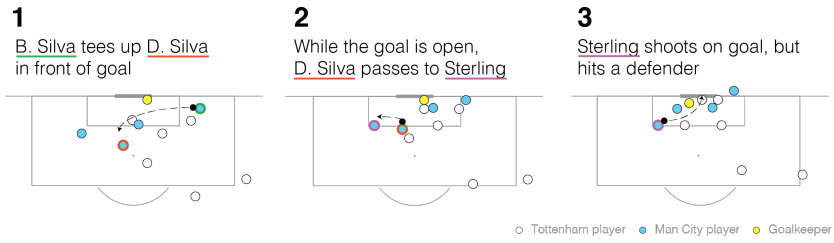


Figure 6.1: During the Tottenham - Manchester City match on October 29th, 2018, David Silva chose to pass to his teammate Raheem Sterling instead of to shoot when he was several meters in front of an open goal. This is the lowest-rated decision for the 2018/2019 Premier League season up until November 11th, 2018.

6.3.1 Contribution Rating

As our contribution rating, we use the VAEP metric introduced by Decroos et al. [63]. It considers 22 different types of actions, such as a shot, pass, or dribble, all of which can either succeed or fail. In either case, the result of an action is that it modifies the game state. The goal of the contribution rating is to measure how valuable this resulting change of game state is. It does so by computing the difference between the game state values before and after an action, where the value of a game state reflects its likeliness of yielding a goal. Formally, the contribution rating for an action a_i with outcome o_i in a game state s_i is:

$$CR(s_i, a_i, o_i) = V(s_j) - V(s_i) \tag{6.2}$$

, where s_j represents the game state resulting from performing action a_i with outcome o_i in game state s_i, and $V(s)$ gives the value for a game state s.The contribution of an action depends on several factors, including its type, location on the pitch, and outcome. Intuitively, the contribution rating is positive if the resulting game state has increased the team's chances of scoring (e.g., the ball moved to a more dangerous area via a successful dribble or pass) or decreased an opponent's chances of scoring (e.g., the keeper saved a shot). In contrast, the contribution rating is negative if the resulting game state has decreased the team's chances of scoring (e.g., a failed cross) or increased an opponent's chances of scoring (e.g., the opponent intercepted a pass).

6.3.2 Decision Rating

In each game state, a player must decide which action to perform from several possibilities. To understand if a player selected a good action to perform, we need to (1) abstract away from the actual result of the action (i.e., we need to consider what happens both when the action succeeds and when it fails), and (2) consider how the chosen action relates to the other possible actions that the game state afforded. To this end, we measure the quality of a player's choice in a game state by comparing the expected contribution rating of the chosen action with the expected contribution rating across all possible actions in a game state. Formally, the decision rating for an action a_i in a game state s_i corresponds to:

$$DR(s_i, a_i) = E(CR|s_i, a_i) - E(CR|s_i) \tag{6.3}$$

The first term is simply the chosen action's expected contribution rating. The second term requires determining other possible actions in a game state, which is challenging because we use event data and thus do not know the precise locations of the players on the pitch. Therefore, we estimate the expected contribution rating across the possible actions by predicting the next action's contribution rating given the current game state.

Computing the expected contribution rating for the chosen action Since an action can be successful or unsuccessful, we compute the expected value of its contribution rating as the weighted sum of the contribution of both outcomes:

$$E(CR|s_i, a_i) = P(o_i^+) \cdot CR(s_i, a_i, o_i^+) + P(o_i^-) \cdot CR(s_i, a_i, o_i^-), \tag{6.4}$$

where $P(o_i^+)$ is the probability that action a_i succeeds, and $P(o_i^-)$ is the probability that it fails. To predict the probability that a given action will be successful, we train a binary classifier on historical match data for each type of action (e.g., pass, dribble or interception). We use a Gradient Boosted Trees model because it produces well-calibrated probability estimates, which is important for this task. Positive examples are successful actions (e.g., passes finding a teammate or shots resulting in a goal), whereas negative examples are unsuccessful actions (e.g., missed shots). The features include an action's start and end locations, the body part used to execute the action, and the start and end locations of previous actions in the sequence. For action types for which the end location might contain information about the success (e.g., for a shot), we exclude this feature.

Estimating the expected contribution rating across all actions We use historical observations of actions performed in highly similar game states to estimate the expected contribution rating across the possible actions in a game state. To this end, we train a Gradient Boosted Trees model that predicts the next action's contribution rating given the current game state. We use the following features to describe the current game state: the current location of the ball, the start and end locations of the previous two actions, the action types of the previous two actions, and the speed of the sequence (consisting of the last two actions) captured in distance traveled and time covered.

6.3.3 Execution Rating

The execution rating attempts to assess whether a player executed an action well or not, regardless of whether the selected action was a good choice. Intuitively,

we want to reward players who successfully perform a difficult action such as completing a through ball or connecting on a long-range shot. Similarly, we want to punish players who flub an easy action such as having a lateral pass to an open teammate under no pressure. To this end, we measure the execution rating by computing the difference between the observed outcome of the action (e.g., did the cross reach a teammate or did the shot go in) and the predicted probability that the action would be successful. As an example, our model assigns a high execution rating to Zlatan Ibrahimović's first goal with the Galaxy against Los Angeles FC, which was a brilliant long-range shot over the keeper. Formally, the execution rating for an action a_i with outcome o_i corresponds to:

$$ER(a_i, o_i) = [o_i^+] - P(o_i^+), \tag{6.5}$$

where $[o_i^+]$ takes the value of one if o_i succeeds and is zero otherwise, and $P(o_i^+)$ is given by the action success predictor from the previous section.

6.4 Measuring Player Performance Under Mental Pressure

To analyze the performances of players under different mental pressure levels, we need to combine our pressure model and three performance metrics. First, we use our pressure model to make a global ranking of the mental pressure level of each situation in all the analyzed matches. We then label each situation's pressure level as follows:

1. High-pressure situations fall in the top 20% of the ranking;
2. Normal-pressure situations fall in the middle 60% of the ranking; and
3. Low-pressure situations fall in the bottom 20% of the ranking.

Second, we rate all actions in all considered matches using each of our three performance metrics. Third, for a given player or team, we aggregate our three performance metrics as a function of the pressure level under which the action was performed. In our analysis, we consider the following aggregations of our performance metrics:

- Average contribution rating per 90 minutes;

- Average contribution rating per 90 minutes per action type;
- Percentile rank contribution rating per 90 minutes;
- Average decision rating;
- Average execution rating per action type;
- Percentile rank execution rating per action type.

Since a player's contribution heavily depends on his number of minutes played, we always report contribution ratings normalized per 90 minutes of play. Since the execution ratings for different types of actions have different distributions (i.e., a shot is typically less likely to result in a goal than a pass is to reach a teammate), we never aggregate execution ratings across action types.

6.5 Use Cases

Our analysis relies on event data from the English Premier League, Spanish La Liga, German Bundesliga, Italian Serie A, French Ligue 1, Dutch Eredivisie, and North-American Major League Soccer (MLS). For each league, we consider the 2016/2017 and 2017/2018 seasons as well as the ongoing 2018/2019 seasons up until November 11th, 2018. Additionally, we include the 2014/2015 and 2015/2016 Premier League seasons to analyze the long-term evolution of players. For each situation in the matches we analyzed, we compute the pressure level as well as all three of our performance metrics for the performed action. We compare a player's average contribution per 90 minutes, as well as his average decision, and execution ratings in low-pressure situations to those in high-pressure situations. We only consider players who played at least 900 minutes in total, including at least 180 minutes in each of high- and low-pressure situations. In the remainder of this section, we demonstrate how our metrics can help clubs, managers and coaches in addressing the following key questions:

1. **Player acquisition:** Does a player perform well under pressure?
2. **Training:** Which are a player's recurring poor decisions in certain tense circumstances, such that they can be addressed during training?
3. **Tactical guidelines:** Which tactical plans are more likely to succeed in high-pressure situations?
4. **Lineups and substitutions:** Which players should a manager line up in anticipation of a crucial game or substitute when a game gets tense?

6.5.1 Player Acquisition: Replacing Riyad Mahrez at Leicester City

In this use case, we step into the shoes of Leicester City's technical director in the summer of 2018. After granting Algerian winger Riyad Mahrez's wish to play for a top-six club, he must bring in a suitable replacement for his star player. Hence, this analysis only considers the data until the end of the 2017/2018 season. As seen in Figure 6.2a, our analysis of Riyad Mahrez' contribution ratings per action type shows that the winger contributed mostly to Leicester City's performances with valuable dribbles, take-ons, shots, and crosses. Furthermore, our analysis of his 986 high-pressure minutes shows that he performs particularly well in tense situations. Namely, by both making better decisions and executing his shots better, Mahrez contributed considerably more to Leicester City's performances in high-pressure than in low-pressure situations. Across Europe's top-five leagues, we identify three candidate replacements under the age of 28 who, like Riyad Mahrez, excel in dribbles, take-ons, shots and passes: Xherdan Shaqiri (Stoke City), Houssem Aouar (Olympique Lyonnais), and Rachid Ghezzal, who eventually joined Leicester City from Monaco on August 5th, 2018. Figure 6.2b shows the total contribution per 90 minutes in low and high-pressure situations for these players. While not contributing as much as Mahrez, both Aouar's (547 high-pressure minutes) and Shaqiri's (1477 high-pressure minutes) contribution improve under pressure. In contrast, Leicester City's signing Ghezzal (358 high-pressure minutes) appears to choke under pressure. Figures 6.2c and 6.2d show how the average decision and shot execution ratings depend on the pressure level. Mahrez and Shaqiri improve their decision making when under high mental pressure, whereas only Mahrez and Aouar improve their execution ratings. When under pressure, Ghezzal both seems to make worse decisions and has difficulty executing shots. We conclude that Ghezzal was probably not the best option for Leicester City to buy as a replacement for Mahrez. We observe that Mahrez was able to deliver in high-pressure situations whereas Ghezzal chokes under high mental pressure. Therefore, our analysis suggests that both Aouar and Shaqiri would have been better replacements for Mahrez as they have similar skills, including the ability to perform in high-pressure situations.

6.5.2 Training: Dom Dwyer's Fouls

If a club knows that a player consistently makes poor decisions in certain tense circumstances, this could be addressed. The manager could give a player insight in how his decisions have a negative impact on the team and coach the player on how to act differently. Furthermore, it may be possible to design

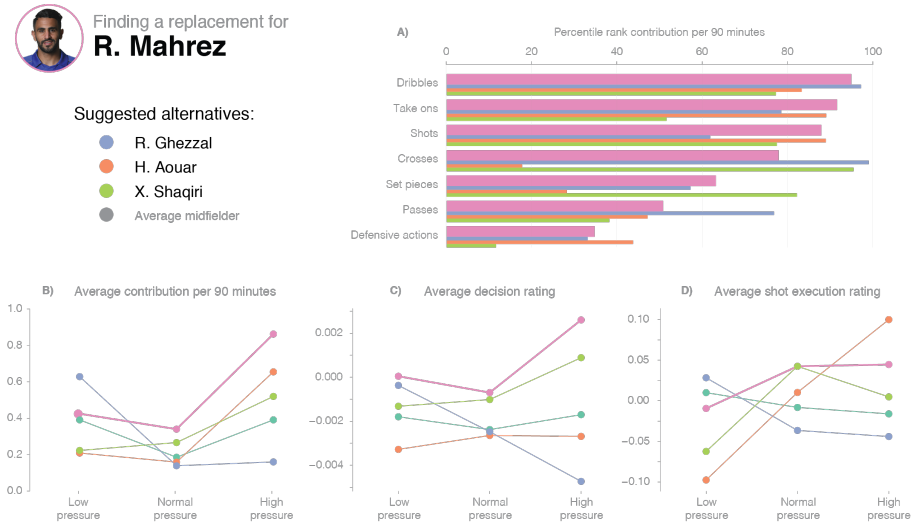


Figure 6.2: A comparison between Riyad Mahrez' and his suggested replacements at Leicester City: Aouar, Ghezzal and Shaqiri. Graph A compares the contribution ratings per action type for these players. Graphs B, C and D compare their performance under pressure.

training sessions tailored towards addressing these weaknesses. To showcase this use case, we compare the fouls committed by Orlando City's central striker Dom Dwyer in low-pressure (916 minutes) and high-pressure situations (2135 minutes). As Figure 6.3 illustrates, Dwyer commits significantly more fouls inside the opponent's penalty box when the pressure is high. Moreover, our metrics indicate that the majority of his fouls in high-pressure situations have a negative decision rating, meaning that they simply hand the possession to the opponent (instead of, for example, blocking a counter attack). His coach could use this insight to work on this aspect of Dwyer's game during training.

6.5.3 Tactical Planning

Manchester United: dribble or cross? Manchester United relies on dribbling in high-pressure situations (1818 high-pressure minutes), with the number of dribbles per 90 minutes more than doubling from 19.41 to 43.88 under pressure. However, as Figure 6.4 illustrates, this is not a very successful strategy as the average contribution from dribbles per 90 minutes declines under pressure for all offensive players except Rashford. Our metrics suggest that crossing the ball would be a more successful strategy. Especially the crosses from players

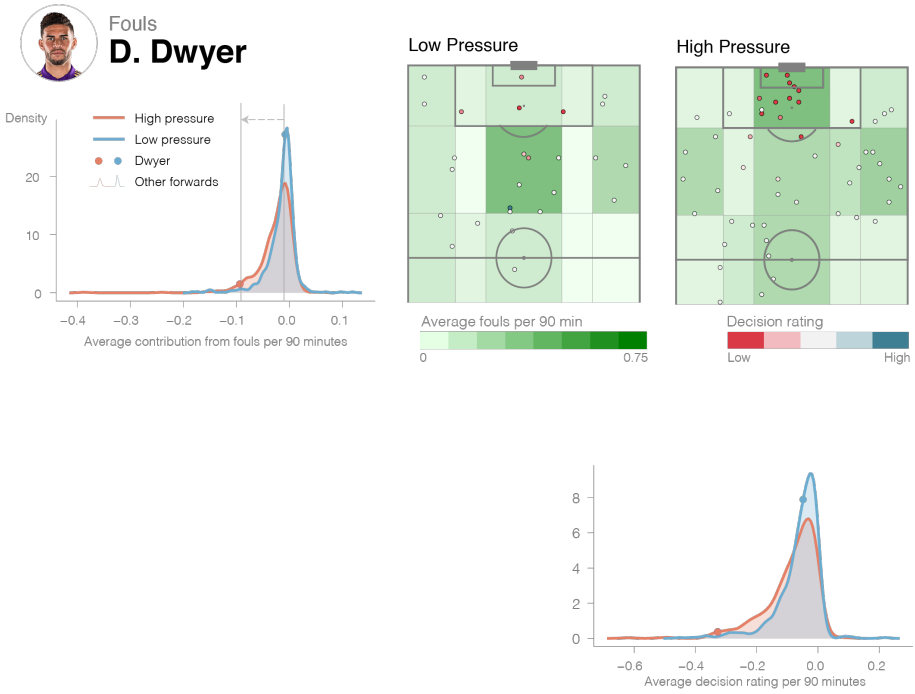


Figure 6.3: A comparison between the fouls committed by Dom Dwyer in low-pressure and high- pressure situations. While under pressure, Dwyer commits significantly more needless fouls, handing ball possession to the opponent.

like Valencia, Pogba and Mata contribute to great scoring opportunities under pressure. Moreover, with players like Lukaku and Fellaini, United has the power to finish off these crosses.

Throw-ins: better decisions under high mental pressure Throw-ins are the action type for which we observe the largest difference between low-pressure and high-pressure situations. For almost all teams, their contribution and decision ratings increase, while their execution ratings remain the same. This is an interesting observation which suggests that teams miss the opportunity to create more danger with throw-ins early in the game. Moreover, Liverpool's manager Jürgen Klopp also recognized the increasing importance of throw-ins in soccer by appointing a dedicated throw-in coach in the summer of 2018.[3] As an example, Figure 6.5 shows the average decision ratings per 90 minutes

[3]http://global.espn.com/soccer/club/liverpool/364/blog/post/3615088/liverpool-sign-throw-in-record-holder-why-jurgen-klopp- thinks-thomas-gronnemark-can-be-key

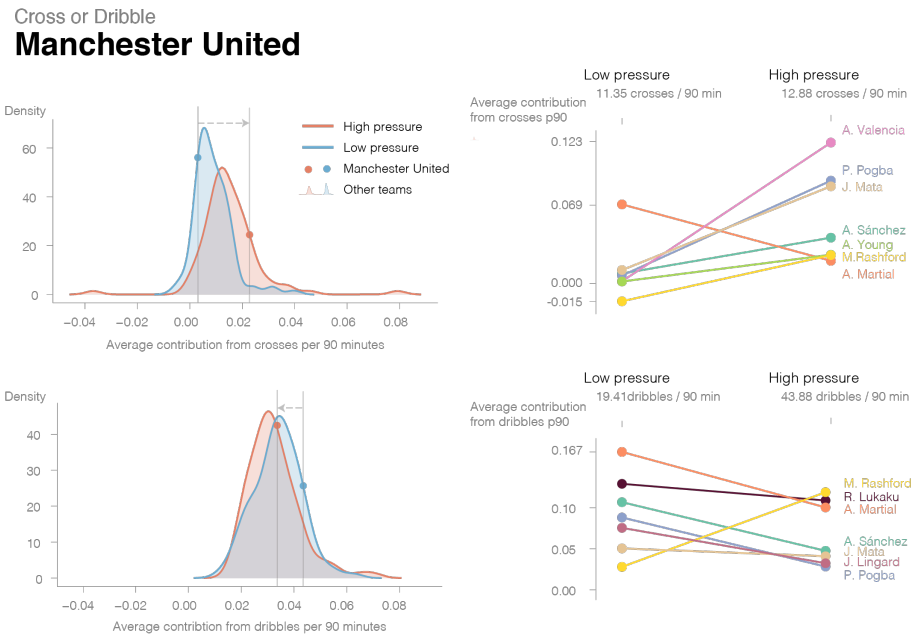


Figure 6.4: Manchester United typically attempts to use dribbles to force a breakthrough in high- pressure situations. Yet, the contribution from these dribbles decreases, while the contribution from crosses increases under pressure. This observation suggests that, in tense games, crossing the ball might be a more valuable strategy than dribbling.

of Leicester City's throw-ins and how the throw-in behavior differs between low-pressure and high-pressure situations. While one might assume that the lower decision ratings in low-pressure situations arise from more backward throws (e.g., to ensure ball possession), this is not the case. Instead, players are more inclined to throw forward in low-pressure situations. Throw-ins from a position near the opponent's goal line are the exception, with more balls thrown inside the penalty box in high-pressure situations. The Leicester City manager could use this information to train his players to apply their high-pressure throw-in behavior throughout the entire game. Similarly, the opposing manager could use these insights in his tactical discussion to highlight how his players should alter their defensive organization against throw-ins in different game situations when playing Leicester City.

6.5.4 Lineups and Substitutions

Sending the right players on the pitch is a key task of the manager. Essentially, this task comes down to predicting which players will perform best in a given situation. A player's expected performance under pressure is therefore an important variable while making these decisions. Our metrics could help inform a manager's decision making in terms of which players to line up in highpressure games or which players to substitute during a tense game. As an example, Figure 6.6 compares the average contribution per 90 minutes of Juventus' central defenders Medhi Benatia (457 high-pressure minutes), Leonardo Bonucci (201 high-pressure minutes), Giorgio Chiellini (470 high-pressure minutes) and Daniele Rugani (278 high-pressure minutes), PSG's offensive midfielders Ángel Di María (451 high-pressure minutes) and Julian Draxler (359 high-pressure minutes), and Chelsea's strikers Olivier Giroud (617 high-pressure minutes at Arsenal and 127 at Chelsea) and Álvaro Morata (477 high-pressure minutes). For each of these cases, our metrics suggest that different players should be selected in high-pressure situations. While all of Juventus' central defenders have a similar average contribution in games with an average pressure, Benatia and Bonucci outperform Chiellini and Rugani in high-pressure situations. Similarly, di María levels up when the pressure mounts, outperforming Draxler. Finally, although Morata has in general a higher contribution than Giroud, he seems to choke under pressure; Giroud, on the other hand, established himself as a clutch-goal game-killing striker at Chelsea. Note, however, that we include data from Giroud's period at Arsenal due to lack of enough high-pressure moments at Chelsea.

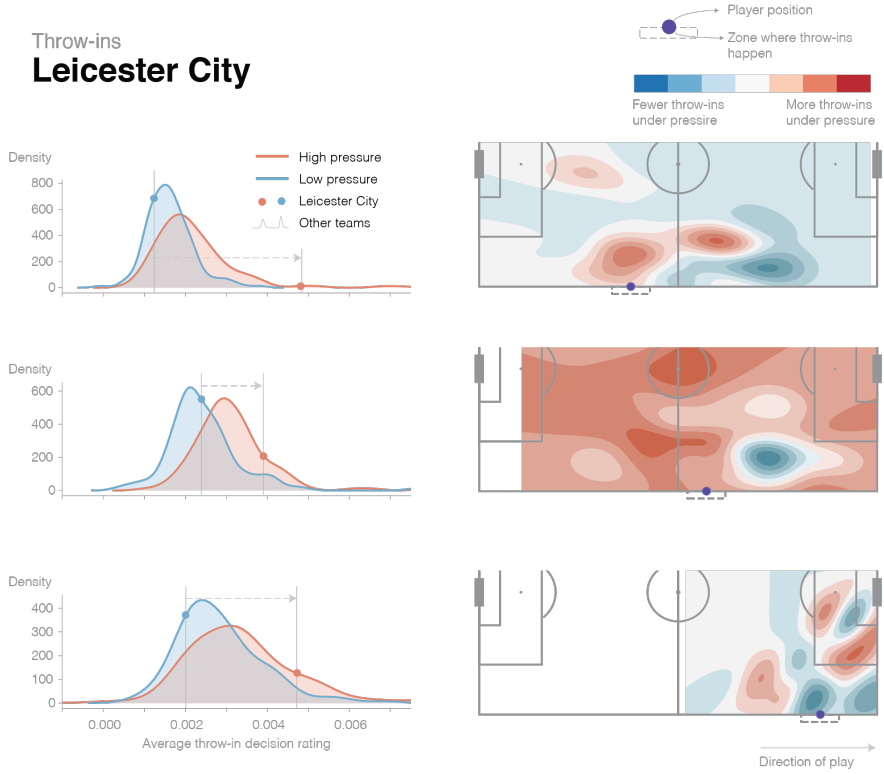


Figure 6.5: Average decision ratings per 90 minutes of Leicester City's throw-ins, compared to the league average (left) and probability maps of where these throw-ins end up (right). Data is modeled as a two-dimensional Gaussian distribution with red zones showing where Leicester is more likely to throw under pressure (compared to when the pressure is low) and blue zones where they are less likely to throw.



Figure 6.6: A comparison of the average contribution per 90 minutes between Juventus' defenders, PSG's offensive midfielders and Chelsea's strikers. These graphs show that it might be worthwhile to select different players in high-pressure situations. For Giroud, we include data from his period at Arsenal due to lack of enough high-pressure moments at Chelsea.

6.6 Observations

This section presents some interesting observations which are based on the analysis of our performance and pressure metrics. First, we show that Neymar's performance declines under mental pressure. Second, we analyze the performance under pressure of three MLS talents. Third, we remark that Liverpool tends to buy players who perform better under high pressure. Finally, we show that our findings go beyond detecting performance changes due to fatigue.

6.6.1 Neymar's Performances Decline Under Mental Pressure

Figure 6.7 shows how the performance of Neymar (Barcelona and PSG) varies between high- and low-pressure situations. Neymar evolves from a player with a very high average contribution per 90 minutes under low pressure (based on 997 low-pressure minutes) to an average player in highpressure situations (based on 472 high-pressure minutes). Further analysis shows that this decrease in performance can be explained by a poor decision making behaviour. Neymar makes more decisions that get a very low or average rating in tense situations, at the cost of fewer decisions with a high rating. In contrast, Neymar's executions do not suffer under pressure as they improve for all action types.

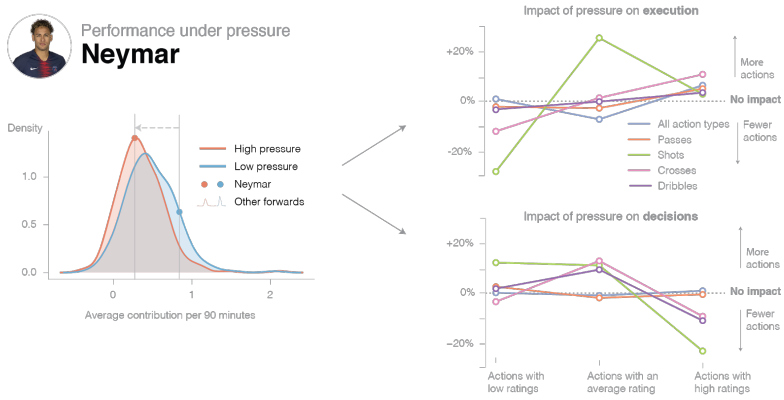


Figure 6.7: Under pressure, Neymar evolves from a player with a very high average contribution per 90 minutes to an average player. This decrease in performance can mainly be attributed to poor decision making in high-pressure situations.

6.6.2 MLS Stars of the Future

Recently, European clubs have been increasingly scouting MLS, as exemplified by Bayern Munich's signing of Alphonso Davies and RB Leipzig's signing of Tyler Adams. Atlanta United's attacking midfielder Ezequiel Barco is another MLS talent that could soon move overseas. These players represent the top three players aged under 20 in terms of average contribution per 90 minutes in the 2018 MLS season. Figure 6.8 shows the percentile rank for the contribution ratings from defensive actions (i.e., interceptions, tackles, clearances and fouls), offensive actions (i.e., crosses, shots, dribbles and takeons) and passes for all three players under different pressure levels for the 2016, 2017, and 2018 seasons. Since the distribution of the contribution per 90 minutes differs per type of action, we show the percentile ranks for each group of action types. Davies (1271 high-pressure minutes) seems not to cope well with pressure: he is among the best players in the league in low-pressure situations whereas he is a below-average player under high pressure. His contribution per 90 minutes from offensive actions, defensive actions and passes drops as pressure increases. In contrast, Adams (1804 high-pressure minutes) and Barco (760 high-pressure minutes) have more of a mixed reaction to pressure: some actions improve and some decline.



Figure 6.8: The percentile rank contribution ratings for MLS talents Alphonso Davies, Tyler Adams and Ezequiel Barco under different pressure levels.

6.6.3 Liverpool Recruitment

Our analysis of Liverpool's signings since the start of the 2017/2018 season reveals that the Reds have a keen eye for buying players who shine under mental pressure. For Alisson (AS Roma), Alex Oxlade-Chamberlain (Arsenal), Andrew Robertson (Hull City), Fabinho (Monaco), Naby Keïta (RB Leipzig), Virgil van Dijk (Southampton) and Xherdan Shaqiri (Stoke City), Figure 6.9 shows each player's average contribution per 90 minutes in low- and high-pressure situations at their previous clubs. Except for Keïta, each of these signings obtained higher contribution ratings in high-pressure than in low-pressure situations. We omitted Mohamed Salah from our analysis since he played too few minutes under high mental pressure at his former club AS Roma.

6.6.4 Fatigue

Mental pressure, like fatigue, often increases during the game. Therefore, we run an analysis to verify that our findings go beyond detecting performance changes due to fatigue. To counter this argument, Figure 6.10 compares each player's performance under mental pressure (measured as the player's average contribution per 90 minutes in high-pressure situations) with his performance under fatigue (measured as the player's average contribution per 90 minutes in the last quarter of each game, excluding the high-pressure situations). Based on the weak Pearson correlation of 0.136, we conclude that our findings reflect something different from the fatigue that sets in towards the end of a game.

6.7 Related Work

The effects of pressure on performance have been studied in other contexts and sports. One approach is to define high-pressure situations by hand, as is done in the NBA (e.g., games within five points with five minutes left) or baseball (e.g., runners in scoring position), and then compute performance metrics (both traditional and advanced) for players in such situations. Some attempts in basketball[4] and baseball[5][6] consider in-game win probability models to identify high-pressure (or "clutch") situations. However, they only consider a limited number of actions (e.g., ignore passes, blocked shots, fouls). Our approach differs

[4] http://www.inpredictable.com/2014/03/measuring-clutch-play-in-nba.html

[5] https://www.baseballprospectus.com/news/article/38398/prospectus-feature-revised-look-clutch-hitting-part-1/

[6] https://www.baseballprospectus.com/news/article/38519/prospectus-feature-revised-look-clutch-hitting-part-2/

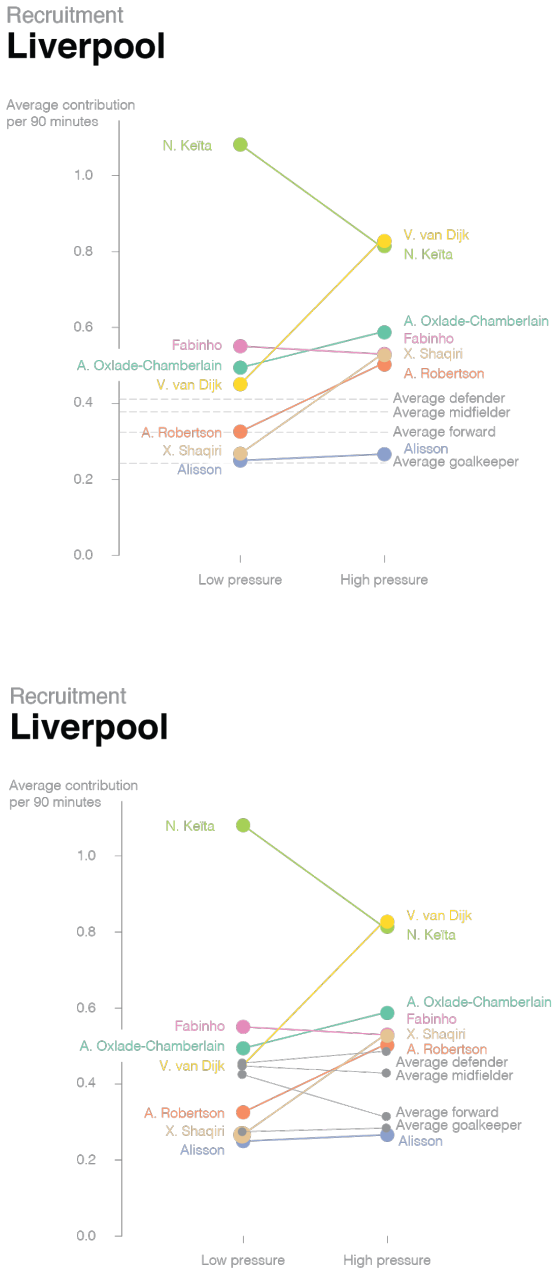


Figure 6.9: Average contribution per 90 minutes of Liverpool's recent signings in low- and high-pressure situations at their former clubs. For comparison, we include the average contributions of an average player (by position) irrespective of the mental pressure situation. Since the 2017/2018 season, Liverpool has almost consistently signed players who perform better in high-pressure situations. Naby Keïta is the only exception.

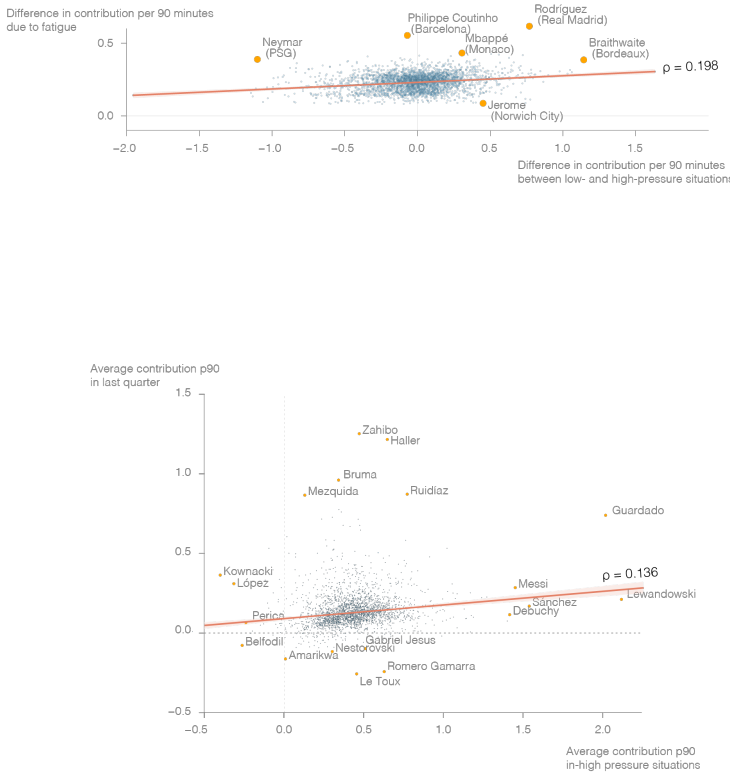


Figure 6.10: A comparison between each player's performance in high-pressure situations and performance under fatigue (i.e., in the last quarter of a game, excluding the high-pressure situations). Performance is measured as the average contribution per 90 minutes. The low Pearson correlation of 0.136 indicates that our pressure metric does not measure fatigue.

in several important ways. First, our pressure model considers both pre-game and in-game factors. Second, it considers a broader set of actions, including both offensive and defensive ones. Third, it rates three different aspects of each action. Beyond that, performance under mental pressure has been extensively studied in the cognitive sciences' literature regarding its mechanisms and moderators [107]. This research is mostly experimental and induces pressure through artificial manipulation, distraction and self-focus [100].

The component parts of our pressure model have also been considered to varying degrees. Pregame pressure has been considered in the context of elimination games in NBA playoffs to research how the threat of severe losses affects performance [146]. This research defines a high-pressure game as one where the team faces elimination in case of a loss. They do not take multiple features into account, as we do. In-game win-probability models have been extensively studied in basketball [92], American football [164], and baseball [194]. In contrast, they have received less attention in soccer [23], probably because its low-scoring nature makes it more challenging to analyze. Additionally, there is a long line of work focused on designing advanced metrics for valuing actions and decisions both in soccer [39, 63, 133] and other sports [17, 50, 159, 137].

6.8 Conclusions

This chapter has taken a step towards trying to objectively understand how high-mental pressure situations affect the performances of soccer players. Given that soccer players are often confronted with such situations, more insight into the link between pressure and performance could help in numerous ways. We illustrated concrete use cases about how it could inform acquiring players, coaching individual players, making tactical decisions, and deciding on lineups or substitutions. Some of our findings may have direct practical implications. For example, our analyses indicate that almost all teams could benefit by adjusting how they treat throw-ins. Furthermore, we made several interesting observations, such as that Liverpool, whether by design or not, has recently targeted players who excel under pressure and that Neymar's performances declined when under pressure.

Chapter 7

Player Chemistry: Striving for a Perfectly Balanced Soccer Team

Whereas chapters 5 and 6 looked at developing player specific evaluation metrics, they ignore the question about how a transfer target's skills will fit in with a team's current players. This chapter explores that question.

The content of this chapter is based on the following publication [38].

L. Bransen and J. Van Haaren. "Player Chemistry: Striving for a Perfectly Balanced Soccer Team". In: *Proceedings of the 14th MIT Sloan Sports Analytics Conference.* 2020

7.1 Introduction

This chapter takes a first step towards objectively providing insight into the question: How well does a team of soccer players gel together? We address this question in both an observational and a predictive setting. In the former setting, we observe the chemistry between players who have actually played together. This setting is relevant for a manager who needs to decide on the best possible line-up for an upcoming match. In the latter setting, we predict the

chemistry between players who have never played together before. This setting is particularly relevant for a scout who needs to assess the fit of a potential signing with the players who are already on the team. In order to do so, we adopt the assumption that two players who have high mutual chemistry perform better than players who have low mutual chemistry, all other things being equal.

In this chapter, we present the following four contributions:

1. We introduce two chemistry metrics that measure the offensive and defensive chemistry for a pair of players. The offensive chemistry metric measures the pair's joint performance in terms of scoring goals, whereas the defensive chemistry metric measures their joint performance in preventing their opponents from scoring goals.
2. We introduce two machine-learned models that predict the offensive and defensive chemistry for a pair of players that has never played together in a match.
3. We propose a Team Builder that automatically assembles a team of eleven players from a set of players that maximizes the mutual chemistries between the players on the team.
4. We present three observations that show the utility of our chemistry metrics as well as three concrete use cases that demonstrate the practical applicability of our chemistry metrics and Team Builder.

We compute our chemistry metrics for 361 seasons in 106 different competitions and show that the partnership between Mohamed Salah and Roberto Firmino in Liverpool's 2017/2018 UEFA Champions League campaign exhibited the highest mutual chemistry between two players. We also show that Mesut Özil's chemistry has rapidly started declining following Alexis Sánchez' departure to Manchester United in 2018. Furthermore, we identify Bayern Munich as the best-suited next destination for Moroccan international Hakim Ziyech in terms of chemistry if the attacker would decide to leave Dutch champions Ajax. Using our Team Builder, we also assemble the maximum-chemistry team for a Bayern Munich side that features Hakim Ziyech as a right winger.

7.2 Dataset

Our dataset comprises Wyscout match event data for a selection of 106 domestic and international competitions since the start of the 2015/2016 season, which describe the on-the-ball actions that the players performed in each match.

While soccer data comes in many different flavors, we use match event data for developing our mutual chemistry metrics. With match event data being available for most professional football competitions around the world, this type of data strikes a good balance between data availability and data granularity. Hence, match event data is ideally suited for developing performance and style metrics that are aimed at player recruitment. We convert our match event data into the SPADL representation to facilitate our analysis [6]. Hence, for each on-the-ball action, our dataset contains a match identifier, a team identifier, a player identifier, the time in the match, the start and end location, the result (e.g., successful or unsuccessful), and the body part that the player used to perform the action. In addition, our dataset contains personal information about the players and coaches who participated in the matches in our dataset. For each person, our dataset contains their age, height, weight, nationality, mother tongue, region (e.g., Americas), subregion (e.g., South America), preferred foot, and position. Furthermore, we enrich our dataset in two ways. First, we obtain ability and playing style indicators for each player from SciSports. We use SciSkill and Potential to reflect the current and potential ability of each player [1]. We use Player Roles to describe the playing style of each player [2]. We use Physical Performance Indicators to capture the duel strength, speed and work rate of each player. Second, for each pair of players, we derive whether they were born in the same region, speak the same language and played in the same competition ahead of each season. We also compute the number of matches that they had played together ahead of each season for both club and country. Table 1 presents the characteristics of our extensive dataset.

7.3 Measuring the Mutual Chemistry for a Pair of Players

We exploit the observation that the mutual chemistry between players is reflected in their performances. Since we cannot observe the mutual chemistry between two players directly, we adopt the assumption that two players who have high mutual chemistry perform better than two players who have low mutual chemistry, all other things being equal. Hence, we need a performance metric that measures the joint performance for a pair of players. Since no such metric exists to the best of our knowledge, we introduce two novel performance metrics that measure the joint performance for a pair of players. We build on top of the VAEP framework for valuing individual on-the-ball actions of players [63], which was inspired by the EPV framework for basketball [50]. The VAEP framework rewards each on-the-ball player action based on its impact on the player's team's chances of scoring and conceding a goal, where the reward can

be positive or negative. We present one metric that captures the joint offensive impact and one metric that captures the joint defensive impact. We now discuss each of both metrics in turn.

7.3.1 Joint Offensive Impact (JOI)

We introduce the JOI metric to capture the Joint Offensive Impact for a pair of players. Intuitively, JOI quantifies the impact of the actions in which both players are involved on increasing the likelihood of scoring a goal. For example, when a pass moves the ball into a more dangerous pitch location, both the sender and the receiver will receive credit for the pass.

To obtain the JOI for a pair of players in a season, we need to address two tasks. First, we extend the notion of the VAEP rating from capturing the impact of a single action to capturing the impact of an interaction, which are two consecutive actions where each of both players performed one action each. For example, when a player passes the ball to a teammate and this teammate subsequently takes on his direct opponent, then the sequence of these two actions is considered an interaction. The sum of the VAEP ratings for the pass and the take-on reflect the impact of this interaction. Second, we compute the JOI for a pair of players in a match by aggregating the VAEP ratings for their interactions in that match.

Computing the offensive impact for interactions

We view a match as a sequence of actions $a_1^p, ..., a_n^p$, where n is the number of actions in the match and $p \in P$ refers to the player who performed the action with P being the set of all players in the match. Each action a has one of the following five types: pass, cross, dribble, take-on, or shot. Formally, we define the jth interaction between two players p and q in a match m as a subsequence of two consecutive actions as follows: $I_j^m(p,q) = (a_i^p, a_{i+1}^q)$, where a_i^p represents the ith action in a match m performed by a player p.

Furthermore, we extend the notion of the VAEP rating from actions to interactions between players. We obtain the VAEP rating for an interaction by summing the VAEP ratings for the two constituting actions. Formally, we define this extension as $VAEP(I_j^m(p,q)) = VAEP(a_i^p) + VAEP(a_{i+1}^q)$.

Computing the joint offensive impact in a season

We compute the joint offensive impact for a pair of players in a match by summing the VAEP ratings for their interactions in that match, regardless of which of both players initiated the interaction. Formally, we define $JOI_m(p, q)$ for a pair of players (p, q) in a match m as follows:

$$JOI_m(p,q) = \sum_k VAEP(I_k^m(p,q)) + \sum_l VAEP(I_l^m(q,p)) \tag{7.1}$$

where k is the number of interactions between players p and q where player p performs the first action and player q performs the second action, and l is the number of interactions between these players where player q performs the first action and player p performs the second action. Furthermore, to allow a fair comparison between pairs of players who spent different amounts of time on the pitch together within a season, we introduce $JOI90(p, q)$ that represents the Joint Offensive Impact per 90 minutes of play for a pair of players (p, q) in that season. Formally, we define $JOI90(p, q)$ as follows:

$$JOI90(p,q) = \sum_m JOI_m(p,q) \cdot \frac{90}{\sum_m MINS_m(p,q)} \tag{7.2}$$

where we sum over all matches m in a season and $MINS_m(p, q)$ represents the number of minutes that players p and q spent together on the pitch in match m.

7.3.2 Joint Defensive Impact (JDI)

We introduce the JDI metric to capture the Joint Defensive Impact for a pair of players. Intuitively, JDI quantifies the impact of the actions in which both players are involved on decreasing the likelihood of conceding a goal. Unfortunately, the match event data only describes the actions that actually happened in the match but not the actions that players prevented from happening, for instance, by smart runs or clever positioning. To overcome this restriction of the match event data, the rationale behind our JDI metric is that when an opponent underperforms their expected offensive impact, then the pairs of players that are responsible for defending this opponent likely have done well. We assume that a pair of players is good at preventing their opponents from making an offensive impact when their opponents structurally underperform their expected offensive impacts. We also assume that multiple pairs of players can be responsible for defending the same opponent. We build upon the notion of the VAEP rating to capture how well a pair of players prevents their direct

opponents from having an offensive impact. That is, we compare the offensive impact that a particular player is expected to have in a match with the offensive impact that this player actually had in the match. We distribute the credit for the difference between the expected and actual offensive impact for each player across the pairs of players that were responsible for preventing the player from being dangerous. For example, a central defender and right back will be responsible for preventing the opponent's left winger to be dangerous in many situations throughout a match. To obtain the JDI for a pair of players, we need to address three tasks. First, we compute the offensive impact above or below expectation for each opponent. Second, we compute the responsibility share for the combination of player pair and opponent. For example, a player pair consisting of a central defender and the right back will have a higher responsibility share than a player pair consisting of a central defender and left back to prevent the opponent's left winger from having an offensive impact. Third, we compute the JDI for a pair of players in a match by combining the offensive impacts above or below expectation and the responsibility shares for that match.

Computing the offensive impact above or below expectation for each opponent

We obtain the offensive impact above or below expectation for each opponent in a match by computing the difference between the expected offensive impact in the match based on earlier matches in the season and the actual offensive impact. An opponent who achieved a higher actual offensive impact than expected offensive impact has performed above expectation. In contrast, an opponent who achieved a lower actual offensive impact than expected offensive impact has performed below expectation. We determine the actual offensive impact for a player in a match by summing the VAEP ratings for the player's passes, crosses, dribbles, take-ons, and shots in that match. Formally, we compute the actual offensive impact for a player p in a match m as follows:

$$OI_m(p) = \sum_k VAEP(a_k^p) \tag{7.3}$$

where each action a_k^p is a pass, cross, dribble, take-on, or shot performed by player p in match m.

We determine the expected offensive impact for a player in a match by computing the player's average offensive impact in the matches in the same season of the same competition that were played earlier. For example, to determine the expected offensive impact of Manchester City's Kevin De Bruyne in the match

against Liverpool on matchday 12 of the 2019/2020 Premier League season, we consider De Bruyne's performances in the first 11 Premier League matches. Formally, we compute the expected offensive impact for a player p in a match m as follows:

$$E[OI_m(p)] = \sum_{1}^{m-1} OI_m(p) \cdot \frac{90}{\sum_1^{m-1} MINS_m(p)} \tag{7.4}$$

where $MINS_m(p)$ represents the number of minutes player p played in match m. However, we apply a Bayesian approach to obtain more robust estimations of the expected offensive impact for players who have played fewer than 700 minutes, for instance at the start of a season or due to an injury. We compute a weighted average between a position-specific prior and the earlier computed expected offensive impact, where the weights are linearly proportional to the share of the required 700 minutes the player has played. For example, for a player who has played exactly 350 minutes, the prior and computed expected offensive impact contribute equally to the final expectation. We determine each position-specific prior by computing the average offensive impact per 90 minutes for players on that position.

Computing the responsibility share for each pair of players and opponent

Each pair of players is responsible for preventing their opponents from having an offensive impact. However, the degree to which a pair of players is responsible for guarding a given opponent depends on the positions of the three involved players. Since we do not know the exact spatial locations of the players at the time of each action, we use the default positions of the players to determine a responsibility share for each combination for a pair of players and an opponent. For example, the pair consisting of the right central defender and the right back will have a large responsibility share in preventing the opponent's left winger from having an offensive impact.

We compute the responsibility share for each combination for a pair of players and an opponent in two steps. First, we compute the individual responsibility share for each individual player and a given opponent by computing the Euclidean distance between their positions. We overlay the pitch with the 5-by-5 grid in Table 7.1 from the perspective of each team and compute the distance between the positions in this grid. The responsibility share is inversely proportional to the Euclidean distance in the grid. The lower the distance is, the higher the responsibility share is. Second, to obtain the responsibility share for a pair, we compute the average of the individual responsibility shares of

Table 7.1: We obtain the responsibility share for each combination of player pair and opponent by computing the Euclidean distance between the positions of the players using this 5-by-5 grid.

Left Wing Forward		Striker		Right Wing Forward
	Left Attacking Midfielder	Second Striker & Attacking Midfielder	Right Attacking Midfielder	
Left Winger	Left Center Midfielder		Right Center Midfielder	Right Winger
Left Wingback	Left Defensive Midfielder	Defensive Midfielder	Right Defensive Midfielder	Right Wingback
Left Back	Left Center Back	Center Back	Right Center Back	Right Back

the two players that constitute the pair. Formally, we define the responsibility share for a pair of players (p, q) for an opponent o in a match m as follows:

$$RESP_m(p, q, o) = \frac{RESP_m(p, o) + RESP_m(q, o)}{2}, \tag{7.5}$$

where $RESP_m(p, o)$ represents the individual responsibility share of player p for defending opponent o in match m based on the Euclidean distance between their positions in the grid.

Computing the joint defensive impact in a season

We compute $JDI_m(p, q)$ for a pair of players (p, q) in a match m by summing the differences between the expected offensive impact and actual offensive impact for each opponent, weighted by the responsibility shares and accounted for the number of minutes spent together on the pitch. Formally, we define $JDI_m(p, q)$ as follows:

$$JDI_m(p, q) = \sum_o [(E[OI_m(o)] - OI_m(o)) \cdot RESP_m(p, q, o) \cdot \frac{MINS_m(p, q, o)}{90}] \tag{7.6}$$

where $MINS_m(p, q, o)$ represents the number of minutes that players p and q, and opponent o spent together on the pitch in match m.

Furthermore, to allow a fair comparison between pairs of players who spent different amounts of time on the pitch together in a season, we introduce $JDI90(p, q)$ that represents the Joint Defensive Impact per 90 minutes of play for a pair of players (p, q) in that season. Formally, we define $JDI90(p, q)$ as follows:

$$JDI90(p, q) = \sum_m JDI_m(p, q) \cdot \frac{90}{\sum_m MINS_m(p, q)}, \tag{7.7}$$

where we sum over all matches m in a season and $MINS_m(p, q)$ represents the number of minutes that players p and q spent together on the pitch in match m.

7.4 Predicting the Mutual Chemistry for a Pair of Players

When two players have spent a sufficient number of minutes on the pitch together, the JOI and JDI metrics provide useful insights in their mutual chemistry. A manager could, for instance, use these insights to decide on the strongest possible line-up for a match. However, the JOI and JDI metrics cannot be used directly for players who have not played together often enough yet. To overcome this limitation, we introduce machine-learned models that predict the joint offensive impact and joint defensive impact for any given pair of players, regardless of whether these players have ever played together. A scout could, for instance, use these predictions to assess whether a potential signing would be a good fit in terms of chemistry with the players who are already on the team.

7.4.1 Representing Players as Feature Vectors

We represent each player in our dataset as a feature vector that captures the characteristics of the player that likely influence the player's chemistry with another player. These characteristics include the player's age, position line (e.g., goalkeeper, defender, midfielder, or striker), height, weight, nationality, region (e.g., Europe), subregion (e.g., Western Europe), mother tongue, Physical Performance Indicators(i.e., ground duel strength, air duel strength, speed, and work rate) and scores for each of the 22 Player Roles. Similarly, we represent each player pair in our dataset as a feature vector that combines the characteristics of the players who constitute the pair. Furthermore, the feature vector captures whether the players have the same nationality, have the same mother tongue,

and originate from the same region and subregion. In addition, the feature vector also captures the number of matches the players have played together before ahead of the season.

7.4.2 Predicting the Joint Impact for a Pair of Players

We train two machine-learned regression models to predict the joint impact for a pair of players in a given season. One model predicts the joint offensive impact, whereas the other model predicts the joint defensive impact. To do so, we split our dataset into a training set, validation set and test set as is common in machine learning. Our training set covers the 2015/2016 through 2016/2017 seasons, our validation set covers the 2017/2018 season and our test set covers the 2018/2019 and 2019/2020 seasons until Tuesday 10 December 2019, for which we report observations and present use cases later[1]. We construct one example for each pair of players that played at least 700 minutes together. As a result, our training set contains 355,671 pairs, our validation set contains 185,927 pairs, and our test set contains 234,408 pairs. We use the feature representation from Section 4.1 to represent the player pairs. We use the JOI90 metric as the label to train the model that predicts the joint offensive impact, whereas we use the JDI90 metric as the label to train the model that predicts the joint defensive impact. We train the models using the CatBoost gradient boosting toolkit [170], which handles categorical features (e.g., region or subregion) in a natural way. We train the models on the training set, tune the hyperparameters on the validation set, and select the appropriate features by optimizing the Root Mean Square Error (RMSE) on the validation set. The model that predicts the joint offensive impact consists of 500 trees with a maximum depth of 7, while the model that predicts the joint defensive impact consists of 1000 trees with a maximum depth of 5. We compare the performance of both models on the test set to a baseline that predicts the average JOI90 or JDI90 in the training set for each example. Our models both outperform this baseline as can be seen in Table 7.2.

7.4.3 Analyzing the Machine-Learned Models

To gain a better understanding of which characteristics contribute most to the chemistry between a pair of players, we investigate the importance of each feature

[1]We include the data for competitions that are on a calendar-year cycle (e.g., MLS) to the same set as the data for competitions that are on a season cycle (e.g., English Premier League) based on the year in which the season ends. For example, the data for the 2017/2018 season and 2018 calendar year are in the same set

Table 7.2: Root Mean Square Error (RMSE) on the test set for both our models and the baselines. In both settings, our models outperform the baseline models.

	RMSE for the baseline	RMSE for our model
Joint Offensive Impact	0.05448	0.04464
Joint Defensive Impact	0.89075	0.88906

in both machine-learned models with respect to making accurate predictions. In both cases, the scores for a number of Player Roles have a considerable impact on the predictions. The scores for the Mobile Striker, Deep-Lying Playmaker and Ball-Playing Defender roles contribute to the predictions of the JOI90 metric, while the scores for the Holding Midfielder and Ball-Winning Defender roles do so for the JDI90 metric. Furthermore, the number of matches that two players had played together ahead of the season has a considerable impact on the predictions of both the JOI90 and JDI90 metric for a pair of players. This effect is most notable for players who have not or have barely played together. However, their mutual chemistry can quickly increase after playing a few matches together. Beyond the mark of 50 matches together (i.e., slightly more than a season in most competitions), this effect is much weaker. Somewhat surprisingly, the cultural features (e.g., whether the players have the same nationality, whether they originate from the same region, or whether they have the same mother tongue), exhibit only limited predictive power in both models. Supposedly, it is more important that two players communicate well with their feet and that they have complementary playing styles.

7.5 Automatically Assembling a Maximum-Chemistry Team

The ability to estimate the offensive and defensive chemistry between any pair of players enables several different applications. One potentially exciting use case for a soccer club's manager or recruitment department is to automatically assemble a maximum-chemistry team from a set of players. In addition to analyzing traditional performance metrics and style indicators, a recruitment department can also investigate the likely chemistry between the players who are already on the team and each of the transfer targets on their shortlist. To facilitate this use case, we introduce our Team Builder which automatically assembles a maximum-chemistry team from a set of players. Inspired by the approach taken in [18], our Team Builder addresses this combinatorial optimization task by solving a mixed-integer programming problem where the

goal is to select precisely eleven players who maximize the sum of their mutual offensive and defensive chemistry[2]. We allow the user to trade-off offensive chemistry for defensive chemistry by providing a weighting parameter in the optimization. Formally, we solve the following mixed-integer programming problem:

$$\begin{array}{ll} \max & \sum_{p\in P}\sum_{q\in P}(\alpha \cdot E[JOI90(p,q)] + (1-\alpha)\cdot E[JDI90(p,q)]) \cdot x_p \cdot x_q \\ \text{subject to} & \sum_{p\in P} x_p = 11, \\ & \sum_{p\in P}(GKP_p \cdot x_p) = 1, \\ & 3 \leq \sum_{p\in P}(DEF_p \cdot x_p) \leq 5, \\ & 3 \leq \sum_{p\in P}(MID_p \cdot x_p) \leq 5, \\ & 1 \leq \sum_{p\in P}(FWD_p \cdot x_p) \leq 3, \end{array} \tag{7.8}$$

where each x_p is a decision variable that indicates whether the player p is selected for the team or not, and the GKP_p, DEF_p, MID_p, and FWD_p indicators indicate whether the player p is a goalkeeper, defender, midfielder, or forward, respectively. The first constraint ensures that precisely 11 players are selected, while the other constraints ensure that a logical formation is constructed.

7.6 Observations

We now present a number of observations that result from computing the Joint Offensive Impact (JOI) and Joint Defensive Impact (JDI) metrics for all seasons in our dataset. First, we present three lists of highest-chemistry player pairs. Second, we analyze the chemistry of the Liverpool team that won the 2018/2019 Champions League. Third, we analyze the mutual chemistry between Mesut Özil and his Arsenal teammates since the start of the 2015/2016 season.

7.6.1 Highest-Chemistry Player Pairs

Table 7.3 shows the top-ten-ranked player pairs in terms of Joint Offensive Impact per 90 minutes (JOI90) across all seasons in our dataset. Mohamed Salah and Roberto Firmino top the ranking with their partnership at Liverpool in the 2017/2018 Champions League. Luis Suárez and Lionel Messi in the 2015/2016 Primera División season rank second. Yuki Nagasato and Sam Kerr complete the top three with their joint performances in the 2019 NWSL season.

[2]We use the *PuLP* Python package: https://github.com/coin-or/pulp.

Table 7.3: Player pairs with the highest Joint Offensive Impact per 90 minutes across all seasons in our dataset with at least 900 minutes played together.

Player pair	**Team**	**Season**	**Competition**	**JOI90**
M. Salah R. Firmino	Liverpool	2017/18	UEFA Champions League	0.7077
L. Suárez L. Messi	Barcelona	2015/16	Primera División (Spain)	0.6497
Y. Nagasato S. Kerr	Chicago Red Stars	2019	NWSL (USA)	0.6096
A. Milevskiy P. Nekhaychik	Dinamo Brest	2017	Premier League (Belarus)	0.6050
O. Mishurenko O. Akymenko	Inhulets	2018/19	Persha Liga (Ukraine)	0.6017
Z. Ibrahimović Á. Di María	PSG	2015/16	Ligue 1 (France)	0.5972
D. Kovacić S. Conteh	St. Pauli	2018/19	Regionalliga (Germany)	0.5873
G. Barbosa G. De Arrascaeta	Flamengo	2019	Serie A (Brazil)	0.5763
L. Sané R. Sterling	Manchester City	2018/19	Premier League (England)	0.5701
D. Tadić Q. Promes	Ajax	2019/20	Eredivisie (Netherlands)	0.5700

Table 7.4: Player pairs with the highest Joint Offensive Impact per 90 minutes in the 2018/2019 season with at least 900 minutes played together or 450 minutes for the Champions League.

#	**German Bundesliga '18/19**	**Men's Champions League '18/19**	**American NWSL '19**	**American MLS '19**	**Mexican Liga MX '18/19**
1	K. Volland K. Bellarabi	K. Mbappé Neymar	Y. Nagasato S. Kerr	J. Altidore A. Pozuelo	L. Zelarayán A-P. Gignac
2	S. Gnabry T. Müller	L. Suárez L. Messi	C. Dunn L. Williams	J. Russell K. Németh	A. Vega I. Brizuela
3	J. Sancho M. Reus	L. Insigne J. Callejón	Debinha L. Williams	C. Gil G. Bou	P. López F. Jara
4	M. Reus M. Götze	H. Ziyech D. Tadić	C. Sinclair T. Heath	C. Pavón Z. Ibrahimović	M. Bolaños F. Castillo
5	L. Alario K. Volland	D. Tadić K. Huntelaar	M. Brian S. Kerr	B. White A. Gamarra	V. Guzmán E. Cardona

Table 7.4 shows the top-five-ranked player pairs in terms of JOI90 in the 2018/2019 and 2019 seasons for a selection of leagues. In the German Bundesliga, right winger Karim Bellarabi gels well with striker Kevin Volland at Bayer Leverkusen. In the Major League Soccer (MLS), Jozy Altidore and Alejandro Pozuelo top the list, which also did not go unnoticed by the media[3]. In the National Women's Soccer League (NWSL), Chicago Red Stars' Yuki Nagasato and Sam Kerr, who led their team to the NWSL final, gel best.

Table 7.5 shows the top-five-ranked player pairs in terms of Joint Defensive Impact per 90 minutes. In the Champions League, Tottenham Hotspurs' right back Kieran Trippier gels well with midfielders Son Heung-Min and Christian Eriksen. In the Mexican Liga MX, Leon's experienced center backs Ramiro González and William Tesillo top the list.

The full results for all seasons in our dataset are available as an online appendix[4].

[3]https://torontosun.com/sports/soccer/mls/altidore-pozuelo-form-great-duo-for-tfc
[4]https://github.com/soccer-analytics-research/player-chemistry

Table 7.5: Player pairs with the highest Joint Defensive Impact per 90 minutes in the 2018/2019 season with at least 900 minutes played together or 450 minutes for the Champions League.

#	**German Bundesliga '18/19**	**Men's Champions League '18/19**	**American NWSL '19**	**American MLS '19**	**Mexican Liga MX '18/19**
1	N. Mukiele M. Halstenberg	H-M. Son K. Trippier	L. Barnes T. Nielsen	Auro O. González	W. Tesillo R. González
2	C. Günter L. Kübler	K. Trippier C. Eriksen	T. Nielsen B. Yanez	M. Figueroa A. DeLaGarza	C. Ortiz J. Basanta
3	M. Veljković N. Moisander	Fabinho T. Alexander-Arnold	M. Oyster L. Barnes	A. Lundqvist A. DeLaGarza	C. Salcedo R. Carioca
4	M. Veljković N. Şahin	G. Chiellini L. Bonucci	M. Oyster T. Nielsen	C. Mavinga O. González	J. Torres R. Carioca
5	J. Gouweleeuw M. Hinteregger	D. Olmo M. Leovac	S. Gorden S. Kerr	J. Meram L. Pírez	J. Salas A. Aldrete

7.6.2 Mutual Chemistry for the 2018/2019 Liverpool Side

Liverpool fans will not quickly forget the 2018/2019 season. The side of manager Jürgen Klopp won the UEFA Champions League after beating Tottenham Hotspur in the final and were long in the running to win their first ever English Premier League title. Moreover, several Liverpool players claimed a spot in the prestigious Ballon d'Or XI. Hence, we analyzed the mutual chemistry between the regular starters for Liverpool in their Premier League campaign. Figure 7.1 shows the mutual offensive chemistry between the Liverpool players in the 2018/2019 English Premier League season, where green lines represent high mutual chemistry. We observe very strong offensive chemistry links among the three attackers as well as between the two full backs, Andrew Robertson and Trent Alexander-Arnold, and the attackers.

Figure 7.2 shows the mutual defensive chemistry between the Liverpool players in the 2018/2019 English Premier League season, where green lines represent high mutual chemistry. We observe strong defensive chemistry links in both midfield and central defense. Interestingly, the attackers also fulfill their defensive duties. In contrast, the defensive chemistry links between the central defenders and

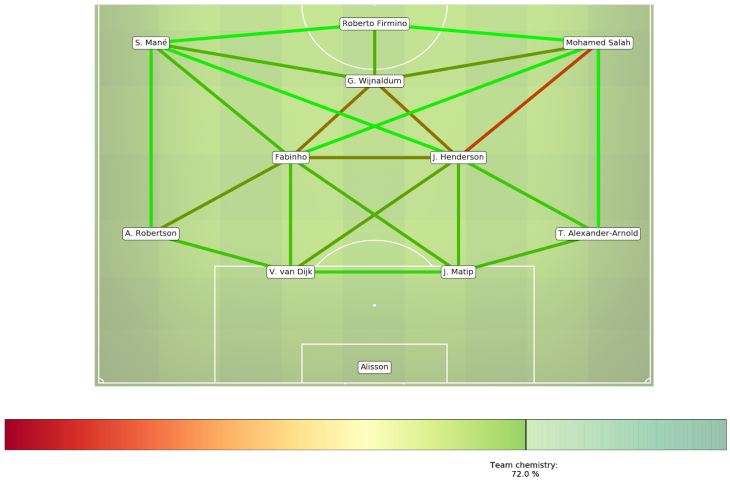


Figure 7.1: The mutual offensive chemistry between the Liverpool players in the 2018/2019 English Premier League season. Green lines reflect high offensive chemistry links, whereas dark red lines reflect low offensive chemistry links.

the full backs are rather weak. The links between Sadio Mané and Andrew Robertson on the left wing and Mohamed Salah and Trent Alexander-Arnold on the right wing have been omitted as these pairs of players have not been responsible for defending the same opponent for long enough. The link between Fabinho and Mohamed Salah has been omitted for the same reason.

7.6.3 Mesut Özil's Chemistry at Arsenal Over Time

Having joined Arsenal from Real Madrid at the start of the 2013/2014 season, Mesut Özil was a key player for The Gunners in the first few seasons after his arrival. However, over time, his performances have increasingly received criticism. Moreover, the German midfielder is even occasionally dropped from Arsenal's starting eleven. Hence, we analyze Özil's Joint Offensive Impact in the 2015/2016 through 2018/2019 seasons. As shown in Figure 7.3, the average JOI90 for the five best-ranked Arsenal player pairs has rapidly declined for both Özil and Arsenal since Alexis Sánchez' departure in 2018. Moreover, in the ongoing 2019/2020 Premier League season, the average JOI90 for the five best-ranked Arsenal player pairs cannot even match that of the average Premier League side. In summary, Arsenal's offensive chemistry has clearly declined over the years with Özil and his teammates making less of an offensive impact together. Furthermore, Figure 7.4 shows the top-ten-ranked Arsenal player

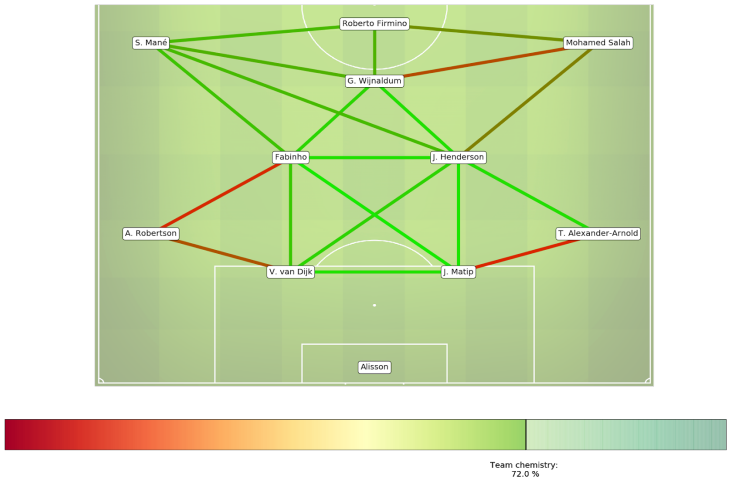


Figure 7.2: The mutual defensive chemistry between the Liverpool players in the 2018/2019 English Premier League season. Green lines reflect high defensive chemistry links, whereas dark red lines reflect low defensive chemistry links.

pairs in terms of Joint Offensive Impact per 90 minutes since the start of the 2015/2016 season. Mesut Özil features in six of these ten player pairs, including the top-four-ranked player pairs. Interestingly, the Özil - Sánchez partnership appears no less than three times in the top ten.

7.7 Use Cases

Our metrics for measuring and estimating offensive and defensive chemistry for pairs of players enable a multitude of use cases for practitioners within soccer clubs. In this section, we present three concrete use cases. First, we investigate which center back Manchester City should acquire taking chemistry with the players on their team into account. Second, we investigate which right winger Real Madrid should field from a chemistry perspective. Third, we investigate at which club Ajax' Hakim Ziyech would find optimal chemistry with the players surrounding him.

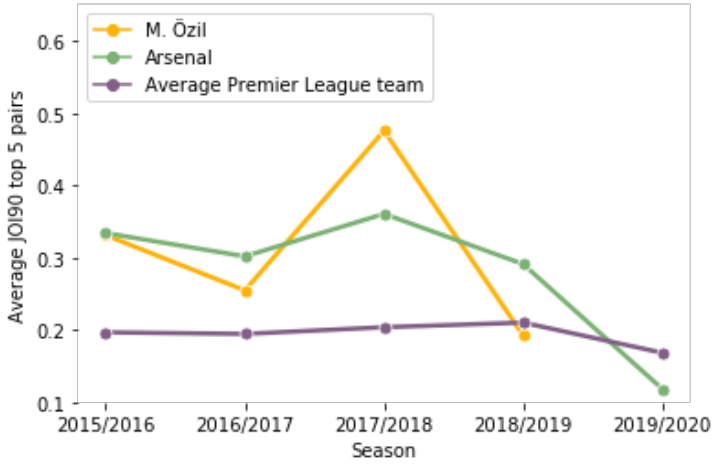


Figure 7.3: The evolution of the average Joint Offensive Impact per 90 minutes for Mesut Özil, Arsenal and the average Premier League team over the course of the past five seasons. Both Özil's and Arsenal's average have rapidly declined since Alexis Sánchez' departure in 2018.

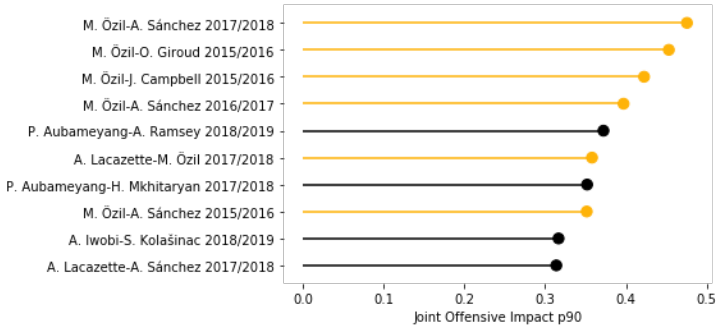


Figure 7.4: The top-ten-ranked Arsenal player pairs in terms of Joint Offensive Impact per 90 minutes since the start of the 2015/2016 season. Mesut Özil features in six of these ten player pairs, including the top-four-ranked pairs. Player pairs involving Mesut Özil are shown in orange.

7.7.1 Which Center Back Should Manchester City Acquire?

Manchester City is left with only two senior options for the center back position in John Stones and Nicolás Otamendi after Vincent Kompany returned to his boyhood club Anderlecht in the summer of 2019 and Aymeric Laporte picked up a long-term knee injury just a few weeks after the summer transfer window in the English Premier League had closed. As a result, manager Pep Guardiola has been forced to play Fernandinho, who was pivotal to City's midfield in their last few campaigns, at center back. Unsurprisingly, several high-profile center backs have been linked with a move to the Etihad Stadium during the first half of the 2019/2020 season, including Tottenham Hotspur's Toby Alderweireld, Napoli's Kalidou Koulibaly, and Internazionale's Milan Škriniar. Hence, we investigate which of these potential targets would be the best fit in terms of chemistry. To address this use case, we predict the defensive and offensive chemistry between each of the three potential targets and a selection of relevant Manchester City players. In terms of defensive chemistry, Koulibaly would gel best with center back Stones as well as defensive midfielders Fernandinho and Rodrigo, whereas Alderweireld would gel best with Otamendi, as can be seen in Figure 7.5. In terms of offensive chemistry, Koulibaly would fit best with all four players under consideration, as can be seen in Figure 7.6. However, with respect to offensive players Kevin De Bruyne, David Silva, Bernardo Silva, Riyad Mahrez, Gabriel Jesus and Sergio Agüero, Alderweireld would be a better choice due to his strong long-passing skills, as can be seen in Figure 7.7. The high offensive chemistry between Alderweireld and De Bruyne is partly due to the 41 matches that they have played together for the Belgian national team. Hence, Koulibaly would be the most obvious pick from a defensive perspective, whereas Alderweireld would be the most obvious pick from an offensive perspective. Although Škriniar would be the least obvious choice overall, he would be a better fit with City's offensive compartment in terms of offensive chemistry than Koulibaly.

7.7.2 Which Right Winger Should Real Madrid Field?

Despite signing elite players such as Eden Hazard, Luka Jovic, Rodrygo, Vinícius Júnior and Mariano Diaz for a combined transfer fee of over 300 million dollar, Real Madrid boss Zinedine Zidane is still struggling to rebuild his side following the departure of Portuguese star player Cristiano Ronaldo to Juventus in the summer of 2018. The French manager has been particularly struggling with filling the spot on the right wing as 30-year-old Gareth Bale's form has been on the decline and Lucas Vázquez picked up an early-season toe injury that has kept him sidelined for several months. Hence, we investigate whether Gareth Bale, Rodrygo or Vinícius Júnior would be the best fit for the spot on the right

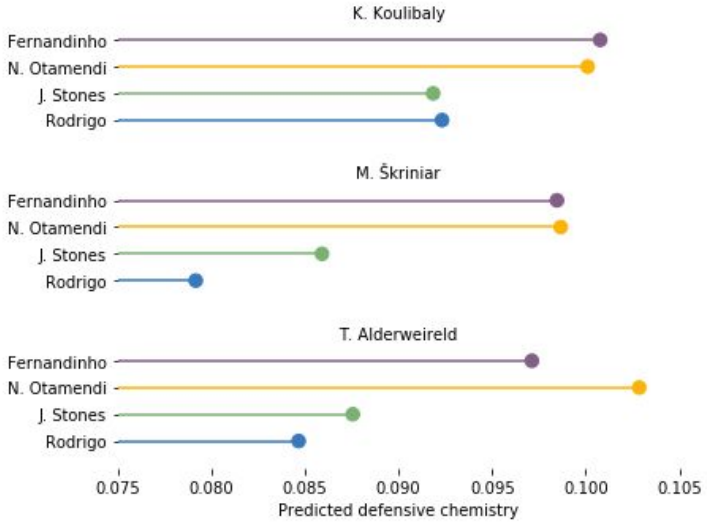


Figure 7.5: Predicted defensive chemistry between potential targets Koulibaly, Škriniar and Alderweireld on one hand and Fernandinho, Otamendi, Stones and Rodrigo on the other hand. In terms of defensive chemistry with the defensive compartment, Koulibaly would be the best fit with all players but Otamendi.

wing from a chemistry perspective, both in the short run and in the long run. To address this use case, we compute the average predicted offensive chemistry between Bale, Rodrygo and Vinícius Júnior on one hand, and all other Real Madrid players on the other hand. As shown in Figure 7.8, Bale had the highest average predicted offensive chemistry at the start of the 2019/2020 season with Rodrygo ranking second and Vinícius Júnior ranking third. However, when we artificially increase the number of matches played with each of the other Real Madrid players over time, the story changes. That is, we expect the average predicted offensive chemistry for Rodrygo and Vinícius Júnior to increase rapidly, whereas the average predicted offensive chemistry for Bale will increase much slower. Vinícius Júnior is expected to overtake Bale around the 20-extra-matches mark, whereas Rodrygo is expected to do the same around the 40-extra-matches mark.

However, when comparing the mutual offensive chemistry between each of Bale, Rodrygo and Vinícius Júnior on one hand, and Benzema, Kroos, Hazard and Isco on the other hand, Bale turns out the have the highest mutual offensive chemistry with all players but latest signing Hazard, as is shown in Figure 7.9. In contrast, the Belgian winger is expected to gel better with both Rodrygo

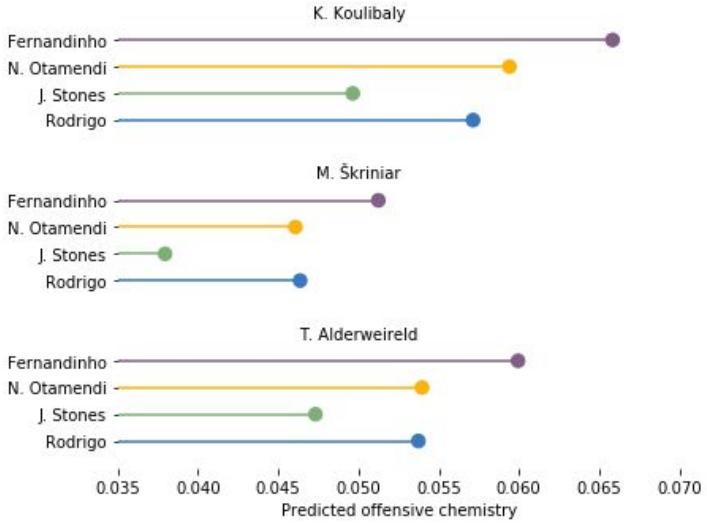


Figure 7.6: Predicted offensive chemistry between potential targets Koulibaly, Škriniar and Alderweireld on one hand, and Fernandinho, Otamendi, Stones and Rodrigo on the other hand. In terms of offensive chemistry with the defensive compartment, Koulibaly would be the best fit.

and Vinícius Júnior. In summary, based on our metrics, we would advise Real Madrid boss Zinedine Zidane to give both Rodrygo and Vinícius Júnior enough playing time in the 2019/2020 season to improve their chemistry with their teammates. Furthermore, as Eden Hazard is set to become Real Madrid's next star after the departure of Cristiano Ronaldo to Juventus, it is essential for him to gel well with his teammates. The Belgian winger seems to gel better with Rodrygo than with Bale.

7.7.3 Which Club Should Hakim Ziyech Move To?

Outclassing the opposition in the Dutch Eredivisie week after week, attacker Hakim Ziyech is set to move to an elite European league in the summer of 2020. After an incredible 2018/2019 season, which saw Ajax reach the Champions League semi-finals, several clubs showed interest in signing the Moroccan international in the summer of 2019. With a multitude of options on the table for the talented attacker, we step in Ziyech's shoes and help him evaluate his options. In particular, we help the attacker find a team where he would experience optimal chemistry with the players surrounding him, while at the

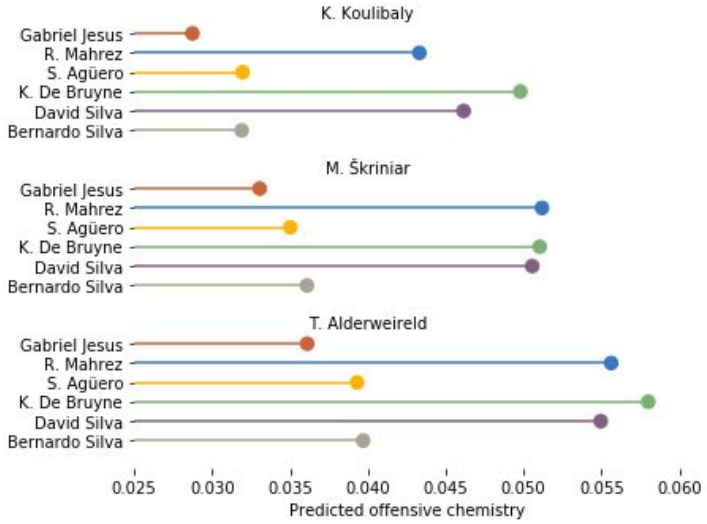


Figure 7.7: Predicted offensive chemistry between potential targets Koulibaly, Škriniar and Alderweireld on one hand, and Jesus, Mahrez, Agüero, De Bruyne, D. Silva and B. Silva on the other hand. In terms of offensive chemistry with the offensive compartment, Alderweireld would be the best fit.

same time being able to collect sufficient playing time as well as being able to further develop himself. To address this use case, we perform two steps. In the first step, we identify the five most appropriate clubs for Hakim Ziyech in terms of playing time, future development and likeliness that an agreement can be reached between club and player. To do so, we first use SciSports' machine-learned models that predict the number of minutes Ziyech would play for each candidate club, by how many points his SciSkill would increase or decrease in the next year at each candidate club, and the likeliness that each candidate club can afford to acquire Ziyech from Ajax and pay his wages. Next, we transform each of the three predictions into a score on a zero-to-one scale and compute the average score for each candidate club. Finally, we sort the candidate clubs according to their average score in descending order and select the top-five-ranked clubs: Internazionale, Roma, Chelsea, Bayern Munich, and Arsenal.

In the second step, we first predict the offensive chemistry for all possible player pairs within the selected teams as well as between Hakim Ziyech and each of the players on the selected teams, and then use our Team Builder to assemble the maximum-chemistry line-up for each of the five selected clubs. We force the

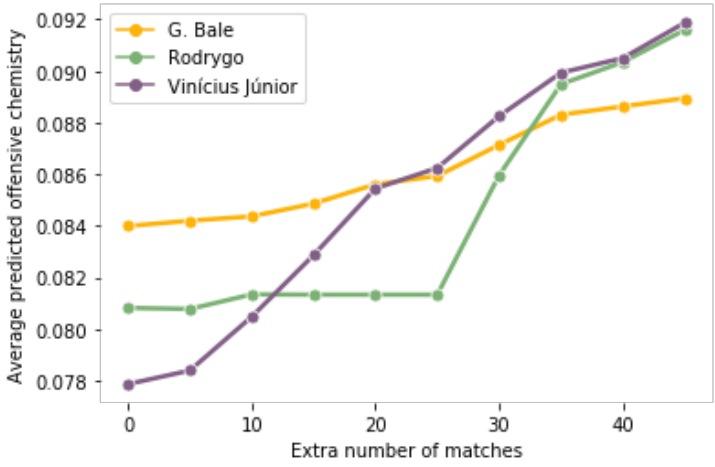


Figure 7.8: The average predicted offensive chemistry between Bale, Rodrygo and Vinícius Júnior on one hand, and all other Real Madrid players on the other hand. While Bale had the highest offensive chemistry at the start of the 2019/2020 season, he is expected to be quickly overtaken by both Rodrygo and Vinícius Júnior as they collect more matches.

Team Builder to include Hakim Ziyech and to select precisely ten players from the squad of each selected club by imposing the following additional constraint:

$$\sum_{p \in P_T} x_p = 10, \tag{7.9}$$

where P is the set of all players in the current squad of candidate club T. Furthermore, we set the T weighting parameter α to 1 to only consider the offensive chemistry.

Given that we are searching for the best-suited club from Ziyech's perspective, we investigate in which of the five optimal-chemistry teams that our Team Builder assembled, Ziyech would experience the highest average offensive chemistry with his teammates. As is shown in Figure 7.10, the Moroccan attacker would achieve the highest offensive chemistry with the players who would be surrounding him at Bayern Munich. Figure 7.11 shows how Bayern Munich's optimal-chemistry side would look like when they would have Ziyech in their ranks.

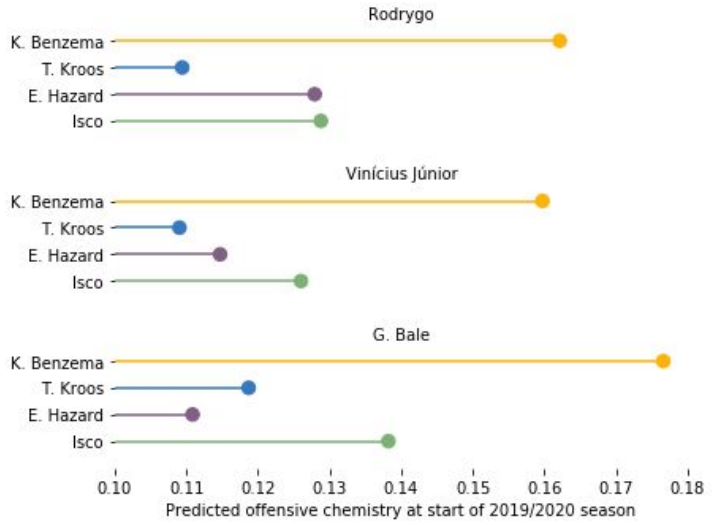


Figure 7.9: The mutual offensive chemistry between each of Bale, Rodrygo and Vinícius Júnior on one hand, and Benzema, Kroos, Hazard and Isco on the other hand. Bale has the highest mutual offensive chemistry with all players but Hazard.

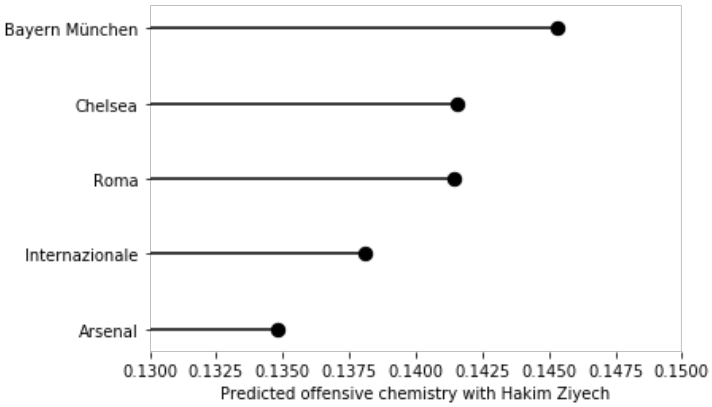


Figure 7.10: Predicted offensive chemistry between Hakim Ziyech and the other ten players in the maximum-chemistry teams of Bayern Munich, Chelsea, Roma, Internazionale, and Arsenal.

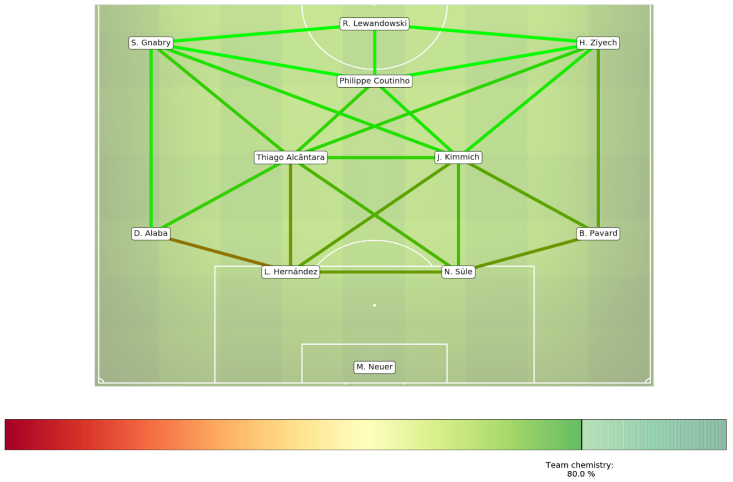


Figure 7.11: The mutual offensive chemistry between the Bayern Munich players for the maximum-chemistry team that includes Hakim Ziyech.

7.8 Conclusions

This chapter has taken a first step towards providing insight into how well a team of soccer players gels together. We presented metrics that capture the mutual offensive and defensive chemistry for a pair of players by quantifying their joint impact on scoring goals and preventing their opponents from scoring goals. We introduced two machine-learned models that predict these metrics for players who have never played together. Furthermore, we introduced a Team Builder that assembles a maximum-chemistry team from a given set of players. We demonstrated how our chemistry metrics and Team Builder can be used to identify an appropriate transfer target for a club, to decide on the best possible line-up for a team, and to provide transfer advice to a player. In the future, we plan to pursue several different avenues to further investigate the chemistry among the players on a soccer team. We aim to expand our metrics and models from capturing the mutual chemistry between pairs of players to capturing the mutual chemistry among groups consisting of more than two players. Furthermore, we aim to incorporate spatio-temporal tracking data to obtain more accurate estimates for our defensive chemistry metric.

Chapter 8

Stop Guessing: Optimizing Goalkeeper Policies for Soccer Penalty Kicks

The previous chapters examined the use of AI for recruitment. This chapter shifts to using AI for match preparation, more specifically using AI to optimize goalkeeper policies for penalty kicks.

The content of this chapter is based on the following paper [33].

L. Bransen et al. "Stop Guessing: Optimizing Goalkeeper Policies for Soccer Penalty Kicks". In: *arXiv preprint arXiv:2505.24629* (2025)

8.1 Introduction

Penalty kicks in soccer often are associated with high-leverage moments. On the one hand, 65% of the penalties are awarded in tied games or when a team is trailing by a goal. Because penalties are converted at substantially higher rates than open-play shots (78% vs. 12%), these represent a significant chance to alter a game's outcome. On the other hand, even though penalty shootouts are only used to decide the winner of games that are tied after extra time in the knockout stages of tournaments such as the FIFA World Cup and the

UEFA Champions League, the low-scoring nature of soccer means that penalty shootouts are surprisingly common. Since 1990, 23% of the men's World Cup knockout games, 27% of the men's Euros knockout games, and 23% of the men's Champions League's finals have ended in a shootout.

Consequently, teams explicitly prepare for penalties. For example, Louis van Gaal famously substituted in goalkeeper Tim Krul prior to the Netherlands' shootout against Costa Rica in the quarterfinals of the 2014 World Cup. Krul saved two kicks and the Netherlands advanced. More recently, at the 2024 European Championship, English goalkeeper Jordan Pickford had a list taped to his waterbottle with instructions for each potential Swiss taker.[1] The instructions for the first taker, Manual Akanji, said to dive left. He did and saved the kick.

Research has investigated penalties from a variety of different perspectives. Psychological studies seek to understand the mental aspects of penalties, particularly related to stress. They aim to develop strategies to prepare players to cope with pressure [9, 119, 221]. Performance analysis studies have looked at how to support the kicker by investigating where a kick should be placed to optimize the chance of converting [108, 5, 14]. These consider aspects such as the kicking technique and speed [111]. Finally, researchers have studied the choices that players make from a game-theoretic perspective [160, 210]. Interestingly, in aggregate, players closely follow a mixed-Nash equilibrium strategy. However, partitioning the takers based on playing styles leads to different recommendations [210].

Unfortunately, the existing work has several important limitations. First, the game-theoretic approaches have greatly simplified the problem by assuming that both players *independently pre-select* their decision [160, 210]. This induces a major misalignment with observed strategies where some takers and goalkeepers *base their decision on each other's movements*. That is, some takers employ a *keeper-dependent* strategy in which they select where to place their kick by reacting to the goalkeeper's behavior [150] whereas some goalkeepers delay their dive in order to base its direction on the kicker's movements [151]. However, we will show that simply incorporating these aspects into a game-theoretic approach does not work because not all players have the same action capacities (i.e., what physical actions a player is able to perform) [74]. That is, only some players are capable of basing their decisions on an opponent's movements and the small sample sizes means that one may have limited insights into an opponent's capacities. Second, little work analyzes the tactical choices available to the goalkeeper such as providing advice about their initial positioning (e.g.,

[1] It is unclear how this list was created and what role data played. For an image, see: `https://news.sky.com/story/england-keeper-jordan-pickfords-not-so-secret-weapon-during-euro-2024-penalty-shootout-victory-over-switzerland-13174555`

should they stand in the center of the goal or shade themselves slightly to one side).

We overcome these weaknesses by proposing a simulation-based approach for analyzing the efficacy of different goalkeeper strategies. Our approach confers several important advantages. First, it considers a rich set of tactical choices that others do not, including whether or not the goalkeeper bases their dive on the taker's movements, and the goalkeeper's initial position. Second, the approach can be configured based on the specific action capacities of the goalkeeper, which can be measured in a testing or training environment. Third, it addresses sample-size issues about takers by learning player-agnostic models. Finally, our study is grounded in a large dataset of 7,872 penalties where penalty experts annotated, e.g., the strategies employed by both takers and goalkeepers. Consequently, the framework allows for a coaching staff to provide tailored advice based on a specific goalkeeper's capabilities. Moreover, it can support making informed decisions (e.g., whether to substitute in a goalkeeper for a penalty shootout). Our experiments show that the best policy for a specific goalkeeper depends on their action capacities and aligning slightly off center increases the chance of saving a kick.

8.2 Preliminaries

Penalty kicks in soccer are taken 11 meters (12 yards) from the center of the goal and only involve the kicker and the goalkeeper. The rules stipulate that (1) **the goalkeeper** must stand with one foot on or behind the goal line until the ball has been kicked; and (2) **the kicker** must not feint or hesitate while running up to kick the ball [84].

When certain games in knockout competitions end in a tie, a penalty shootout determines the winner. Penalty shootouts start with a coin flip to decide which team kicks first and on which goal the kicks will be taken. The teams alternate taking penalties. Initially, a shootout involves 10 penalties and the teams the converts the most penalties wins. If the score is tied after 10 penalties, the shootout enters sudden death: in each round a player from both teams takes a kick. This continues until there is a round in which one team scores and the other does not. A player can only take a second penalty kick after all their teammates have taken a kick. Teams typically determine the order of takers prior to the shootout. Often, experienced penalty takers attempt the first kicks and less-experienced players take later kicks (see Appendix A.1.4).

8.2.1 Penalties from a Takers' Perspective

Since players tend to kick the ball with the inside of their foot, the kicking motion makes it more *natural* and thus more common for right-footed (left-footed) players to kick toward the left (right) [160]. Thus, instead of referring to the left and right corners of the goal, the convention is to refer to the natural and non-natural corners, based on the kicker's dominant foot.

When taking a penalty, kickers tend to employ one of two strategies:

Keeper-independent kickers predetermine the location they will aim the kick toward. Thus, the direction is selected independent of the goalkeeper's movements.

Keeper-dependent kickers watch the goalkeeper's movements and select the kick direction based on those movements [126].

Most players follow the independent strategy. The dependent approach is harder and can degrade performance because the kicker has a limited amount of time to modify their kicking motion [212]. Moreover, its success is also tied to the goalkeepers' temporal decision-making [151].

8.2.2 Penalties from a Goalkeepers' Perspective

The goalkeeper's primary decisions revolve around where and when to dive. The short distance between the ball and the goal makes it challenging for a goalkeeper to react to the kick [68, 191] as ball flight times vary between 500ms and 600ms [69, 227]. Beyond this, a goalkeeper's dive range (i.e., how far they can dive in a set amount of time) depends on their physical characteristics. This range varies among goalkeepers.

We distinguish between two different strategies a goalkeeper can employ:

Early dive goalkeepers guess a direction and start moving toward one side of the goal shortly before the shooter kicks the ball.

Late dive goalkeepers base their dive direction on the kickers movement by delaying their movement to collect more and more accurate information from the kicker to increase their chances of diving to the correct location [110].

There is a trade-off between these two strategies. The early dive strategy affords the goalkeeper more time to move, meaning that they have a larger dive range

(i.e., more likely to reach the corners) but have less information about the kicker's intentions. A late dive allows the goalkeeper to glean information from the kicker's run up that might help them dive toward the correct corner, but they will have a shorter dive range due to having less time. The footnote links to illustrative videos showing this trade-off.[2]

8.2.3 Data

We had access to two data sources that provided disjoint information. First, a penalty expert annotated kicks of players likely to take part in the 2022 FIFA Men's World Cup. Most importantly, this data contains information about the kicker's strategy (keeper-independent or keeper-dependent). It also recorded the goalkeeper's dive strategy for about half of the kicks. Second, we had event data as collected by commercial data providers about most matches. This does not record the kicker's or goalkeeper's strategy. However, it does record three important things not present in the first dataset: (1) the end coordinates (x,y,z) of the penalty kick, either where it was stopped by the goalkeeper or where it passes the goal line, (2) information about what every player did during the game (e.g., position played, actions executed), and (3) the order of penalty takers in a shootout, which cannot be derived from the first dataset since it only includes kicks involving a player likely to participate in the 2022 World Cup.

Because they contain unique information needed for analyzing penalties, we merge the datasets. However, this requires matching penalties between the two datasets. This is not straightforward as they were collected in different ways by different organizations. Hence, relevant information such as team names (e.g., Athletic Bilbao versus Athletic Club), player names (e.g. Sami Al Naji versus Halil Al Naji Sami), and times (e.g., timezone used, representation of the date) are recorded differently. Moreover, the datasets do not contain the same set of matches. Appendix A.1.3 describes our procedure to merge the data.

After merging, we had 7,872 kicks, including 1,201 taken in a shootout. These penalty kicks involve 722 unique penalty takers and 161 distinct goalkeepers. To highlight the sample size issues, only 59 goalkeepers and 28 takers were involved in at least 25 kicks. The average conversion rate of penalty kicks in our dataset is 77.9%. Players place the kick on target 93.5% of the time (i.e., 6.5% of the kicks are wide or high). In terms of taker's strategy, 20.6% kicks were taken

[2] An early dive `https://youtu.be/OjAUTduO_Fk?t=70` and a late dive `https://youtu.be/OjAUTduO_Fk?t=142` vs. a keeper-dependent kick. When the goalkeeper dives early, the kicker chooses the other corner. On the late dive, the goalkeeper picks the correct corner, but fails to reach the well-placed kick.

Table 8.1: The empirical payoff matrix for the Penalty Kick on our dataset.

	GK N	**GK Late**	**GK NN**
Kick N	0.615	0.785	0.939
Kick C	0.846	0.273	0.865
Kick NN	0.947	0.785	0.556
Kick Dep	0.846	0.773	0.846

keeper-dependently and 79.4% were taken keeper-independently. We define a goalkeeper's save percentage as the number of kicks saved over the number of kicks on target. On average goalkeepers save 16.8% of the kicks on target. The goalkeeper's dive timing was initially not recorded and is only available for 3,997 kicks. In these kicks, goalkeepers use a late dive on 38.5% of penalties and an early dive 61.5% of the time.

8.3 Penalty Game Theory with Player Tactics

To motivate our approach, we will start with the typical game-theoretic approach that views a penalty kick as a zero-sum game between the kicker and the goalkeeper [160, 210]. Whereas prior work simplifies the setting by assuming that all takers use the independent strategy and all goalkeepers dive early, we will consider a richer setting where each player has a novel choice: the goalkeeper can decide to dive late and the kicker can take the kick keeper-dependently. We will see that the game-theoretic approach is less well suited to modeling these strategies due to differences in players' action capacities.

In our augmented setting, the kicker has four choices: kick to natural corner (N), non-natural corner (NN), center (C) and kick dependently (Dep). The goalkeeper has three choices: diving early toward the kicker's natural corner, diving toward the non-natural corner or diving late.[3] The players' choices should be made simultaneously.

Table 8.1 shows the payoff matrix for the Penalty Kick game as estimated empirically from our dataset. We used the Minimax theorem [148] to find the optimal mixed strategies for both players:

- **Goalkeepers: (N, Late, NN)** = (6.9%, 87.1%, 6.0%)

[3]An early dive to the center is not considered a choice because this is indistinguishable from staying in the center on a late dive when the kicker does not aim for a corner.

- **Kickers: (N, C, NN, Dep)** = (43.1%, 0.0%, 35.7%, 21.1%)

Goalkeepers should dive late 87.1% of the time and dive early 12.9% of the time (6.9% to the natural corner and 6.0% to the non-natural corner). Kickers should employ the keeper-dependent strategy 21.1% of the time and the independent strategy 78.9% of the time (natural corner 43.1%, center 0%, and non-natural corner 35.7%). Note that one should not perform a keeper-independent kick to the center: if a taker wants to aim in this zone, then they should use the dependent strategy to ensure that the goalkeeper dives towards a corner (i.e., vacates the central zone) prior to aiming there.

In contrast to what has been observed in the simplified setting [160, 210], we see a stronger divergence from the optimal mixed strategy with players empirically following:

- **Goalkeepers: (N, Late, NN)** = (35.9%, 38.5%, 25.6%)
- **Kickers: (N, C, NN, Dep)** = (39.5%, 11.5%, 28.4%, 20.6%)

In particular, we see two substantial deviations as (1) goalkeepers dive late much less often in reality than the optimal strategy suggests they should (38.5% vs. 87.1%), and (2) takers aim a keeper-independent kick to the center 11.5% of the time, whereas the optimal strategy is never to do this.

The mismatch between observed behavior and optimal behavior likely arises due to the differing action capacities of the players. Not all goalkeepers have the action capacity to execute the late dive strategy [110] and even if they can, their range on a late dive will vary [227]. Similarly, not all takers have the action capacity to take goalkeeper-dependent kicks [212]. Extending the analysis is challenging, because there is substantial uncertainty about players' action capacities due to the small sample sizes and players' evolving skill sets. Moreover, it is possible that the empirical conversion rates are biased because, e.g., we have more observations about players with these capacities. Therefore, these values may not generalize, which impacts situations such as shootouts that are likely to involve less-experienced takers. Consequently, a more nuanced approach is needed that can consider both action capacities and contextual factors in a more player-agnostic way.

8.4 Modeling Goalkeeper Policies for Penalty Kicks

We propose a simulation framework that makes use of learned models and allows us to better account for the variability of goalkeepers' action capacities.

A goalkeeper's strategy involves three major components. First, the goalkeeper can decide where to lineup with respect to the goal line. For example, they may lineup in the center or may want to shade themselves slightly toward one corner. Second, they can decide to dive early or late. Third, they need to decide which direction to dive, which depends on the previous point. We refer to a setting for each of these choices as a policy π. Formally, we address the following task:

Given: A goalkeeper's capacities, a policy π, and historical penalties

Do: Compute what percentage of on-target penalties the goalkeeper would stop when following π.

We address this task by estimating the probability of saving penalty kick k as:

$$p(s|k, \pi, gk) = p(s|k, c, \pi, gk) \times p(c|k, \pi, gk) \tag{8.1}$$

where $p(c|k, \pi, gk)$ is the probability to choose the correct corner for penalty k when goalkeeper gk uses policy π, and $p(s|k, c, \pi, gk)$ is the probability to stop penalty k given that you dive to the correct corner c using policy π for goalkeeper gk. Instead of trying to simulate an end location of a penalty, we will evaluate policies on the actual end location of historical penalties (i.e., the x,y,z coordinate), though care will be taken when a kicker employs the keeper-dependent strategy.[4] Next, we describe how to design informed policies and compute these two probabilities.

8.4.1 Specifying Educated Policies with Learned Models

When specifying dive timing and dive direction policies, a natural question is whether it is possible to use learned models to make a more informed decision. Such models could be used in two important ways within a policy. First, having a better indication of what zone the ball will be directed toward can allow a goalkeeper to make an educated guess when executing an early dive. Second, knowing how far away from the goalkeeper's position the ball will be placed can help a goalkeeper decide whether to dive early or late: if a kick is likely to be out of a late goalkeeper's reach, then they should dive early.

These are challenging problems to solve due to small sample sizes of historical kicks for a given taker and the fact that game context features can affect these choices. Therefore, we use generic information about the takers and the game context. Moreover, to remain flexible, these models do not depend on knowing

[4]Given that keeper-dependent kickers choose their direction based on the goalkeeper's movements, we cannot use the actual x,y,z coordinate to evaluate goalkeeper policies.

the goalkeepers diving range, which is typically only known to their team. Specifically, we describe each penalty using 47 features that can be broken down into five categories: contextual features (e.g., time in the game, goal difference), player characteristics (e.g., preferred foot, position), penalty taking experience (e.g., number of taken penalties), penalty taker's history (e.g., percentage of previous kicks to natural and non-natural corners, average distance that ball is placed from the center of the goal), and shootout features (e.g., result and direction of the last penalty kick taken by the opponent and by the kicker's own team). These are described in more detail in Appendix A.2.

We train two models:

Penalty Direction Model predicts whether the kick will be aimed toward the (1) center, (2) the kicker's natural corner or (3) the kicker's non-natural corner. Specifically, we train an XGBoost [51] multi-class probablistic classifier on the keeper-independent kicks in our dataset.

Penalty Distance Model predicts the distance of the end location of the kick to the center of the goal. Specifically, we train an XGBoost regressor.

We consider four different policies for dive timing and direction. In the **late** policy, the goalkeeper always dives late. In the **early** policy, the goalkeeper always dives early. The direction is selected according to the observed probability of diving to each zone (i.e., this policy does not use a learned model). In the **early educated** policy, the goalkeeper always dives early but dives toward each zone with a weighted probability as predicted by the learned Penalty Direction Model. Finally, in the **mixed educated** policy the goalkeeper bases the decision about whether to dive early or late on the learned Penalty Distance Model. If the predicted distance is within the goalkeeper's late dive range, then they will dive late. Otherwise, they use the early educated policy.

8.4.2 Computing $p(c|k, \pi, gk)$

The probability to choose the correct corner depends on both the goalkeeper's dive type and the taker's strategy. When a goalkeeper's policy involves an early dive against a keeper-dependent taker, the kicker will choose an opposing corner to aim for. Hence, the goalkeeper is unlikely to guess the correct corner.[5] Therefore, we need to learn a separate probability for a combination of the four considered goalkeeper policies and the two taker strategies.

[5]The taker could mishit the ball, leading to a correct guess.

For the **late** and **early** policies, we estimate the probabilities of guessing correctly empirically (i.e., by counting) from the merged dataset. For the **educated** policies, the direction of the early dive is determined by the Penalty Direction Model. Specifically, we sample a direction according to the distribution of the zones that are predicted by the model. Therefore, when facing an independent taker, the probability of guessing the correct corner when using an educated early dive is the probability predicted by the Penalty Direction Model for the kick's true location (i.e., observed end coordinates in the data). Dependent kickers base their penalty direction on the goalkeeper's movement. Therefore, for dependent kicks, we cannot use the actual direction of the kick. Hence, we use the emperical value of choosing the correct corner when facing a dependent kick. This is computed over all dependent kicks in the dataset.

8.4.3 Computing $p(s|k, c, \pi, gk)$

To stop a shot, a goalkeeper must be able to *reach* the point where the ball crosses the goal line *prior* to it fully crossing the line. What locations a goalkeeper can reach depend on a number of factors such as when they start moving, how fast they move, and the speed of the ball being kicked. We will focus on the first factor, as when they start moving depends on their timing strategy. We use $r(\pi, gk)$ to denote the goalkeepers dive range for the dive type (early or late) selected by policy π for a specific kick.

Because we use historical kicks with known end coordinates, we could naively assume that $p(s|k, c, \pi, gk) = 1$ if the goalkeeper guesses the correct corner and $d(loc_s, k_e) < r(\pi, gk)$ and 0 otherwise. Here, $d(loc_s, k_e)$ is the Euclidean distance between the goalkeepers starting location loc_s[6] and the location k_e where the ball crosses the goal plane.[7] However, two issues arise: a) a goalkeeper will not stop all penalty kicks within reach because they may make an error in their technique or not be able to deflect the ball out of the goal, and b) the end coordinates are annotated manually which will make them less accurate. Therefore, we introduce some uncertainty and compute the probability that goalkeeper gk saves penalty kick k when they dive toward the correct corner as

$$p(s|k, c, \pi, gk) = \begin{cases} 0, & d(loc_s, k_e)(\pi) > r(\pi, gk) + \mu \\ \rho, & d(loc_s, k_e)(\pi) < r(\pi, gk) - \mu \\ \rho(\frac{1}{2} - \frac{d(loc_s, k_e)(\pi) - r(\pi, gk)}{2\mu}), & \text{otherwise} \end{cases} \quad (8.2)$$

[6] Taken as the point on the goalline that is equidistant between their two feet
[7] Both locations are three dimensional

If the kick ends within μ meters of the keeper's reach, the probability of stopping the ball linearly increases as ball's end location nears the goalkeeper initial position. ρ represents the probability that a goalkeeper saves a penalty that they reach (e.g., they may fail due to a technique error). Figure A.2 in Appendix A.3 shows the influence of ρ and μ on the save probability.

8.5 Evaluation

The evaluation's goal is to (1) explore the efficacy of each policy for different diving ranges given the standard center of the goal initial alignment, (2) investigate the effect of the goalkeeper's initial position, (3) evaluate the quality of the individual learned models, and (4) show illustrative practical use cases.

For the first two questions, we simulate each penalty in our dataset and compute the expected save percentage using Equation 1. We set $\mu = 0.7m$ and $\rho = 0.7$; these values were derived empirically as described in Appendix A.3. We compare the four policies defined in Section 4.1: the **late**, **early**, **early educated** and **mixed educated** policy. Because we are unaware of other approaches that evaluate goalkeeper policies for soccer penalties, we compare our approach to the policy of following the game-theoretic optimum as derived in Section 8.3. We also report the empirical save percentage.

8.5.1 Evaluating Goalkeeper Policies

We evaluate the expected save percentage for each policy as a function of a goalkeeper's dive ranges. Based on estimates from experts, we vary the late range from $2.6m$ to $2.9m$ and the early range from $3.0m$ to $3.2m$ in increments of $0.1m$. Figure 8.1 shows four plots, each showing how the expected save percentage varies as function of the early dive range for a fixed late dive range.

The best policy depends on a goalkeeper's capacities. If the goalkeeper is capable of diving late, then the **mixed educated** policy is always the optimal policy. However, if the goalkeeper does not have the capacity to dive late, the **early educated** policy outperforms the **early** policy. When a goalkeeper has a lower late dive range, the game-theoretic optimum is outperformed by the **early educated** policy. This again illustrates the importance of incorporating action capacities into a framework in order to provide meaningful advice. Finally, the game-theoretic optimum closely tracks the late dive policy, as it states that the goalkeeper should dive late 87.1% of the time.

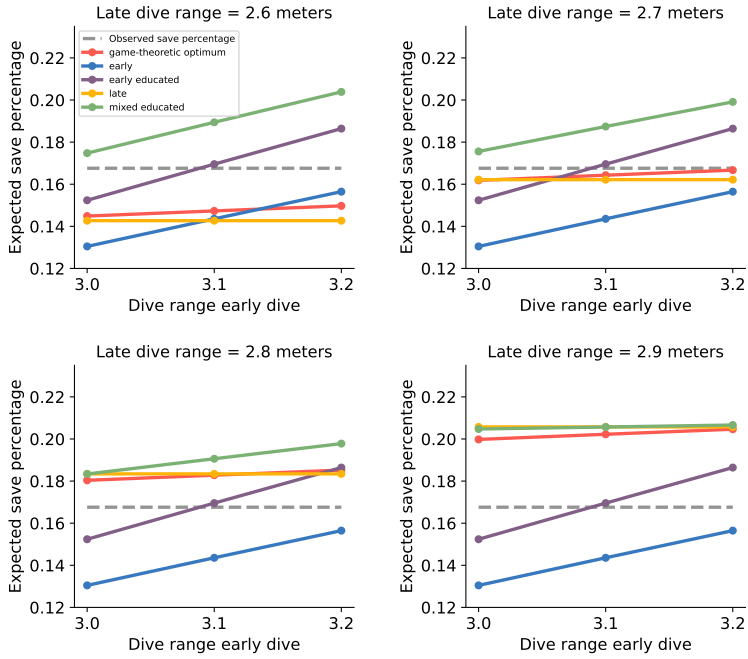


Figure 8.1: Each plot shows how the expected save percentage varies as function of the early dive range for a fixed late dive range.

8.5.2 Initial Goalkeeper Position

Over time, the number of keeper-independent kicks placed close to the corners of the goal has increased: In 2021, 10.8% of 1,140 such kicks were placed within 25cm of a post compared to only 5.0% of such 524 kicks in 2016. Such kicks are simply out of range (i.e., impossible to save) for any realistic early dive range *if* the goalkeeper's starting position is in the middle of the goal. Hence, we investigate a policy where the goalkeeper's initial alignment is slightly closer to one corner. Research has indicated that player's are able to shoot more forcefully (i.e., higher ball velocity) when aiming toward their natural corner [130]. Moreover, the natural corner is the most common placement location [150]. Therefore, we evaluate the efficacy of the goalkeeper lining up $0.1m$, $0.2m$, and $0.3m$ closer to the natural corner than the non-natural one.

Figure 8.2 shows the estimated save percentage for a goalkeeper with an early dive range of $3.1m$ and a late dive range of $2.7m$ for each of our four policies as a function of their initial position. The results indicate that shading toward the

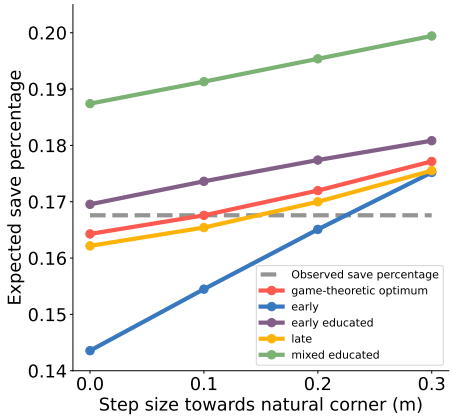


Figure 8.2: Expected save percentage for on-target penalties as function of moving x meters toward the kicker's natural corner for all goalkeeper policies.

natural corner can potentially increase the expected save percentage. However, players may change their penalty direction based on the positioning of the goalkeeper. Research shows that this might especially be the case when the goalkeeper is (clearly) not positioned in the center of the goal before the player's run up to the ball [136, 219, 152]. Therefore, it may be advisable to alter one's alignment during the kicker's run toward the ball. As Figure 8.2 shows, this policy could increase the chance of saving a penalty by between 0.5 and 3 percentage points depending on which policy is used and how much one moves towards the kicker's natural corner.

8.5.3 Evaluating the Learned Models

We use nested 5-fold cross-validation to train and evaluate our direction predictor and distance estimator models. We ensure that all penalties taken by the same player are in the same fold. We tune the max tree depth, learning rate, number of estimators and the minimum child weight (see Appendix A.4.1).

Predicting the penalty direction. We compare our learned classifier to the base model of just predicting the empirical distribution of aiming to each of the three considered zones. As we are interested in how well calibrated the probability estimates are, we report the logloss. On all penalties, our model obtains a logloss of 0.954 ± 0.011 vs. 0.969 ± 0.014 for the base model. When only considering shootout penalties, our model obtains a logloss of 0.941 ± 0.032

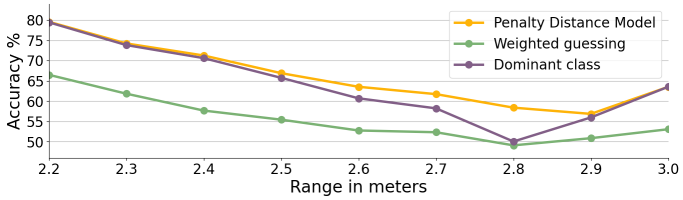


Figure 8.3: Accuracy per range threshold for our Penalty Distance Model and baselines. Most shots are placed within 2.5 and 3 meters from the center, and this is also the area in which our model slightly outperforms the baselines.

whereas the base model's is 0.988 ± 0.034. The model is well calibrated; see Figure A.4 in Appendix A.4.2.

Predicting penalty distance. We will use this model by thresholding its predicted value based on a goalkeeper's maximum range such that they can decide whether to dive early or late. Therefore, we vary this range and compute each model's accuracy. A prediction is correct if both the predicted and actual distance fall on the same side of the threshold. For each threshold, we compare our model's performance to two baselines: 1) the training set's average distance to the goalkeeper, and 2) randomly guessing whether the shot is within reach based on the training set's base rate for the specific threshold. Figure 8.3 shows how the accuracy varies as a function of ranges. Our model slightly outperforms the baselines in the ranges between 2.5 to 2.9 meters from the center of the goal.

Appendix A.4.3 provides insights in what our models learned from the data and which feature combinations impact the predictions.

8.5.4 Illustrative Use Cases

We show three ways in which our proposed methodology could aid practitioners. In the first two use cases, we compare three goalkeepers with different action capacities. **GK1** has an average late diving reach ($2.8m$), average early diving reach ($3.1m$), and an average correct corner guess percentage when diving late ($p = 0.59$). **GK2** has an above average early dive ($3.1m$) range, but is not able to dive late. Finally, **GK3** has lower diving ranges ($2.7m$ late range and $3.0m$ early range), but has a strong skill in choosing the correct corner based on the kicker's movement when diving late ($p = 0.7$). The third use case discusses how our model could be used to provide player-specific advice to a goalkeeper.

Table 8.2: The expected save percentage for U1 (left) and UC2 (right) for goalkeepers GK1, GK2 and GK3 following the **early educated**, **late** and **mixed** strategies.

	Use Case 1: In-game on-target kicks			**Use Case 2: Shootout on-target kicks**		
	early educated	late	mixed	early educated	late	mixed
GK1	0.170	0.186	**0.193**	0.169	0.170	**0.177**
GK2	**0.186**	-	-	**0.187**	-	-
GK3	0.153	**0.194**	0.191	0.150	0.175	**0.178**

UC1: How do action capacities affect a goalkeeper's optimal policy? For each of these goalkeepers we want to find their optimal policy for in-game penalty kicks (so excluding shootout kicks) given their capacities. Table 8.2 shows the expected save percentage for on target in-game penalty kicks for each goalkeeper for the three policies: **early educated**, **late** and **mixed**. In this use case we do not employ the tactic of taking a step to the side. We observe that each goalkeeper has a different optimal policy which is due to the differences in their action capacities. GK2 should employ the **early educated** policy, GK3 should employ the **late** policy and GK1 should employ the **mixed** strategy. When these goalkeepers employ their best strategy, GK3 (19.4%) and GK1 (19.3%) are expected to save more penalties than GK2 (18.6%).

UC2: Which goalkeeper is best for penalty shootouts? When participating in a tournament, a team needs to select three goalkeepers. Ideally, one of them is a strong penalty stopper that could be used in a shootout. This analysis only considers shootout penalty kicks. Table 8.2 shows the results. Interestingly, for shootout penalty kicks, GK2 now has the highest expected save percentage amongst the three goalkeepers. Following the **early educated** policy, GK2 is expected to stop 18.4% of the kicks aimed on target. Both for GK1 and GK3 the optimal policy is now to follow the **mixed educated** policy which will let them save 17.4% (GK1) and 17.5% (GK2) of the penalties.

GK2 scores best in shootout kicks which might be related to the fact that in shootouts players are less likely to take their kicks keeper-dependently (21.2% in-game vs 15.1% in shootouts). In shootouts, there are fewer keeper-dependent kicks. Hence, the fact that GK2 cannot dive late is less important than their higher early dive range.

UC3: How can the framework be used in a game? One potential real-life scenario is to provide a list of taker-specific penalty advice to a goalkeeper in a specific game situation, similar to the list Jordan Pickford had at Euro 2024 that was mentioned in the introduction. Providing such advice requires knowing a goalkeeper's traits. First, it requires the early and late dive ranges, and the correct corner guess percentage when diving late. These values are typically collected by teams themselves from experiments in training and from game data if the goalkeeper has faced a sufficiently large number of penalties. Next, when facing a particular penalty taker at a specific moment of the match[8], the framework will provide the probabilities of saving the kick for each of the policies. Sampling (rolling a weighted dice) from these probabilities will then determine what policy to use. This yields instructions that are tailored to the goalkeeper's traits. Research [113] has shown that there is no significant difference in diving toward the right versus left side for elite goalkeepers. Furthermore, following such an instruction will remove the pressure from the goalkeeper as it will not be the goalkeeper's fault if the wrong choice was made. This will avoid the well-known action bias when goalkeepers face penalty kicks [15].

8.6 Related Work

Multiple game theoretic analyses have modeled an individual penalty kick as a 2-player zero sum game where each player has two actions: shooting or diving to the natural or non-natural side [12, 53, 160, 210]. Palacios-Huerta [160] analyzed 1,417 penalty kicks taken between 1995-2000 and showed that kickers and goalkeepers follow a mixed-strategy Nash equilibrium. Tuyls et al. [210] confirmed this on dataset containing 12,399 penalty kicks taken between 2011 and 2017. Moreover, they also performed a more fine-grained analysis where players are clustered by playing style. Their findings indicate that Nash-derived recommendations depend on players' style of play.

Sports science and performance analysis studies have investigated placement from the perspective of a kicker. Here, multiple studies analyzed between 300-1000 penalty kicks from top-level games and concluded that kicks should be directed high, ideally toward the top corners [108, 5, 14]. Still, there are many low kicks in penalty shootouts. Hunter et al. [111] analyzed a small number of kicks taken by amateur and semipro players in a controlled setting. They used a hand-crafted model to analyze a shooter's choice of aim, kicking technique, and speed when taking a keeper-independent kick. It is unclear how well this

[8]The models include game state features; these could also be omitted to pre-compute probabilities.

would translate to game environments and professional players. Finally, several approaches have tried predicting the direction of a penalty kick [42, 31, 196].

8.7 Conclusion

We analyzed a large dataset of nearly 8,000 penalties that were annotated by penalty experts. To be able to take into account action capacities, we proposed a novel simulation framework to evaluate the efficacy of different policies that goalkeepers could employ. These policies exploited novel, player-agnostic learned models to make more informed decisions. Our results indicate that following the advice provided by the models would result in saving more penalties than using more generic information. We showed how the best policy depends on the goalkeeper's capacities, that aligning slightly off center increases the expected save percentage, and that diving late is beneficial if you expect the kick to be within reach. We also discussed several practical use cases for our framework.

Chapter 9

Conclusion

This chapter summarizes the main contributions of this dissertation. It also provides possible directions for future work.

9.1 Summary

The introduction of this thesis identified three gaps in the literature. This section summarizes the progress made along each avenue in this dissertation.

9.1.1 Understanding Women's Soccer Using Data

There is a lack of data-driven analysis on women's soccer, with most existing work that uses advanced computational techniques exploring data from men's football. For example, prominent metrics such as xG were originally developed for the men's game and it is unclear if these are directly applicable to women's football. At the same time, the women's game is evolving rapidly, yet these developments have not been studied at larger scale. These developments make cross-league comparisons more challenging than in the men's game, where the relative strength of the top leagues has long been established. To address these gaps, we applied data analytics and AI techniques and made the following three contributions.

Chapter 3 explored developing an xG model tailored specifically to women's football. We found that while expected goals models developed using data from

men's (women's) matches are generally applicable to shots from a women's (men's) match, there are notable differences in how these models value certain types of shots. For example, shots from the corner of the penalty box and headers following a set piece are assigned lower values by the model trained on women's data compared to the men's model. On the other hand, tap-ins following a cross are assigned higher values by the women's model. Moreover, we observed that women tend to take shots with higher xG-values than men.

Chapter 4 showed that the women's game has developed significantly in both tactical behaviors and technical skills. In this chapter, we analyzed an extensive longitudinal dataset containing match event data for 3,510 matches played over eight years in the top women's leagues. By leveraging AI models such as xT and xG, we investigated how the game has evolved in recent years. Tactically, we observed longer possession sequences and fewer long-distance shots. Technically, we observed that players are completing more difficult and valuable passes at a higher rate, and putting a higher percentage of shots on target, with a particular emphasis on aiming for the corners of the goal. These findings could aid football practitioners to account for the developments of the game.

Chapter 5 introduced a novel Pace of Play (PoP) metric that measures the amount of time a player takes to pass the ball after gaining control, either by receiving a pass or intercepting the ball. We assessed the utility of the PoP metric by analyzing five seasons of match event data about six top women's leagues. We made the following three findings. First, our PoP metric is positively correlated with team strength as measured by Elo ratings. Second, our PoP metric is correlated with the amount of playing time following interleague transfers: players that exhibit a faster pace appear in a greater percentage of games in a new league. Third, as a measure of face validity, the PoP metric identified players who subsequently exhibited strong (e.g., prize winning) performance. Although this analysis focused on women's soccer, the metric could also be valuable in the player recruitment process in men's soccer.

9.1.2 Capturing More Context in Data-Driven Recruitment

Player performance metrics such as xG, xT and VAEP ignore the broader context in which they appear. Player performance also relies on external factors such as the player's teammates and the mental pressure in a game. Chapters 6 and 7 showed how such context could be captured by using AI techniques for data-driven recruitment. As I am one of the developers of the VAEP metric [62], both chapters built on this metric.

Chapter 6 introduced a framework for evaluating player performance under mental pressure. This framework is based on two components: measuring mental

pressure and measuring player performance. A novel mental pressure model was developed using two machine learning models to measure both the pre-game and in-game mental pressure levels. Additionally, two novel performance metrics were introduced. An action's Decision Rating values the quality of the decision by evaluating the chosen action compared to the most likely action. An action's Execution Rating values the execution by using a machine learning model to predict the likelihood of success for the chosen action. By integrating the mental pressure levels with these performance metrics, we showed that our framework could support clubs in their recruitment process, coaching individual players, and lineup selection and substitutions under high mental pressure scenarios.

Chapter 7 introduced two chemistry metrics that measure the offensive and defensive chemistry between a pair of players using an extension of the VAEP metric [62]. ML techniques were employed to value game states in a manner similar to the VAEP metric. These game state values were then used to value offensive and defensive chemistry between player pairs. Using historical match event data, these metrics could be computed for any pair of players who played enough minutes together. For any unseen pair of players, the chemistry was predicted using two machine-learned models. These ML models took into account aspects such as the players' position line, nationality, mother tongue, playing style and physical performance indicators. Moreover, Chapter 7 introduced a Team Builder using combinatorial optimization techniques that automatically selects the best team of eleven players from a given set of players, maximizing the mutual chemistry of all player pairs in the line-up.

9.1.3 Optimizing Goalkeeper Policies for Penalty Kicks

Chapter 8 introduced a novel framework to optimize goalkeeper policies for penalty kicks. We showed that current literature simplifies the penalty game, as it does not take into account the goalkeeper's and taker's action capacities.

First, we analyzed a large dataset of over 8,000 penalties that were annotated by penalty expert. These annotations included aspects strategies employed by both the kicker and goalkeeper. Second, we developed a player-agnostic simulation framework based on several models learned from the data for evaluating the efficacy of different goalkeeper tactics. Third, we showed how our framework can be used to optimize goalkeeper policies in real-world situations. More specifically, we showed that following the policy based on player-agnostic machine-learned models would result in stopping more penalty kicks. Furthermore, we showed that taking a small step to the side could increase the goalkeeper's chances of saving the kick.

9.2 Critical Reflections

I will briefly reflect on the thesis by outlining lessons learned and potential limitations.

Lessons Learned

I will highlight some general messages and insights that I acquired during my research which I believe may be useful for machine learners working with experts in other domains.

We do not always need the most complex solution As a data scientist it is often tempting to search for complex solutions to problems. However, end users are not always interested in maximizing predictive performance. While these less-complex solutions may score lower on performance metrics, they can be more valuable in practice due to a variety of factors such as their ease of use and interpretability. For example, the PoP metric introduced in Chapter 5 is a simple yet effective metric. We showed that it can aid soccer practitioners to assess player skills in the recruitment process. This work actually generated more engagement from soccer practitioners than the more sophisticated Execution and Decision Ratings from the mental pressure work in Chapter 6. This could be due to the intuitive nature of the PoP metric. It is easy to define and the importance of being able to move the ball quickly is a notion that makes sense to practitioners. This makes it easier to understand and integrate into existing workflows.

Presentation is key Related to the previous point, for metrics to be adopted by soccer practitioners, it is crucial that the end user understands both what the metric captures and what it does not capture. This requires clearly presenting the methods and results. I have learned that soccer practitioners, and end users in general, understand AI models better if you use their language to relate them to examples they are familiar with. In soccer, video clips are commonly used to illustrate relevant concepts. For instance, when explaining the Decision Ratings from Chapter 6, providing videos illustrating high- and low-valued actions from players the practitioner is familiar with can help explain the concept and build trust. More broadly, this relates to the concept of face validity [58], which has been useful in many of the works. For example, the fact that players with the highest PoP values included Ballon D'Or winners provides an intuitive validation, helping to build trust among soccer practitioners. This lesson is applicable in any domain where data science is used to solve problems.

Domain knowledge is essential When building data science solutions it is essential to incorporate domain knowledge from experts in the field. For instance, this was very useful in the penalty simulation framework introduced in Chapter 8. Through discussions with soccer experts, I learned about the different diving strategies that goalkeepers employ and that not all goalkeepers are capable of diving late. This information significantly changed how we designed our simulation framework. It made us change the framework such that it would be player-agnostic and could be run for goalkeepers with different capabilities. Domain knowledge also plays a role in feature engineering for the different models in this thesis. For instance, the Penalty Direction and Penalty Distance models from Chapter 8 use information from previous kicks, the player's preferred foot and their penalty experience which use domain knowledge. Similarly, the features designed for the Player Chemistry Predictor in Chapter 7 use features guided by domain knowledge, such as the Player Roles and physical indicators. These examples illustrate a broader point: regardless of the application domain, engaging with domain experts is essential when building data science solutions.

9.2.1 Limitations

I will describe some of the main limitations of this thesis.

Analysis on match event data Chapters 3 through 7 presented work based on traditional match event data as opposed to the more granular data that some vendors now provide. While we showed that such data can yield valuable insights for soccer practitioners, this type of data does have limitations. Most notably, it lacks information about the positions of all players not involved in the on-the-ball action. For example, for the xG models in Chapter 3 this means that we could not incorporate features about the position of the goalkeeper or other defenders. Similarly, the analyses in Chapters 4 and 6 focus solely on on-the-ball actions neglecting the defensive aspect of soccer. Chapter 5 focuses on on-the-ball pace of play and ignores contextual factors such as whether the player was under pressure when receiving the ball as this information was availability in our data. Chapter 7 does address defensive contribution. However, this approach was specifically introduced to overcome the limitations of the match event data. These limitations offer opportunities for future work, which will be discussed in the next section.

Data partitioning Across the different chapters we used different ways of splitting the datasets in training, validation and test sets. In all chapters we split chronologically to avoid using future information in model training.

However, this setup is not without limitations. The data is inherently non-stationary: team level, playing styles and player skills change over time [58]. Moreover, as highlighted in Chapter 4, the women's game in particular has evolved rapidly. For example, if the technical skills of female players generally increase over time, training performance models such as xG and VAEP on older data may lead to inaccurate values when applied to newer data. Newer actions, executed by players with better skills in general, will then likely come from a different distribution than the data the models were trained on. Similarly, if team playing styles evolve, certain actions that were common may occur less frequently and be more difficult to execute. This will impact all performance metrics introduced in Chapter 6: the Contribution, Execution and Decision Ratings. Therefore, future work could explore the impact of using chronological splits and how to best handle the non-stationarity of the data.

9.3 Future Work

This dissertation has shown how ideas from machine learning and artificial intelligence can help analyze soccer. Nonetheless, there remain many open challenges. This section describes possible directions for future work.

9.3.1 Player Recruitment

Chapter 7 introduced a Team Builder that selects the best lineup given the mutual chemistry between all player pairs in the lineup. The downside of our approach is that it does not scale. Ideally, one would be able to measure and predict the chemistry between a set of eleven players, instead of just two players. Given recent developments in the world of AI, this limitation could be addressed using graph-based learning [18]. One could represent the players (from both teams) as nodes and add node features such as the player's position, player roles, nationality, and age. The edges between players would indicate whether players belong to the same team. Next, the learning task could be to predict match outcomes given the lineups of both teams. This approach would allow the model to learn the value of a given lineup and estimate the impact of replacing individual players.

Furthermore, clubs typically maintain a squad of around 25 players that they use during a season. Extending the Team Builder to a Squad Builder would therefore be an interesting next step. Composing a strong and balanced squad is challenging. Maximizing the on-pitch performance of the squad requires considering various factors including player qualities, playing styles, player

development (i.e., younger players improving), tactical plans, injuries, and suspensions. AI techniques could assist in several areas, such as estimating player development of younger players, evaluating the effects of tactical plans, and predicting injuries and suspensions during the season. In addition to on-pitch performance, squad composition also involves financial considerations. In fact, clubs must contend with financial constraints that affect which players they can sign. Moreover, clubs typically aim to make revenue and thus need their squad to contain players with increasing transfer values. AI models could be developed to estimate transfer values and even to predict future transfer values based on expected player development. To conclude, AI could thus play an important role in composing the best squad [91, 162].

Decision Making Timing An important skill for soccer players is to make the right decision at the right time. Chapter 5 introduced the Pace of Play (PoP) metric that identifies players that quickly move the ball after receiving it. However, this metric ignores whether it was the best decision to quickly move the ball. For example, sometimes when possessing the ball, it is beneficial to lure the opponent towards you to create space for your teammates and create a better passing option than when passing it quickly. To identify players with such decision-making skills, one could combine an action's PoP metric with the decision ratings to train a model to find the optimal timing of the action. This would require tracking data to be able to observe the passing options available to the player.

9.3.2 Match Preparation

There is also room for future research in the area of match preparation. Most existing metrics focus on what happens with the ball, leaving opportunities to analyze what could have happened. Also, Chapter 8 introduced a novel framework to optimize goalkeeper policies for penalty kicks. While this was a first step in this direction, there still remains a lot to be explored regarding penalty kicks in soccer.

Unexposed Weaknesses Every team has its weaknesses and automated approaches for identifying them would help analysts charged with match preparation and opposition scouting. The work in this dissertation focuses on actions that *did* happen on the pitch. However, it ignores what *could* have happened. For example, an unexposed weakness might be a large gap between a team's defensive midfielder and its center backs during attacking phases. This space could be exploited during a counter-attack, as it has been

shown that compactness in defense is important [88]. However, if opposing teams exploit such a weakness, it might be hard to detect this weakness using current techniques in literature. Especially in the women's game, being able to automatically find such weaknesses would be very powerful. Due to the bigger differences in team quality and fewer international club matches, weaknesses might remain unexposed. For instance, FC Barcelona may have a potential weakness when their defensive midfielder loses a duel when in a low block. This might remain unexposed because they typically dominate domestic league matches and are thus not often playing in a low block, and their defensive midfielder is often stronger than most direct opponents in the league. However, if they play English-side Chelsea with physically strong midfielders in the Champions League, this weakness could become relevant to Chelsea.

Identifying such unexposed weaknesses using a what-if-framework could be an interesting direction of future work and would require the development of AI techniques. Such a framework should be able to simulate scenarios like:

- **Better decisions.** What if the midfielder uses space behind the defense to launch the center forward instead of passing backwards?
- **Better timing.** What if the player passed it 2 seconds earlier, or waited a bit longer to lure the opponent to create space?
- **Higher running speed.** What if the player is X m/s faster with/without the ball?
- **Better first touch.** What if the first touch is faced forwards instead of backwards?
- **Stronger in duels.** What if the player won the midfield duel? Is the team weak if they lose the duel or is the defensive structure then still good?

To be able to play out these scenarios, AI techniques should be developed that can faithfully capture the players and their interactions. These models should allow for modification of player skills and behaviors, enabling the simulation of different scenarios and the identification of potential weaknesses.

Optimizing Kicker Strategies for Penalty Kicks The work presented in this dissertation focuses on goalkeeper policies. However, a similar framework could be built from the viewpoint of the kicker. Given the goalkeeper's previous faced penalty kicks, one could optimize the kicker's strategy whether to kick keeper-independent or keeper-dependent, and if early, which direction is most likely

to yield a goal. This framework could then also incorporate the kicker's action capacities such as the keeper-dependent kick skills, and the kick's accuracy and speed when aiming towards the natural and non-natural corner.

Mental Pressure on Penalty Kicks In the work in Chapter 8 we largely ignored the mental aspects of penalty kicks which are clearly important. It could be very interesting to investigate from a Sport Science perspective what happens with kickers if they are told beforehand where to aim the kick. Would this remove some pressure as they are not to blame if the goalkeeper chooses the correct corner? Or would it add extra pressure as they might feel less comfortable kicking it towards that side?

Appendix A

Stop Guessing: Optimizing Goalkeeper Policies for Soccer Penalty Kicks

A.1 Data

We have access to two different datasets about penalty kicks taken during the course of the game and in shootouts: the World Cup Penalty Dataset and the Event Stream Dataset. Because they contain unique information needed for analyzing penalties, we merge the datasets. However, this requires matching penalties between the two datasets which is not straightforward as they were collected in different ways by different organizations. Hence, relevant information such as team names (e.g., Athletic Bilbao versus Athletic Club), player names (e.g. Sami Al Naji versus Halil Al Naji Sami), times (e.g., timezone used) and dates (e.g., swapping the order of month and day) are recorded differently in each dataset. Moreover, the datasets cover a disjoint set of matches. Next, we describe each dataset and the process used to align them.

A.1.1 World Cup Penalty Dataset

Starting from January 1st 2012, all penalties kicked or faced by the players likely to be included in a 2022 Men's World Cup squad were annotated. This includes penalties from friendly games, cups (e.g., Champions League, FA Cup),

domestic league games and international games. A trained penalty analyst annotated the penalties using a video tool and recorded for every annotated kick the person involved (either the goalkeeper or kicker), kick direction (left, center or right, and low vs high), dive direction (left, center or right), result (goal, save or wide), score and time, shootout identifier, foot used (left or right), kicker strategy (keeper-independent or keeper-dependent) and goalkeeper timing (early or late dive). The goalkeeper's dive timing was initially not recorded and is only available for about half of the kicks.

Soccer experts also provided a definition to annotate the pressure level of the kick. High pressure situations arise in (a) a shootout or (b) penalties taken after the 80th minute when your team is tied or trailing by a goal. Normal pressure situations arise when penalties are taken prior to the 80th minute when your team is tied or trailing by a goal. Anything else is low pressure.

A.1.2 Event Stream Dataset

This is the standard event data that records information about all on-the-ball events. Such data is collected by commercial providers about most (semi)professional matches. Besides containing more penalties, this dataset records different information than the World Cup dataset. It does not annotate the kicker's strategy (independent vs. dependent) or information about the goalkeeper's tactics (dive direction, early vs. late, distraction). However, it does record three important characteristics not present in the World Cup Data. Firstly, it records the end coordinates (x,y,z) of the penalty kick, either where it was stopped by the goalkeeper or where it passes the goal line (in goal or outside the posts). Secondly, it contains information about what every player did during the game (e.g., time on field, position played, actions executed). Finally, as all penalties in a game are annotated, it provides the order of penalty takers in a shootout, which is not necessarily contained in the World Cup dataset because a kick is only annotated if it involves a player likely to participate in the World Cup.

A.1.3 Merging the Data

We align the penalties contained in each dataset using the following three step matching procedure.

First, we map games. We start by mapping team names. For each pair of team names, we assign a similarity score as the maximum of the Ratfcliff-Obershelp

similarity[1] and the Levenshtein similarity score computed as $1 - \frac{\delta}{w}$ where δ is the Levenshtein distance between the two team names and w is the maximum number of words in both team names. Given one team name, we find its two most similar matches and automatically map them if the similarity score is higher than 80% for most similar one and lower than 80% for the second most similar one. This results in matching 89% of the teams. The remaining teams are mapped manually. Using these team mappings, we then map games based on the involved teams and the date. If there is no exact date match we allow 1 day difference or a swapped day and month. This results in 8,613 game mappings.

Secondly, we match the penalties within a game. We use the same approach as for team names, except we exploit the fact that we have mapped games to restrict the search for similar player names to those players participating in the same game. This matched all but nine players, who were mapped manually (e.g., a manually matched player was Jorginho, who was identified as Jorge Luiz Frello Filho in one dataset).

Finally, we map penalties within the same game. We look for penalties around the same time in the game but allow for slight differences (e.g., different time zones, extra time handled differently). One challenge is when there are multiple penalty kicks from the same player in a game, which are not differentiated in the World Cup data. In these cases, we match based on the result of the penalty, the direction of the kick and foot used. This enables mapping all but nine penalties. We checked video footage to map these.

The merged data contains 7,872 penalties.

A.1.4 Order of Takers

Figure A.1 shows the distribution on the number of prior penalties taken a player based on their position in the penalty shootout order. While there is variance, the average number of prior kicks drops around the fifth position. Moreover, there are few takers in these positions who have previously taken a large number of kicks.

[1]We use the implementation in the difflib Python library.

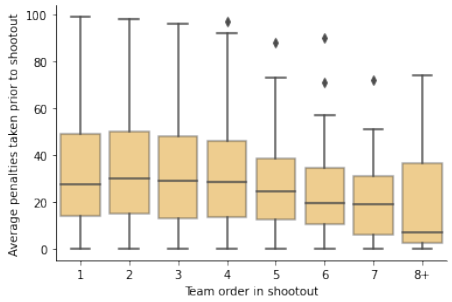


Figure A.1: The distribution of penalties (both shootout and non-shootout) taken by players prior to a shootout, per order in the shootout. We observe that the further we progress into the shootout, the less experienced players get.

A.2 Features for the Penalty Distance and Direction Models

These features can be categorized as follows:

Contextual features There are seven contextual features which record the time in match when the penalty was taken, a Boolean variable indicating shootout kicks, the goal difference (positive if kicker's team is leading), number of penalties taken in shootout, number of penalties taken by own team, and two Boolean variables indicating whether a penalty miss would lead to a loss or whether a scored penalty would lead to winning the shootout.

Penalty taker and goalkeeper features There are three general penalty taker features which record the player's preferred foot (left or right), position line (goalkeeper, defender, midfielder or striker) and age. Furthermore, we add a feature describing the height of the goalkeeper in centimeters.

Penalty taker penalty experience features There are four features describing the penalty taker's experience of taking penalties (under pressure). These are the total number of penalties taken, number of penalties scored, number of penalties taken under normal pressure and the number of penalties taken under high pressure.

Penalty taker preference features There are 18 features describing the penalty taker's potential preference for certain corners. These are the

percentage of penalties to the (non-)natural corner and the percentage of penalties to the (non-)natural corner scored. The anchoring effect bias is a cognitive bias in which people tend to put a lot of weight on certain information, often being the first anchor point. Therefore we include information on the player's first penalty taken recording whether the first penalty was a goal, was saved or missed the goal, and booleans describing whether the first penalty was aimed toward the center, natural corner or non-natural corner. Similarly, people's availability bias relates to the fact that decisions are influenced by information that is directly available to a person. In our case this could be the fact that player's might be biased by their most recent penalty's result. Therefore we add the same information (goal, save or miss, and corner) for the player's most recent penalty kick.

Penalty taker distance features There are 2 features describing the player's penalty kick placement in previous penalty kicks: 1) the average distance from the center of the goal and 2) the number of kicks placed within 50 centimeters of one of the posts.

Shootout features Other information that is readily available to a player are the penalties taken by their teammates and the opponents in the ongoing shootout. These might influence their decision in their kick due to the availability bias. For this, we include 12 Boolean features. These describe the result (goal, save or miss) and the direction (center, non-natural, natural (from kicker's perspective)) of the last penalty kick taken by the opponent and by the kicker's own team. When the kicker takes the first penalty in the shootout or for the team these features values are set to NaN.

Some features require a player to have taken a penalty before. If this is not the case the value for features related to previous penalties are all set to NaN.

A.3 Learning the Uncertainty Values for Equation 2

Our simulation framework computes the probability of stopping penalty k given policy π. This probability depends on the probability of diving toward the correct corner ($p(c|k, \pi, gk)$) and stopping the penalty given that you are in the correct corner ($p(s|k, c, \pi, gk)$). Equation 2 describes how we compute the probability that the goalkeeper chose the correct corner. This Equation depends on four values; 1) the Euclidean distance between the goalkeeper and the location where the ball crosses the line, 2) the goalkeeper's range following

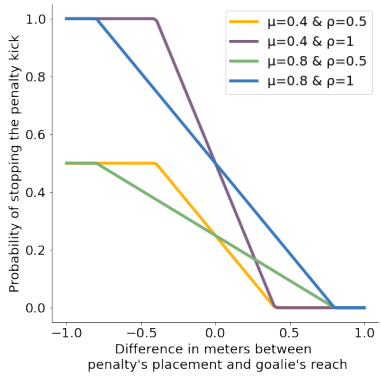


Figure A.2: The influence of within-reach save probability ρ and uncertainty factor μ on the save probability.

policy π, 3) the tolerance factor μ, and 4) the within-reach save probability ρ. In this section we show that this Equation makes sense and learn the uncertainty values from our dataset.

We evaluate this Equation by comparing our predicted probability using the Equation with the actual outcome for all penalty kicks on target for which the goalkeeper was in the correct corner. We can thus view this as a probabily estimation task and hence want our probabilities to be well calibrated. That is, if we expect the stop probability to be 0.8 for a specific group of penalties, we expect 80% of these penalty kicks to be stopped by the goalkeeper.

We assume that goalkeepers stand in the center of the goal and derive the location where the ball crosses the line from our dataset. Next, we learn the most likely values for the goalkeeper's ranges for early and late dives, μ and ρ using a grid search. For μ and π we considered values in the range between 0.5 and 1 with a step size of 0.1. For $r(\pi_e)$ and $r(\pi_l)$ we considered values in the range between 2.5 and 3.5 with a step size of 0.1. The objective is to minimize the Brier score of the computed values for $p(s|k, c, \pi, gk)$ and the actual results of the penalty kicks (saved or goal). Empirically, the values that minimize the Brier score are $r(\pi_e, gk) = 3.1$, $r(\pi_l, gk) = 2.8$, $\mu = 0.7$, $\rho = 0.7$. We will therefore use these values for μ and ρ in our experiments.

Figure A.3 shows the calibration plot for these settings. We observe that we quite accurately predict the save probability with Equation 2.

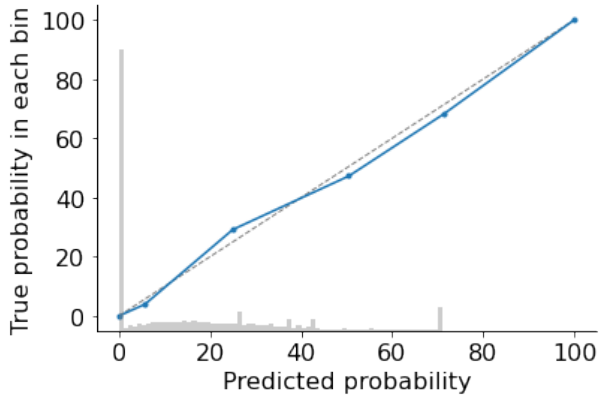


Figure A.3: Calibration plot for predicted probabilities of saving a penalty given that you are in the correct corner using $r(\pi_e, gk) = 3.1$, $r(\pi_l, gk) = 2.8$, $\mu = 0.7$, $\rho = 0.7$.

A.4 Model Learning

This section provides extra information on the model learning process for the Penalty Direction Model and the Penalty Distance Model. We also provide insights into what the models learned.

A.4.1 Hyperparameter Tuning

For both the Penalty Direction Model and the Penalty Distance Model, we tune the hyperparameters of the models using a grid search. We vary the learning rate in $\{0.01, 0.05, 0.1\}$, maximum tree depth in $\{3, 4, 5, 6\}$, and the number of trees in $\{50, 100, 250\}$. Tables A.1 and A.2 show the optimal hyperparameters per fold.

A.4.2 Calibration Plots for the Penalty Direction Model

Figure A.4 shows calibration plots for each of our classes for the Penalty Direction Model. These show that our classifier is well calibrated for all three classes.

Table A.1: Optimal hyperparameters per fold for the Penalty Direction Model

Fold	Learning rate	Maximum depth	Number of trees
1	0.1	5	50
2	0.1	3	50
3	0.1	4	50
4	0.05	4	100
5	0.05	4	100

Table A.2: Optimal hyperparameters per fold for the Penalty Distance Model

Fold	Learning rate	Maximum depth	Number of trees
1	0.1	4	100
2	0.1	4	100
3	0.1	4	100
4	0.1	4	100
5	0.1	4	100

A.4.3 Insights in Feature Combinations for Learned Models

We inspect the predictions that our Penalty Direction Model made on the dataset and inspect predictions made for different feature combinations. We especially focus on the predictions for the natural corner as that is the most common corner. We observe that players aim for their natural corner less often when (a) they take their first ever penalty, (b) when a miss will result in losing a shootout, or (c) the previous kick from the player's team was saved in the player's natural corner, with this effect being stronger when a goal will result in a shootout win. Finally, the first kick for each team in a shootout is aimed less often to the natural corner, whereas a team's 5th kick is aimed more toward the natural corner, especially if the taker is a forward.

We also inspect the estimations that our Penalty Distance Model made for different feature combinations. We observe that the importance of the kick in the shootout and the player's penalty experience impact the player's kick placement. The model learned that players who happen to take their first penalty kick in an official match ever in a penalty shootout are more likely to place the ball closer to the goalkeeper when a goal will result in a win. Such

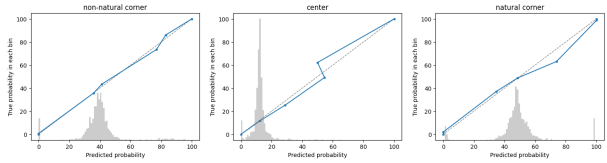


Figure A.4: Calibration plots for the predictions by the Penalty Direction Model for each of our classes (non-natural, center, natural).

players might be afraid to shoot wide and therefore place the ball closer to the goalkeeper in such a situation. On the other hand, experienced pressure penalty takers (at least 5 penalties taken under normal or high pressure), seem not to be influenced by this and will place the ball further away from the goalkeeper if a goal will result in a win.

Bibliography

[1] B. Aalbers. *SciSkill Index - Why and How*. 2016. URL: https://www.scisports.com/sciskill-index-why-and-how/.

[2] B. Aalbers and J. Van Haaren. "Distinguishing between roles of football players in play-by-play match event data". In: *Machine Learning and Data Mining for Sports Analytics: 5th International Workshop, MLSA 2018, Co-located with ECML/PKDD 2018, Dublin, Ireland, September 10, 2018, Proceedings 5*. Springer. 2019, pp. 31–41.

[3] D. Adams, R. Morgans, J. Sacramento, S. Morgan, and M. D. Williams. "Successful short passing frequency of defenders differentiates between top and bottom four English Premier League teams". In: *International Journal of Performance Analysis in Sport* 13.3 (2013), pp. 653–668.

[4] A. Ali. "Measuring soccer skill performance: a review". In: *Scandinavian journal of medicine & science in sports* 21.2 (2011), pp. 170–183.

[5] C. Almeida, A. Volossovitch, and R. Duarte. "Penalty kick outcomes in UEFA club competitions (2010-2015): The roles of situational, individual and performance factors". In: *International Journal of Performance Analysis in Sport* 16 (Aug. 2016), pp. 508–522.

[6] A. S. Analysis. *Goals added: Introducing a new way to measure soccer*. 2020. URL: https://www.americansocceranalysis.com/home/2020/4/22/37ucr0d5urxxtryn2cfhzormdziphq.

[7] G. Anzer, K. Arnsmeyer, P. Bauer, J. Bekkers, U. Brefeld, J. Davis, N. Evans, M. Kempe, S. J. Robertson, J. Wyatt Smith, and J. Van Haaren. "Common Data Format (CDF): A Standardized Format for Match-Data in Football (Soccer)". In: *arXiv preprint: arxiv:2505.15820* (2025). URL: https://arxiv.org/abs/2505.15820.

[8] G. Anzer and P. Bauer. "A goal scoring probability model for shots based on synchronized positional and event data in football (soccer)". In: *Frontiers in sports and active living* 3 (2021), p. 624475.

[9] J. Apesteguia and I. Palacios-Huerta. "Psychological Pressure in Competitive Environments: Evidence from a Randomized Natural Experiment". In: *American Economic Review* 100.5 (Dec. 2010), pp. 2548–64.

[10] V. Armatas, A. Yiannakos, and P. Sileloglou. "Relationship between time and goal scoring in soccer games: Analysis of three World Cups". In: *International Journal of Performance Analysis in Sport* 7 (May 2007), pp. 48–58. DOI: 10.1080/24748668.2007.11868396.

[11] L. Arrondel, R. Duhautois, and J.-F. Laslier. "Decision under Psychological Pressure: The Shooter's Anxiety at the Penalty Kick". In: *Journal of Economic Psychology* 70 (Oct. 2018).

[12] O. Azar and M. Bar-Eli. "Do soccer players play the mixed-strategy Nash equilibrium?" In: *Applied Economics* 43.25 (2011), pp. 3591–3601.

[13] A. Baouan, S. Coustou, M. Lacome, S. Pulido, and M. Rosenbaum. "Crediting football players for creating dangerous actions in an unbiased way: the generation of threat (GoT) indices". In: *arXiv preprint arXiv:2304.05242* (2023).

[14] M. Bar-Eli and O. Azar. "Penalty kicks in soccer: An empirical analysis of shooting strategies and goalkeepers' preferences". In: *Soccer and Society* 10 (Mar. 2009).

[15] M. Bar-Eli, O. Azar, I. Ritov, Y. Keidar-Levin, and G. Schein. "Action Bias among Elite Soccer Goalkeepers: The Case of Penalty Kicks". In: *Journal of Economic Psychology* 28 (Oct. 2007), pp. 606–621.

[16] E. Baron, N. Sandholtz, D. Pleuler, and T. C. Chan. "Miss it like Messi: Extracting value from off-target shots in soccer". In: *Journal of Quantitative Analysis in Sports* 20.1 (2024), pp. 37–50.

[17] B. S. Baumer, S. T. Jensen, and G. J. Matthews. "openWAR: An open source system for evaluating overall player performance in major league baseball". In: *Journal of Quantitative Analysis in Sports* 11.2 (2015), pp. 69–84.

[18] R. Beal, N. Changder, T. Norman, and S. Ramchurn. "Learning the value of teamwork to form efficient teams". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 05. 2020, pp. 7063–7070.

[19] R. Beal, T. J. Norman, and S. D. Ramchurn. "Artificial intelligence for team sports: a survey". In: *The Knowledge Engineering Review* 34 (2019), e28.

[20] J. Bekkers. *An Intuitive Measure for Pressing*. 2024. URL: https://unravelsports.github.io/2024/12/12/pressing-intensity.html.

[21] R. Bellman. “Dynamic programming”. In: *science* 153.3731 (1966), pp. 34–37.

[22] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, and I. Matthews. “Win at home and draw away: Automatic formation analysis highlighting the differences in home and away team behaviors”. In: *Proceedings of 8th annual MIT sloan sports analytics conference*. Citeseer. 2014, pp. 1–7.

[23] J. Boice. *How Our 2018 World Cup Predictions Work*. 2018. URL: https://fivethirtyeight.com/features/how-our-2018-world-cup-predictions-work/.

[24] B. H. Boon and G. Sierksma. “Team formation: Matching quality supply and quality demand”. In: *European Journal of Operational Research* 148.2 (2003), pp. 277–292.

[25] P. S. Bradley, C. Carling, A. G. Diaz, P. Hood, C. Barnes, J. Ade, M. Boddy, P. Krustrup, and M. Mohr. “Match performance and physical capacity of players in the top three competitive standards of English professional soccer”. In: *Human movement science* 32.4 (2013), pp. 808–821.

[26] P. S. Bradley, A. Dellal, M. Mohr, J. Castellano, and A. Wilkie. “Gender differences in match performance characteristics of soccer players competing in the UEFA Champions League”. In: *Human movement science* 33 (2014), pp. 159–171.

[27] P. S. Bradley, A. Dellal, M. Mohr, J. Castellano, and A. Wilkie. “Gender differences in match performance characteristics of soccer players competing in the UEFA Champions League”. In: *Human Movement Science* 33 (2014), pp. 159–171.

[28] L. Bransen. “Valuing Passes in Football Using Ball Event Data”. MA thesis. Erasmus University Rotterdam, 2017. URL: https://thesis.eur.nl/pub/41346.

[29] L. Bransen and J. Davis. “A data-driven analysis of the technical and tactical evolution of elite women’s football”. In: *International Journal of Sports Science & Coaching* 19.6 (2024), pp. 2449–2459.

[30] L. Bransen and J. Davis. “Measuring player pace of play in elite female association football”. In: *International Journal of Sports Science & Coaching* (2025), p. 17479541251334022.

[31] L. Bransen and J. Davis. “Predicting football penalty directions using in-match performance indicators”. In: *Proceedings of the Sports Tomorrow Congress 2021*. 2021.

[32] L. Bransen and J. Davis. “Women’s football analyzed: Interpretable expected goals models for women”. In: *Proceedings of the AI for Sports Analytics (AISA) Workshop at the International Joint Conference on Artificial Intelligence*. 2021.

[33] L. Bransen, T. Janssen, and J. Davis. “Stop Guessing: Optimizing Goalkeeper Policies for Soccer Penalty Kicks”. In: *arXiv preprint arXiv:2505.24629* (2025).

[34] L. Bransen, P. Robberechts, J. Davis, T. Decroos, and J. Van Haaren. “How does context affect player performance in football?” In: *Football Analytics 2021: The Role of Context in Transferring Analytics to the Pitch* (2020).

[35] L. Bransen, P. Robberechts, J. Van Haaren, and J. Davis. “Choke or shine? Quantifying soccer players’ abilities to perform under mental pressure”. In: *Proceedings of the 13th MIT Sloan Sports Analytics Conference*. 2019.

[36] L. Bransen and J. Van Haaren. “Measuring football players’ on-the-ball contributions from passes during games”. In: *Proceedings of the Machine Learning and Data Mining for Sports Analytics ECML/PKDD 2018 workshop*. 2018.

[37] L. Bransen and J. Van Haaren. “Measuring football players’ on-the-ball contributions from passes during games”. In: *Machine Learning and Data Mining for Sports Analytics: 5th International Workshop, MLSA 2018, Co-located with ECML/PKDD 2018, Dublin, Ireland, September 10, 2018, Proceedings 5*. Springer. 2019, pp. 3–15.

[38] L. Bransen and J. Van Haaren. “Player Chemistry: Striving for a Perfectly Balanced Soccer Team”. In: *Proceedings of the 14th MIT Sloan Sports Analytics Conference*. 2020.

[39] L. Bransen, J. Van Haaren, and M. van de Velden. “Measuring soccer players’ contributions to chance creation by valuing their passes”. In: *Journal of Quantitative Analysis in Sports* 15.2 (2019), pp. 97–116.

[40] G. W. Brier. “Verification of forecasts expressed in terms of probability”. In: *Monthly weather review* 78.1 (1950), pp. 1–3.

[41] S. Brown. “A PageRank model for player performance assessment in basketball, soccer and hockey”. In: *arXiv preprint arXiv:1704.00583* (2017).

[42] B. Buscà, R. Hileno, B. Nadal, and J. Serna. “Prediction of the penalty kick direction in men’s soccer”. In: *International Journal of Performance Analysis in Sport* 22 (2022), pp. 571–582.

[43] M. Bush, C. Barnes, D. T. Archer, B. Hogg, and P. S. Bradley. "Evolution of match performance parameters for various playing positions in the English Premier League". In: *Human movement science* 39 (2015), pp. 1–11.

[44] M. Caley. "Premier League Projections and New Expected Goals". Cartilage Free Captain. 2015. URL: https://cartilagefreecaptain.sbnation.com/2015/10/19/9295905/premier-league-projections-and-new-expected-goals.

[45] M. Caley. *Premier League Projections and New Expected Goals*. http://cartilagefreecaptain.sbnation.com/. 2015. URL: http://cartilagefreecaptain.sbnation.com/2015/10/19/9295905/premier-league-projections-and-new-expected-goals (visited on 12/15/2015).

[46] M. Cardoso de Araújo, C. Baumgart, C. T. Jansen, J. Freiwald, and M. W. Hoppe. "Sex Differences in Physical Capacities of German Bundesliga Soccer Players". In: *Journal of Strength and Conditioning Research* 34 (2020), pp. 2329–2337.

[47] C. Casal, R. Maneiro, A. Ardá, and J. L. Losada. "Gender differences in technical-tactical behaviour of Laliga Spanish football teams". In: *Journal of Human Sport and Exercise* 16 (2021), pp. 37–52.

[48] C. Casal, J. Stone, I. Iván-Baragaño, and J. Losada. "Effect of goalkeepers' offensive participation on team performance in the women Spanish La Liga: a multinomial logistic regression analysis". In: *Biology of Sport* 41.1 (2024), pp. 29–39.

[49] M. Cavus and P. Biecek. "Explainable expected goal models for performance analysis in football analytics". In: *2022 ieee 9th international conference on data Science and advanced analytics (DSAA)*. IEEE. 2022, pp. 1–9.

[50] D. Cervone, A. D'amour, L. Bornn, and K. Goldsberry. "Pointwise: Predicting points and valuing decisions in real time with nba optical tracking data". In: *Proceedings of the 8th MIT Sloan Sports Analytics Conference, Boston, MA, USA*. Vol. 28. 3. 2014.

[51] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2.

[52] T. Chen and C. Guestrin. "XGBoost: A scalable tree boosting system". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 785–794.

[53] P. Chiappori, S. Levitt, and T. Groseclose. "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer". In: *American Economic Review* 92.4 (2002), pp. 1138–1151.

[54] C. M. Clapp and R. Henderson. *Speed Kills? The Effect of the Hurry-Up-No-Huddle Offense on Injuries in College Football.* 2018. URL: https://cclapp.github.io/ChrisClapp.org/Files/Manuscripts/Clapp%20&%20Henderson%20-%20Speed%20Kills-The%20Effect%20of%20HUNH%20Offenses%20on%20Injuries%20in%20College%20Football.pdf.

[55] L. Csató. "Club coefficients in the UEFA Champions League: Time for shift to an Elo-based formula". In: *International Journal of Performance Analysis in Sport* 24.2 (2024), pp. 119–134.

[56] G. Dantas and M. Ritt. "Squad Optimizing: Integrating Tactics into Football Team Composition Problem via Integer Programming". In: *Environment Systems and Decisions* 45 (June 2025). DOI: 10.1007/s10669-025-10033-2.

[57] J. Davis, L. Bransen, T. Decroos, P. Robberechts, and J. Van Haaren. "Assessing the performances of soccer players". In: *International symposium on computer science in sport*. Springer. 2019, pp. 3–10.

[58] J. Davis, L. Bransen, L. Devos, A. Jaspers, W. Meert, P. Robberechts, J. Van Haaren, and M. Van Roy. "Methodology and evaluation in sports analytics: challenges, approaches, and lessons learned". In: *Machine Learning* 113.9 (2024), pp. 6977–7010.

[59] J. Davis, L. Bransen, L. Devos, W. Meert, and P. Robberechts. "Evaluating Sports Analytics Models: Challenges, Approaches, and Lessons Learned." In: *EBeM at the International Joint Conference on Artificial Intelligence*. 2022.

[60] J. Davis and P. Robberechts. "Biases in Expected Goals Models Confound Finishing Ability". In: *arXiv preprint arxiv:2401.09940* (2024).

[61] J. Davis and P. Robberechts. "Biases in expected goals models confound finishing ability". In: *arXiv preprint arXiv:2401.09940* (2024).

[62] T. Decroos, L. Bransen, J. Van Haaren, and J. Davis. "Actions Speak Louder Than Goals: Valuing Player Actions in Soccer". In: *Proc. of 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019, pp. 1851–1861.

[63] T. Decroos, L. Bransen, J. Van Haaren, and J. Davis. "Actions speak louder than goals: Valuing player actions in soccer". In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 1851–1861.

[64] T. Decroos, L. Bransen, J. Van Haaren, and J. Davis. "VAEP: An objective approach to valuing on-the-ball actions in soccer". In: *International Joint Conference on Artificial Intelligence*. 2020, pp. 4696–4700.

[65] T. Decroos and J. Davis. "Interpretable Prediction of Goals in Soccer". In: *AAAI 2020 Workshop on Artificial Intelligence in Team Sports*. 2020.

[66] T. Decroos, J. Van Haaren, V. Dzyuba, and J. Davis. "STARSS: A Spatio-Temporal Action Rating System for Soccer." In: *Proceedings of the Fourth Workshop on Machine Learning and Data Mining for Sports Analytics*. 2017, pp. 11–20.

[67] U. Dick, D. Link, and U. Brefeld. "Who can receive the pass? A computational model for quantifying availability in soccer". In: *Data mining and knowledge discovery* 36.3 (2022), pp. 987–1014.

[68] M. Dicks, C. Button, and K. Davids. "Availability of Advance Visual Information Constrains Association-Football Goalkeeping Performance during Penalty Kicks". In: *Perception* 39 (Aug. 2010), pp. 1111–24.

[69] M. Dicks, C. Button, and K. Davids. "Examination of Gaze Behaviors under in situ and Video Simulation Task Constraints Reveals Differences in Information Pickup for Perception and Action". In: *Attention, Perception & Psychophysics* 72 (Apr. 2010), pp. 706–20.

[70] J. W. Dipple, L. Bruce, and D. B. Dwyer. "Identifying the optimal characteristics of ball possession and movement in elite women's soccer". In: *International Journal of Performance Analysis in Sport* 22.4 (2022), pp. 594–603.

[71] N. E. Dona and T. Swartz. "A causal investigation of pace of play in soccer". In: *Statistica Applicata-Italian Journal of Applied Statistics* 1 (2023).

[72] B. Dure. *Europe's leagues are overtaking the NWSL. That's a win for women's soccer*. 2023. URL: `https://www.theguardian.com/football/2023/aug/23/nwsl-soccer-wsl-overtaken-womens-soccer`.

[73] A. Elo. *The Rating of Chessplayers, Past and Present*. 1978.

[74] B. R. Fajen. "Affordance-Based Control of Visually Guided Action". In: *Ecological Psychology* 19 (2007), pp. 383–410.

[75] J. Fernández, L. Bornn, and D. Cervone. "Decomposing the Immeasurable Sport: A Deep Learning Expected Possession Value Framework for Soccer". In: *MIT Sloan Sports Analytics Conference*. 2019.

[76] J. Fernandez-Navarro, L. Fradua, A. Zubillaga, P. R. Ford, and A. P. McRobert. "Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams". In: *Journal of sports sciences* 34.24 (2016), pp. 2195–2204.

[77] FIFA. *FIFA benchmarking report women's football*. 2020. URL: https://digitalhub.fifa.com/m/3ba9d61ede0a9ee4/original/dzm2o61buenfox51qjot-pdf.pdf.

[78] FIFA. *FIFA Global Transfer Market Report 2023*. 2023. URL: https://digitalhub.fifa.com/m/114622e4e17cf6a8/original/FIFA-Global-Transfer-Report-2023.pdf.

[79] FIFA. *FIFA Professional Football Report 2023. Men's Football*. 2023. URL: https://digitalhub.fifa.com/m/2a5dc95026d9cf8a/original/FIFA-Professional-Football-Report-2023.pdf.

[80] FIFA. *FIFA Survey: approximately 250 million footballers worldwide*. 2006. URL: https://web.archive.org/web/20060915133001/http://access.fifa.com/infoplus/IP-199_01E_big-count.pdf.

[81] FIFA. *FIFA Women's World Cup Australia New Zealand 2023. Global Engagement Audience Report*. 2023. URL: https://digitalhub.fifa.com/m/efa90ed1ddbe3bf/original/FIFA-Women-s-World-Cup-Australia-New-Zealand-2023-Global-Engagement-Audience-Executive-Summary.pdf.

[82] FIFA. "FIFA Women's World Cup 2019™ watched by more than 1 billion". FIFA. 2019. URL: https://inside.fifa.com/tournaments/womens/womensworldcup/france2019/news/fifa-women-s-world-cup-2019tm-watched-by-more-than-1-billion.

[83] FIFA. *FIFA World Cup Qatar 2022. Global Engagement Audience Report*. 2022. URL: https://digitalhub.fifa.com/m/336c30db79dafa93/original/FIFA-World-Cup-Qatar-2022-Global-Engagement-Audience-Executive-Summary.pdf.

[84] FIFA. "Laws of the Game 2022/2023". In: (2022). URL: https://www.theifab.com/laws-of-the-game-documents/.

[85] FIFA. *Women's Football: Member Associations survey report 2023*. 2023. URL: https://digitalhub.fifa.com/m/28ed34bd888832a8/original/FIFA-Women-s-Football-MA-Survey-Report-2023.pdf.

[86] FIFA. *Women's transfer spending on the rise, bucking global trend*. 2021. URL: https://www.fifa.com/legal/news/women-s-transfer-spending-on-the-rise-bucking-global-trend.

[87] L. Forcher. "Success Factors in Soccer Defense–Match Analysis in Soccer based on Positional Tracking Data". PhD thesis. Institut für Sport und Sportwissenschaft (IfSS), 2024.

[88] L. Forcher, L. Forcher, S. Altmann, D. Jekauc, and M. Kempe. "Is a compact organization important for defensive success in elite soccer?–Analysis based on player tracking data". In: *International Journal of Sports Science & Coaching* 19.2 (2024), pp. 757–768.

[89] L. Forcher, L. Preine, L. Forcher, H. Wäsche, D. Jekauc, A. Woll, T. Gross, and S. Altmann. "Shedding some light on in-game formation changes in the German Bundesliga: Frequency, contextual factors, and differences between offensive and defensive formations". In: *International Journal of Sports Science & Coaching* 18.6 (2023), pp. 2051–2060.

[90] P. Furley. "An evolutionary approach to emotions in sports". In: *Feelings in Sport*. Routledge, 2020, pp. 27–36.

[91] P. Galaz-Cares. "Optimizing Football Squad Planning: Balancing Competitive Success and Financial Sustainability". In: *StatsBomb Conference 2024* (2024).

[92] S. Ganguly and N. Frank. "The problem with win probability". In: *Proceedings of the 12th MIT Sloan Sports Analytics Conference*. 2018.

[93] M. Garnica-Caparrós and D. Memmert. "Understanding gender differences in professional European football through machine learning interpretability and match actions data". In: *Scientific reports* 11.1 (2021), p. 10805.

[94] F. Goes, M. Kempe, L. Meerhoff, and K. Lemmink. "Not every pass can be an assist: a data-driven model to measure pass effectiveness in professional soccer matches". In: *Big data* 7.1 (2019), pp. 57–70.

[95] F. Goes, E. Schwarz, M. Elferink-Gemser, K. Lemmink, and M. Brink. "A risk-reward assessment of passing decisions: comparison between positional roles using tracking data from professional men's soccer". In: *Science and Medicine in Football* 6.3 (2022), pp. 372–380.

[96] S. Green. *Assessing The Performance of Premier League Goalscorers*. `https://www.statsperform.com/resource/assessing-the-performance-of-premier-league-goalscorers/`. 2012.

[97] S. Gregory, S. Robertson, R. Aughey, B. Spencer, and J. Alexander. "Assigning goal-probability value to high intensity runs in football". In: *Plos one* 19.9 (2024), e0308749.

[98] J.-F. Gréhaigne, P. Godbout, and D. Bouthier. "The foundations of tactics and strategy in team sports". In: *Journal of teaching in physical education* 18.2 (1999), pp. 159–174.

[99] J. Griffin, B. Larsen, S. Horan, J. Keogh, K. Dodd, M. Andreatta, and C. Minahan. "Women's football: an examination of factors that influence movement patterns". In: *The journal of strength & conditioning research* 34.8 (2020), pp. 2384–2393.

[100] D. Gucciardi and J. Dimmock. "Choking under pressure in sensorimotor skills: Conscious processing or depleted attentional resources?" In: *Psychology of Sport and Exercise* 9.1 (2008), pp. 45–59.

[101] L. Gyarmati, H. Kwak, and P. Rodriguez. "Searching for a Unique Style in Soccer". In: *arXiv preprint arXiv:1409.0308* (2014).

[102] L. Gyarmati and R. Stanojevic. "QPass: a Merit-based Evaluation of Soccer Passes". In: *CoRR* abs/1608.03532 (2016). URL: http://arxiv.org/abs/1608.03532.

[103] A. Harkness-Armstrong, K. Till, N. Datson, N. Myhill, and S. Emmonds. "A systematic review of match-play characteristics in women's soccer". In: *PloS one* 17.6 (2022), e0268334.

[104] C. Harpur. *Women's football scouting is budget-constrained – but there are still savvy ways to thrive*. 2024. URL: https://www.nytimes.com/athletic/5158707/2024/01/02/womens-football-scouting-wsl/.

[105] N. Harris. "Number-crunched: The rise and rise of the Women's World Cup". Sporting Intelligence. 2015. URL: https://www.sportingintelligence.com/2015/06/01/number-crunched-the-rise-and-rise-of-the-womens-world-cup-010601/.

[106] J. Hewitt and O. Karakuş. "A machine learning approach for player and position adjusted expected goals in football (soccer)". In: *Franklin Open* 4 (2023), p. 100034.

[107] D. M. Hill, S. Hanton, N. Matthews, and S. Fleming. "Choking in sport: A review". In: *International Review of Sport and Exercise Psychology* 3.1 (2010), pp. 24–39.

[108] M. Horn, S. de Waal, and W. Kraak. "In-match penalty kick analysis of the 2009/10 to 2018/19 English Premier League competition". In: *International Journal of Performance Analysis in Sport* 21 (2020), pp. 139–155.

[109] M. Hughes and I. Franks. "Analysis of passing sequences, shots and goals in soccer". In: *Journal of sports sciences* 23.5 (2005), pp. 509–514.

[110] A. Hunter, M. Angilletta, and R. Wilson. "Behaviors of shooter and goalkeeper interact to determine the outcome of soccer penalties". In: *Scandinavian Journal of Medicine & Science in Sports* 28 (Aug. 2018).

[111] A. Hunter, T. Pavlic, M. Angilletta, and R. Wilson. "Identifying the best strategy for soccer penalty success: A predictive model for optimising behavioural and biomechanical trade-offs". In: *Journal of Biomechanics* 141 (June 2022), p. 111208.

[112] L. M. Hvattum and H. Arntzen. “Using ELO ratings for match result prediction in association football”. In: *International Journal of forecasting* 26.3 (2010), pp. 460–470.

[113] R. Ibrahim, I. Kingma, V. de Boode, G. Faber, and J. Van Dieen. “Kinematic and kinetic analysis of the goalkeeper’s diving save in football”. In: *Journal of Sports Sciences* 37 (July 2018), pp. 1–9.

[114] S. IJtsma. “A Close Look at My New Expected Goals Model”. 11tegen11. 2015. URL: `http://11tegen11.net/2015/08/14/a-close-look-at-my-new-expected-goals-model/`.

[115] I. Iván-Baragaño, R. Maneiro, J. L. Losada, and A. Ardá. “Multivariate analysis of the offensive phase in high-performance Women’s soccer: a mixed methods study”. In: *Sustainability* 13.11 (2021), p. 6379.

[116] A. Z. Jacobs and H. Wallach. “Measurement and fairness”. In: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 2021, pp. 375–385.

[117] P. Jones, N. James, and S. D. Mellalieu. “Possession as a performance indicator in soccer.” In: *International Journal of Performance Analysis in Sport* 4.1 (2004), pp. 98–102.

[118] L. M. de Jong, P. B. Gastin, L. Bruce, and D. B. Dwyer. “Teamwork and performance in professional women’s football: A network-based analysis”. In: *International Journal of Sports Science & Coaching* 18.3 (2023), pp. 848–857.

[119] G. Jordet and M. Elferink-Gemser. “Stress, Coping, and Emotions on the World Stage: The Experience of Participating in a Major Soccer Tournament Penalty Shootout”. In: *Journal of Applied Sport Psychology* 24 (Jan. 2012), pp. 73–91.

[120] G. Jordet, M. Elferink-Gemser, K. Lemmink, and C. Visscher. “The "Russian roulette" of soccer?: Perceived control and anxiety in a major tournament penalty shootout”. In: *International Journal of Sport Psychology* 37 (Apr. 2006), pp. 281–298.

[121] T. Knutson. *StatsBomb Release Expected Goals with Shot Impact Height*. `https://statsbomb.com/2020/07/statsbomb-release-expected-goals-with-shot-impact-height/`. 2020.

[122] S. Kovalchik. “Extension of the Elo rating system to margin of victory”. In: *International Journal of Forecasting* 36.4 (2020), pp. 1329–1341.

[123] S. Kovalchik. “Player tracking data in sports”. In: *Annual Review of Statistics and Its Application* 10.1 (2023), pp. 677–697.

[124] J. Kubatko, D. Oliver, K. Pelton, and D. T. Rosenbaum. "A starting point for analyzing basketball statistics". In: *Journal of quantitative analysis in sports* 3.3 (2007).

[125] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. Blei. "Automatic differentiation variational inference". In: *Journal of machine learning research* 18.14 (2017), pp. 1–45.

[126] W. Kuhn. "Penalty-kick strategies for shooters and goalkeepers". In: *Science and football* (1988), pp. 489–492.

[127] C. Lago. "The influence of match location, quality of opposition, and match status on possession strategies in professional association football". In: *Journal of sports sciences* 27.13 (2009), pp. 1463–1469.

[128] T. Lawrence. *Expected Saves*. 2015. URL: `https://deepxg.com/2015/10/20/expected-saves/`.

[129] T. Lawrence. *Introducing xGChain and xGBuildup*. 2018. URL: `https://www.hudl.com/blog/introducing-xgchain-and-xgbuildup`.

[130] A. Lees and L. Nolan. "The Biomechanics of Soccer: a Review." In: *Journal of sports sciences* 16 3 (1998), pp. 211–34.

[131] D. Link, S. Lang, and P. Seidenschwarz. "Real time quantification of dangerousity in football using spatiotemporal tracking data". In: *PloS one* 11.12 (2016), e0168768.

[132] G. Liu, Y. Luo, O. Schulte, and T. Kharrat. "Deep soccer analytics: learning an action-value function for evaluating soccer players". In: *Data Mining and Knowledge Discovery* 34 (2020), pp. 1531–1559.

[133] P. Lucey, A. Bialkowski, M. Monfort, P. Carr, and I. Matthews. "Quality vs Quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data". In: *Proceedings of the 9th MIT Sloan Sports Analytics Conference*. 2015.

[134] P. Madrero Pardo. "Creating a Model for Expected Goals in Football using Qualitative Player Information". MA thesis. Universitat Politècnica de Catalunya, 2020.

[135] J. Magee. *Women, football and Europe: Histories, equity and experience*. Vol. 1. Meyer & Meyer Verlag, 2008.

[136] R. Masters, J. Kamp, and R. Jackson. "Imperceptibly Off-Center Goalkeepers Influence Penalty-Kick Direction in Soccer". In: *Psychological Science* 18 (Apr. 2007), pp. 222–3.

[137] P. McFarlane. "Evaluating NBA end-of-game decision-making". In: *Journal of Sports Analytics* 5.1 (2019), pp. 17–22.

[138] I. G. McHale and Ł. Szczepański. "A mixed effects model for identifying goal scoring ability of footballers". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 177.2 (2014), pp. 397–417.

[139] L. A. Meerhoff, F. R. Goes, A. W. De Leeuw, and A. Knobbe. "Exploring successful team tactics in soccer tracking data". In: *Joint european conference on machine learning and knowledge discovery in databases*. Springer. 2019, pp. 235–246.

[140] T. Mendes-Neves, L. Meireles, J. Mendes-Moreira, et al. "Estimating player performance in different contexts using fine-tuned large events models". In: *arXiv preprint arXiv:2402.06815* (2024).

[141] T. Mendes-Neves, L. Meireles, and J. Mendes-Moreira. "Towards a foundation large events model for soccer". In: *Machine Learning* 113.11 (2024), pp. 8687–8709.

[142] C. Merhej, R. J. Beal, T. Matthews, and S. Ramchurn. "What happened next? Using deep learning to value defensive actions in football event-data". In: *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 2021, pp. 3394–3403.

[143] P. Mesquita, B. Silva, M. Rodrigues, N. Maurício, F. M. Clemente, M. Camões, and R. Lima. "Analysis of goal-scoring in an elite European women's football teams". In: *Sustainability and Sports Science Journal* 1.1 (2023), pp. 16–24.

[144] K. Michalczyk. *An attempt to extend xG gain*. 2018. URL: `https://kubamichalczyk.github.io//2018/11/18/An-attempt-to-extend-xG-gain.html`.

[145] M. Mitrotasios, J. G. Ródenas, V. Armatas, and R. A. Malavés. "Creating goal scoring opportunities in men and women UEFA Champions League soccer matches: Tactical similarities and differences". In: *Retos: Nuevas tendencias en educación física, deporte y recreación* 43 (2022), pp. 154–161.

[146] E. Morgulev and Y. Galily. "Choking or delivering under pressure? The case of elimination games in NBA playoffs". In: *Frontiers in Psychology* 9 (2018), p. 979.

[147] S. Narayanan and N. D. Pifer. "An xG of Their Own: Using Expected Goals to Explore the Analytical Shortcomings of Misapplied Gender Schemas in Football". In: *Journal of Sport Management* 38.2 (2023), pp. 92–109.

[148] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1947.

[149] Nielsen. *The 2022 World Football Report*. 2022. URL: https://www.nielsen.com/insights/2022/discover-what-world-cup-fans-really-want/.

[150] B. Noël, P. Furley, J. Kamp, M. Dicks, D. Memmert, and T. Francis. "The development of a method for identifying penalty kick strategies in association football". In: *Journal of Sports Sciences* (June 2014).

[151] B. Noël, J. van der Kamp, and S. Klatt. "The Interplay of Goalkeepers and Penalty Takers Affects Their Chances of Success". In: *Frontiers in Psychology* 12 (2021).

[152] B. Noël, J. van der Kamp, and D. Memmert. "Implicit Goalkeeper Influences on Goal Side Selection in Representative Penalty Kicking Tasks". In: *PLOS ONE* 10 (Aug. 2015).

[153] H. Nori, S. Jenkins, P. Koch, and R. Caruana. "InterpretML: A Unified Framework for Machine Learning Interpretability". In: *arXiv preprint arXiv:1909.09223* (2019).

[154] NWSL. *NWSL Competition Manual*. 2025. URL: https://images.nwslsoccer.com/image/private/t_q-good/prd/bqso2kywvrxygd3ppsru.pdf.

[155] C. F. Observatory. *Spatial analysis of transfer revenues streams (2014-2023)*. 2023. URL: https://football-observatory.com/MonthlyReport92.

[156] K. Okholm Kryger, A. Wang, R. Mehta, F. M. Impellizzeri, A. Massey, and A. McCall. "Research on women's football: a scoping review". In: *Science and Medicine in Football* 6.5 (2022), pp. 549–558.

[157] S. B. Olthof, W. G. Frencken, and K. A. Lemmink. "When something is at stake: Differences in soccer performance in 11 vs. 11 during official matches and training games". In: *The Journal of Strength & Conditioning Research* 33.1 (2019), pp. 167–173.

[158] D. C. Oruç, L. Cascioli, L. Stradiotti, M. Van Roy, P. Robberechts, and J. Davis. "An Analysis of the Influence of Game Context on Team Playing Style". In: *International Workshop on Machine Learning and Data Mining for Sports Analytics*. Springer. 2024, pp. 57–68.

[159] F. Outsiders. *Methods To Our Madness*. URL: https://footballoutsiders.com/info/methods.

[160] I. Palacios-Huerta. "Professionals Play Minimax". In: *Review of Economic Studies* 70 (Feb. 2003), pp. 395–415.

[161] L. Pappalardo, A. Rossi, M. Natilli, and P. Cintia. "Explaining the difference between men's and women's football". In: *PLoS one* 16.8 (2021), e0255407.

[162] W. Pauwels. “Supporting Transfer and Wage Decisions in Football Using a Player’s Expected On Field Value and Future Transfer Fee”. MA thesis. KU Leuven, 2022.

[163] A. V. Pedersen, I. M. Aksdal, and R. Stalsberg. “Scaling Demands of Soccer According to Anthropometric and Physiological Sex Differences: A Fairer Comparison of Men’s and Women’s Soccer”. In: *Frontiers in Psychology* 10 (2019), p. 762.

[164] K. Pelechrinis. “iwinrnfl: A simple, interpretable & well-calibrated in-game win probability model for nfl”. In: *arXiv preprint arXiv:1704.00197* (2017).

[165] S. Perform. *Introducing a Possessions Framework*. 2019. URL: `https://www.statsperform.com/resource/introducing-a-possessions-framework/`.

[166] D. Pleuler. *On the topic of Expected Goals and the repeatability of finishing skill*. Opta, Mar. 2014. URL: `http://www.optasportspro.com/about/optapro-blog/posts/2014/on-the-topic-of-expected-goals-and-the-repeatability-of-finishing-skill.aspx`.

[167] R. Pollard and C. Reep. “Measuring the effectiveness of playing strategies at soccer”. In: *Journal of the Royal Statistical Society Series D: The Statistician* 46.4 (1997), pp. 541–550.

[168] P. Power, J. Hobbs, H. Ruiz, X. Wei, and P. Lucy. “Mythbusting Set-Pieces in Soccer”. In: *MIT Sloan Sports Analytics Conference*. 2018.

[169] P. Power, H. Ruiz, X. Wei, and P. Lucey. “Not All Passes Are Created Equal: Objectively Measuring the Risk and Reward of Passes in Soccer from Tracking Data”. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017, pp. 1605–1613.

[170] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. “CatBoost: unbiased boosting with categorical features”. In: *Advances in neural information processing systems* 31 (2018).

[171] P. Rahimian and L. Toka. “A data-driven approach to assist offensive and defensive players in optimal decision making”. In: *International Journal of Sports Science & Coaching* 19.1 (2024), pp. 245–256.

[172] P. Rahimian, J. Van Haaren, T. Abzhanova, and L. Toka. “Beyond action valuation: A deep reinforcement learning framework for optimizing player decisions in soccer”. In: *16th MIT sloan sports analytics conference*. Vol. 3. 2022.

[173] P. Rahimian, J. Van Haaren, and L. Toka. “Towards maximizing expected possession outcome in soccer”. In: *International Journal of Sports Science & Coaching* 19.1 (2024), pp. 230–244.

[174] E. Rampinini, F. M. Impellizzeri, C. Castagna, A. J. Coutts, and U. Wisløff. "Technical performance during soccer matches of the Italian Serie A league: Effect of fatigue and competitive level". In: *Journal of science and medicine in sport* 12.1 (2009), pp. 227–233.

[175] A. Rathke. "An examination of expected goals and shot efficiency in soccer". In: *Journal of Human Sport and Exercise* 12.2 (2017), pp. 514–529.

[176] L. Raudonius and T. Seidl. "Shot analysis in different levels of german football using expected goals". In: *International Workshop on Machine Learning and Data Mining for Sports Analytics*. Springer. 2022, pp. 14–26.

[177] T. Reilly, A. M. Williams, A. Nevill, and A. Franks. "A multidisciplinary approach to talent identification in soccer". In: *Journal of sports sciences* 18.9 (2000), pp. 695–702.

[178] R. Rein and D. Memmert. "Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science". In: *SpringerPlus* 5 (2016), pp. 1–13.

[179] P. Robberechts and J. Davis. "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, Proceedings 7*. Springer. 2020, pp. 17–27.

[180] P. Robberechts, J. Van Haaren, and J. Davis. "Who Will Win It? An In-game Win Probability Model for Football". In: *arXiv preprint arXiv:1906.05029* (2019).

[181] P. Robberechts, M. Van Roy, and J. Davis. "un-xPass: Measuring Soccer Player's Creativity". In: *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*. 2023, pp. 4768–4777.

[182] A. Rovshitz and R. Puzis. "Transformer-Based Framework for Versatile Analysis of Events Data in Soccer". In: *International Workshop on Machine Learning and Data Mining for Sports Analytics*. Springer. 2024, pp. 80–92.

[183] S. Rudd. "A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains". In: *New England Symposium on Statistics in Sports*. 2011. URL: `http://nessis.org/nessis11/rudd.pdf`.

[184] H. Ruiz, P. Power, X. Wei, and P. Lucey. "" The Leicester City Fairytale?" Utilizing New Soccer Analytics Tools to Compare Performance in the 15/16 & 16/17 EPL Seasons". In: *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 2017, pp. 1991–2000.

[185] M. Russell and M. Kingsley. “Influence of exercise on skill proficiency in soccer”. In: *Sports Medicine* 41 (2011), pp. 523–539.

[186] M. Russell, G. Rees, and M. I. Kingsley. “Technical demands of soccer match play in the English championship”. In: *The Journal of Strength & Conditioning Research* 27.10 (2013), pp. 2869–2873.

[187] A. Sahasrabudhe and J. Bekkers. “A graph neural network deep-dive into successful counterattacks”. In: *Proceedings of the 17th annual MIT SLOAN sports analytics conference*. 2023.

[188] D. Sakellaris. “The In-Game Comparison Between Male and Female Footballers”. Statathlon. 2017. URL: https://statathlon.com/the-in-game-comparison-between-male-and-female-footballers/.

[189] J. Sampaio and V. Maçãs. “Measuring tactical behaviour in football”. In: *International journal of sports medicine* 33.05 (2012), pp. 395–401.

[190] H. Sarmento, R. Marcelino, M. T. Anguera, J. CampaniÇo, N. Matos, and J. C. LeitÃo. “Match analysis in football: a systematic review”. In: *Journal of sports sciences* 32.20 (2014), pp. 1831–1843.

[191] G. Savelsbergh, J. van der Kamp, A. Williams, and P. Ward. “Anticipation and visual search behaviour in expert soccer goalkeepers”. English. In: *Ergonomics* 48 (2005), pp. 1686–97. ISSN: 0014-0139.

[192] M. Scanlan, C. Harms, J. Cochrane Wilkie, and F. Ma’ayah. “The creation of goal scoring opportunities at the 2015 women’s world cup”. In: *International Journal of Sports Science & Coaching* 15.5-6 (2020), pp. 803–808.

[193] A. Scholtes and O. Karakuş. “Bayes-xG: player and position correction on expected goals (xG) using Bayesian hierarchical approach”. In: *Frontiers in Sports and Active Living* 6 (2024), p. 1348983.

[194] A. Schwarz. *The numbers game: Baseball’s lifelong fascination with statistics*. Macmillan, 2004.

[195] SciSports. “SciSports and Sports Interactive will bring next-gen analytics to life in Football Manager”. SciSports. 2020. URL: https://www.scisports.com/scisports-and-sports-interactive-will-bring-next-gen-analytics-to-life-in-football-manager/.

[196] B. Secco Faquin, L. A. Teixeira, C. R. Coelho Candido, D. Boari Coelho, J. Bayeux Dascal, and V. H. Alves Okazaki. “Prediction of ball direction in soccer penalty through kinematic analysis of the kicker”. In: *Journal of Sports Sciences* 41.7 (2023), pp. 668–676.

[197] J. Serra-Olivares, F. M. Clemente, and S. González-Víllora. “Tactical expertise assessment in youth football using representative tasks”. In: *Springerplus* 5.1 (2016), p. 1301.

[198] L. Shaw and S. Gopaladesikan. "Routine Inspection: A Playbook for Corner Kicks". In: *MIT Sloan Sports Analytics Conference*. 2021.

[199] E. Shen, S. Santo, and O. Akande. "Analyzing pace-of-play in soccer using spatio-temporal event data". In: *Journal of Sports Analytics* 8.2 (2022), pp. 127–139.

[200] R. M. Silva, J. Davis, and T. B. Swartz. "The evaluation of pace of play in hockey". In: *Journal of Sports Analytics* 4.2 (2018), pp. 145–151.

[201] I. Simpson, R. J. Beal, D. Locke, and T. J. Norman. "Seq2event: Learning the language of soccer using transformer-based match event prediction". In: *Proceedings of the 28th ACM sigkdd conference on knowledge discovery and data mining*. 2022, pp. 3898–3908.

[202] K. Singh. *Introducing expected threat (XT)*. 2019. URL: `https://karun.in/blog/expected-threat.html`.

[203] W. Spearman. "Beyond expected goals". In: *Proceedings of the 12th MIT Sloan Sports Analytics Conference*. 2018, pp. 1–17.

[204] W. Spearman, A. Basye, G. Dick, R. Hotovy, and P. Pop. "Physics-based modeling of pass probabilities in soccer". In: *Proceeding of the 11th MIT Sloan Sports Analytics Conference*. Vol. 1. 2017.

[205] M. Stöckl, T. Seidl, D. Marley, and P. Power. "Making offensive play predictable-using a graph convolutional network to understand defensive performance in soccer". In: *Proceedings of the 15th MIT sloan sports analytics conference*. Vol. 2022. 2021.

[206] Ł. Szczepański and I. McHale. *Beyond completion rate: evaluating the passing ability of footballers*. 2016.

[207] TheComeOnMan. *xPo*. 2020. URL: `https://thecomeonman.github.io/xPo/`.

[208] T. Tureen and S. Olthof. ""Estimated Player Impact"(EPI): Quantifying the effects of individual players on football (soccer) actions using hierarchical statistical models". In: *StatsBomb Conference Proceedings*. StatsBomb. 2022.

[209] O. Tütüncü, S. Mehta, and D. Memmert. "Zones where corner kicks are most effective in the penalty box: Is there a sweet spot?" In: *International Journal of Sports Science & Coaching* 19.4 (2024), pp. 1706–1722.

[210] K. e. a. Tuyls. "Game Plan: What AI Can Do for Football, and What Football Can Do for AI". In: *J. Artif. Int. Res.* 71 (Sept. 2021), pp. 41–88. ISSN: 1076-9757.

[211] UEFA. "UEFA Women's Champions League lays foundations for more sustainable future". UEFA. 2021. URL: https://www.uefa.com/insideuefa/news/0268-122415a0fc64-78d08826a35b-1000--uefa-women-s-champions-league-lays-foundations-for-more-sustain/.

[212] J. Van Der Kamp. "A field simulation study of the effectiveness of penalty kick strategies in soccer: late alterations of kick direction increase errors and reduce accuracy". In: *Journal of Sports Sciences* 24.05 (2006), pp. 467–477.

[213] J. Van Haaren, P. Robberechts, T. Decroos, L. Bransen, and J. Davis. "Analysing performance and playing style using ball event data". In: *Football Analytics: Now and Beyond. A deep dive into the current state of advanced data analytics* (2019).

[214] M. Van Roy, P. Robberechts, T. Decroos, and J. Davis. "Valuing On-the-Ball Actions in Soccer: A Critical Comparison of xT and VAEP". In: *2020 AAAI Workshop on AI in Team Sports*. 2020.

[215] M. Van Roy, P. Robberechts, W.-C. Yang, L. De Raedt, and J. Davis. "A Markov framework for learning and reasoning about strategies in professional soccer". In: *Journal of Artificial Intelligence Research* 77 (2023), pp. 517–562.

[216] M. Van Roy, P. Robberechts, W.-C. Yang, L. De Raedt, and J. Davis. "Leaving goals on the pitch: Evaluating decision making in soccer". In: 2021.

[217] M. Waldron and P. Worsfold. "Differences in the game specific skills of elite and sub-elite youth football players: Implications for talent identification". In: *International journal of performance analysis in sport* 10.1 (2010), pp. 9–24.

[218] Z. Wang, P. Veličković, D. Hennes, N. Tomašev, L. Prince, M. Kaisers, Y. Bachrach, R. Elie, L. K. Wenliang, F. Piccinini, et al. "TacticAI: an AI assistant for football tactics". In: *Nature communications* 15.1 (2024), p. 1906.

[219] M. Weigelt, D. Memmert, and T. Schack. "Kick it like Ballack: The effects of goalkeeping gestures on goal-side selection in experienced soccer players and soccer novices". In: *Journal of Cognitive Psychology* 24 (Dec. 2012).

[220] J. Williams. *Women's Football, Europe and Professionalization*. 2011.

[221] G. Wood, G. Jordet, and M. R. Wilson. "On winning the "lottery": Psychological preparation for football penalty shoot-outs". In: *Journal of Sports Sciences* 33.17 (2015), pp. 1758–1765.

[222] T. Worville. “Explained: How different are men’s and women’s football”. The Athletic. 2020. URL: https://theathletic.com/1756375/2020/04/20/womens-vs-mens-football-the-trends/.

[223] Y. Xu, L. Bretzner, T. Wang, and A. Maki. “Skor-xG: SKeleton-ORiented Expected Goal Estimation in Soccer”. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, pp. 5967–5977.

[224] C. Yeung and K. Fujii. “A strategic framework for optimal decisions in football 1-vs-1 shot-taking situations: An integrated approach of machine learning, theory-based modeling, and game theory”. In: *Complex & Intelligent Systems* 10.5 (2024), pp. 5989–6008.

[225] C. Yeung, K. Ide, and K. Fujii. “Autosoccerpose: Automated 3d posture analysis of soccer shot movements”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, pp. 3214–3224.

[226] D. Yu, C. Boucher, L. Bornn, and M. Javan. “Playing Fast Not Loose: Evaluating team-level pace of play in ice hockey using spatio-temporal possession data”. In: *Proceedings of the 13th MIT Sloan Sports Analytics Conference*. 2019.

[227] R. Zheng, C. Reus, and J. Kamp. “Goalkeeping in the soccer penalty kick: The dive is coordinated to the kicker’s non-kicking leg placement, irrespective of time constraints”. In: *Human Movement Science* 76 (Apr. 2021), p. 102763.

Statement on the use of Generative AI

I did not use generative AI assistance tools during the research/writing process of my thesis, except for mere language assistance.
The text, code, and images in this thesis are my own (unless otherwise specified). Generative AI has only been used in accordance with the KU Leuven guidelines and appropriate references have been added. I have reviewed and edited the content as needed and I take full responsibility for the content of the thesis.

Curriculum Vitae

Lotte Bransen was born in Utrecht, the Netherlands, on March 4th, 1993. She obtained her Bachelor of Science in Mathematics at Utrecht University in 2014 and her Master of Science in Econometrics at Erasmus University Rotterdam in 2017. She received the Football Science Groningen Award 2018 for her master thesis on "Valuing passes in football using ball event data".

Prior to starting her doctoral studies at KU Leuven in January 2020 she worked at football data analytics company SciSports. Lotte has published papers at both computer science and sport science conferences and journals. She gave a keynote talk at MathSports 2022 in Reading and an invited talk at Data Day MX 2023 in Mexico City.

Publications

Journal Papers

L. Bransen and J. Davis. "Measuring player pace of play in elite female association football". In: *International Journal of Sports Science & Coaching* (2025), p. 17479541251334022

L. Bransen and J. Davis. "A data-driven analysis of the technical and tactical evolution of elite women's football". In: *International Journal of Sports Science & Coaching* 19.6 (2024), pp. 2449–2459

J. Davis et al. "Methodology and evaluation in sports analytics: challenges, approaches, and lessons learned". In: *Machine Learning* 113.9 (2024), pp. 6977–7010

L. Bransen et al. "Measuring soccer players' contributions to chance creation by valuing their passes". In: *Journal of Quantitative Analysis in Sports* 15.2 (2019), pp. 97–116

Book Chapters

L. Bransen et al. "How does context affect player performance in football?" In: *Football Analytics 2021: The Role of Context in Transferring Analytics to the Pitch* (2020)

J. Van Haaren et al. "Analysing performance and playing style using ball event data". In: *Football Analytics: Now and Beyond. A deep dive into the current state of advanced data analytics* (2019)

Conference Papers

J. Davis et al. “Evaluating Sports Analytics Models: Challenges, Approaches, and Lessons Learned.” In: *EBeM at the International Joint Conference on Artificial Intelligence*. 2022

L. Bransen and J. Van Haaren. “Player Chemistry: Striving for a Perfectly Balanced Soccer Team”. In: *Proceedings of the 14th MIT Sloan Sports Analytics Conference*. 2020

T. Decroos et al. “VAEP: An objective approach to valuing on-the-ball actions in soccer”. In: *International Joint Conference on Artificial Intelligence*. 2020, pp. 4696–4700

L. Bransen et al. “Choke or shine? Quantifying soccer players’ abilities to perform under mental pressure”. In: *Proceedings of the 13th MIT Sloan Sports Analytics Conference*. 2019

J. Davis et al. “Assessing the performances of soccer players”. In: *International symposium on computer science in sport*. Springer. 2019, pp. 3–10

T. Decroos et al. “Actions speak louder than goals: Valuing player actions in soccer”. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 1851–1861

Workshop Papers

L. Bransen and J. Davis. “Women’s football analyzed: Interpretable expected goals models for women”. In: *Proceedings of the AI for Sports Analytics (AISA) Workshop at the International Joint Conference on Artificial Intelligence*. 2021

L. Bransen and J. Davis. “Predicting football penalty directions using in-match performance indicators”. In: *Proceedings of the Sports Tomorrow Congress 2021*. 2021

L. Bransen and J. Van Haaren. “Measuring football players’ on-the-ball contributions from passes during games”. In: *Proceedings of the Machine Learning and Data Mining for Sports Analytics ECML/PKDD 2018 workshop*. 2018

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF COMPUTER SCIENCE
Celestijnenlaan 200A box 2402
B-3001 Leuven

