METHODOLOGY INFORMING THE ASSIGNMENT OF MATCH REFEREES IN PROFESSIONAL SOCCER

Matthew Hawkey

Thesis submitted for the fulfilment of the requirements for the degree of

Doctor of Philosophy

Victoria University, Australia
Institute for Health and Sport

November 2023

ABSTRACT

Match officials, a term used to refer to a referee or assistant referee, are responsible for enforcing the laws of the game of soccer set forth by the International Football Association Board and the International Federation of Association Football for each match domestic or international governing bodies have assigned them. Defining the performance of match officials is not clearly established; however, selecting the correct match official, on the right match, on the right day, to deliver high performance is always the objective of the assignment. Although a soccer match consists of three match officials, this project focused on the preparation, recovery, match performance, and assignment of referees. Operationalizing on the abundance of data collected from match officials yields an opportunity to inform decision making and apply a decision support system to define, understand, and forecast performance better. Current evaluation and determination of a high-performing match official are unclear; therefore, this project implemented a methodology whereby a complete view of the match official on and off the field was explored to define better the optimal data required to design a decision-making framework. Although potentially useful for the day-to-day practitioner, findings in Studies 1 and 2 would require a different approach to link into the project's overall objective. Profiling match officials, players, and teams in Study 3 provided a foundation for determining physical demands during match play. Study 4 explained further the relationship and influence of fixtures, teams, and players on match officials. Based on findings from Studies 3 and 4, the fifth and final capstone study provided an analytical framework and potential for informing the decision-making process of assigning match officials. The methodology explored in this project has potential application across sports and the areas where talent selection is required.

STUDENT’S DECLARATION

Doctor of Philosophy

“I, Matt Hawkey declare that the PhD thesis entitled

Methodology Informing the Assignment of Match Referees in Professional Soccer

is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references, and footnotes. To the best of my knowledge, this thesis contains no content of material published by any other person except where acknowledgement has been made. This thesis contains no content or material which has been accepted for the award of any other degree or qualification in any university”.

“I have conducted my research in alignment with the Australian Code for the Responsible Conduct of Research and Victoria University’s Higher Degree by Research Policy and Procedures.”

Signature: [redacted]

Date: 18-11-2023

Name: Matt Hawkey

ACKNOWLEDGMENTS

This thesis would not have been possible without several people supporting me throughout this process. While my list could extend for several pages, I sincerely thank everyone who has helped make this possible.

First, to my supervisors, Professor Sam Robertson, Dr. Grant Duthie, and Dr. Patrick Ward. Without question, the level of guidance, feedback, and expertise you have shared across not only this project but within each conversation has been sensational. I look forward to many years of collaboration and, most importantly, friendship. Sam, your patience, and ability to guide this project is deeply appreciated. While I won't partake in any hot pepper guidance, I will always say my biggest regret was the geographical difference that prohibited me from not being able to sit daily and learn from you. Your leadership in this field is the gold standard, and I am fortunate to have had the opportunity to work on this project with your guidance. Grant, the respect you showed this project and your ability to help me keep a grounded perspective and direction has been a privilege. Thank you for your ability to teach, stretch my boundaries, and provide a consistent example of how to keep things high-level and straightforward simultaneously. Patrick, I can never say enough how often I feel so fortunate to have had your local presence throughout this process. Thank you once more. The consistency of your guidance, willingness to brainstorm, and readiness to deliver constructive feedback throughout this process have been tremendous. Your ability to make me walk away from every conversation motivated me to dig deeper into each topic and helped me create a learning space that I now crave. Again, a massive thanks to the three of you, and I look forward to the years ahead.

Finally, the most enormous thanks go to my family. Holly, your love, support, and, most importantly, patience has been incredible. Your williness to sacrifice your time, ability to listen to my complaints, deal with my stress, and keep me grounded on the best path forward has helped me in many ways. I can't thank you enough. Evangeline and Lilah, thank you. My days of having you wait to play while I typed away are over, and I appreciate the motivation from both of you to keep my perspective on the importance of keeping things light. Again, thanks to the three of you. We did this together.

LIST OF SYMBOLS AND ABBREVIATIONS

°	degree
=	equals
>	greater than
<	less than
≤	less than or equal too
-	minus
%	percent
+	plus
±	plus or minus
AU	Arbitrary unit
ACM	Attacking center midfielder
BIP	Ball in play
CD	Central defender
CL	Confidence limits
CM	Central midfielder
CMJ	Counter movement jump
CR10	Category ratio anchored at the number 10
CT	Concurrent training
CV	Coefficient of variation
DM	Defensive midfielder
DSS	Decision support system
FA	Football Association
FIFA	Football
HIIT	High intensity interval training
HR	Heart rate
HSR	High speed running
Hz	Hertz
i.e.	In other words
km	Kilometer
$km \cdot h^{-1}$	Kilometers per hour

m	Meters
min	Minute
MLS	Major League Soccer
$m\text{-}min^{-1}$	Meters per minute
$m\cdot s^{-1}$	Meters per second
$m\text{-}s^{-2}$	Meters per second squared
OM	Outside midfielder
REF	Referee
RPE	Rating of perceived exertion
RPE_{load}	Rating of perceived exertion load
s	Second
sec	Second
SD	Standard deviation
sRPE	Session rating of perceived exertion
TD	Total distance
TL	Training load
VAR	Video assistant referee
WD	Wide defender
WNG	Winger

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Background

Successful performance of match officials can be attributed to making decisions fast and accurately as well as positively influencing professional soccer matches. However, the drivers for understanding and determining why high performance occurs need to be clarified. Historical definition of referee performance was focused predominantly on physical performance (Castagna et al., 2007; Krustrup et al., 2009; Mallo et al., 2009; Weston, Drust, et al., 2011). Although very little published work has defined contextualized influences on match officials, previous work has identified positioning, anticipation, and ability to adapt to the varying demands of a game as key indicators of high performance (Jiang et al., 2022; Johansen & Erikstad, 2021; Riiser et al., 2019). Since 2012, professional match officials in North America have been provided with an organizational structure that has increased overall expectations and demands for high performance. Domestic soccer federations have increased resources continually to aid in match officials' education and preparation, aiming to improve on-field performance. Consequently, increased support staff, structured fitness programs, and a formal sports science program have been employed to determine best practices regarding physical training, between-match recovery, and in-match performance.

The increased resources and support, however, have not led to implementing a decision support system—specifically, addressing the assignment problem. Consequently, the process is highly variable across domestic leagues and international competitions. Within Major League Soccer (MLS), the process includes a combination of recent subjective performance assessments, known historical relationships to teams playing in the match, and subject matter experts (SMEs) making assignment decisions based on subjective opinions or feelings. Optimization issues are also prevalent based on the volume of various data streams collected, which could all assist in understanding the demands and determinants of professional referees' performance. Such data are specific to daily wellness, airline travel, physical training, and match indicators; however, there needs to be an understanding and methodology beyond day-to-day data collection, analysis, and implementation to produce relevant applied research. Consequently, these data present a challenge in determining the link between training and preparation to explain match officials' performance. Although sports teams often follow a decision process when selecting formation, personnel, and game strategy, best practice is not clear concerning the assignment of match officials. Generally, governing bodies follow an unstructured selection criterion that can be prone to bias due to the decision being based on the experience of subjective observations about the official, which leads to a selection process missing objective classification of the appropriate match official for a specific fixture based on a systematic lack of criteria.

Based on these statements, a challenge is presented to define and determine the crucial aspects to evaluate a match official accurately. Talent identification, evaluation, and development of match officials is an important task for the game and for local, domestic, and international governing bodies. Predicting the outcome of a match officials' performance and determining the key components that drive performance requires physical and technical attributes of the individual but also of the match; therefore, a data-driven solution is in demand.

Accurately defining the criterion for high performance is necessary and will help inform future decision-making regarding assignments. Machine learning can be used widely across many disciplines, including sports. The challenge of selecting a match official for a match could benefit from describing demands and applying machine learning techniques implemented as a decision support system.

Machine learning modeling is data driven; the algorithms build models from the observed input–output data pairs (Sarker, 2021). In classical machine learning, each input–output pair is represented by a set of features and a label. These features (e.g., velocity, distance, acceleration) can be computed or measured. Features also can be engineered using logarithmic transformation of measured values or general estimating equations (Feng et al., 2014). The quality of data-driven predictive models can be estimated in computational experiments by dividing observed input–output pairs repeatedly into training and validation subsets, deriving a model from the training subset, and then applying the trained model to predict outputs for the validation subset. Because the true labels in the validation subset are known, a performance metric (e.g., accuracy or an F1 score) can estimate the usefulness of the trained model by computing the percentage of correctly predicted labels in the validation set. In situations where data labels are not available, different types of learning algorithms can be used to partition or cluster the data into distinct groups based on the similarity of their features. These approaches defined as either supervised or unsupervised learning, have been applied successfully in different domains, including sports science. These applications vary in their intended use and the type of machine learning. For instance, predictive models have been developed to identify youth talent (Barron et al., 2020; Pion et al., 2017; Sieghartsleitner et al., 2018; Zhou et al., 2020), forecast outcomes of sports games (Hubáček et al., 2019; Li, 2020; Thabtah et al., 2019), and analyze relationships between performance and training regimens (Weaving et al., 2018; Wiig et al., 2020). Utilizing match data from professional soccer matches, in this work, I applied supervised and unsupervised machine learning approaches to understand relationships and demands as well as to predict match-official and match types.

1.2 Objective

The overarching goal of this thesis was to construct a framework for an assignment tool determined by a clearly understood evaluation of referee and match characteristics operationalized on data collected from match officials. In summary, the main contributions of this thesis are:

1. To determine the extent to which training data and recovery behaviors of match officials would deem useful metrics for determining the suitability of match officials for a specific assignment.
2. To describe the physical match demands of match officials and players.
3. To develop an understanding of the relationships and similarities between match fixtures, teams, positions, players, and match officials.
4. Evaluate the utility of predicting match type and referee performance using previous match data.

1.3 Thesis Outline and Overview

To deliver these research aims, the remainder of this thesis is organized as follows. Chapter 2 provides an overview of the relevant research with a focus on match officials. In Chapter 3, Study 1 provides a detailed quantification of training and match demands across a season. Volume, intensity, and training types were classified into three categories: matches, physical training, and rest. Ratings of Perceived Exertion (sRPE, CR10 scale) and duration were collected to provide RPE_{load}. Effects of session type, weekday, and time of season on RPE_{load} were analyzed using magnitude-based inferences. RPE_{load} was greatest for Matches, with large differences compared to Speed, Repeated Sprints, Endurance, and Strength sessions, and moderate differences were found when compared to High-Intensity Intervals and any session where both strength and a running session were performed concurrently. Findings supported the logical conclusion that RPE_{load} was highest for Matches and the distribution of RPE_{load} was relatively consistent across training weeks and mesocycles. Although this study provided reasoning for training decisions with regard to volume, intensity, and type, consistencies across the season, regardless of referee, would not suggest a supporting tool for decision making with regard to assignment.

Study 2 in Chapter 4 aimed to determine the effects of RPE_{load} on match officials' morning wellness scores. Findings identified limited correlation between RPE_{load} and morning wellness scores. RPE_{load} does not appear to have a substantial effect on the perceptual daily wellness scores of referees. Although RPE_{load} was greatest on Tuesdays and Saturdays, wellness scores remained relatively stable throughout the week and similarly across mesocycles. When comparing within subjects, small to moderate correlations were found when comparing RPE_{load} and overall wellness, tiredness, energy, and sleep quality. These results suggest prior-day training load may influence the individual responses referees have for wellness on the subsequent day.

Although a daily practitioner could use these findings to inform daily decisions on training, preparation, or recovery strategies, the link to understanding match performance was unclear and required in-game objective measures to guide assigning decisions.

With the intention of establishing match profiles in Chapter 5, Study 3 presented a detailed description of the physical match activity profiles of match officials and players where the ball was in play and across whole-match play. Accounting for both ball in play (BIP) and ball out of play, defensive midfielders recorded the highest mean speed at 236.5 m-min^{-1} for the 1-min window; central defenders and central referees indicated 205 m-min^{-1} peak demands across the same timeframe. BIP was 48 secs, 71% of all team possessions were 60 > s, mean maximum speed for the 0–15 s BIP window was 5.95 m-s^{-2}, mean max speeds dropped to 3.9 m-s^{-2} for the 60 s window, and player and match official BIP total distance covered ranged from 7,676 m to 9,945

m. Defensive midfielders covered the most distance, and central forwards covered the least. Central forwards and attacking central midfielders showed the highest mean 5-sec BIP speeds of 8.26 and 8.24 m-s^{-2}, respectively. Compared to other domestic leagues, these findings present higher mean max speed distances yet similar degradation across similar windows. These summaries and findings presented a foundation to build specific profiles and understand physical performances of match officials more clearly.

In Chapter 6 and 7, the final two studies investigated the relationship between match officials and match characteristics. Study 4 used unsupervised analysis techniques to initialize an integrated approach to explaining referee physical performance. Implementing both K-means and hierarchical clustering, I conducted three separate analyses: (a) physical outputs only by club and referee, (b) physical outputs by position, and (c) a combination using both event data and physical data. Highlighting the importance of determining appropriate selection of machine learning techniques, similarities between two separate clustering techniques were found when only comparing physical data; however, disagreements occurred when integrating the physical and technical data sets. Application of this approach could be deemed useful when explaining performance from a within and across cluster perspective. Determining the influence of players, clubs, or match fixtures on how a match referee performs physically led to the final paper where both unsupervised and supervised machine learning were applied. Findings included classification accuracy at 0.79 or higher for accuracy as well as balanced accuracy and F1 scores for all supervised learning techniques. The framework was implemented with the intention of creating a match assignment tool. Lastly, Chapter 8 provides a summary of findings, potential future projects, and suggestions for practical applications of using a decision support system.

CHAPTER 2: LITERATURE REVIEW

2.1 Match Performance

2.1.1 Referee Movement

The movement of soccer officials is a vital component of their roles. Characteristics of movement of referees as they move around the field have been defined as linear, diagonal, and often times non-linear (International Football Association Board [IFAB], 2020). As early as 1891 the International Federation of Association Football (FIFA) suggested the ideal system of movement to ensured optimal views of the ball and the players was on a diagonal pathway. This recommendation advised Referees to move in a diagonal fashion across the field typically along a path from corner to corner putting the action along their right side. However, depending on the action of the game, a referee may also take linear pathway. Along diagonal and opposite sidelines, two linesman or more recently referenced as assistant referees, take position. Utilizing the markings of the field, the assistant referees move between the endline and halfline. With an understanding of the guidance to match officials, physical demands must consider these directives.

During competitive soccer games, total distance comprises a majority focus in previous research on physical demands of match officials. Despite large age differences of over 10 years, senior match officials must keep up with younger players (Weston et al., 2010). Average distance covered by referees during a competitive game was shown to be 9.4 km (Catterall et al., 1993). More than nine years later, (Helsen and Bultynck (2004) reported total distances of 10.2 ± 0.90 km covered by the referee. Mallo et al. (2009) suggested no differences between the first and second half and reported total distance covered during matches was 10,218 ± 643 m for international match play. While domestic fixtures presented greater distances covered at 11 km when reported by Weston et al. (2004).

During an entire season across 254 games in the English Premier League (EPL), Weston et al. (2004) implemented an analysis on data from professional referee's match performance and focused on match official's running performance at high-speed. To match the demands of play, a regular occurrence of short, explosive bouts classified as high-intensity were demonstrated by match officials. Weston, Drust, et al. (2011) examined the variability from match-to-match mobility of referees officiating in professional leagues in England from 2003 to 2008. Variations between-matches (CV) were very high the number of fouls within the match, sprint count, high intensity sprints, recovery time between actions, and distance. Top speed, distance to both fouls and the ball, along with the total distance covered regardless of speed presented smaller CVs. Speed-related categories reflecting a percentage of total distance were reported to determine physical demands imposed on soccer referees (Castagna et al., 2007). Distances at low velocities (< 4 m·s^{-1}) comprises many actions performed during matches (Krustrup & Bangsbo, 2001); however, similar to players, high-speed activities have been reported as an indicator of physical demand (Krustrup & Bangsbo, 2001; Mallo et al., 2007). When comparing studies, high-speed running distance is difficult to determine due to the variability of reported thresholds. A threshold range of 3.6 to 5.5 m·s^{-1} has been reported in the literature; consequently, high-intensity running, presented as a total percentage of distance covered, varies from 7% to 17% in referees (Krustrup & Bangsbo, 2001; Mallo et al., 2007; Mallo et al., 2009; Weston, Drust, & Gregson, 2011). Differences in the methodologies the authors used to classify physical demands make direct comparisons between previous reported physical demands no possible. Consistently, arbitrary speed thresholds fail to determine individualized match activities or physical capacity relative to each referee's movement speeds or fitness capacity (Krustrup & Bangsbo, 2001). Weston et al. (2007) proposed match speed thresholds are classified according to the individual referee's physical capabilities. Given the high variability across the literature, presenting demands as percentages of individual capabilities would allow future work to achieve higher accuracy and consistency when determining physical demands.

While the referee's positioning during match play, specifically concerning distance from foul decisions, is noted as a required technical component of match officials' research, limited research exists. Most examinations of match officials' ball-tracking ability are directed at the physical ability to match the demands of match play. They are not focused on the influence of positioning on correct or incorrect decisions (Elsworthy et al., 2014; Gilis et al., 2008; Oudejans et al., 2000). Early research on referee positioning from Krustrup and Bangsbo (2001) suggested that distance from play hinders the referee's ability to view the entirety of a game situation when too close or being susceptible to error when too far away. Consequently, the distance from the match official to the game event could impact decision quality and the game's result.

When examining the location on the playing field, Krustrup et al. (2002) observed variability in distances from the match event and the referee in the center of the field of play; however, increased distances were noted when evaluating attacking areas in the lateral parts of the field. The attacking parts of the field present critical challenges when incidents are obstructed by players or simply because of the referee distance to the incident. Contextual components that also may impact referees' positioning and subsequent decision-making during an incident include score, weather, time of the match, or the fitness capability of the match official.

2.1.2 Player Movement

Players movement characteristics in match play have been shown to influence the movement patterns of match officials (Weston et al., 2011). Therefore, it is important to understand and explore players movement characteristics. Studies examining the application of computerized tracking technology from multiple cameras have quantified the physical demands of soccer. Historically, distance covered is the most analyzed and reported metric. A complete examination of performance metrics by position by Di Salvo et al. (2007) looked at two seasons, three hundred players, and 30 games from the top division of Spain, along with Champions League games captured from a multi-camera system. Over 90 minutes of match play time and regardless of position, the mean total distance was 11,393 ± 1016 m (Di Salvo et al., 2007). Further examination by position presented that central and wide midfielders covered more distances than defender groups and attackers (12,027 m and 11,990 m compared to 10,627 m, 11,410 m, and 11,254 m). More recently, these findings were supported by midfielders covering the highest total distances (Bojkowski et al., 2015; Bradley et al., 2009; Duk et al., 2011; Mallo et al., 2015). A clear understanding of positional categorization is critical when comparing physical performance analysis. When placed into the same categorization, Bradley et al. (2009) reported that an EPL midfielder's mean distance was 11,459 m during a 90-min game; however, Mohr et al. (2003) and Rampinini et al. (2007) recorded outside or wide defenders and attackers as the positions covering the most significant distance.

Examination of high-speed running, change of direction, and explosiveness in relation to playing positions have also been reported. Similar to total distance, Di Salvo et al. (2007) reported that central midfielders ran the most significant distances at a high-speed band of 11.1–19 km·h-1. Further analysis of sprint performance by Di Salvo et al. (2010) presented differences by position during Champions League play. A multiple-camera system captured physical performance from four seasons and 67 matches. This analysis of 717 outfield players focused on sprinting actions (>25.2 km·h-1). Overall sprint count from outside midfielders was 35.8 ± 13.4, while attackers, outside defenders, and central midfielders performed 30.0 ± 12.0, 29.5 ± 11.7, and 29.5 ± 11.7, respectively. Interestingly, central defenders performed significantly less at 17.3 ± 8.7. Findings from three domestic seasons in the English Premier League presented similar findings, with outside midfielders performing the highest number of sprints (41 ± 13), central defenders the least (20 ± 9), outside defenders and attacking players similar (34 ± 13, 34 ± 12) while central midfielders were slightly less (30 ± 13). To highlight differences between positional profiles, high-speed running, and total distances covered were also demonstrated (Currell & Jeukendrup, 2008). Furthermore, these findings agree with previous findings (Bradley et al., 2009; Di Salvo et al., 2007; Leventer et al., 2016). Reilly and Gilbourne (2003) indicated a link between playing position and sprinting performance. Furthermore, high-speed running profiles have shown a differentiation in player capacity and level of play (Abrantes et al., 2004; Bangsbo, 2014; Bangsbo et al., 1991; Mohr et al., 2003). Utilizing camera-based tracking systems over seven seasons in the EPL, Barnes et al. (2014) presented an increase in high-speed running and sprint distance (890 ± 299 m in 2006–2007 to 1151 ± 337 m in 2012–2013; 232 ± 114 m in 2006–2007 to 350 ± 139 m in 2012–2013, respectively). This longitudinal study, while only profiling one domestic league, portrays key increases in demands of players and furthermore requires practitioners and coaches to adjust training, recovery, and scheduling.

A multi-camera time motion system captured and quantified movement patterns by position in the EPL (Bloomfield et al., 2007). Forward running, backward peddling, lateral shuffling, and diagonal movements were notated. Midfielders were highest for forward and diagonal movements, while defenders demonstrated the highest back and lateral or side-to-side actions. Positional demands reflect the frequency of specific actions, such as turning, creating space, changing directions, or dodging opponents in short time spans due to the nature of unpredictability in the sport (Nicholas et al., 2000; Wragg et al., 2000). Supporting these findings, Akenhead et al. (2016) found that the backward actions of outside defenders are higher than those of central defenders, portraying notable differences even among a similar position type. Interestingly, while the mean total turns in a match was 727 ± 203, Bloomfield et al. (2007) reported a dissimilarity between higher total distance and the number of turns from midfielders. Bloomfield et al. (2007) also presented key takeaway physical training concepts for practitioners based on these turning data. They suggested defenders and attackers utilize change of direction and explosive action type exercises, while high-speed running with intervals would be beneficial Bloomfield et al. (2007). Performance metrics such as high-speed running, changing direction, or count and types of turns allow for developing a profile of players and portray positional requirements. While essential to match analysis, interpretation of the validity of these demands is limited, and only recently have these internal and external data been made during elite-level play.

Maintaining a high level of physical and technical performance throughout a soccer game is a key indicator of the successful physical development of an elite soccer player. Overseeing the mechanisms of decline in fatigue levels, customarily defined as monitoring, has been researched extensively to further understand and assess performance degradation during a match. Utilizing high-intensity actions such as accelerations or decelerations has declined over the first and second halves (9 ± 8%, respectively). When comparing the start to each half, Mohr et al. (2003) indicated that the initial 5 mins of the first half presented higher performance than the comparable period to start the second half. Additionally, Mohr et al. (2003) analyzed 18 players throughout a season and determined a second-half decline in high-speed running. Furthermore, Barnes et al., 2014, showed an increase in capacity by indicating an increase in acceleration production over several seasons. This study further described physical threshold increases in metrics such as total distance, count, or frequency, along with high-speed running volume. Key points from these studies require a more focused examination of the physical requirements and subsequent development of physical characteristics of elite soccer players.

2.1.3 Match Performance Indicators

A performance indicator refers to a set of variables, features, or metrics combined to define high and low performance (Hughes & Bartlett, 2002). When defining the ultimate performance profile at either the individual or team level, these indicators aid in predicting future performances (O'Donoghue, 2005). Through research, the characteristics of successful teams are often imitated within domestic competitions as they help to determine high-performing teams or players (Hughes & Bartlett, 2002). Relationships between performance indicators from either a technical, tactical, or combined definition are often determining influences of successful teams (Harrop & Nevill, 2014; Lago-Ballesteros & Lago-Peñas, 2010; Liu, Gómez, et al., 2016; Liu, Hopkins, & Gómez, 2016). Early research on determining winning teams indicated the non-physical components of the game, such as technical capabilities of the players involved in the match or team tactical strategies devised by coaches, were more useful (Di Salvo et al., 2009); however, more recently, winning a match vs. losing was determined by physical aspects irrespective of the ball location, such as sprints, and high-speed running (Aquino, Martins, et al., 2017; Aquino, Vieira, et al., 2017).

Context of the match, including match location (e.g., home, away, or neutral), opponent quality describe by current positioning or ranking in the table, and match status described as the standing of both teams involved in the match, and historical reputation of the match have been shown to be the main factors influencing performance (i.e., physical, player technical or team tactical) and even the outcome during the competition (Aquino, Martins, et al., 2017; Lago et al., 2010; B. J. Taylor et al., 2010; J. B. Taylor et al., 2008).

2.1.4 Referee Decision Making

During the Confederations Cup 2009, Mallo et al. (2012) analyzed mobility's possible effects on professional referees' decision-making. Distance from play of 380 fouls, captured from 15 matches, were analyzed. Notation consisted of two main categories further divided into two components: lateral or central area of the field and distance from play separated arbitrarily into 5-meter groups. With an average of 25 fouls per game, Mallo et al.'s results were summarized with the following: the lowest error rate for foul detection occurred from distances in the 11-15m group, central parts of the field contained three-quarters of events while wide areas were more closely related to the positioning of the assistant referee, the average distance from the match official to the event was 17 m and interestingly no distinct differences for call accuracy when considering the distance, while distance did not present changes between halves incorrect decisions did increase from 8 to 17%; similarly central areas also demonstrated an increase, and lastly accuracy was impacted negatively the further from the event. These findings from Mallo et al., reference a physical distance to the play. This distance would reflect the physical demands of the match, as the referee would to need to move according to the flow of play and achieve these distances. Decision accuracy as referenced above (Elsworthy et al., 2014; Gilis et al., 2008; Oudejans et al., 2000), obtaining ideal positioning would reflect the physical demands required to achieve. Although this research is vital to understanding decision making of match officials, the subjective nature of making correct versus incorrect decisions needs further research. Anecdotal experiences among those involved would lend an argument to how these correct versus incorrect decisions are evaluated. Post-incident review—Video Assistant Referee (VAR)—has become commonplace in the sport, and further research could offer a more objective view of the best outcome for all decisions made by match officials.

2.2 Referee Physical Training and Response to Training

2.2.1 Physical Training

Webb et al. (2016) found a focused increase on fitness training for match officials was potentially explained by the requirement for match officials to obtain good positioning for key decisions. Further, referees' movement has shown similarities in high-speed and sprinting distances covered correlated with players' similar efforts; at the same time, referees' total distance volume was greater (Weston et al., 2010). Analyses of the training loads of other team sport athletes have been researched extensively (J. J. Malone et al., 2015; Oliveira et al., 2019; Ritchie et al., 2016); however, despite this increased focus, knowledge of elite soccer referees' fitness training loads is limited to a small number of training studies (Krustrup & Bangsbo, 2001; Weston et al., 2004), deliberate practice observations (Catteeuw et al., 2009; MacMahon et al., 2007), and a comprehensive account of the training loads undertaken by a single referee (Weston, Gregson, et al., 2011). Moreover, research on referees' training practices demonstrated most training sessions were physical (Weston et al., 2012), yet quantification of in-season physical loads has eluded the literature.

2.2.2 Response to Training

Training or match load has historically been monitored using internal or exercise-induced adaptation or relative stress imposed and external measures expressed as the actual work performed by athletes (Impellizeri et al., 2004). However, published literature has suggested a multitude of approaches and tools to monitor fatigue, readiness, and the current status of athletes (Saw et al., 2015b). Current status captured from self-report tools in various sports such as Australian football (Gallo et al., 2016), rugby (Gastin, Meyer, & Robinson, 2013), and soccer have been examined (Moalla et al., 2016; Thorpe et al., 2015). Further, self-reported questionnaires capturing wellness in response to training or matches have also been investigated (Gallo et al., 2016). While challenging to conduct daily, objective performance measures (e.g., countermovement jump, intermittent recovery test) of players are the ultimate indicator of performance (Currell & Jeukendrup, 2008). The challenge to practitioners, therefore, requires an examination of different ways to monitor responses to training, matches, and seasonal variations in demand. Research on match officials' response to training, matches, or longitudinal monitoring is minimal, and, similar to players, the selection of which objective and subjective measures are most beneficial for referee monitoring remains ambiguous. Furthermore, when analyzing internal loads such as RPE, when compared to external loads, the dose-response relationship becomes more unclear due to a lack of variability and sensitivity (Weston et al., 2011).

Monitoring athletes' fatigue and over-training indicators has commonly been conducted via self-reported questionnaires (K.-L. Taylor, 2012). Highlighted by the simplicity of collecting valuable insights, these noninvasive tools are often implemented in a team sport setting. However, due to the length of questions, the time requirement has been reported to hinder daily collection (Twist et al., 2012). The impractical time demand has led to the creation of modified, shortened questionnaires to aid practitioners in monitoring daily well-being (Buchheit et al., 2013; Coutts & Reaburn, 2008; Gastin, Meyer, & Robinson, 2013; Hooper et al., 1995; Saw et al., 2015a). The modified self-report monitoring questionnaires have focused on limiting the number of questions to a maximum of 5. Typical questions selected ratings that measured sleep, stress, fatigue, and soreness while focusing on the overall goal of understanding what is impacting performance.

Within the literature, subjective rating of perceived exertion (RPE) is a commonly reported tool that measures exercise intensity in team and individual sports and rehabilitation settings. Furthermore, RPE has been evaluated to investigate managing training loads acutely and over long time periods (Eston, 2012; R. Robertson & Noble, 1997). RPE measures are often coupled with heart rate responses because of the similar indices of physical exertion being performed during exercise (Borg, 1982). Further suggestions from Borg (1982) simplified the rating to 10 (CR10), allowing for easier comprehension for the user. The implementation of session RPE (sRPE) provides subjective training load responses whether collected in team or individual-based sport. Modification of wellness questionnaires and the relationship of responses to training or games in athletes at the elite level is limited yet has been discussed in the EPL and AFL (Buchheit et al., 2013; Gallo et al., 2016; Gastin, Fahrner, et al., 2013; Haddad et al., 2013; Moalla et al., 2016; P. G. Montgomery & Hopkins, 2013; Thorpe et al., 2015).

2.3 Referee Testing

FIFA Fitness Test

Match officials must successfully complete the FIFA fitness test to establish eligibility in a FIFA-regulated competition. The FIFA fitness tests include two separate measures of referee abilities. The first test is a repeated sprint test of 40 m with rest interval 90 sec between, followed by a series of 75 m high-speed runs interspersed with 25 m of walking. The timing of these runs is 15 seconds and 18 seconds, respectively. Notably, Cerqueira (2011) determined that fitness test outcomes do not accurately reflect match demands. Weston et al. (2004) and later Mallo et al. (2007) found that FIFA fitness test outcomes did not correlate with movement demands on soccer officials during matches. Recognizing the high rate of activity changes and change of direction (COD) actions presented by Krustrup and Bangsbo (2001), Castagna et al. (2011) examined fitness tests and the relationship to physical match demands of officials and suggested a component of the fitness test should incorporate a measure of COD. Moreover, performance outcomes in the 50 m sprint and 200 m high-intensity portions of the test correlated poorly with the total distance demands from matches. Similarly, Mallo et al. (2009) examined cardiovascular test results of elite match officials and demonstrated a low correlation with heart-rate results and total distance or high-intensity distances. Results demonstrated elite match-officials display fatigue at variable time points throughout a game however were able to successfully complete the FIFA test. In order to improve the prediction of physical fitness capacities and enhance performance from both a quality of decision quality and physical perspective, amendments to FIFA fitness test protocols are required. Current outcomes are not deemed valid measures or reflect physical performance within the match, and further research is required to simulate match officials' movement and positioning within a game (Cerqueira et al., 2011).

2.4 Referee Demographic Profile

Age and Experience

While often highlighted as an essential component of referee performance, limited research exists on the relationship between the quality of decisions quality and years of experience. The performance of 22 officials and the effect of experience on physical attributes were explored by Weston et al. (2010). Despite an increase in referee age, a reduction in physical performance did not appear to impact the requirements to maintain an optimal distance from play. Notably, there were fewer explosive actions, less high-speed running, and less total distance, yet the more experienced match officials displayed similar distances from play as their younger counterparts. Referee experience potentially enables ideal positional strategies by becoming accustomed to situational movements from players.

Beyond the physical performances of match officials, an exploration of 59 match officials' decision-making variables, such as the number of incidents or fouls, was presented by Weston, Drust, et al. (2011). Match official experience was not indicative of variability of performance across matches. A positive correlation was presented by Catteeuw et al. (2009) between officiating techniques, years of officiating games, and weekly time spent training decision-making.

2.5 Referee Travel

Jet Lag and Performance

Referee travel has yet to be investigated, yet travel could impact physical and cognitive performance based on the concept that the referee is always the “away team”. Typical match assignments for officials included in this project required 10,000-15,000 miles per month, often crossing multiple time zones upward of 30 times in the same period. With jet lag a constant issue that match officials need to deal with in the profession the influence on performance perhaps decision-making ability should be discounted. Jet lag has been shown to contribute to a reduction in performance (Antal, 1975; Forbes-Robertson et al., 2012; Lee & Galvez, 2012; Loat & Rhoades, 1989; Samuels, 2012; Winget et al., 1985). Chapman et al. (2012) investigated and reported on skeleton athletes’ variation in jump performance post-travel. After an initial deterioration in peak velocity, mean velocity, and jump power, an increase was reported 3 days post travel. These findings may suggest airline travel affects jump performance and neuromuscular control over the first 1–2 days after travel. Performance degradation following travel crossing single or multiple time zones has been illustrated within soccer (Fowler et al., 2017; Lastella et al., 2019). Throughout 10 Major League Baseball seasons, data were analyzed to determine the effect of travel on physical performance attributes (Winter et al., 2009). Teams with a 3-hour advantage won more games than teams with 1- and 2-hour advantages over their opponents. It appears crossing more time zones correlates with diminished performance.

Many studies have shown travelers flying eastbound tend to experience more marked symptoms of jet lag that persist longer, requiring lengthier time for resynchronization, than those of westbound travelers, due to the body’s ability to adjust by phase delay more rapidly (Ayala et al., 2021; Fowler et al., 2017; Loat & Rhoades, 1989; Reilly & Edwards, 2007; Reilly et al., 2009). Following prior studies, Lemmer et al. (2002) found jet lag symptoms after westbound flights were most pronounced through the first 3 days postflight, and symptoms after eastbound flights were more severe and persisted up to 7 days after arrival.

2.6 Ecological Dynamics

2.6.1 Ecological Dynamics and Match Officials

While momentum in the application of ecological dynamics has been witnessed across sports, the application of ecological dynamics to match officials' physical or decision-making performance within the literature is limited. Dynamic systems theory presents that systems comprise multiple levels of analysis that diversify across time or space. Match officials operating in a match or training environment are congruent with select elements of this theory based on the fundamental principle that they must continually adapt and organize their positioning to the situational movements of players and contextual components of a match, with the sole intent of optimizing for correction decisions (Davids et al., 2015).

Dynamic systems theory provides for a complex level of analysis, including the process by which players and match officials are organized in a game. From a match official perspective, the spatial properties of opposing teams, the interaction of players within the playing space, the physical qualities of players' speed or explosiveness, and decision-making characteristics of specific positions or players all interact to drive the positioning behaviors of match officials. Furthermore, the coordination of the team of officials, including the assistant referees and 4th officials, all interact on the constraints within the observed physical and decision performances of the referee (Davids et al., 2015).

Explaining how behavior occurs is explained best by the principles of constraints within the ecological dynamics systems theory (Araújo et al., 2006; Araújo et al., 2010; Davids & Araújo, 2013; Davids et al., 2012; Davids et al., 2015; Dicks et al., 2010; Renshaw et al., 2010; Travassos et al., 2012). Performance adaptations, displayed through physical movement or decision-making of the match official, are contained non-linearly and dynamically by individual skill, physiological mechanisms, and environmental conditions on a consistent situational basis. The theory of affordance further influences the relationship between movement, positioning, and decision-making. Dissimilar to constraints, an individual's physical profile or the situational environment is not described as affordance. However, affordance is identified as moments provoked by actions and decisions of the match that incite the physical behavior of the performer (Jones, 2003). The reciprocated interaction between the performer and the environment is viewed as a guide to positioning and movement, with the convergence of the two shaping the end result. Three constraints are classified as the task (application of rules of the match, positioning for best view angle, reaction to team coordination and player behaviors), environment (temperature, wind, humidity, media exposure and expectations, playing surface, match participants including the match official, spatiotemporal characteristics, match score, and match importance), and individual (height, weight, physiological capacities, motivation, age, and experience). (Araújo et al., 2006; Araújo et al., 2007; Araújo et al., 2010). Constraints are constant, and awareness of the perturbations in optimal movement or decisions must be considered when evaluating the match official's performance. Minimal application of contextual evaluation of match official performance conducted in the form of ecological dynamics is currently found within the literature.

2.6.2 Environmental Factors

A simple definition of crowd factors may influence referees' decisions indirectly. To understand the influence of the home team crowd on decision-making, Nevill et al. (2002) asked English Premier League match officials to evaluate tackles involving players from both the home and away teams. Time after each tackle was allotted for the official to determine and note if an infraction occurred and, if necessary, which team the foul should have been given. Furthermore, the match officials were divided into two groups: one that listened to stadium noise and a second that made decisions sans noise. Notably, the match officials working with noise in the background awarded more decisions to the visiting team. This work was further explored by Balmer et al. (2007), looking at the consistency of referee decisions and the influence of crowd noise.

Notwithstanding crowd noise, stadiums and team playing strength acquired fewer fouls or decisions than their opponents. The literature demonstrates that host teams are more successful in the win column, receive fewer yellow and red cards, and see fewer free kicks (Balmer et al., 2007; Nevill et al., 1997; Nevill et al., 1996; Nevill et al., 2002). Throughout five international club seasons, Dawson et al. (2010) supported the home team favoritism by demonstrating increased count of decisions against the visiting team.

2.6.3 Decision Support Systems

A simple definition for a decision support system (DSS), as Janakiraman et al. (2008) provided, is "an interactive, computer-based system which supports managers in making unstructured decisions" (p. 26). Mathematical models to determine optimal referee assignment have been researched (Alarcón et al., 2014; Mancini & Isabello, 2014) and more recently (Alarcón et al., 2017; Durán, 2021; Durán et al., 2022; Linfati et al., 2019). These approaches evaluated the assignments from a balance of perspective and did not use performance variables of the individual referee. Although computers are exceptionally good at making many decisions rapidly and accurately, some cases exist where the steps required to reach a decision cannot be quantified in any manner a computer can understand (Janakiraman et al., 2008). DSSs have been developed to tackle these unstructured decisions. In the case of referee assignment using performance metrics, unstructured decisions would be difficult for computers to handle because they do not have a set sequence of events that must be undertaken to reach the goal. Schelling and Robertson (2020) proposed a detailed framework for the evaluation and implementation of a DSS within a sporting environment. Instead of trying to replace human decision makers, use of this framework would, for the purposes of assignment, suggest a DSS take on a supplemental role by providing users with information and perspectives that improve their decision-making ability. Turban et al. (2004) provided a complete definition for a DSS:

> A DSS is an interactive, flexible, and adaptable computer-based information system that utilized decision rules, models, and a model base coupled with a comprehensive database and the decision maker's insights, leading to a specific, implementable decision in solving problems that would not be amenable to management science models per se. Thus, a DSS supports complex decision making and increases its effectiveness. (p. 68)

This definition is significantly more precise and hints at some important characteristics of DSSs. A DSS relies upon the information it contains, but its application as a tool for content experts to use has not been investigated. Most often, decision makers in sports, and specifically within domestic professional association or federations tasked with match assignment, are content experts and have a great deal of hard-earned knowledge that must be used (Fry & Ohlmann, 2012). A DSS that takes control and forces decision makers down a path without allowing expert input is not achieving its goal. A DSS can be used in many environments to solve various problems, ranging from military applications (Ben-Bassat & Freedy, 1982; Riedel & Pitz, 1986) to financial and business-oriented software (Janakiraman et al., 2008). One of the most popular application areas is medicine (Bahill et al., 1995; A. A. Montgomery et al., 2000; Raghavan et al., 2005). Some DSSs are used in environments allowing for the use of heuristics to reach a solution quickly, which may not be optimal but is generally good enough.

Similar to sports where potential contextual factors influence the decision process, medical applications also often have a time constraint, where answers must be made with a high degree of accuracy and as quickly as possible (Raghavan et al., 2005). Other medical DSSs focus on optimizing use of expensive or limited resources (e.g., medications, equipment; A. A. Montgomery et al., 2000). If the context of the DSS's use influences the tool's design, then this influence also should be reflected in how the tool is evaluated. Further, recommendation systems fit into the DSS concept and are used widely on the internet to help customers find products or services to fit their preferences. Current implementations reduce the amount of information available and generate a personalized suggestion of objects (e.g., books, movies). However, thus far, limited findings have occurred in the sports arena.

Using recommendation systems for recruitment and athlete selection is limited. Since Resnick and Varian (1997) established the term "recommender system," researchers have improved the quality and scalability by various means. Researchers have evaluated content-based filtering, where a system suggests items familiar to those preferred or liked in the past (Melville et al., 2002); a collaborative system suggests a user item people with similar preferences liked in the past (Ungar & Foster, 2000). Each system has its strengths and weaknesses, whereas a hybrid system compensates for the shortcomings of both approaches.

DSSs have been designed and developed mainly for the business sector, depending on the company's or organization's perspectives or stakeholder objectives and needs; however, more recent traction has seen the proposal and development of DSSs in the sports arena. S. Robertson et al. (2016) explored the use of a DSS in sports and suggested implementing a commonly used traffic light system to guide decision making. Schelling et al. (2021) proposed a DSS to address the problem of scheduling to maximize health and performance in a league or organization. The operating environment can significantly influence how a DSS is used and how it should be built. Gong and Ling (2009) presented a study to improve athlete performance using a DSS with a multi-agent system for table tennis players to analyze the psychological structure of players and scientifically strengthen their psychological adjustment for them. In comparison, Kostuk and Willoughby (2012) developed a DSS that can produce several schedule versions for the Canadian Football League instead of manually creating the league schedule.

CHAPTER 3: DEFINING THE REQUIREMENTS OF MATCH OFFICIALS WITH RESPECT TO PHYSICAL TRAINING

Quantification of Referee Training Loads Across a Season

3.1 Abstract

The purpose of this study was to provide a detailed quantification of a multidimensional training program for professional soccer referees' training loads. To understand the total physical demands from a fitness prescription implemented across a season, total volume, intensity, and specific training types of 18 soccer referees from Major League Soccer were observed during this season-long study. Three main categories were used to quantify activity: (a) matches, (b) physical training, and (c) rest. To understand referees' physical training loads further, sessions were divided into six categories, with four categories representing days when referees performed concurrent training. Following each session, referees reported session Ratings of Perceived Exertion (sRPE, CR10 scale), with this score multiplied by duration to provide RPE_{load}. Effects of session type, weekday, and time of season on RPE_{load} were analyzed using magnitude-based inferences. RPE_{load} was greatest for Matches, with large differences compared to Speed, Repeated Sprints, Endurance, and Strength (range 314 to 413 AU) and moderate differences compared to High-Intensity Intervals and all Concurrent Training (CT) sessions (154 to 251 AU). Using the regular match day (i.e., Saturday) as reference, there was a small difference in RPE_{load} compared to Sunday (-123 AU; 90% CL ± 23 AU) with trivial differences for all other days (range -54 to 23 AU). RPE_{load} was greatest in Mesocycle 5, with small differences compared to all other mesocycles (148 AU to 272 AU). In conclusion, RPE_{load} was highest for Matches, and distribution of RPE_{load} was relatively consistent across training weeks and mesocycles. As a result, when using RPE_{load} to understand

referee training loads, reference to match days should assist in the planning and prescription of specific physical typologies.

3.2 Introduction

Soccer refereeing is a physically demanding activity with referees covering more total distance and a similar amount of high intensity running compared to players (Weston, Drust, & Gregson, 2011). Such a high match physical demand requires appropriate fitness to ensure referees can keep up with play and obtain good positions when making key decisions (Weston, 2015), which somewhat explains why the area of referee training that has seen the most rapid development is fitness (Webb et al., 2016). However, although extensive analyses on the training loads of other team-sport athletes have been published (J. J. Malone et al., 2015; Ritchie et al., 2016), knowledge of elite soccer referees' fitness training loads has been limited to a small number of training studies (Krustrup & Bangsbo, 2001; Weston et al., 2004), deliberate practice observations (Catteeuw et al., 2009; MacMahon et al., 2007), and a comprehensive account of training loads undertaken by a single referee (Weston, Gregson, et al., 2011). Therefore, a detailed quantification of referee training loads has yet to be undertaken.

The most vital aspect of referees' match performance is decision making (Helsen & Bultynck, 2004), yet research into this area has been very limited (Weston et al., 2012; Weston, 2015). Although not tied to this project, this aspect of performance should not be ignored within the context of this work or future investigations of referee performance. Improvements in refereeing performance and decision making have been traced to the employment of full-time referees who have embraced the advice provided to them from an enhanced level of support, including sports scientists (Webb et al., 2016). Major League Soccer (MLS) in North America followed countries such as England and Italy and employed referees on a full-time basis, providing them with comprehensive sports science support. Despite an established support program in place, however, a detailed account of referee training practices could help inform and advance prescription of training type and volume as well as improve referee decision making.

Given the high volume yet inconsistent schedule of matches and physical training sessions performed each season (MacMahon et al., 2007; Weston, Drust, & Gregson, 2011), Weston (2015) recommended practitioners pay close attention to the monitoring and prescription of referees' physical training loads. This program should be delivered with the intention to achieve high fitness levels and reduce fatigue by manipulating the intensity and volume of training across weeks, months, and seasons. Player training has been shown to include a variety of training components to address the strength and energy system demands of a professional match (Ritchie et al., 2016); however, to date, no training analysis of referees exists for North American soccer match officials. The intention here was to provide a detailed analysis of MLS referees' training loads and the type of training across their first season as full-time employed referees to advance an understanding of elite soccer referees' physical training.

3.3 Methods

To examine the influence of activity type, training day, and time of season on referees' internal training loads, the study used a single group, observational design. Referees' internal training loads were measured using the session RPE method (Foster et al., 2001). Here, using the CR10 scale, all referees provided a session RPE for each match and training session. This score was multiplied by session duration to obtain the RPE load (RPE_{load}) score, a composite score reflecting both exercise intensity and duration. Session RPE data were entered within 30 mins after each related match or training session through VisualCoaching Pro (Australia).

3.3.1 Subjects

Data were collected from 18 North American match officials (mean age ± SD 35.9 ± 9.9 years) from within MLS, with the data collected as part of the ongoing sports science support provided to match officials. MLS is the highest level of competition played in North America, and the data reflected one season of competition.

3.3.2 Procedures

Three main categories were used to quantify referees' activities: (a) matches, (b) physical training, and (c) rest. Match-related activities included any assignment as match official or involvement in the game as the fourth official. Rest was recorded when referees were assigned a day off or complete rest, and recovery was marked when an activity was performed with the sole purpose of aiding the recovery process. To match descriptions of the fitness prescription used within the organization while also further understanding referees' training loads, physical sessions were divided into six distinct categories (i.e., Match, Speed, Repeated Sprints, High Intensity Intervals, Endurance, Strength) with four additional categories representing days when referees performed concurrent training (CT; Speed + Strength [CT1], Repeated Sprints + Strength [CT2]), High Intensity Intervals + Strength [CT3], Endurance + Strength [CT4]). With daily prescription delivered by the sport science support staff, training schedules were determined by the upcoming match assignment; therefore, referees followed an inconsistent training prescription and required detailed planning of appropriate daily training plans.

Referees' season consisted of 34 matches played over 30 weeks, during which there was a 2-week break when no matches were played (Weeks 11 and 12). To examine the effect of time of season on referees' internal training loads, the 30 weeks were divided into five mesocycles of equal duration (e.g., 6 weeks). Although J. J. Malone et al. (2015) used 6-week cycles, the methodology of 6-week training mesocycles was implemented to reflect the time period available for the preseason training cycle. For this study, Mesocycle 1 commenced with Week 1 leading into the first round of scheduled matches. Matches were played on a weekly basis, with variation in number of days between matches depending on individual referee match assignments. In total, 3,670 session training loads were analyzed across the five mesocycles of equal duration.

3.3.3 Statistical Analysis

All data are presented as means ± standard deviation. For all analyses, a magnitude-based inference approach was used (Batterham & Hopkins, 2006; Hopkins et al., 2009). The general and generalized mixed linear model (R, Version 3.6.1) was used for analysis of continuous and count data, respectively. To examine differences between matches and the nine types of training sessions, RPE_{load} was entered as a fixed effect with a random slope and intercept for session (i.e., unstructured covariance matrix) to account for the hierarchical nature of our design (e.g., repeated measurements from the same referees). Match was used as the reference category for all comparisons. As the fourth official, rest, and recovery categories did not represent any substantial load, they were omitted from this aspect of the analysis. Examination of the effect of training day (Sunday to Saturday) and Mesocycle 1 to 5 on RPE_{load} was undertaken using the same approach (e.g., training day, mesocycle as the fixed effect). Here, the regular match day (Saturday) and the final mesocycle of the season (Mesocycle 5) were used as the reference category. The uncertainty in the estimates is expressed as 90% confidence limit (CL). Magnitude-based inferences were applied subsequently with inferences based on standardized thresholds for small, moderate, and large differences of 0.2, 0.6, and 1.2 of the pooled between-subject standard deviations (*SDs*;

Hopkins et al., 2009). The chance of the difference being substantial or trivial was interpreted using the following scale: 25%–75% = possibly, 75%–95% = likely, 95%–99.5% = very likely, and > 99.5% = most likely (Batterham & Hopkins, 2006). If the 90% CL overlapped the thresholds for the positive and negative effects, the effect was deemed unclear (Hopkins et al., 2009). The within-subject variability (expressed as a *SD*) in RPE_{load} by session and by training day also was derived from the mixed linear model, with the *SD* doubled to interpret its magnitude (Smith & Hopkins, 2011). Using the initial three referee activity categories of match, physical training, and rest, I examined the effect of day (Sunday to Saturday) and Mesocycle 1 to 5 on the frequency of activity in each of these categories using a generalized mixed linear model (Poisson loglinear), with activity category entered as the fixed effect. Here, the regular match day (Saturday) and the final mesocycle of the season (Mesocycle 5) were used as the reference category, and the uncertainty in our estimates was expressed as 90% CL. Magnitude-based inferences were applied using thresholds of 1.11, 1.43, 2.0, 3.3, and 10 for small, moderate, large, very large, and extremely large, respectively, and the inverses of 0.9, 0.7, 0.5, 0.3, and 0.1, respectively (Hopkins et al., 2009).

3.4 Results

3.4.1 Match and Training Session RPE_{load}

Figure 1 presents the overall RPE_{load} and RPE_{load} for matches and the nine types of training sessions. RPE_{load} was greatest for Matches, with most likely large differences compared to Speed (-413 AU; 90% CL ± 18 AU), Strength (-391 AU; ± 18 AU), Endurance (-365 AU; ± 15 AU), and Repeated Sprints (-314 AU; ± 25 AU) as well as most likely moderate differences compared to High Intensity Intervals (-251 AU; ± 18 AU), CT4 (-214 AU; ± 18 AU), CT1 (-213 AU; ± 18 AU),

CT2 (-158 AU; ± 20 AU), CT3 (-154 AU; ± 18 AU). The magnitude of within-subject variability in RPE_{load} was moderate for Endurance, Repeated Sprints, Strength, Speed, and Matches (range 102 AU; ± 9 AU to 137 AU; ± 8 AU) and large for High Intensity Intervals, CT1, CT4, CT2, CT3 (152 AU; ± 13 AU to 176; ± 16 AU).

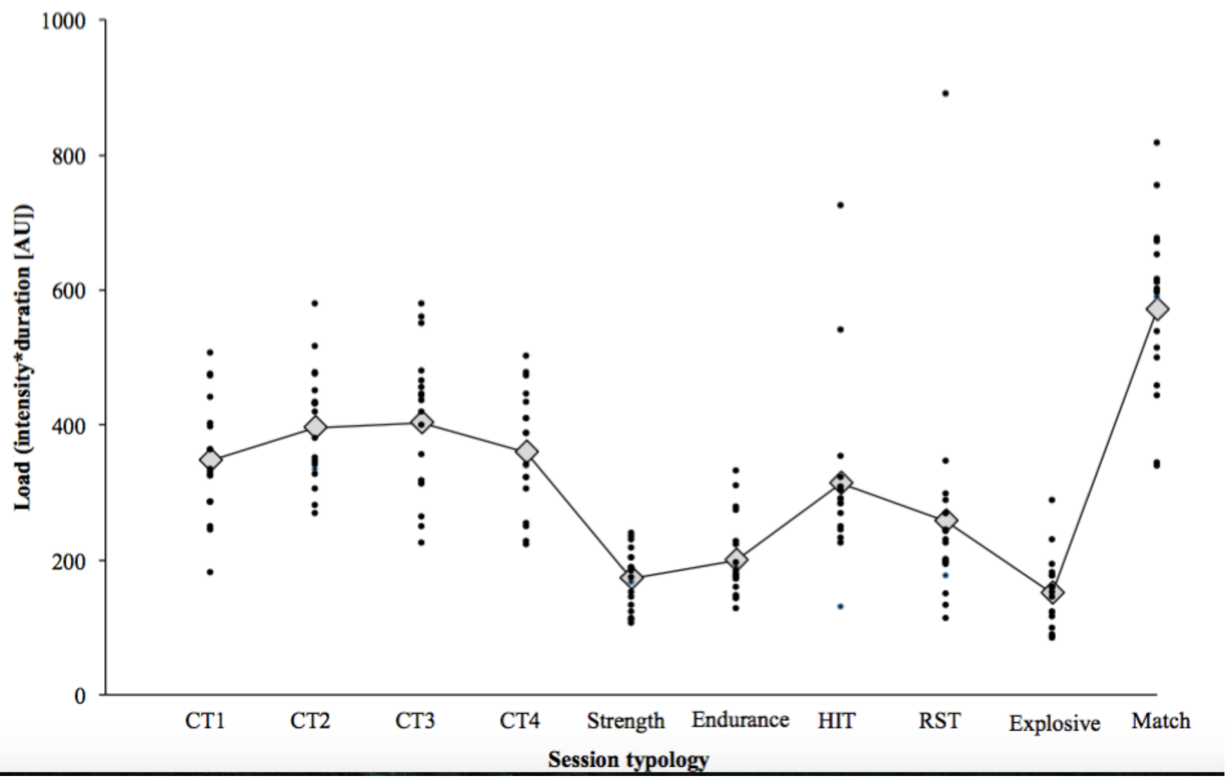


Figure 1. Mean RPE_{load} and Referees' Mean RPE_{load} for Matches and Nine Types of Training Sessions. *CT1 = Speed+Strength, CT2 = Repeated Sprints+Strength, CT3 = High Intensity Intervals+Strength, CT4 = Endurance+Strength. SD bars omitted for figure clarity.*

3.4.2 Daily RPE_{load} and Activity Breakdown

Weekly RPE_{load} was 1365 ± 225 AU. Figure 2 presents the mean daily RPE_{load} and weekly activity breakdown. Using the regular match day (Saturday) as the reference category, a most likely small difference in daily RPE_{load} was observed compared to Sunday (-123 AU; ± 23 AU) with trivial differences for all other days (range -54 AU; ± 22 AU to 23 AU; ± 23 AU). Within-subject variability in daily RPE_{load} was large for Sunday to Friday (204 AU; ± 11 AU to 225 AU; ± 12

AU) and very large for Saturday (268 AU; ± 14 AU). Expressed as a percentage of total activity, the mean frequency of matches, rest, and training activities was 14 ± 15%, 38 ± 18%, and 48 ± 24%, respectively. Across the week differences in daily activity frequency ranged from extremely large to large, Exp(B) 0.08 to 0.32, for matches, moderate to large, Exp(B) 0.65 to 2.05, for rest, and small to very large for training, Exp(B) 0.65 to 3.8.

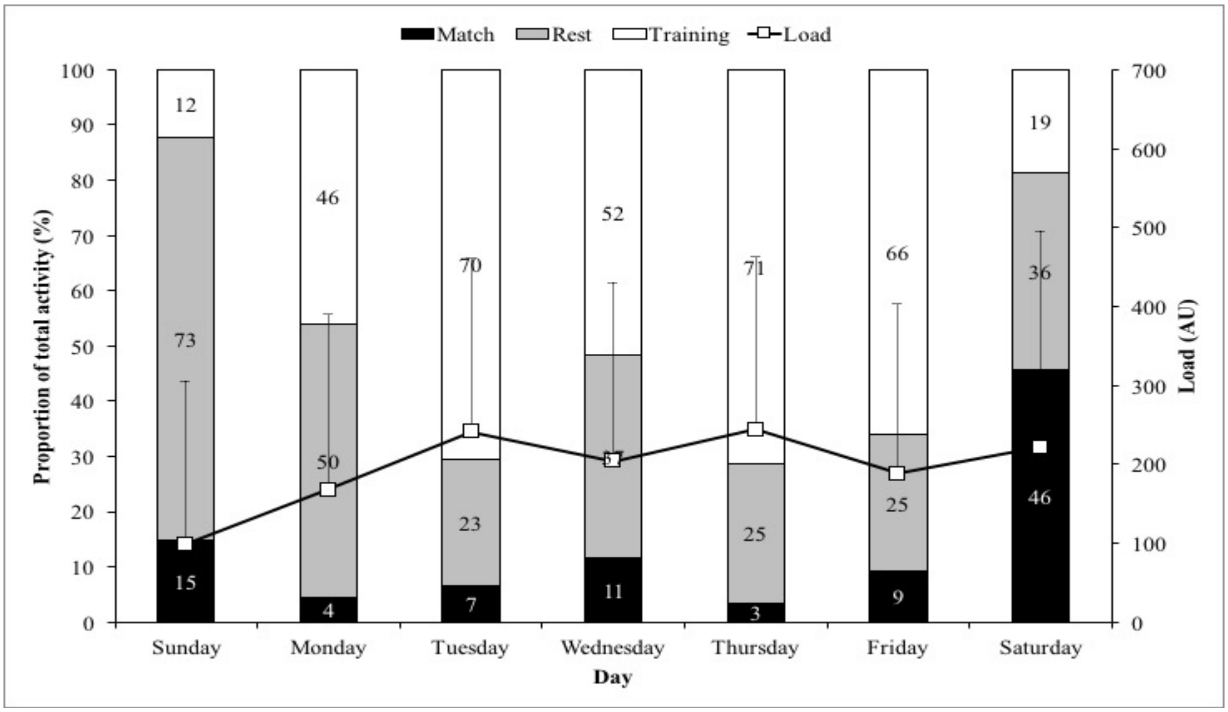


Figure 2. *Mean Referee Daily RPE_{load} and Weekly Activity Breakdown as Percentage of Total Daily Activity.*

3.4.3 Mesocycle RPE$_{load}$ and Activity Breakdown

Figure 3 shows the RPE$_{load}$ for mesocycle and activity. RPE$_{load}$ was greatest in Mesocycle 5, with very likely small differences compared to Mesocycle 2 (272 AU; ± 123 AU), likely for Mesocycle 3 (209 AU; ± 123 AU) and Mesocycle 4 (180 AU; ± 123 AU), and possible for Mesocycle 1 (148 AU; ± 123 AU). Compared to Mesocycle 5, differences in daily activity frequency ranged from trivial to small for matches, Exp(B) 0.87 to 1.01, and rest, Exp(B) 1.05 to 1.31, and small for training, Exp(B) 1.15 to 1.36.

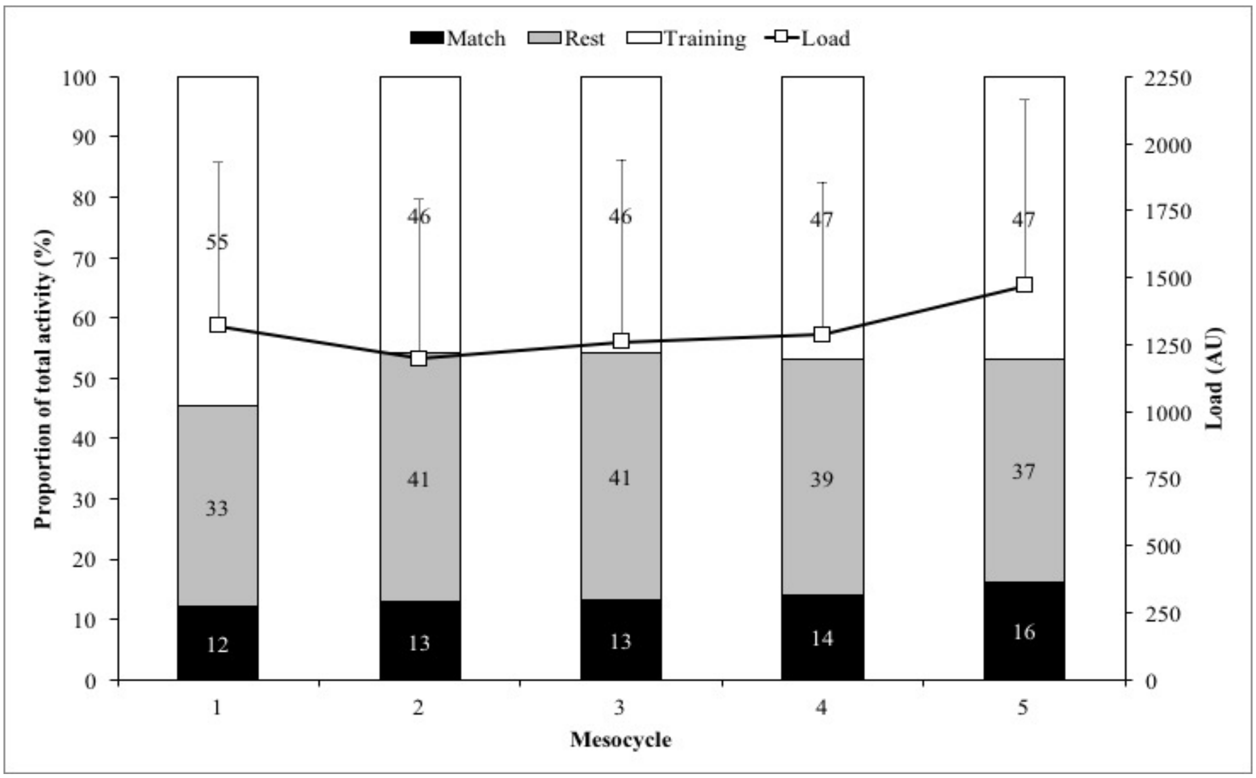


Figure 3. *Mean Mesocycle RPE$_{load}$ and Activity Breakdown as Percentage of Total Mesocycle Activity.*

3.5 Discussion

This study presented the training practices of an elite group of professional soccer referees. By doing so, these data represent the most detailed account to date of soccer referees' training loads. Through analysis of referees' match and training RPE_{load} scores, matches represented the highest training load of the week, and training load was relatively consistent, at least at a group level, across training weeks and mesocycles.

The high match physical demand imposed on soccer referees (Weston, Drust, et al., 2011) requires good physical fitness to ensure referees can keep up with play and in turn take up good positions when making key decisions (Weston, 2015). As such, a holistic approach to referee fitness training is recommended, one addressing all relevant components of fitness (Weston et al., 2012). For example, a training routine of the same five training typologies used in the present study (i.e., Speed, Repeated Sprints, High Intensity Intervals, Endurance, and Strength) helped an elite soccer referee progress from part-time status to refereeing the 2010 FIFA World Cup and Union of European Football Associations Champions League finals, with only two missed matches due to injury (Weston, Gregson, et al., 2011). Despite being a single referee case study, this information can help to inform the training prescription of elite soccer referees and educate on the long-term impact of a professionalized training environment on referees' physical performance. Physical training forms the foundation for development of an athlete's physical, physiological, and performance characteristics (Impellizzeri et al., 2004); therefore, replicating the types of training presented for MLS referees represents an ideal position for the sports science support program.

Match day RPE_{load} represents the highest load of the week and thereby shows competitive soccer matches contribute a large percentage of the overall weekly dose of activity (Weston, 2014). The absence of any detailed quantification of soccer referees' match and training RPE_{load} make it difficult to reconcile this finding; however, the fact that matches represented the highest internal training load is consistent with soccer (Impellizzeri et al., 2005; Thorpe et al., 2016b) and AFL players (16, 17). The mean match RPE_{load} was ~80 AU lower than that reported by Weston et al. (2010) for professional English soccer referees. This difference could be explained by the observation that match activity profiles appear to be dependent upon competition level (Castagna et al., 2007), which could then impact upon referees' internal loads. However, comparisons of the match activity profiles of MLS and English soccer referees are not possible given the discrepancy or lack of comparable technology used in previously noted research.

In this group of referees, the mean internal training load for the five isolated training sessions (i.e., Speed, Repeated Sprints, High Intensity Intervals, Endurance, and Strength) was 219 AU, a training load consistent with the 219 AU and 229 AU reported for all training sessions in soccer (Gaudino et al., 2015) and AFL players (Lovell et al., 2013). Compared to matches, training load scores for all combined training sessions (CT1–CT4) and High Intensity Intervals were moderately lower, whereas all remaining isolated training sessions were largely lower. These moderate to large differences in RPE_{load} between matches and training reflect specificity in the programming of different exercise session typology durations and intensities.

In the present study, mean weekly RPE_{load} was 1365 AU, a load below that previously reported in other professional team sports. For example, higher in-season weekly training loads have been reported for soccer (1980 AU and ~1600 AU; 11, 19) and AFL (1852 AU and ~1651 AU; Ritchie et al., 2016; Rogalski et al., 2013). Given that the physical demands imposed upon elite soccer referees are consistent with those experienced by players (Weston, Drust, & Gregson, 2011), soccer referees could be expected to have training loads commensurate with those undertaken by players. The supposition is strengthened by the similarity in actual training session

RPE_{load} between the MLS referees and those previously reported for elite soccer players (Gaudino et al., 2015; Thorpe et al., 2016b). However, the training practices of soccer referees shows a prevalence of physical training sessions over skill practice (Catteeuw et al., 2009; MacMahon et al., 2007). This emphasis on physical training was also evident in the present study because all training sessions referees performed were physical in nature. As such, additional load from skill training sessions was not observed for MLS referees. This may account for the weekly RPE_{load} disparity, given that skill training session load in team sports can be more than 300 AU (Lovell et al., 2013; Ritchie et al., 2016), which helps to explain the relatively high percentage (38%) of rest days in referee training schedules.

The prescription and distribution of training within professional soccer is influenced heavily by competition frequency (Akenhead et al., 2016). The manipulation of training load, called periodization, is undertaken to promote physiological adaptations and limit the negative effects of fatigue; consequently, the prescription of training load in elite soccer players is lower on the day closest to matchday (Impellizzeri et al., 2004; J. J. Malone et al., 2015; Thorpe et al., 2016a). In the present study, despite small to very large differences in training activity frequency across the week, analysis of referees' RPE_{load} showed an absence of the within-week periodization that is apparent for soccer players. Absence of any clear differences in training load across the week, except for Sunday, is possibly a consequence of moderate to large within-subject variability for training session RPE_{load} combined with large to very large within-subject variability for daily RPE_{load}. Such variability quantifies that training practices are indeed different on an individual level (Ritchie et al., 2016), the consequence of training structured around individual referee assignments. As such, there was no apparent group-level weekly periodization. To gain a further understanding of distribution of referees' training load, the season was broken down into five mesocycles. Here, RPE_{load} was greatest in the final mesocycle of the season, which is no doubt reflective of this phase of the season containing the highest frequency of matches. Trivial to small differences in RPE_{load} and activity frequency across the five mesocycles is a finding consistent with J. J. Malone et al. (2015) who reported soccer players' training load showed limited in-season variation.

This study provides and in-depth detail on elite soccer referees' training practices. In doing so, it presents not only the internal training loads undertaken by referees but also assessed the effect of activity type, training day, and time of season on internal training load. The data reveal matches represent by far the highest training load imposed on this group of elite soccer referees; although individual training session loads are comparable with elite team sport players, weekly loads are slightly lower by comparison. Substantial within-subject training session and daily RPE_{load} variability make it difficult to draw any meaningful conclusions about referees' periodization strategies, suggesting any future analyses should be performed on an individual level.

3.5.1 Practical Applications

These findings are specific to referees in the United States, and the data reflect implementation of training regimens and match assignment during only one season. Although the types of training referees performed were consistent with those reported previously (Weston, Gregson, et al., 2011), it was not possible to present any fitness test or injury incidence data to confirm the training's effectiveness. Nonetheless, these data provide valuable practical information for a comprehensive understanding of the internal training loads obtained over referees' full season. Quantification of referees' training loads was confined to internal loads because, at the time, no external load data (e.g., GPS) were available; therefore, it was not possible to investigate the relationship between internal and external load, which would have enabled a more accurate prescription and monitor of training load (Bartlett et al., 2017). Furthermore, substantial within-subject variability training session and daily RPE_{load} impair a more informative appraisal of referees' weekly periodization practices. As soccer grows in popularity in the United States, professional matches are increasingly televised, and, as television rights expand, matches are played on more nights of the week. Therefore, these fluctuating schedules impact the training plan for each referee, which may render any group-level analysis of periodization irrelevant.

CHAPTER 4: DEFINING THE REQUIREMENTS OF MATCH OFFICIALS WITH REGARD TO PHYSICAL TRAINING AND RESPONSE

Match Officials Response to a Season Long Professional Soccer Season

4.1 Introduction

Monitoring athletes through objective and subjective measures, with the goal of increased performance, has been integrated into the preparation of both individual and team sports in recent years (Buchheit, 2014; Buchheit et al., 2013; Meeusen et al., 2013; Saw et al., 2015b; Saw et al., 2016; Thorpe et al., 2015; Thorpe et al., 2016b). Objective approaches to quantifying recovery status have been investigated in individual and team sport athletes. Complex and invasive collection measurements looking at the variation of creatine kinase samples post-match (A. Scott, 2016), as well as testosterone and cortisol levels (Rowell et al., 2016) post training, are commonplace among team and individual sport settings. Furthermore, neuromuscular measurements such as counter movement jump height (Rowell et al., 2016) are used but can require expensive equipment. Resting heart rate (Buchheit, 2014; Buchheit et al., 2013; Lazic et al., 2017; Vesterinen et al., 2016) and heart rate variability (HRV; Buchheit, 2014; Esco et al., 2016; Flatt & Esco, 2014, 2015; Plews et al., 2012, Plews et al., 2013; Stanley et al., 2013) are both simple, noninvasive tools used to measure recovery and fatigue. More recently, subjective questionnaires have been used based on the low-cost point and ease of use for both practitioners and athletes (Gallo et al., 2016, 2017; Saw et al., 2015a; Saw et al., 2016; Thorpe et al., 2015). Subjective measures which are often reported by the athletes using self-assessment surveys, include wellness profiles to quantify physical, mental, and emotion state, sleep quality, energy levels and fatigue, and measures of perceived effort during training. Subject measures have been reported to be more sensitive and consistent than objective measures (Saw et al., 2016). For example, during the competition phase, professional soccer players indicated high correlations between self-reported fatigue and total high intensity running distance (Thorpe et al., 2015), and Buchheit et al. (2013) found the previous day's training load influenced the summation of wellness scores the following morning. Considering this evidence, subjective measures have been purported to enable early detection of athletes who may respond negatively to training (Meeusen et al., 2013).

To date, limited research exists on the dose-response relationship of training and matches on referees' perceived response and recovery to training through subjective wellness questionnaires. As Burgess (2017) noted, these perceived responses are important for practitioners because a validated and efficient wellness questionnaire is indicative of an efficient and effective sport science program. As a result, athlete responses can be interpreted more easily and accurately, leading to better organizational data-based decision making. While it is widely known that teams and organizations implement customized subjective wellness questionnaires, one of the greatest challenges is developing validated, reliable, and rigorously tested psychometrically sound questionnaires. Simultaneously, these data collected need to achieving the organizational goals and purpose for collection. Despite these challenges, subjective morning wellness questionnaires are appropriate for a remote training environment based on ease of data collection. Armed with an abundance of subjective questionnaires available for the practitioner to use (Gallo et al., 2015) and implementing guidelines for the development of wellness questionnaire's (Saw et al., 2016), the aim of this paper was to understand the variation of morning wellness responses to physical training of professional referees in North America over the duration of a full season.

4.2 Methods

4.2.1 Participants

Data were collected from 21 North American match officials (mean age ± *SD* 40 ± 9.5 years) from within Major League Soccer (MLS). MLS is the highest level of competition played in the United States, and data reflected in-season competition.

4.2.2 Research Design

To examine retrospectively the influence of training loads on morning wellness subjective feedback over a season, a single-group observational design was used. The 2017 referee season consisted of 39 weeks, commencing with Week 1 of the preseason period. Training schedules were determined by upcoming match assignment; therefore, referees followed an inconsistent training prescription. In total, 10,475 daily wellness and session RPE (sRPE) training loads were analyzed across the seven training cycles. Morning wellness scores were analyzed as a response to the daily training session from the previous day.

4.2.3 Methodology

Morning Wellness Questionnaires. Each morning, within 30 mins of waking, referees completed a wellness questionnaire administered via a web-based data application (i.e., VisualCoaching Pro, Australia). Referees for this study had utilized this tool for the previous season, therefore the familiarization had taken place the previous season. Introduction to the tool was conducted over a 2-week in-person pre-season. Specific definitions and reasoning for each question were provided to each referee and further explanation as to how the collected metrics were utilized in daily practice was explained. Multiple subjective tools, such as POMS and the Recover-Stress Questionnaire, have been examined extensively for both validity and reliability; however, the time requirements and lack of specificity to sporting environments has posed problems in applied settings (Saw et al., 2015b). Therefore, to ensure compliance while also utilized the guidelines and recommendations for developing a wellness questionnaire (Saw et al., 2016), the number of questions was limited to 7 to reduce respondents' time requirement. Questions were selected to reflect physical and psychological components common in the psychological tools used to assess for training imbalances in the literature (Gastin, Meyer, & Robinson, 2013; Gaudino et al., 2015). A 7-question wellness questionnaire was designed to generate ratings of tiredness, motivation, stress, mood, energy, sleep quality, and appetite. All questions were scored on a 10-point Likert scale, where measures of motivation, mood, energy, sleep quality, and appetite were scored as 10 = *optimal*/1 = *poor*, and tiredness and stress were scored as 1 = *optimal*/10 = *poor*. Referees were instructed to respond on how they felt related to the aforementioned factors. An overall daily wellness score was generated by summing the total of the seven items.

Training Load. Training loads were measured using the session RPE, as Foster et al. (2001) described. Using the CR10 scale, all referees provided a session RPE for each activity, and this score was then multiplied by session duration to obtain a RPE load (RPE_{load}) score, reflecting both intensity and duration. Individuals provided their session RPE using the same web application used to collect morning wellness ratings (i.e., VisualCoaching Pro, Australia). Individuals were asked to provide their session RPE within 30 mins of completing each match or training session. To represent the training load and its effect on morning wellness, scores were offset by 1 day, which was performed by moving the training load score forward to the next day for analysis.

4.2.4 Statistical Analysis

All data are presented as means ± SD. All analyses were performed on the raw untransformed data with homoscedasticity confirmed via visual inspection of the residual plots. A general mixed linear model (R, 3.6.1) was used to examine the effect of RPE_{load} (fixed effect) on the following day wellness score, represented here as a *z* score. To account for the repeated measures nature of this experimental design, a random slope and intercept was included in all the mixed models. The modifying effect of RPE_{load} also was calculated as the effect of two standard deviations (i.e., the difference between a typically low value and a typically high value; Hopkins et al., 2009) on the following day wellness *z* score. I expressed uncertainty in our estimates as 90% CL, and magnitude-based inferences were subsequently applied with inferences for a change in wellness *z* score based on a change of 1.15, which equates to a probability of a likely change of 0.75 (Hopkins et al., 2009). Mixed modelling also was used to examine the fixed effect of day (Monday to Sunday, with Sunday used as the reference category) on daily RPE_{load} and the following day wellness *z* score and also the effect of Mesocycles 1 to 7 (with Mesocycle 7 used as the reference category) on daily RPE_{load} and wellness *z* score. Magnitude based inferences again were applied subsequently with the threshold for wellness *z* score change as described previously; whereas, for RPE_{load} standardized thresholds for small, moderate, and large differences of 0.2, 0.6, and 1.2 of the pooled between-subject standard deviations (Hopkins et al., 2009). Here, the chance of the difference being substantial or trivial was interpreted using the following scale: 25%–75% = *possibly*, 75%–95% = *likely*, 95%–99.5% = *very likely*, > 99.5% = *most likely* (Batterham & Hopkins, 2006). Effect magnitude was evaluated mechanistically, whereby if the 90% CL overlapped the thresholds for the smallest worthwhile positive and negative effects, the effect was deemed unclear (Hopkins et al., 2009).

Within-referee correlations were used to determine if high RPE_{load}'s were associated with lower overall wellness and the seven individual wellness subscales (i.e., tiredness, motivation, stress, mood, energy, sleep, appetite). This method is appropriate because it permits analysis of within-referee relations by removing between-referee variability (Bland & Altman, 1995). CL (90%) for within-referee correlations were calculated per Altman and Bland (2011). The following scale of magnitudes (Hopkins et al., 2009) was used to interpret the magnitude of the correlation coefficients: 0.1 = *trivial*, 0.1–0.3 = *small*, 0.3–0.5 = *moderate*, 0.5–0.7 = *large*, 0.7–0.9 = *very large*, and .0.9 = *nearly perfect*.

4.3 Results

The mean ± SD daily referee RPE_{load} and following day overall wellness score was 175AU ± 205 AU and 50 AU ± 4.7 AU, respectively. Fixed effect analysis revealed a most likely trivial effect of RPE_{load} on the following days wellness *z* score (0.001; ±90% CL 0.001). Using two standard deviations of the predictor variable, the difference between a typically low and typically high daily RPE_{load} was a 0.42 ± 0.03 change in wellness *z* score (most likely trivial).

4.3.1 Daily RPE_{load} and Wellness Scores

The mean ± *SD* daily referee RPE_{load} scores were 153 ± 180 AU, 233 ± 95 AU, 205 ± 203 AU, 179 ± 63 AU, 159 ± 181 AU, 208 ± 272 AU, and 89 ± 189 AU for Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday, respectively. For daily referee wellness scores, the mean ± *SD z* scores were -0.07 ± 1.05, 0.12 ± 0.93, 0.02 ± 0.98, -0.04 ± 0.98, 0.05 ± 0.94, 0.13 ± 0.95, and -0.20 ± 1.12 for Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday, respectively (see Figure 4).

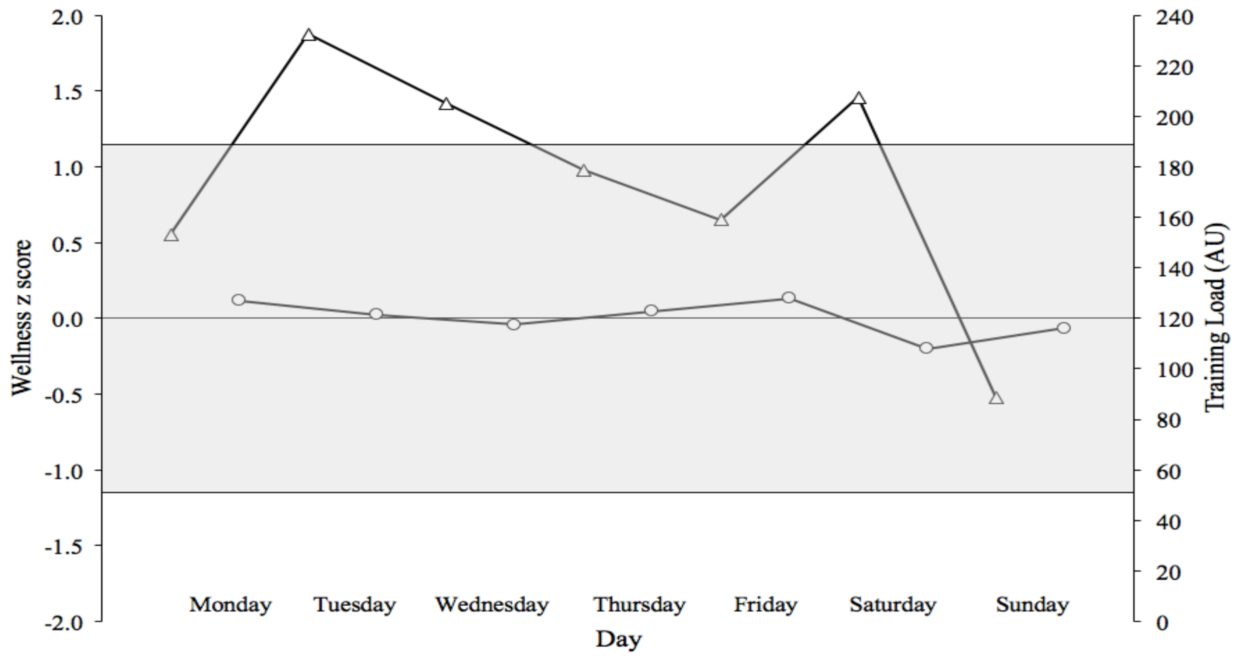


Figure 4. Daily Referee RPE_{load} and Wellness Scores. *Using a threshold of 1.15 for a meaningful change in wellness z score, the shaded grey area represents the trivial change zone for daily wellness sores. Mean scores are presented with SD bars omitted for figure clarity.*

Using the final day of the week as the reference category (Sunday), fixed effect analysis for the effect of day on RPE_{load} revealed substantial changes across the week. Compared to Sunday, a moderately higher daily RPE_{load} was most likely on a Tuesday (145 AU, ± 90% CL 12 AU) and a small higher daily RPE_{load} was most likely on a Saturday (118 AU, ± 11.7 AU), Wednesday (117 AU, ± 12 AU), Thursday (91 AU, ± 12 AU), Friday (70 AU, ± 12 AU), and Monday (64 AU, ± 12 AU). However, fixed effect analysis of referees' daily wellness scores revealed the magnitude of between-day changes, compared to Sunday, was most likely trivial (range 0.13, ±0.1 to 0.33, ±0.1).

4.3.2 Mesocycle RPE_{load} and Wellness Scores

Referees' mean ± SD mesocycle RPE_{load} scores were 188 ± 200 AU, 177 ± 194 AU, 175 ± 201 AU, 171 ± 212 AU, 177 ± 214 AU, 176 ± 210 AU, and 152 ± 200 AU for Cycles 1, 2, 3, 4, 5, 6, and 7, respectively. For referee wellness, the mean ± *SD* cycle *z* scores were 0.03 ± 1.03, 0.05

± 1.03, 0.06 ± 0.99, 0.04 ± 0.93, -0.03 ± 0.97, -0.07 ± 0.99, and -0.12 ± 1.07 for Cycles 1, 2, 3, 4, 5, 6, and 7, respectively (see Figure 5).

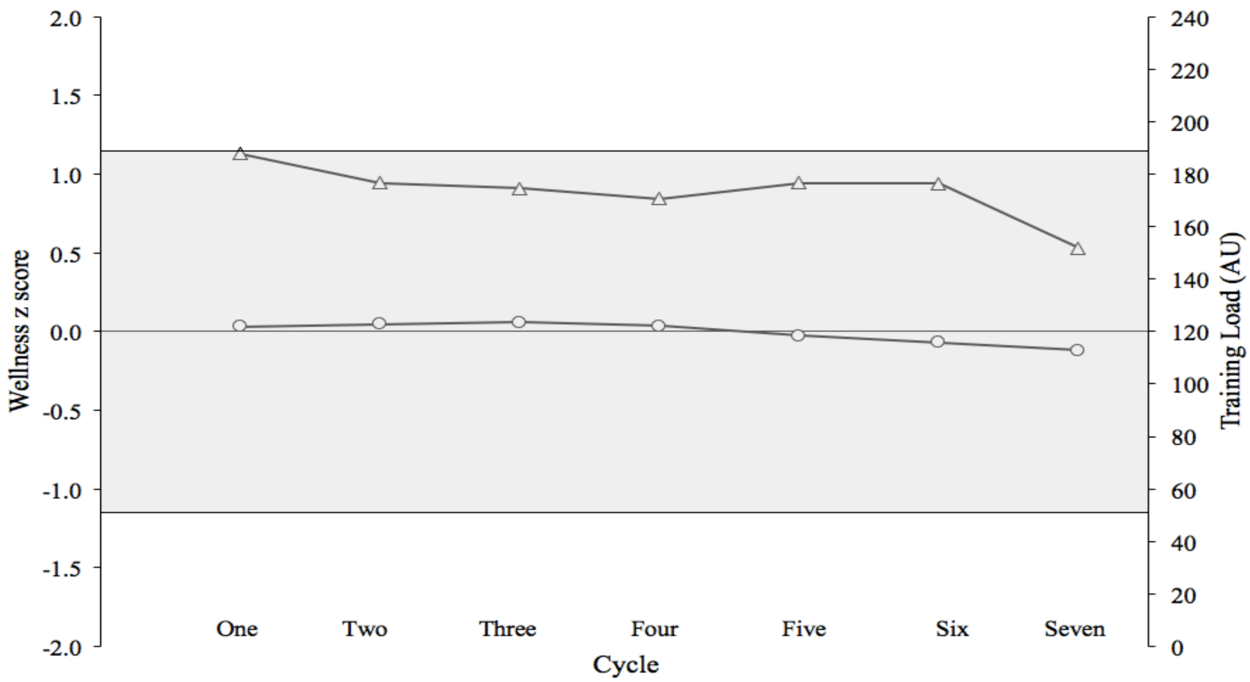


Figure 5. Mesocycle Referee RPE_{load} and Wellness Scores. *Using a threshold of 1.15 for a meaningful change in wellness z score, the shaded grey area represents the trivial change zone for daily wellness sores. Mean z scores are presented with SD bars omitted for figure clarity.*

Fixed effect analyses using the final mesocycle of the season (Mesocycle 7) as the reference group revealed possibly to most likely trivial effect on RPE_{load} (range 20 AU, ± 14 AU to 37 AU, ± 14 AU) and most likely trivial effect on referee wellness (range 0.05, ± 0.07 to 0.18, ± 0.07).

4.3.3 Relationship Between RPE_{load} and Wellness Scores

Overall, within-referee correlation between RPE_{load} and following day wellness score was most likely small. When examining the strength of the associations on an individual referee level, there were 3 moderate, 16 small, and 3 trivial correlations (see Table 1). Figure 6 includes the mean individual referee and overall referee mean for the seven wellness subscales. Overall, within-referee correlations between RPE_{load} and following day wellness subscales scores (see Table 1) ranged from most likely small (tiredness, energy), likely small (sleep quality), very likely trivial (motivation), and most likely trivial (sleep quality, appetite, mood). When examining the magnitude of within-referee correlations on an individual referee level, the strongest associations with RPE_{load} were for tiredness (13 small and 4 moderate), energy (11 small and 5 moderate), and sleep quality (13 small and 1 moderate), whereas, for motivation, stress, mood, and appetite, there were 11, 5, 3, and 3 clear correlations (e.g., small or greater) with RPE_{load}, respectively.

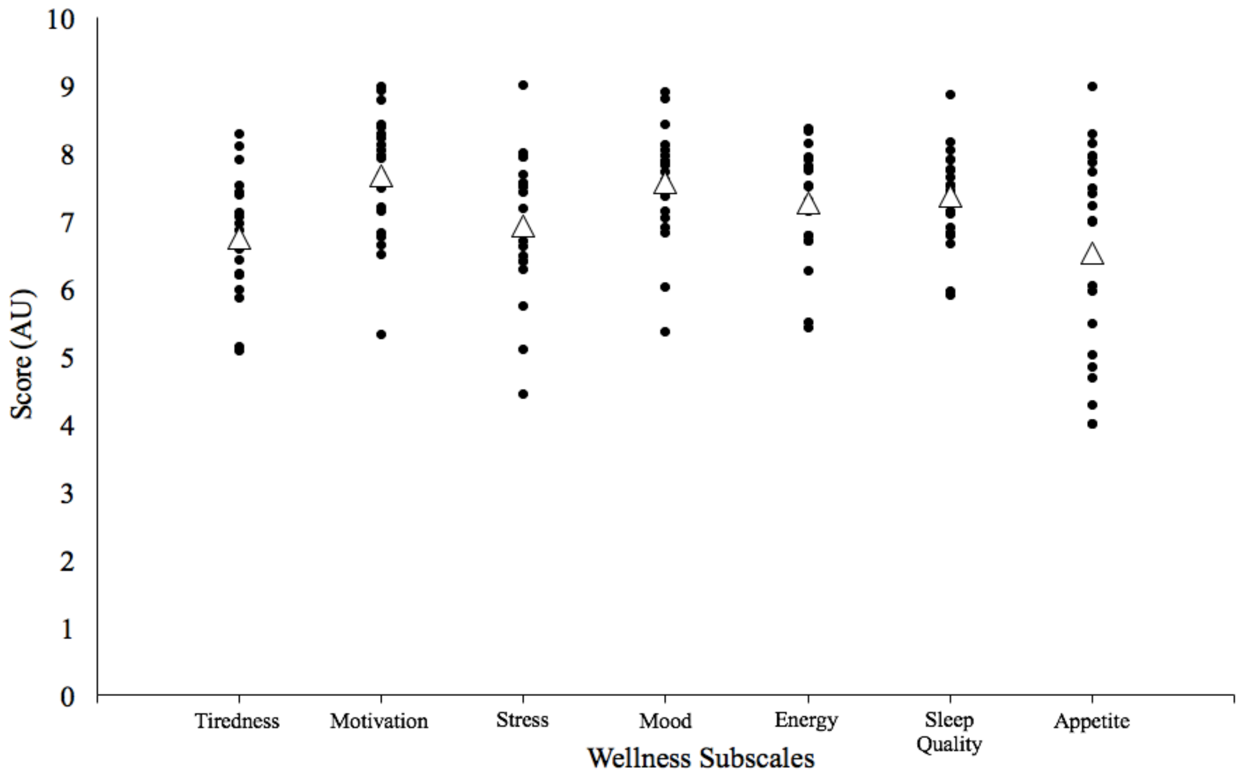


Figure 6. *Mean Individual Referee (Closed Circles) and Overall Referee Mean (Open Triangle) for Wellness Subscales*

Table 1

Within-Referee Correlations Between Referee RPE$_{load}$ With Overall Wellness and Wellness Subscales

Referee	RPE$_{load}$ and wellness	RPE$_{load}$ and stress	RPE$_{load}$ and tiredness	RPE$_{load}$ and motivation	RPE$_{load}$ and mood	RPE$_{load}$ and energy	RPE$_{load}$ and sleep quality	RPE$_{load}$ and appetite
1 (*n* = 474)	0.23 (****Small)	0.05 (**Trivial)	0.38 (****Moderate)	0.02 (***Trivial)	0.01 (***Trivial)	0.34 (**Moderate)	0.03 (***Trivial)	0.06 (**Trivial)
2 (*n* = 510)	0.19 (***Small)	0.00 (***Trivial)	0.21 (***Small)	0.21 (***Small)	0.02 (***Trivial)	0.21 (***Small)	0.20 (***Small)	0.15 (**Small)
3 (*n* = 517)	0.16 (**Small)	0.09 (*Trivial)	0.24 (***Small)	0.04 (**Trivial)	0.10 (*Trivial)	0.34 (**Moderate)	0.21 (***Small)	0.04 (***Trivial)
4 (*n* = 457)	0.11 (*Small)	0.04 (**Trivial)	0.10 (*Small)	0.10 (*Small)	0.05 (**Trivial)	0.11 (*Small)	0.12 (*Small)	0.03 (**Trivial)
5 (*n* = 512)	0.11(*Small)	0.02 (**Trivial)	0.12 (*Small)	0.07 (**Trivial)	0.05 (**Trivial)	0.07 (*Trivial)	0.16 (**Small)	0.04 (**Trivial)
6 (*n* = 528)	0.19 (***Small)	0.25 (****Small)	0.21 (***Small)	0.08 (*Trivial)	0.06 (**Trivial)	0.21 (***Small)	0.25 (****Small)	0.05 (**Trivial)
7 (n = 375)	0.03 (**Trivial)	0.07 (*Trivial)	0.01 (***Trivial)	0.04 (**Trivial)	0.06 (**Trivial)	0.13 (*Small)	0.16 (**Small)	0.07 (*Trivial)
8 (*n* = 437)	0.24 (***Small)	0.03 (**Trivial)	0.37 (***Moderate)	0.12 (*Small)	0.04 (**Trivial)	0.25 (****Small)	0.20 (***Small)	0.08 (*Trivial)
9 (*n* = 503)	0.24 (****Small)	0.09 (*Trivial)	0.21 (***Small)	0.11 (*Small)	0.03 (**Trivial)	0.14 (**Small)	0.20 (***Small)	0.02 (***Trivial)
10 (*n* = 480)	0.34 (**Moderate)	0.09 (*Trivial)	0.34 (***Trivial)	0.36 (***Moderate)	0.02 (***Trivial)	0.29 (****Small)	0.27 (****Small)	0.00 (***Trivial)
11 (*n* = 506)	0.16 (**Small)	0.03 (**Trivial)	0.02 (***Trivial)	0.22 (****Small)	0.16 (****Small)	0.22 (****Small)	0.07 (**Trivial)	0.05 (**Trivial)
12 (*n* = 462)	0.13 (*Small)	0.02 (***Trivial)	0.08 (*Trivial)	0.16 (**Small)	0.03 (**Trivial)	0.11 (*Trivial)	0.21 (***Small)	0.09 (*Trivial)
13 (*n* = 495)	0.16 (**Small)	0.02 (***Trivial)	0.09 (*Trivial)	0.21 (***Small)	0.04 (**Trivial)	0.24 (****Trivial)	0.23 (****Small)	0.04 (**Trivial)
14 (*n* = 522)	0.24 (****Small)	0.02 (*Small)	0.16 (**Small)	0.05 (**Trivial)	0.01 (**Trivial)	0.37 (****Moderate)	0.14 (**Small)	0.08 (**Trivial)
15 (*n* = 482)	0.02 (***Trivial)	0.06 (**Trivial)	0.13(*Small)	0.1 (*Trivial)	0.13 (*Trivial)	0.14 (**Small)	0.08 (****Trivial)	0.06 (**Trivial)
16 (*n* = 467)	0.41 (****Moderate)	0.18 (***Small)	0.27 (****Small)	0.39 (***Moderate)	0.32 (*Moderate)	0.42 (****Moderate)	0.31 (*Moderate)	0.03 (****Trivial)
17 (*n* = 439)	0.10 (*Small)	0.03 (**Trivial)	0.46 (****Moderate)	0.02 (***Trivial)	0.06 (**Trivial)	0.16 (**Small)	0.06 (**Trivial)	0.04 (****Trivial)
18 (*n* = 507)	0.15 (**Small)	0.13 (**Small)	0.24 (****Small)	0.03 (**Trivial)	0.00 (***Trivial)	0.15 (**Small)	0.18 (***Small)	0.03 (**Trivial)
19 (*n* = 518)	0.14 (**Small)	0.11 (*Small)	0.16 (**Small)	0.09 (*Trivial)	0.06 (**Trivial)	0.09 (*Trivial)	0.07 (**Trivial)	0.09 (*Trivial)
20 (*n* = 513)	0.38 (***Moderate)	0.05 (**Trivial)	0.30 (*Moderate)	0.25 (****Small)	0.07 (**Trivial)	0.37 (***Moderate)	0.38 (***Moderate)	0.10 (*Small)
21 (*n* = 474)	0.05 (**Trivial)	0.04 (**Trivial)	0.10 (*Small)	0.00 (***Trivial)	0.01 (***Trivial)	0.01 (****Trivial)	0.03 (**Trivial)	0.06 (**Trivial)
22 (*n* = 268)	0.26 (****Small)	0.23 (***Small)	0.19 (**Small)	0.13 (*Small)	0.14 (*Small)	0.18 (***Trivial)	0.15 (**Small)	0.18 (** Small)
Overall	0.15 (****Small)	0.02 (****Trivial)	0.19 (****Small)	0.08 (***Trivial)	0.00 (****Trivial)	0.18 (****Small)	0.17 (**Small)	0.01 (****Trivial)

Note. * possibly; ** likely, *** very likely, **** most likely

4.4 Discussion

The primary aim of this study was to examine the effects of RPE_{load} on match officials' morning wellness scores over the course of a single North American professional soccer season. Our findings identified limited correlation between RPE_{load} and morning wellness scores. From our analysis, it would appear RPE_{load} does not have a substantial effect on referees' perceptual daily wellness scores. Across the week, I found RPE_{load} was greatest on Tuesdays and Saturdays, although wellness scores remained relatively stable throughout the week. Across the seven different mesocycles, training loads were constant, and similar to our weekly analysis findings, I found no substantial change in daily wellness scores. A slightly stronger correlation was found when comparing the effects of higher or lower RPE_{load} on individual wellness responses. Within-referee correlations were small to moderate when comparing RPE_{load} and overall wellness, tiredness, energy, and sleep quality. While limited, these results suggest prior-day training load may influence the individual responses referees have for wellness on the subsequent day.

Daily wellness questionnaire responses have been examined in a similar manner to previous studies using daily wellness to understand perceptual feelings (Buchheit et al., 2013; Coutts & Reaburn, 2008; Gallo et al., 2015; Gallo et al., 2017; Gastin, Meyer, & Robinson, 2013; Saw et al., 2015a; Thorpe et al., 2016a) within various sport settings. Most commonly, these survey responses are recorded on ordinal scales. Previous studies looked at the average of a set of ordinal wellness response variables and treated the average as a continuous response variable (Gallo et al. 2017). This approach has potential shortcomings in that there is a reduction in multiple metrics to one value, there is an assumption of equal importance of each variable, and the averaging of wellness metrics does not identify potential important relationships between training and wellness. Within this study, small correlations between morning wellness and the RPE_{load} from the previous day were found, which was unexpected given previous research demonstrated a stronger relationship between training loads and subjective wellness. Research also has examined relationships between internal–perceptual feelings and external training loads (McLean et al., 2010; B. R. Scott, 2013), and strong correlations were shown between internal responses: sRPE and external loads (B. R. Scott, 2013). Low between-match variability was shown when comparing sRPE and external load measure (S. Malone, Owen, et al., 2017; McLaren et al., 2016), and several studies have demonstrated match intensities influenced reported measures of fatigue and soreness (Carling et al. (2016, Thorpe et al., 2015). Contrary to available research, our findings indicate the RPE_{load} did not influence wellness scores when summing the total value of all questions to create a daily score.

This analysis of North American referees showed relatively constant RPE_{load} and wellness score responses across seven seasonal mesocycles. In North America, matches are played every day of the week, which creates a need to understand individual responses to training and matches from not only within the week but also across longer time frames. Our findings present an opportunity to understand how referees cope with structured physical training, not only within a single day's training load but also across training loads consisting of multiple days, weeks, and months. With a constant RPE_{load} across mesocycles, I did not find a similar sensitivity to RPE_{load} as demonstrated in previous research (Buchheit et al., 2013; S. Malone, Owen, et al., 2017; Thorpe et al., 2016a), where higher training loads presented lower wellness scores over extended periods of time. This contrary finding is interesting considering the RPE_{load} values were comparable to previous research and theoretically would have instigated fatigue and created a fluctuation in morning wellness. Despite collecting our data over the season, questions should be raised regarding the sensitivity within this population and may not represent a true response to the fluctuation in RPE_{load} across mesocycles.

Individuals' interpretations of the questions and potential outcomes may also have greatly influenced their responses. I found only a select number of the morning wellness questions had a small to moderate correlation to RPE_{load}, which should guide practitioners toward using appropriate questions that provide useful feedback. With the reported correlations, it would be appropriate to use tiredness, energy, and sleep quality, as well as a summation of the three, as the most effective morning wellness responses. The remaining four questions used in this questionnaire reflected stronger correlations between select questions and individual match officials' RPE_{load}, which suggest individual questionnaires may be an appropriate approach. These findings could be the most substantial for practitioners. To determine which data are most valuable for representing morning wellness factors, the practitioner must solicit different data with the highest correlations from each participant.

4.4.1 Limitations

As Weinberg et al. (2001) reported, one distinct disadvantage of using self-report psychological assessment is potential dishonesty in answers. Similar to research from Schwarz et al. (1991), match officials may not have reported their perceived wellness status accurately or with full honesty. They may not have reported indications of fatigue following a match due to the risk of the organization possibly interpreting this as an inferior fitness level or of not being assigned a game the following week. Moreover, match officials may have felt reluctance to report lower scores due to the risk of losing a game assignment or receiving a change in training prescription.

A custom questionnaire, modified from multiple versions of previous research (Gastin, Meyer, & Robinson, 2013; Gaudino et al., 2015), was implemented for this study without performing previous validity or reliability studies. The lack of rigorous testing of specific questions, scaling techniques, or placement of verbal anchors for this population could have led to a lack of meaningful results across the group. Cox et al.'s (2003) study provided an example of successful monitoring of overall athletes' physical and psychological health by using a subjective questionnaire that was modified to ensure compliance and honest answers within specific performance environments. Perhaps the questions were ambiguous, unclear, or irrelevant; therefore, the results reflect futile findings. Krosnick and Fabrigar (1997) provided suggestions on evaluating data quality, scaling, and labeling, yet no definitive answers regarding correct delivery of survey questionnaires was provided. Further, the repetitive nature of our data collection on a daily basis may have promoted an environment of monotony and thus reflect the referee simply supplying a number based on necessity rather than providing an insightful response. Akin to Dickinson and Hanrahan (2009), officials' responses to questions may be reflective of a positive or negative influence of alternate factors not necessarily related to training or match performance (e.g., environmental, travel, or personal family matters).

4.4.2 Practical Applications

Speed, ease of use, and ability to collect data quickly has made perceptual well-being assessment in professional sport a useful daily monitoring tool. Although there are several questionnaires generally used in elite sport settings (Gallo et al., 2015), further data collection from match officials should follow Dickinson and Hanrahan's (2009) recommendations emphasizing using simple, short, and more practical psychological measurement tools to maximize compliance. Further research and assessment on the selection of appropriate questions generated by the questionnaire—and delivery of questionnaires—should be conducted to understand best practices for match officials. Subjective information on potential improvement or regression in daily wellness may help match official organizations evaluate and modify training or match assignment accordingly, which could enable match officials to achieve optimal performance. According to these findings, these decisions would be difficult to establish. Increased education directed toward match officials' understanding of the potential value and insight they can provide on a day-to-day basis with thoughtful and honest responses to the wellness questions also should be implemented. Further, organizations must create an environment where honest feedback will be rewarded and guarantee honest responding will be analyzed appropriately within the scale from absence of fitness to presence of fatigue. Finally, additional analysis is needed to determine which individual wellness metrics within the daily score provide the greatest levels of reliability and validity across match officials.

4.5 Conclusion

The aim of this paper was to evaluate the effects of RPE_{load} on perceptual morning wellness responses of North American referees. Throughout our findings, morning wellness responses remained relatively constant across weeks and mesocycles, and the importance of individual analysis became apparent. I conducted our investigation across a single season; therefore, our findings are only descriptive, and I would need to conduct further analysis across multiple seasons to provide further validation of our findings. More in-depth evaluation of both internal and external responses from individuals, along with inspecting the types of training, would provide additional insight into the nature of our findings. Despite changes in RPE_{load} across the week, a relatively steady daily wellness response was reported. Findings within this study indicate match officials respond individually to specific questions despite differences in training and match schedules; therefore, organizations should favor individual versus group analysis. Understanding the influences of RPE_{load} could lead to a more detailed analysis of how referees cope with the demands of training and matches, allowing organizations to deliver more appropriate training and match assignments. Our findings showed a relatively consistent wellness score regardless of RPE_{load}, which should initiate further investigation into the type of questions asked in wellness questionnaires as well as the influence of appropriate education on the reliability and validity of referees' responses. With increased exposure to daily wellness data collection procedures, improved question selection, education of match officials on the importance of honest answers, and a shift toward a more supportive organizational environment, data-based decision making will improve. Further understanding of individual responses to specific types of training, along with consideration of external environmental influences, will aid in the reliability and validity of data to assist in organizational decision making.

CHAPTER 5: MATCH PHYSICAL DEMANDS IN MAJOR LEAGUE SOCCER

Examination of the Physical Match Profiles of Professional North American Soccer Players and Referees

5.1 Abstract

Identifying the physical match activity profiles of soccer players and officials can help to design appropriate training as well as in talent identification. The purpose of this study was to describe the physical match activity profiles of match officials and players where the ball was in play and across whole match play within the top tier of professional soccer in North America. Data were obtained from 617 matches played over the course of two seasons in Major League Soccer. Velocity, acceleration, deceleration, impulse, and ball-in-play (BIP) time were all obtained. Mean BIP was 48 s, with a maximum of 280 s. Seventy-one percent of all team possessions were 60 > s. An average of 16 (± 5.3) and 18.3 (± 5) BIP occurrences were under 30 s. The average maximum speed for the 0–15 s BIP window was 5.95 m-s^{-2}, and mean max speeds dropped to 3.9 m-s^{-2} for the 60 s window. Average distance covered for the entire match ranged from 9,788 m to 12,379 m, and BIP distances ranged from 7,676 m to 9,945 m. Defensive midfielders covered the most distance, and central forwards covered the least. Central forwards and attacking central midfielders showed the highest 5-s BIP speeds of 8.26 and 8.24 m-s^{-2} respectively. Accounting for both BIP and ball out of play, defensive midfielders recorded the highest mean speed at 236.5 m-min^{-1} for the 1-min window, and central defenders along with central referees indicated 205 m-min^{-1} peak demands across the same timeframe. Compared to other domestic leagues, these findings present higher mean max speed distances yet similar degradation across the similar windows.

5.2 Introduction

Physical performance match profiles of soccer players have been researched extensively across domestic and international competitions (Barnes et al., 2014; Bradley & Noakes, 2013; Di Salvo et al., 2010). However, in the top professional league in North American, Major League Soccer (MLS), on-field performance research has been limited to tactical (González-Rodenal et al., 2017a, 2017b) as well as injury rates and treatment (Calloway et al., 2019; Erickson et al., 2013; Farber et al., 2014). While Weston et al. (2011) described the physical demands of match officials as similar to those of outfield players, further exploration of the influence of metrics beyond total distance or high-speed running is warranted. The presentation of more precise features including formation or tactics that may influence official match performances could provide a more detailed understanding of match demands. Defining the physical performance match profiles of match officials and players specific to the top-tier domestic level could provide more precise insight into recruitment strategies, particularly compared to other leagues, and consequently outline practitioners' guidance to individual and team physical training preparation approaches.

Physical match profiles in soccer typically have been defined through repetitive multidirectional and linear movements at a variety of intensities and velocities (Barnes et al., 2014; Carling & Dupont, 2011; Carling et al., 2012; Varley et al., 2013). Across two first division seasons in Spain, a mean total distance run of ~10.9 km over a 90-min match was reported (Gomez-Piqueras et al., 2019), although a detailed investigation into first division Italian elite outfield players presented total distance run between 9 and 14 km, covering 22%–24% of total match distance at speeds above 15 $km \cdot h^{-1}$, 8%–9% at velocities above 20 $km \cdot h^{-1}$, and 2%–3% at speeds above 25 $km \cdot h^{-1}$ (Rampinini et al., 2007). When expressed relatively, average speeds range from

104 metre per minute ($m \cdot min^{-1}$) to 130 $m \cdot min^{-1}$ (Carling et al., 2010; Varley et al., 2013). Variation in total distance and high-speed running distance has been demonstrated across athlete age (Buchheit et al., 2010), across competition level (Jennings et al., 2012), and across multiple seasons (Barnes et al., 2014). Highlighting an acceleration-based approach in defining workload and energy expenditure, time spent and distances covered in acceleration and deceleration zones have been reported (Akenhead et al., 2013; Cummins et al., 2013; Davies et al., 2013; Gaudino et al., 2014; Varley et al., 2013). Furthermore, average accelerations (Bradley et al., 2010) and count of total accelerations (Aughey, 2011; Bradley et al., 2010) also have been included in defining more comprehensive work demands of intermittent sports. Additionally, match demands may be underreported as they have been summarized by half or full time and do not reflect the maximal values required within a game (Delaney et al., 2017). In an intermittent sport, these periods also could reflect the effective playing time to understand peak values while only analysing time spent with the ball in play.

With consideration to potential contextual factors influencing match profiles, investigation of attacking styles, formations, and position-specific demands also have been presented. A recent study reported 4-3-3 formations perform more high intensity runs than 4-4-2 formations (Aquino, Viera, et al., 2017), but defensive formations (i.e., 4-5-1) perform more high-intensity runs when not in possession compared to other offensive formations (i.e., 4-3-3; Bradley et al., 2011). To understand better the physical demands based on roles and responsibilities of outfield players, differences in match running outputs also exist between players of different positions (Bush et al., 2015; Di Salvo et al., 2007). However, regardless of formation or playing position, match-activity data suggest with consistency that elite players tend to perform most of the soccer game at slower velocities and consequently cover shorter distances as speeds increase (Barrett, 2017; Di Salvo et al., 2007). Furthermore, although the research has been somewhat limited, domestic and international soccer matches have shown BIP time, also described as effective playing time, average ~55 mins (Lago-Peñas, 2012a; Siegle & Lames, 2012). Representing physical demands more precisely within effective playing time, where physical demands are analysed only while the ball is on the playing field, has shown second half total distance run was greater compared to the first half (Lago-Peñas, 2012b), and variance in physical performance was related more closely to effective playing time than fatigue (Carling & Dupont, 2011).

Defining match profiles serves to highlight positional demands where the ball is in play and consequently deliver more heightened knowledge of the physical requirements to assist in the planning and preparation during training. To our knowledge, no research has been conducted on the positional requirements of domestic competition within the top division of the United States and Canada. Therefore, the purpose of this study was to describe the match activity profiles of match officials and players within top-tier professional soccer in North America.

5.3 Methods

The observational design of this study was derived from player movement data collected via optical tracking. Tracking data provided by Metrica Sport (Amsterdam, NL) were combined with event data to allow analysis of ball in and out of play time. The validity of this data has been previously reported (Pappalardo et al., 2019; Renkin et al., 2022). Data were obtained from 617 matches played over the 2017–2018 seasons within MLS. Matches were played on a weekly basis, with a variation in number of days between matches depending on individual team schedules. Only outfield athletes were analyzed, and positions were classified as center referee (REF; *n* = 617), attacking center midfielder (ACM; *n* = 577 players), central defender (CD; *n* = 617), central forward (CF; *n* = 617), central midfielder (CM; *n* = 526), defensive midfielder (DM; *n* = 533), wide defender (WD; *n* = 611), outside midfielder (OM; *n* = 392), and winger (WNG; *n* = 277). Only players who completed 45 mins or greater were included in the analysis.

Upon completion of each match, event data along with one 25 Hz comma separated value file for each team and referee were exported from proprietary software (Metrica Play, Amsterdam, NL) and further analyzed using the statistical software R (Version 3.6.1). Data were cleaned to exclude half-time breaks, and match time was calculated from start to end of each half. BIP was calculated as the total time the ball was in play, with the provided event data used to denote “out of play” (substitutions, injuries, kick-off, throw-in, goal-kick, corner-kick, free-kick, goals). For each match, every BIP duration was extracted to establish the average, mean, minimum, and maximum values. Furthermore, across both seasons, not all games were captured due to logistical or staffing shortages.

For the 25 Hz movement data, a low-pass, second-order Butterworth filter with a cutoff at a .04 sampling rate was applied, and displacement covered by each athlete was calculated using Euclidean distance, defined as the shortest distance between two points. Velocity was then calculated as displacement divided by the sampling rate (.04). Any data point where velocity was greater than 10 $m\text{-}s^{-1}$ was removed because it was considered an error (Delaney et al., 2017). Acceleration and deceleration were calculated as the absolute rate of change of speed ($m\text{-}s^{-2}$). Impulse was used as a load measure of acceleration and deceleration, which was calculated by multiplying the instantaneous absolute value by body mass and then multiplying by the 0.04 s (sampling rate). The sum of impulse accumulated within each percentage of the maximal mean acceleration was then established to reflect the volume of acceleration or deceleration performed at different intensities (Johnston et al., 2022). A moving average technique then was applied to each output speed variable using 10 different durations (i.e., 1–10 min), with the peak value achieved throughout each match for each variable recorded (Delaney et al., 2017).

5.4 Results

Overall match mean BIP was 49 minutes, mean BIP was 48 s, with a maximum of 280 s. Average count and percentage of total time of each respective time window is presented in Figure 7. Seventy-one percent of all possessions were 60 > s. An average of 16 (± 5.3) and 18.3 (± 5) BIP occurrences were under 30 s. Average max speed for the 0- to 15-s window was 5.95 m-s^{-2} while mean max speeds dropped to 3.9 m-s^{-2} for the 60-s window.

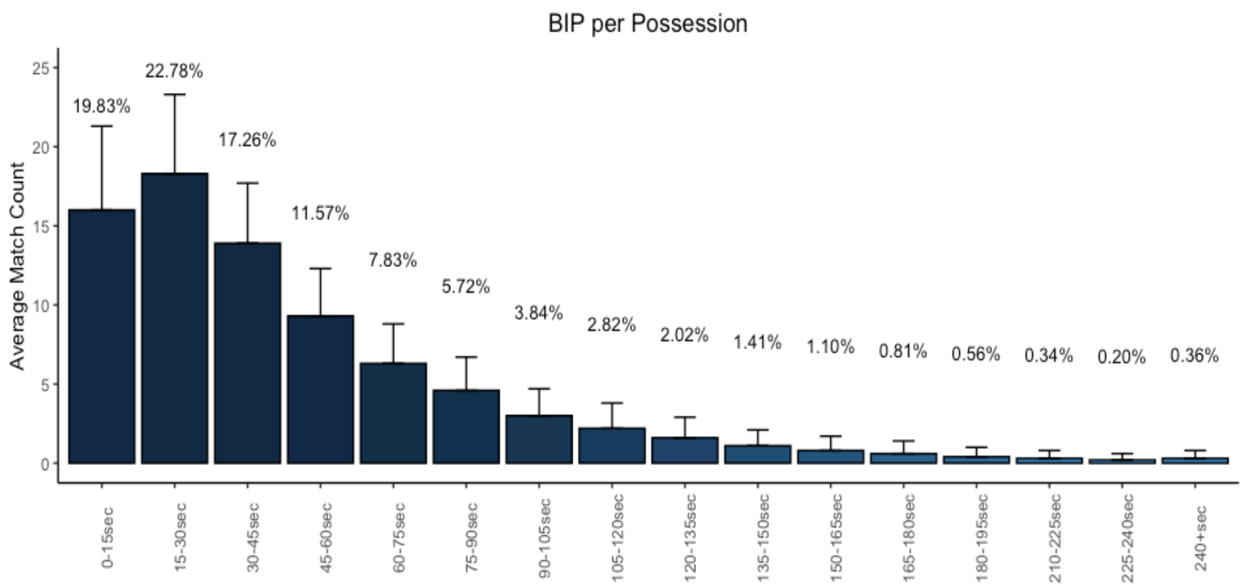


Figure 7. Count and Percentage of Total BIP Time by 15-s Window Increments. *Data are presented as the mean ± SD in reference to the entirety of the match.*

Average distance covered for the entire match ranged from 9,788 m to 12,379 m, and BIP distances ranged from 7,676 m to 9,945 m. Both in-play and out-of-play scenarios indicated defensive midfielders covered the most distance, and central forwards covered the least. Average max speeds for each 5-s window over the first 60 s are shown in Table 2. Central forwards and attacking central midfielders showed the highest 5-s BIP speeds of 8.26 and 8.24 m-s^{-2}, respectively (see Figures 8–11).

Table 2

Total Distance and Average Max Speeds Described by Position

Position	Outside back	Central defender	Outside midfielder	Defensive midfielder	Central midfielder	Attacking central midfielder	Winger	Central forward	Center referee
Distance per match (m)	10654 ± 1488	10626 ± 1015	8838 ± 2982	11890 ± 1890	10222 ± 2791	8548 ± 2516	9153 ± 2465	8902 ± 1839	11364 ± 1042
BIP distance per match (m)	8263 ± 1225	8277 ± 914	6891 ± 2358	9564 ± 1620	8199 ± 2330	6706 ± 2040	7137 ± 1983	6975 ± 1495	8836 ± 965
BIP avg Speed 5 s ($m\text{-}s^{-2}$)	8.09	8.02	7.86	8.04	8.04	8.24	8.14	8.26	6.77
BIP avg Speed 10 s ($m\text{-}s^{-2}$)	6.91	6.75	6.7	6.8	6.85	7.04	6.89	7.03	5.79
BIP avg Speed 15 s ($m\text{-}s^{-2}$)	6.05	5.91	5.92	6.06	6.1	6.23	6.04	6.19	5.09
BIP avg Speed 20 s ($m\text{-}s^{-2}$)	5.52	5.32	5.4	5.57	5.6	5.69	5.46	5.64	4.69
BIP avg Speed 25 s ($m\text{-}s^{-2}$)	5.14	4.89	5.04	5.21	5.24	5.3	5.06	5.2	4.37
BIP avg Speed 30 s ($m\text{-}s^{-2}$)	4.83	4.58	4.75	4.94	4.96	4.99	4.76	4.91	4.11
BIP avg Speed 35 s ($m\text{-}s^{-2}$)	4.58	4.35	4.52	4.73	4.75	4.77	4.53	4.68	3.92
BIP avg Speed 40 s ($m\text{-}s^{-2}$)	4.38	4.16	4.36	4.56	4.57	4.59	4.34	4.48	3.77
BIP avg Speed 45 s ($m\text{-}s^{-2}$)	4.21	4.01	4.2	4.42	4.42	4.42	4.18	4.33	3.64
BIP avg Speed 50 s ($m\text{-}s^{-2}$)	4.07	3.88	4.06	4.3	4.3	4.29	4.04	4.19	3.52
BIP avg Speed 55 s ($m\text{-}s^{-2}$)	3.96	3.77	3.95	4.2	4.19	4.17	3.92	4.07	3.43
BIP avg Speed 60 s ($m\text{-}s^{-2}$)	3.85	3.67	3.85	4.11	4.09	4.07	3.82	3.96	3.34

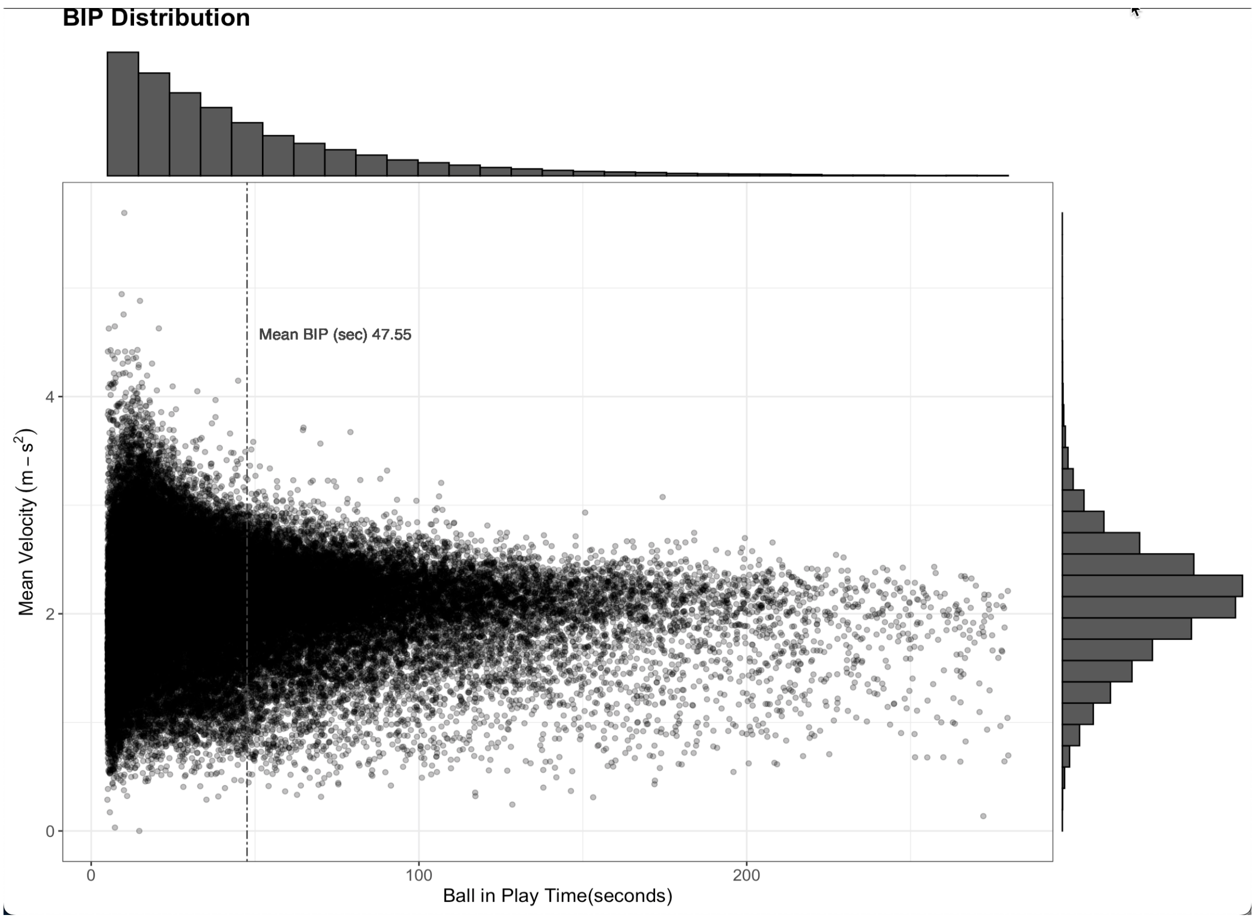


Figure 8. Relationship Between BIP and Max Mean Speed. *Distribution of max mean velocities and BIP with reference to the mean BIP of all matches.*

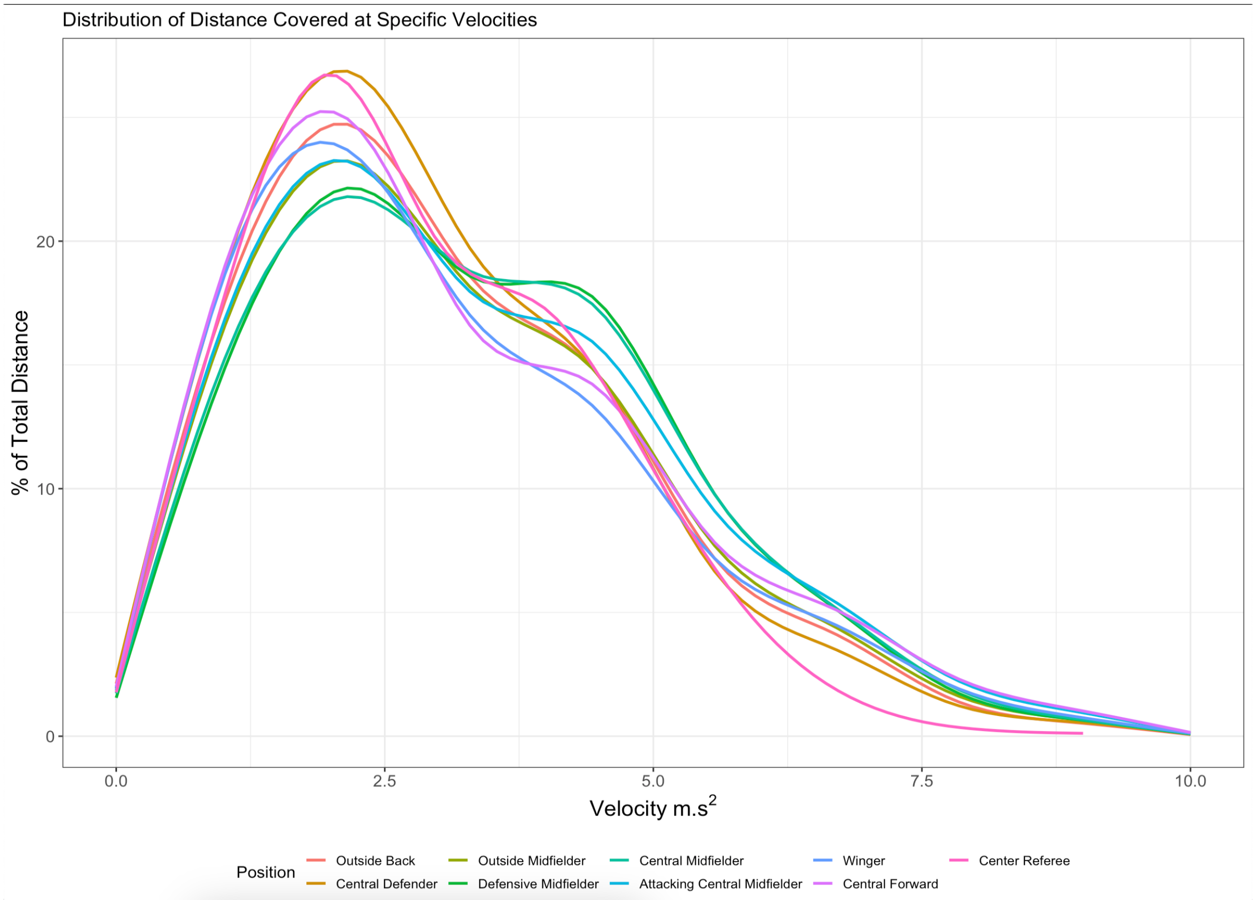


Figure 9. Distribution of Distance Covered at Specific Velocities. *Data are displayed as a percentage of total distance covered at specific Velocities. The highest percentage of mean velocity was spent at ~2.3 m-s^{-2} across all positions.*

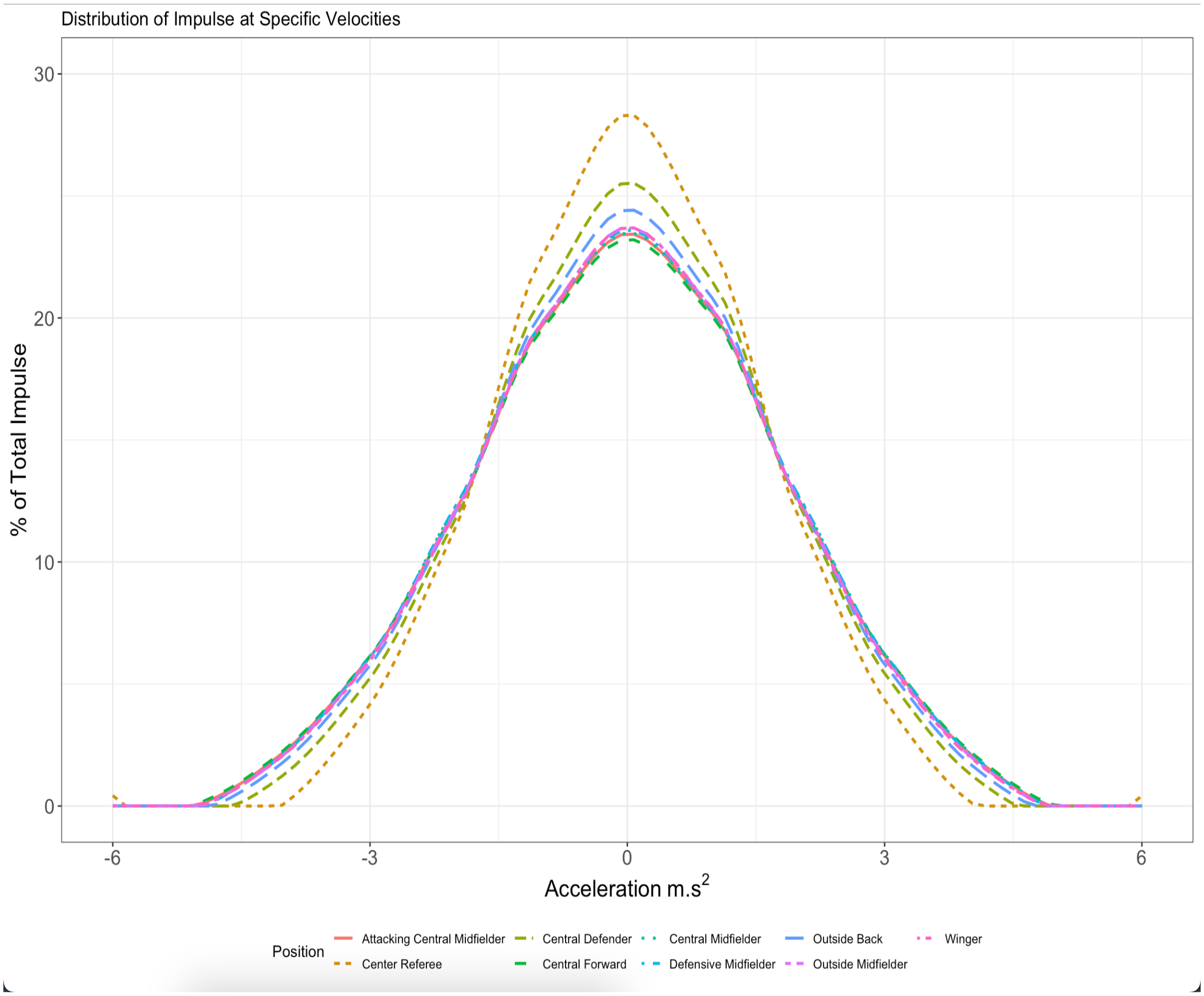


Figure 10. Distribution of Impulse at Specific Velocities. *Greatest proportion of impulse accumulated within acceleration/deceleration of 1 to -1 m-s^{-2}*

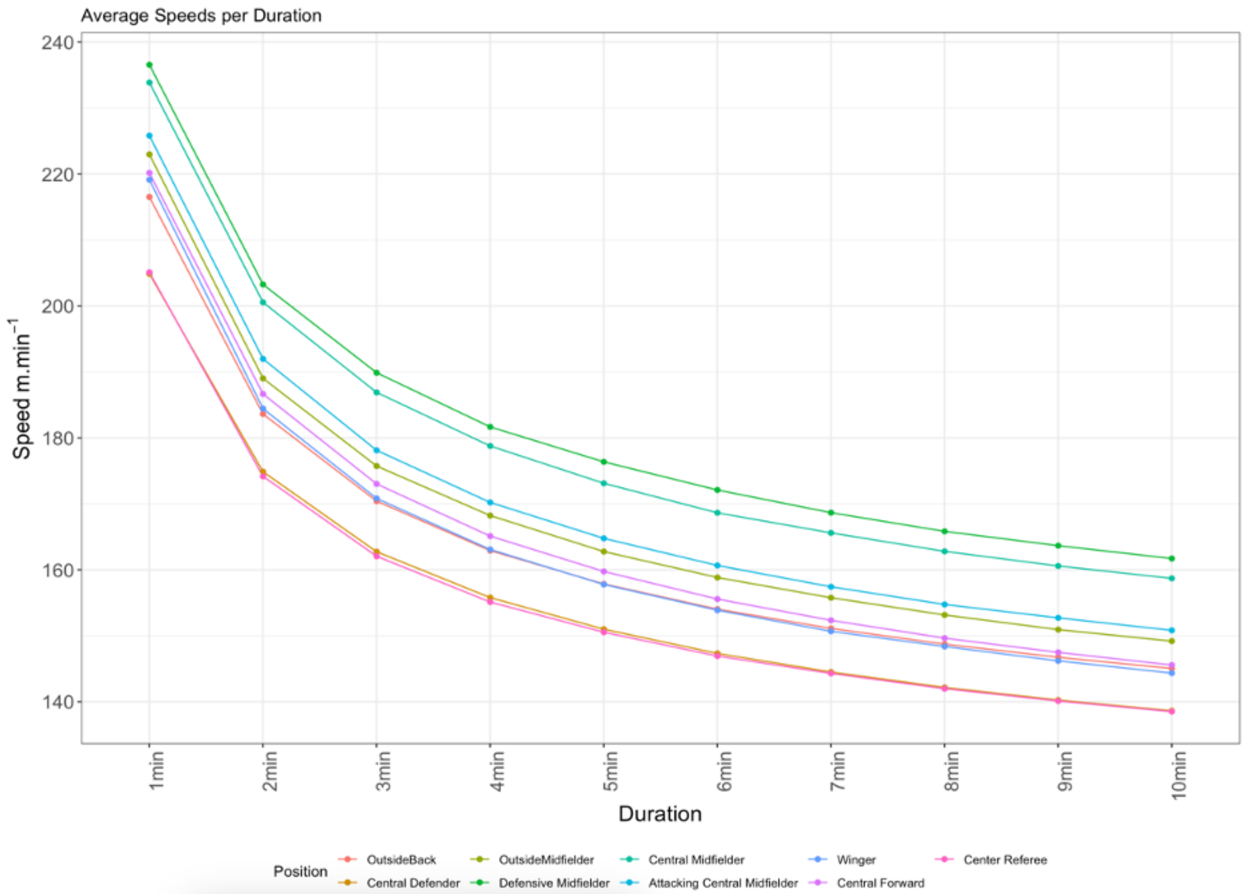


Figure 11. Whole-Match Mean Max m-min^{-1} Across 1- to 10-Min Windows. *Defensive midfielders recorded the highest mean speed at 236.5 m-min^{-1} for the 1-min window, and central defenders and central referees indicated 205 m-min^{-1} peak demands across the same timeframe. Peak 10-min speeds ranging from ~138-161. M-min^{-1}, a ~15% drop occurred between the 1st and 2nd min, and an overall average 33% drop in max peak intensities was observed across all positions.*

5.5 Discussion

The study aimed to define the match profiles of match officials and players in MLS through analysis of BIP time, whole match, and the responding physical demands. These findings are unique to top-tier soccer in North America but also present a novel finding for referees regardless of geography. Results represented a mean BIP time of 48 s, and more than 70% of BIP durations were less than 60 s. Additionally, a confirmation of running intensity reduction occurred as BIP time increased. Whole-match distribution of speed and impulse indicated mean speeds of 2.3 m·s^{-}

^{1}, and most acceleration and decelerations occurred between 1 and -1 $m \cdot s^{-2}$ respectively. Positional differences were represented when looking at windows of 1–10 mins, with defensive midfielders indicating the highest peak values.

Consistent with Bradley and Noakes (2013), Carling and Dupont (2011), and Castellano et al. (2011), these findings suggest a practical utility for understanding BIP time and informing preparation and monitoring strategies used by practitioners. Although Castellano et al. presented BIP time by half, Carling and Dupont used discrete time frames and found ~58–61 s per possession across each window. For this study, a 48-s BIP average was reported; however, 60% of BIP timeframes were 45 s or less, and 19.8% were 15 s or less. These time periods are not trivial, and although the data presented indicate a lower mean physical demand across a match, a seemingly large proportion of the BIP windows suggests these shorter time periods should be emphasized.

These findings are similar to previous research where intensity dropped as time increased when comparing values from the first to the second halves: 5-, 10-, 15-, and 20-s windows indicated drops of 4%–6% regardless of position (Delaney et al., 2017). To define peak demands more accurately, identifying intensities while the ball is in play is required. With an average 53% drop in velocity, the distance covered at 5 to 60 s was presented to indicate further the importance of understanding BIP demands. Although total distance is not a recommended metric to define match physical demands, the paper demonstrated the differences when comparing BIP distance to whole-match distance to showcase the total volume of work players perform. Referee specific total distance performances reflect similar findings from Castagna, Mallo, and Weston in top tier leagues with Italy, Spain, and England respectively (Castagna et al. 2007, Mallo et al. 2009; Weston, Drust, & Gregson, 2011)When designing training programs, these data could be useful to understand total volume of work, but it is more useful to understand the demand differences of both BIP and whole-match short and long windows.

Akenhead et al. (2013) and Russell et al. (2016) highlighted the need for quantification of both acceleration and deceleration metrics in addition to those relating to velocity and displacement. As depicted previously in Figure 10, the greatest proportion of impulse was accumulated at acceleration or decelerations of 1 to -1 m-s^{-2}. Positional differences were presented; however, only the center referee indicated a substantial difference from other positions. Regardless of position, more than 20% of all acceleration and deceleration actions occurred at a seemingly low intensity and should not be discounted when defining physical output. Although this finding aids in understanding the overall acceleration and deceleration demands across a match, further research on the positional impulse response to contextual variables, such as events or the location of these output areas on the field, is required.

Lastly, findings related to peak running demands across 1- to 10-min windows align with others who have used rolling windows (e.g., Mohr et al., 2003), with a similar decrease in running intensities to what Bradley and Noakes (2013), Carling et al. (2016), Delaney et al. (2017) observed. By comparison, these data have presented a similar 1-min average max hierarchy of position demands; however, the distances are higher across all positions, further emphasizing the need to investigate the domestic demands (Fereday et al., 2020; Oliva-Lozano et al., 2020). While confirming previous findings, the difference is unique to the domestic competition, and with an understanding of the BIP and whole-match demands, practitioners have a more complete perspective of the physical requirements in North America.

5.6 Conclusion

This study presents the physical demands of whole-match and BIP scenarios at the elite tier of North American soccer. These findings look at the average BIP time, demands across varying BIP windows, speed and acceleration/deceleration demands across whole matches, and peak demands across rolling windows. As expected, the data indicate differences across matches and within individual positions. Collectively, these findings could be deemed as useful because they are unique to the domestic top tier in North America; however, further understanding of the influences of these data are required. Pairing the findings with the game contextual variables (e.g., location on field, current match score, team tactics) would provide even further insight. Nonetheless, the current findings can aid in decision making for physical preparation and a clearer understanding of the variability of individual's schedule within a group setting. The data provide insight into the typical match play BIP time and related physical demands, which can help provide coaches, fitness coaches, and sport scientists the ability to play training sessions accordingly.

CHAPTER 6: RELATIONSHIP BETWEEN MATCH OFFICIALS PROFILES AND MATCH CHARACTERISTICS

Similarities to Match Characteristics: A Clustering Approach

6.1 Abstract

Quantifying soccer match physical demands provides insight into individual player and referee response capabilities and informs decisions on training prescription, selection and assignment, or strategy. Integrated definitions of performance have been presented more recently through combining event and tracking data; however, to our knowledge, this approach has not been used for match official performance. Therefore, this exploratory study used unsupervised analysis techniques to initialize an integrated approach to explaining referee physical performance and to determine if player and referee tracking and event data could be clustered into homogenous subgroups. Full event and tracking data, along with match event and descriptive data, were collected across two seasons. Implementing both K-mean and hierarchical clustering, I conducted three separate analyses: (a) physical outputs only by club and referee, (b) physical outputs by position, and (c) both event data and physical data. When only comparing physical data, homogeneous subgroups were found regardless of clustering approach. When using both event data and physical output data, despite correlations existing within referee outputs and event data, disagreements occurred when implementing K-mean and hierarchical clustering. To validate an evaluation of referee performance, hierarchical and nonhierarchical methods are appropriate yet require logical feature and clustering method selection. Our findings suggest referee performance evaluators should adopt an integrated approach in the grading criterion; however, further research in this area is required when applying unsupervised methods.

6.2 Introduction

Quantifying soccer match physical demands provides insight into individual player and referee response capabilities and potentially informs decisions on training prescription, selection and assignment, or strategy. Notated as an integrated approach, recent emphasis has been placed on seeking enhanced ways to define performance metrics by combining player movement with event or tactical approaches in soccer (Bradley & Ade, 2018; Gregory et al., 2022; Ju, Doran, et al., 2022; Ju, Lewis, et al., 2022; Liu, Gómez, et al., 2016; Paraskevas et al., 2020). Focus on specific running performance has been associated with variables such as match location or quality with higher distances covered for home matches compared to away (Liu, Hopkins, & Gómez, 2016; Paul et al., 2015). Variations to physical demands placed on referees when looking at levels of play, age, competition level, and the influence of player physical intensity have attempted to define a reason behind the match demands (Castillo et al., 2018; Weston et al., 2010; Weston, Drust, et al., 2011, Weston et al., 2007). When taking an integrated approach, however, the relationship between match events, player physical performance, and referees is still undefined.

Match officials are tasked with keeping up with match activity, governing players by applying laws of the game, and closely observing match play. Early research by Mallo et al. (2007) reported 1,400 m of high-intensity running with roughly 30 sprints per match, and further exploration indicated referees typically ran between 10 and 12 km per game, with 10%–15% of the distance covered at speeds above 5 $m \cdot s^{-1}$ (Krustrup et al., 2009; Mallo et al., 2012). Referees' running distance is associated with the distance of the ball (Mallo et al., 2009) and is similar to that of field players, who were found to run 11 km per game on average (Weimar & Wicker, 2017). Running is essential because referees must be well positioned on the field to improve the accuracy of their decisions (Mallo et al., 2012). For example, existing research showed referees have an attention window they must reach to make more accurate decisions, with differences in spatial attention also affected by physical activity levels (Hüttermann & Memmert, 2017). Referees' resulting positioning affects the quality of their refereeing. Referees' position on the field is also described in the official laws of the game provided by the International Football Association Board (n.d.). These guidelines, to which all FIFA member organizations adhere, indicate the relationship between player and referee running by urging match officials to adapt to tactics, game events, etc. To allow for gameplay to occur between the main referee and the respective assistant referee on the sideline, referees are generally advised to move in a flexible diagonal between the opposite corners of the penalty areas. Ideally, the referee maintains the same vertical position with respect to the player controlling the ball. Hence, it is fair to assume more linear distance covered by players correlates with more distance covered by the referee. Moreover, the work focused on the physical performance of match officials demonstrated the influence of players physical performance on referee demands (Mallo et al. 2007; Weston et al.; Weston, Drust, & Gregson, 2011, Weston et al., 2007). Players, however, also cover ground along the horizontal axes. Although referee running demands for this movement are less, referees are instructed to adapt in case gameplay is shifted toward the sidelines. Namely, they defer from their diagonal and move closer to the gameplay, which leads to more running demands. Collectively, these factors suggest referees' running varies with different playing and running styles of the respective players on the field.

Although match performance and outcomes can be deemed chaotic or unstable, analysis of match event and player tracking data may present a novel approach to explaining how referee movements transpire. Spatiotemporal data allows analysts to learn the game's underlying mechanics, such as style or player movements. As more teams and leagues collect spatiotemporal data, more work has leveraged the information the data sets provide. Bloomfield et al. (2007) used player tracking information to investigate the physical demands of soccer and the work rates of different players, and Gyarmati et al. (2014) leveraged ball-event data and passing sequences to identify the playing styles of different teams. They described a passing sequence by the number of unique players involved in the sequence and showed different teams use different sequences at different rates. Memmert et al. (2017) explored several applications of analyses of team coordination based on collective variables (e.g., surface area, centroid). The influence of these metrics on the physical output of the referee has yet to be explored, yet intuitively their physical behavior will in some way be tied to that of players during the match. Extracting the mechanisms and similarities between match officials' movement to players, positions, and team profiles could further inform profiling and training programs and guide assignment decisions based on the teams involved in a match.

Match officials at the domestic and international levels rely on subjective input to define performance. These subjective evaluations from subject matter experts undoubtedly have value and are deemed applicable; however, availability of objective player tracking, and event data present the opportunity to analyze performance using data-driven insights. The influence of contextual variables from the teams, players, and events could facilitate more objective insight when describing referee performance.

Clustering is often used as an initial means to learn structure in complex data, being successful in the fields of medicine and finance. Previous application of unsupervised learning techniques uncovered homogeneous subgroups by discovering underlying patterns and associations within multidimensional data sets allowing practitioners to base decisions on selected important features (Datta & Datta, 2003; Gomez et al., 2018. Clustering algorithms select similar data features, which associates to a cluster and can help evaluators or practitioners understand the influence of contextual variables on match officials' physical performance. However, before determining which clustering method best suits the application, practitioners must examine the performance and sensitivity of selected clustering approaches when determining feature selection and relevance to defining performance. Using a clustering approach as a first step in adopting an integrated approach, where player and team tactical behaviors assist in defining match official performance, is novel in nature and has yet to be examined.

This two-part exploratory study aimed to initialize an integrated approach to explain referee physical performance further and incorporates two unsupervised machine learning techniques to determine the degree to which player and referee tracking and event data cluster into homogenous subgroups.

6.3 Methods

Full event and tracking data from Metrica Sport (Amsterdam, NL), along with Statsperform data (OPTA, product of Statsperform, Chicago, IL) match event and descriptive data were collected from 559 matches, 24 teams, 711 players, and 25 match officials throughout the 2017–2018 seasons within MLS. Matches were played weekly, with a variation in the number of days between matches depending on individual team schedules. Only outfield athletes were analyzed, and positions were classified as center referee, attacking center midfielder, central defender, central forward, central midfielder, defensive midfielder, left back left midfielder, left winger, right back, right midfielder, and right winger. Referees do not have substitutions, so only players who completed 90 min or greater were included in the analysis.

6.3.1 Data Extraction

Upon completion of each match, event data along with one 25 Hz comma-separated value file for each team and referee were exported from proprietary software (Metrica Play, Amsterdam, NL) and analyzed further using the statistical software R (Version 3.6.1). R packages used for analysis included tidyverse (Version 1.3.2), factoextra (Version 1.0.7), and corrplot (Version 0.92). Data were cleaned to exclude half-time breaks, and match time was calculated from start to end of each half. BIP was calculated as the total time ball was in play, with the provided event data used to denote "out of play" (substitutions, injuries, kick-off, throw-in, goal-kick, corner-kick, free-kick, goals). Furthermore, not all games (71) were captured across both seasons due to logistical or staffing shortages.

For the 25 Hz movement data, a low-pass, second-order Butterworth filter was applied with a cutoff at .04 sampling rate, and displacement covered by each athlete was calculated using Euclidean distance. Velocity was then calculated as displacement divided by the sampling rate (.04). Any data point (46 in total) where velocity was greater than 10 $m\text{-}s^{-1}$ was removed because it was considered an error (Delaney et al., 2017). Acceleration was calculated as the absolute rate of change of speed ($m\text{-}s^2$). Impulse was used as a load measure of acceleration, which was calculated by multiplying the instantaneous absolute value acceleration by body mass and then multiplying by the 0.04 s (sampling rate) for each player in the match. The sum of impulse accumulated within each percentage of the maximal mean acceleration was then established to reflect the volume of acceleration or deceleration performed at different intensities (Johnston et al., 2022). A moving average technique was then applied to each output speed variable, using 10 different durations (i.e., 1–10 min), with the peak value achieved throughout each match for each variable recorded (Delaney et al., 2017).

6.3.2 Analysis

Data were separated into three data frames for analysis: (a) physical running metrics placed at the club level, where each center referee was considered their own club; (b) physical running metrics for each of the 12 positions; and (c) combined physical and event metrics for each club. To demonstrate differences of clustering approaches, separate K-mean and agglomerative hierarchical clustering techniques were applied with each analysis. In addition to the clustering techniques applied, I displayed the relationship between physical and event metrics using Pearson's correlation coefficient. This approach demonstrates a holistic perspective of the relationship of the physical and events data when describing referee performance. Additionally, clusters can provide valuable insights from patterns within a collection of different variables. Thus, providing a select number of groups with similar qualities can help to differentiate performances, or in this case, the differences amongst the three proposed analyses.

K-means clustering is an unsupervised algorithm that can be explained simply because the number of clusters is unknown, and the goal is to discover similar patterns in the data. The goal of K means is to minimize the sum of square differences within groups while maximizing the sum of square differences between groups (Sweeting et al., 2017; Wagstaff et al., 2001). The algorithm aims to partition observations into a set of "*k*" clusters. Subsequently, each observation belongs to the cluster with the nearest mean. The following steps summarize the operations of K means. In practice, this can be done by randomly selecting the k center; however, the choice of k is usually influenced by prior knowledge of the nature of the data or by using clustering validity measures (Alashwal et al., 2019). Relying on previous knowledge of the data through daily work and observations within this environment I was able to combine this perspective and further utilized agreements from both the silhouette and WSS methods to determine our number of clusters.

Hierarchical clustering technique (Banfield & Raftery, 1993) can be applied to determine homogeneous groups. Hierarchal agglomerative clustering is a bottom-up approach, so each data point begins in a separate cluster, and pairs of clusters at the bottom are merged as they move up the hierarchy. Unlike K-means, the number of clusters is not predetermined; generally, smaller clusters are generated. Metrics for every player and center referee were scaled to all available data across both seasons, and Euclidean distance was used as the dissimilarity measure (Shirkhorshidi et al., 2015; Singh et al., 2013). In the initial data exploration of the agglomerative hierarchical cluster technique, four methods (i.e., complete, single, average, and Ward and Hook's (1963) minimum variance) were applied. To determine the appropriate number of clusters (K), gap, silhouette, within-cluster sum of squares (WSS), and the Calinski and Harabasz index (variance ratio criterion) were evaluated. I chose our number of clusters again by using agreements from both the WSS and silhouette methods. Lastly, dendrograms and descriptive statistics of each cluster were plotted to explore the determined clusters further.

6.4 Results

6.4.1 Club and Referee – K-Means

Using only physical data, a model of four clusters was suggested by the aforementioned silhouette method. Therefore, a K-means cluster analysis was performed to create two clusters comprised of only the match officials and two were only clubs (see Figures 12 and 13). The means of physical variables according to cluster are shown in Table 3. Alternatively, five clusters were assigned when only looking at match officials. Interestingly, only only one cluster represented more than 5 referees.

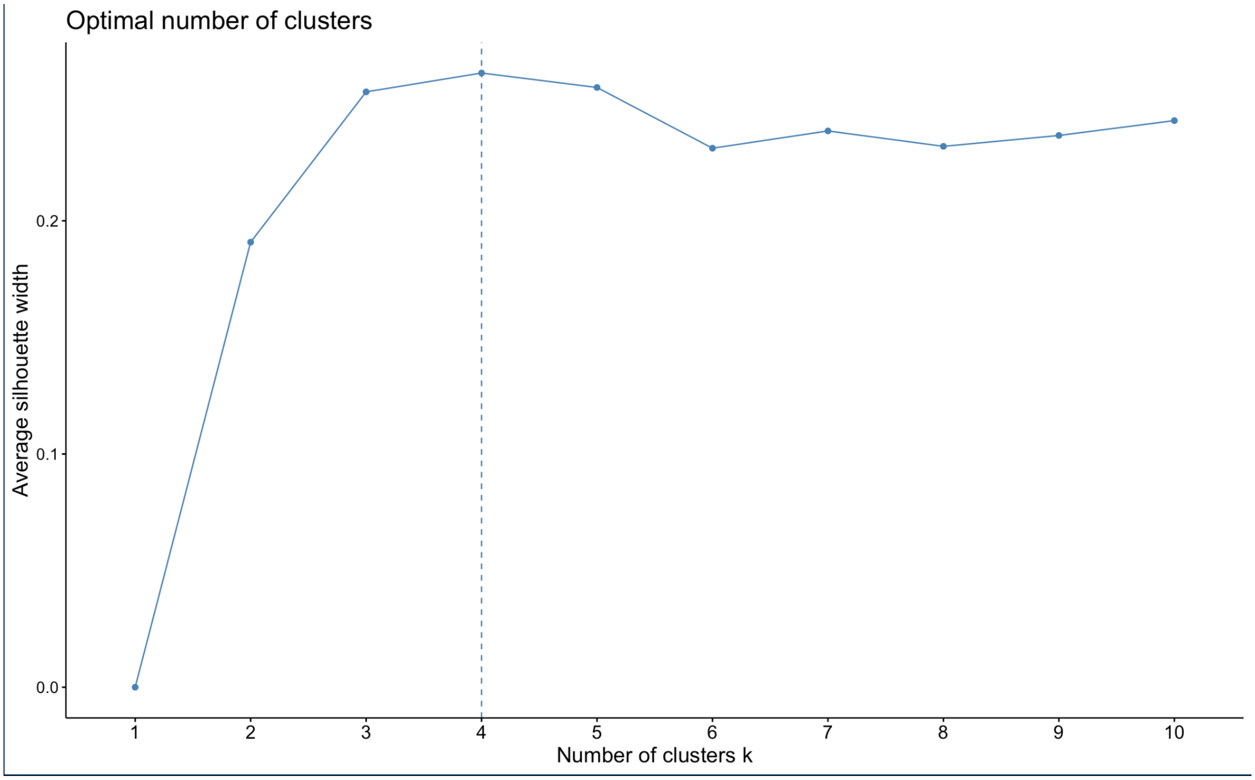


Figure 12. Silhouette Method Results for Determining Optimal Number of Clusters. Utilizing only physical data, four total clusters were suggested - two clusters for clubs and two clusters for match officials

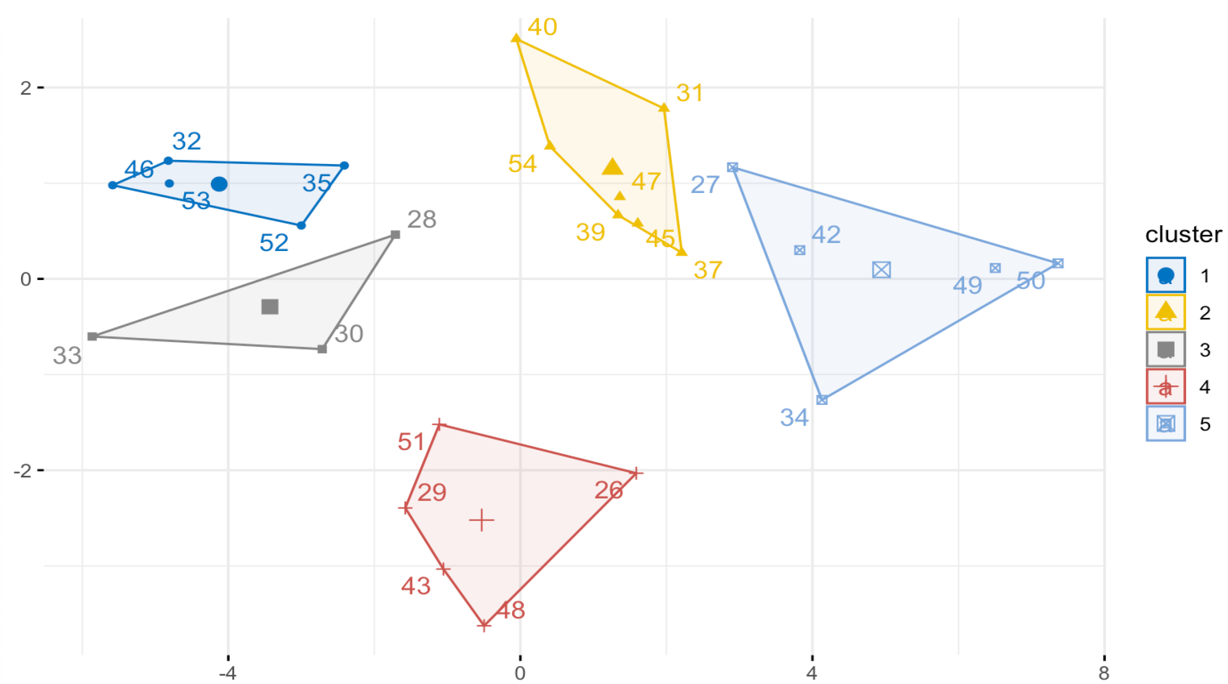


Figure 13. Results from the K-*Means Cluster* Algorithm *for Referees. Five distinct clusters were represented with no less than 3 referees per cluster. The x and y axis provide the distance measurement*

Table 3 *Mean Value and (Standard Deviation) of Physical Metrics by Cluster*

Cluster	Mean total distance (m) (SD)	Mean HSR (m) (SD)	Max accell (m/s) (SD)	Mean max impulse (SD)	Mean impulse by event (SD)	Mean m/s (SD)
1	10242 ± (424)	255 ± (208)	7.02 ± (.08)	197344 ± (2651)	419 ± (354)	1.76 ± (0.44)
2	10552 ± (414)	285 ± (231)	7.0 ± (.07)	208152 ± (3819)	455 ± (366)	1.81± (0.59)
3	9778 ± (393)	170 ± (111)	6.05 ± (1.02)	182156 ± (1964)	445 ± (341)	1.68 ± (0.29)
4	10683 ± (437)	224 ± (188)	6.02 ± (1.09)	203589 ± (3017)	498 ± (382)	1.83 ± (0.66)

Cluster 2 displayed the highest values for each metric presented, and Clusters 1 and 2 were related closely across all metrics. Notably, total distance was higher for the referee cluster, Cluster 4, than all other clusters.

6.4.2 Club and Referee – Hierarchical

A hierarchical analysis was used to identify subgroups with homogeneous physical outputs. A hierarchical cluster tree is displayed in Figure 14. Four total groups were identified, with match officials comprising two and the clubs representing the other two clusters.

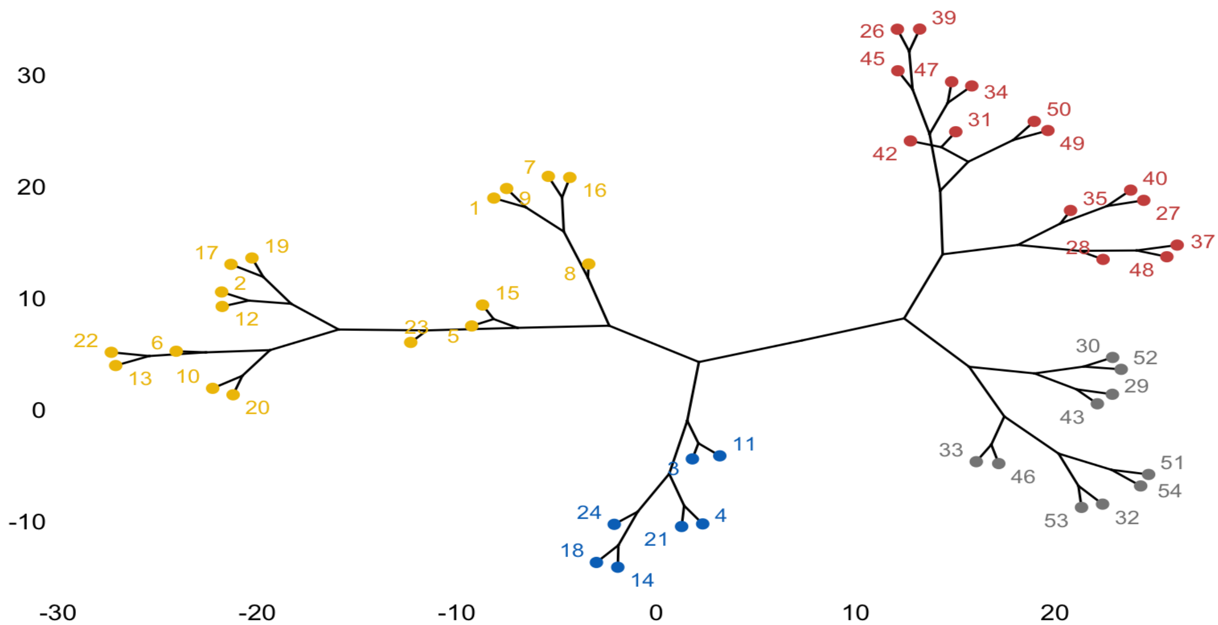


Figure 14. Hierarchical Cluster Breakdown by Club. *Numbers represent the club and individual referee number. Yellow blue comprised of teams, while red and grey are match officials. Distance values along the x and y-axis.*

6.4.3 Club and Referee – K-Means and Hierarchical

Plotting both clustering approaches, mean and spread are displayed for each scaled physical output. Identified K-means and hierarchical clusters are notated along the *x* axis and color, respectively. Both methods returned similar finding with regard to cluster assignment (see Figure 15).

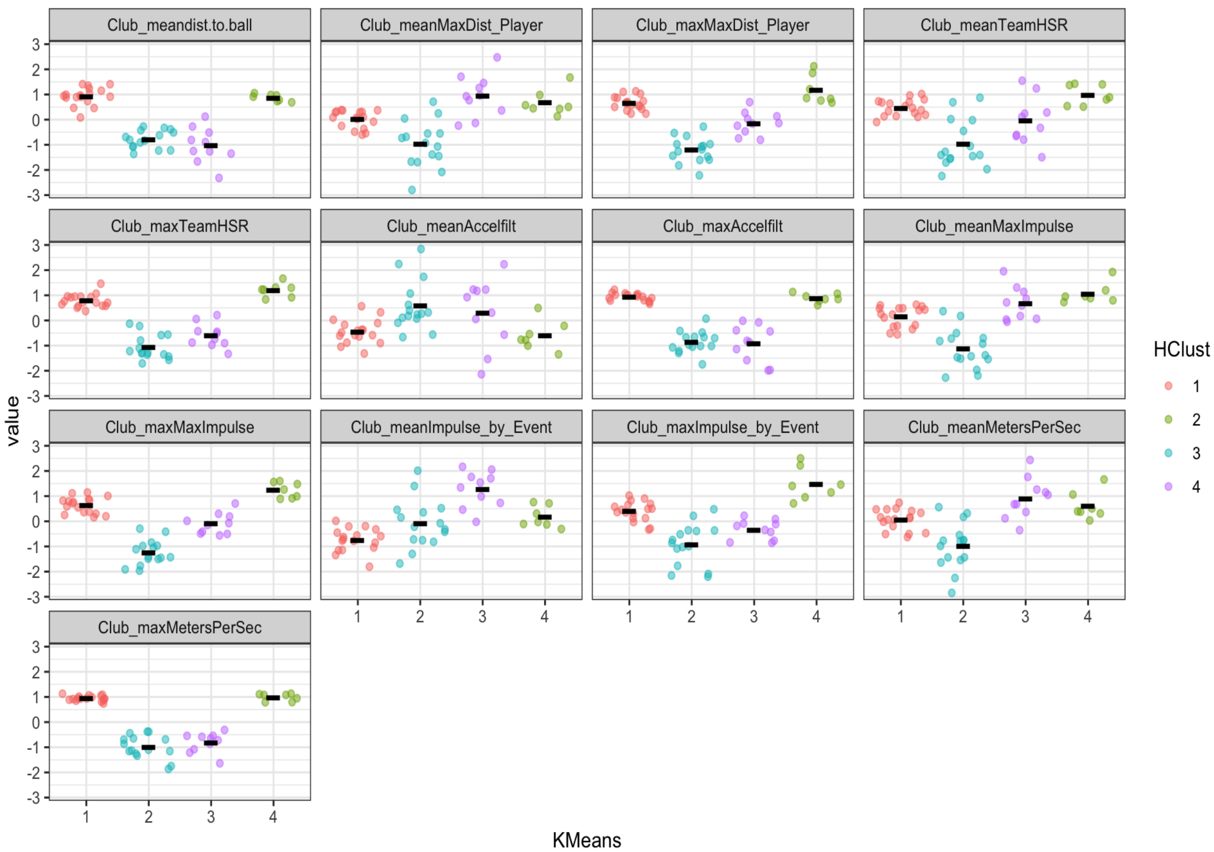


Figure 15. Plot of Both K-Means and Hierarchical Cluster Approaches Together. *Hierarchical cluster represented by color, K-Means cluster number is along the x-axis and distance on the y-axis.*

6.4.4 Referee – K-Means and Hierarchical

Five clusters were assigned for referee physical data, and no disagreements of groups were seen when plotting clustering approaches together. Clusters were comprised of a range of 2 to 8 referees. (see Figure 16).

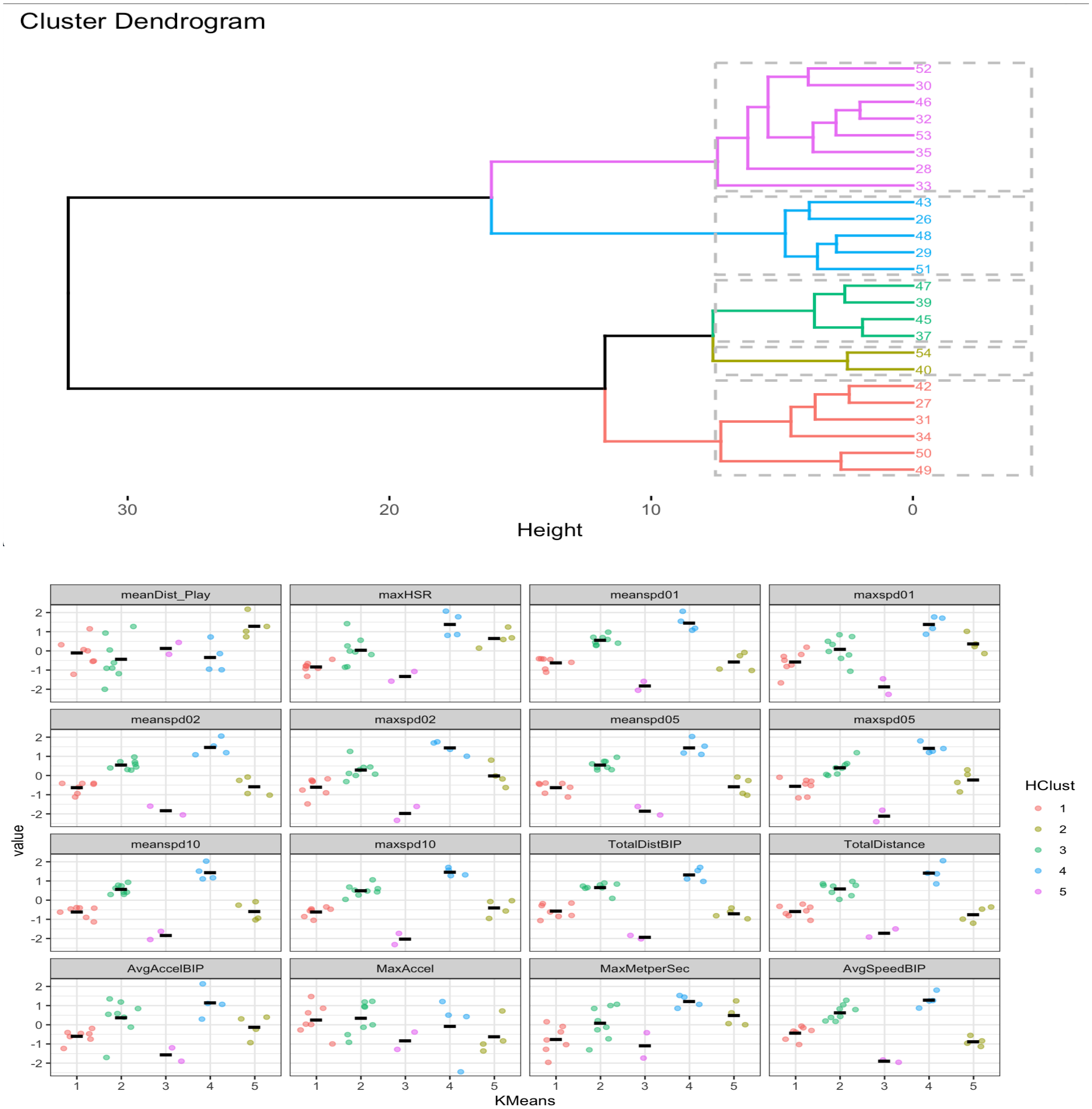


Figure 16. Dendrogram of the Hierarchical Groups and Combined Approaches. *Utilizing a hierarchical cluster algorithm (top), linkage distance on the x-axis and the individual match officials on the y- axis. Comparison on Assigned Cluster with Both Approaches(bottom)*

6.4.5 Position and Referee – K-Means

Using K-means, three distinct clusters were determined when comparing positions league wide, with the referee grouped with the central forward, right wingers, and central defenders (see Figure 17).

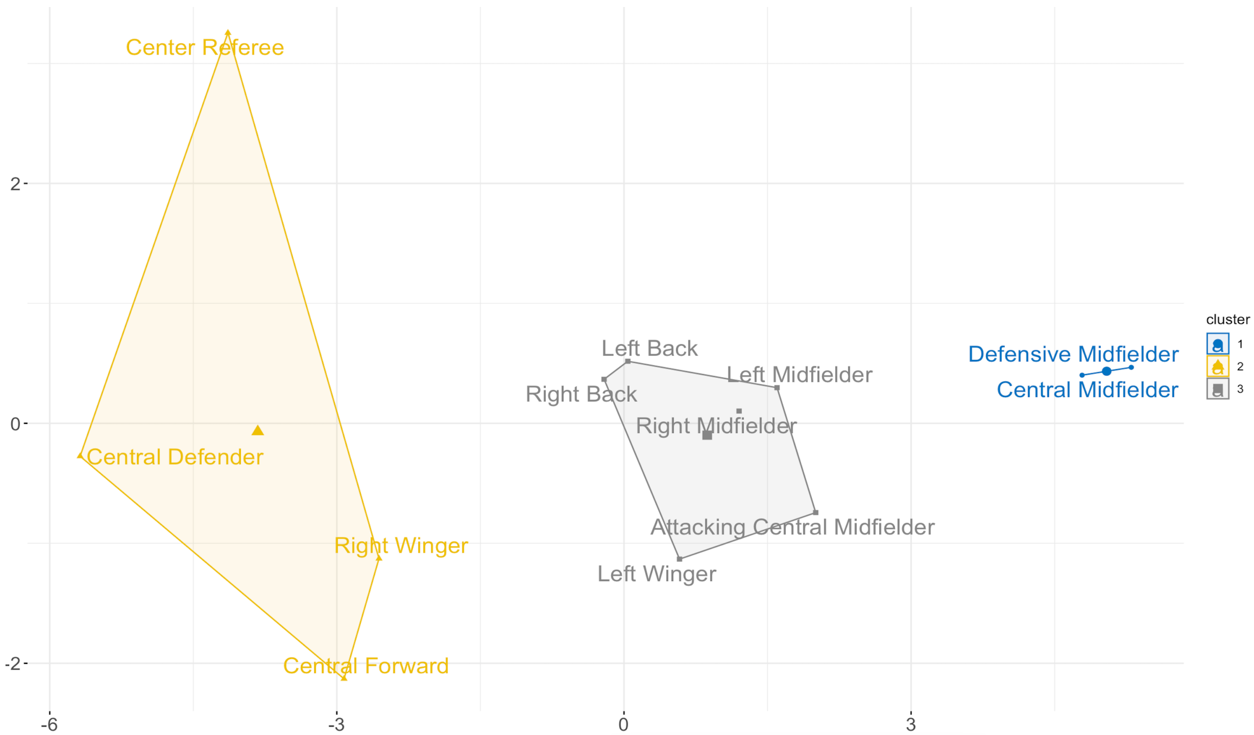


Figure 17. K-Mean Cluster by Position. *Three clusters, Yellow include center referees, central defender, central forward, and right wingers, Grey include right and left back, right and left midfielders, left wingers, and attacking central midfielders. Blue includes two positions, defensive and central midfielders.*

6.4.6 Position and Referee – Hierarchical

Similar groupings and positions for each grouping were found when applying hierarchical clustering (see Figure 18).

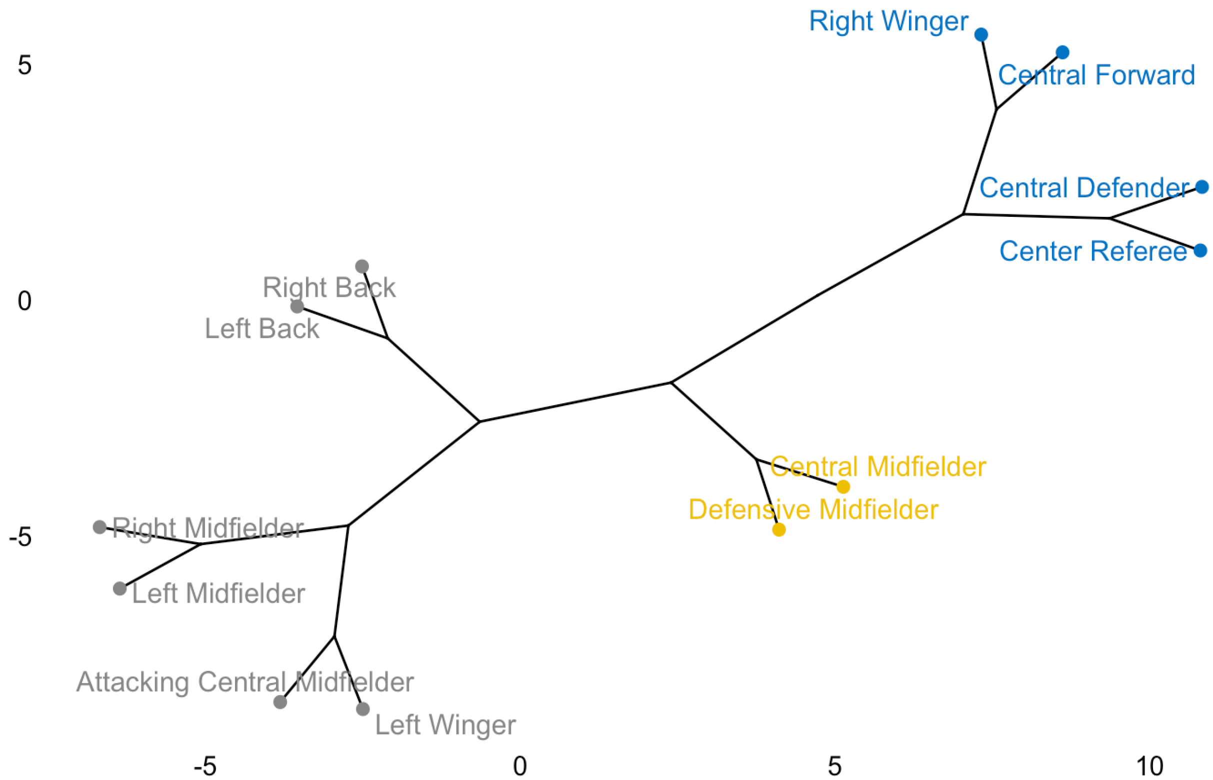


Figure 18. Hierarchical Cluster by Position. *Three clusters with blue including center referees, central defender, central forward, and right wingers. Grey including right and left backs, right and left midfielders, attacking Central Midfielders and Left Wingers. Yellow included central and defensive midfielders.*

6.4.7 Position and Referee – K-Means and Hierarchical

Both K-mean and hierarchical positional clusters returned similar findings when separated by physical trait (see Figure 19).

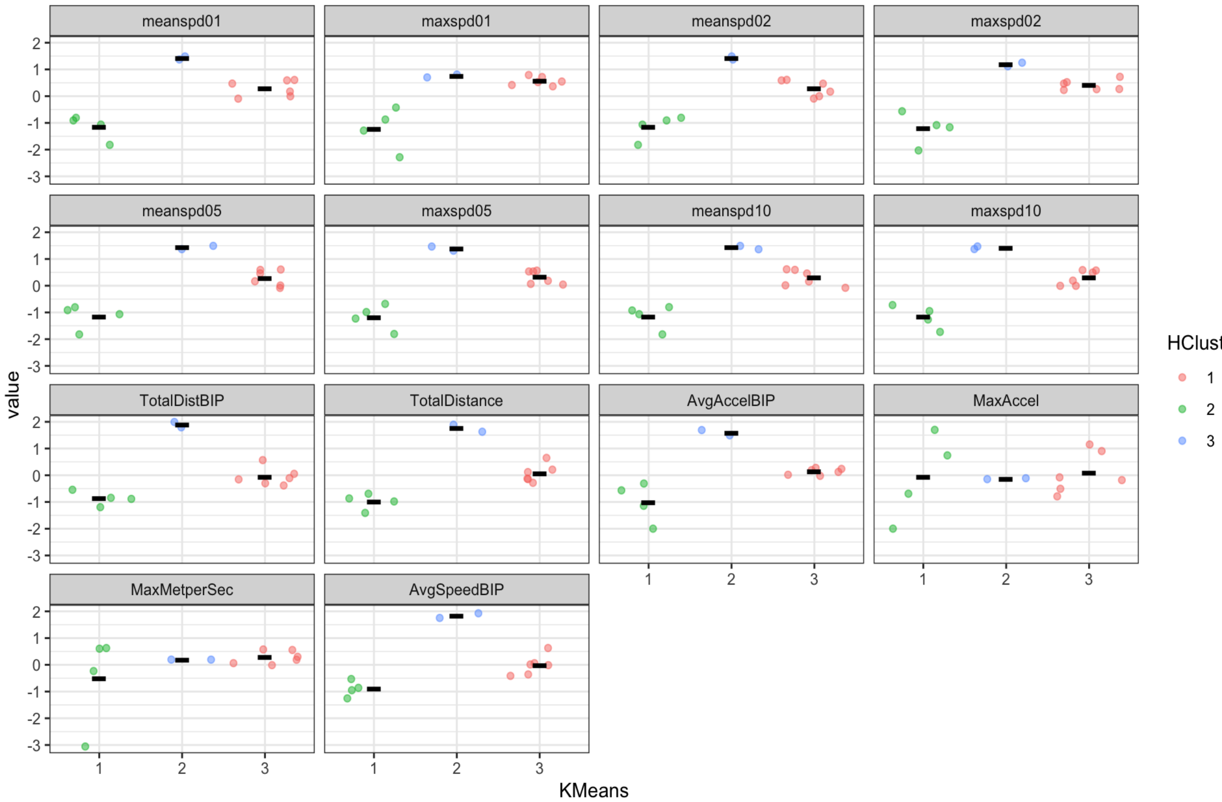


Figure 19. K-Means and Hierarchical Cluster by Physical Trait. *Numbers along the x-axis represent K-Means cluster grouping*

6.4.8 Position by Club and Referee – K-Means

Four clusters for positional physical demands when comparing position by club are shown in Figure 20.

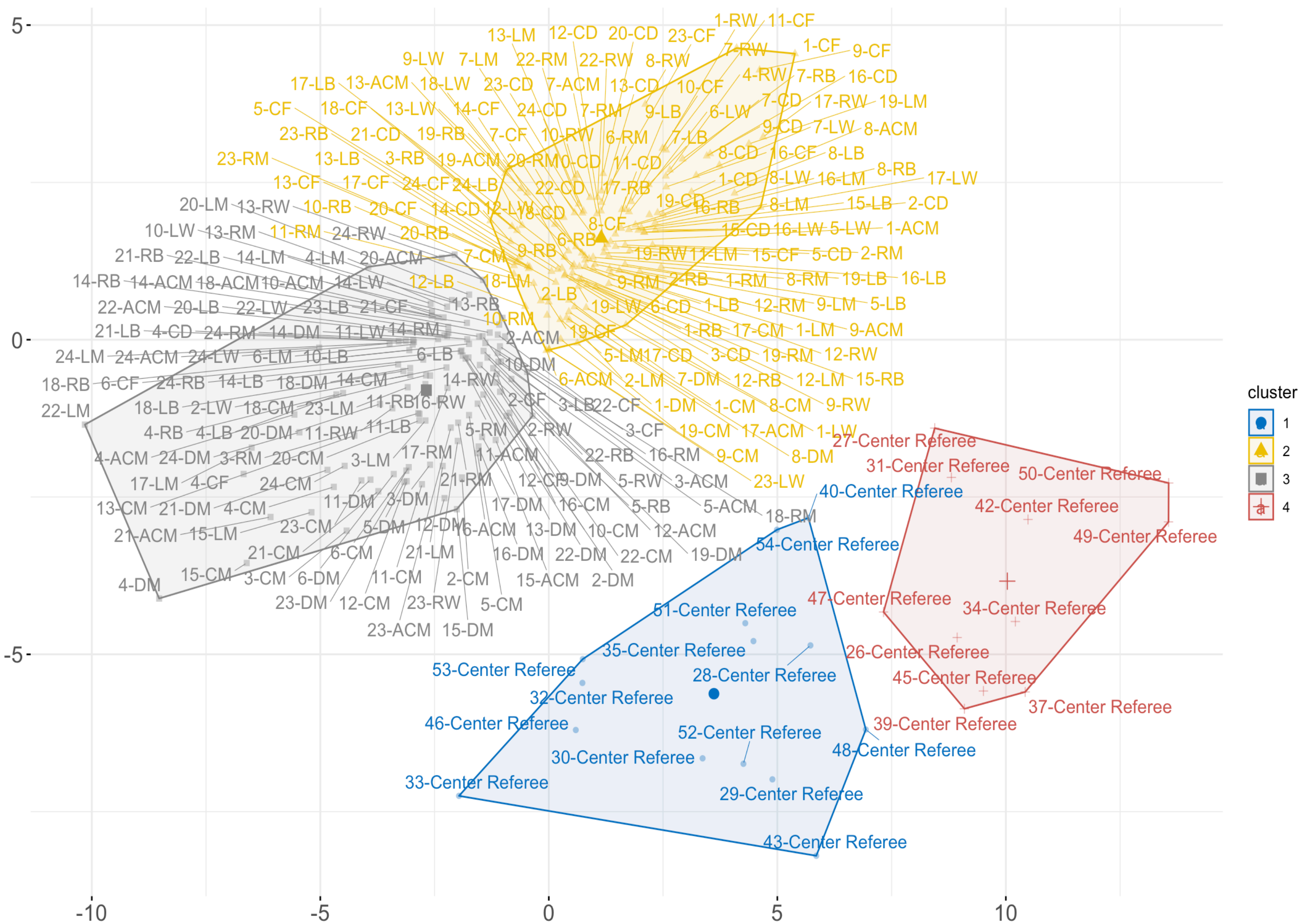


Figure 20. K-Mean Cluster by Position by Club. *Red and blue represent center referees individual ID numbers, while grey and yellow are field players by each club ID number and position. CD – central defender, LB – left back, RB – right back, LM – left midfielder, RM – right midfielder, ACM – attacking central midfielder, CM – central midfield, DM- defensive midfielder, CF – central forward, LW – left winger, RW – right winger*

6.4.9 Position by Club and Referee – Hierarchical

Despite similar naming conventions, clustering presented different clustering by position and team (see Figure 21).

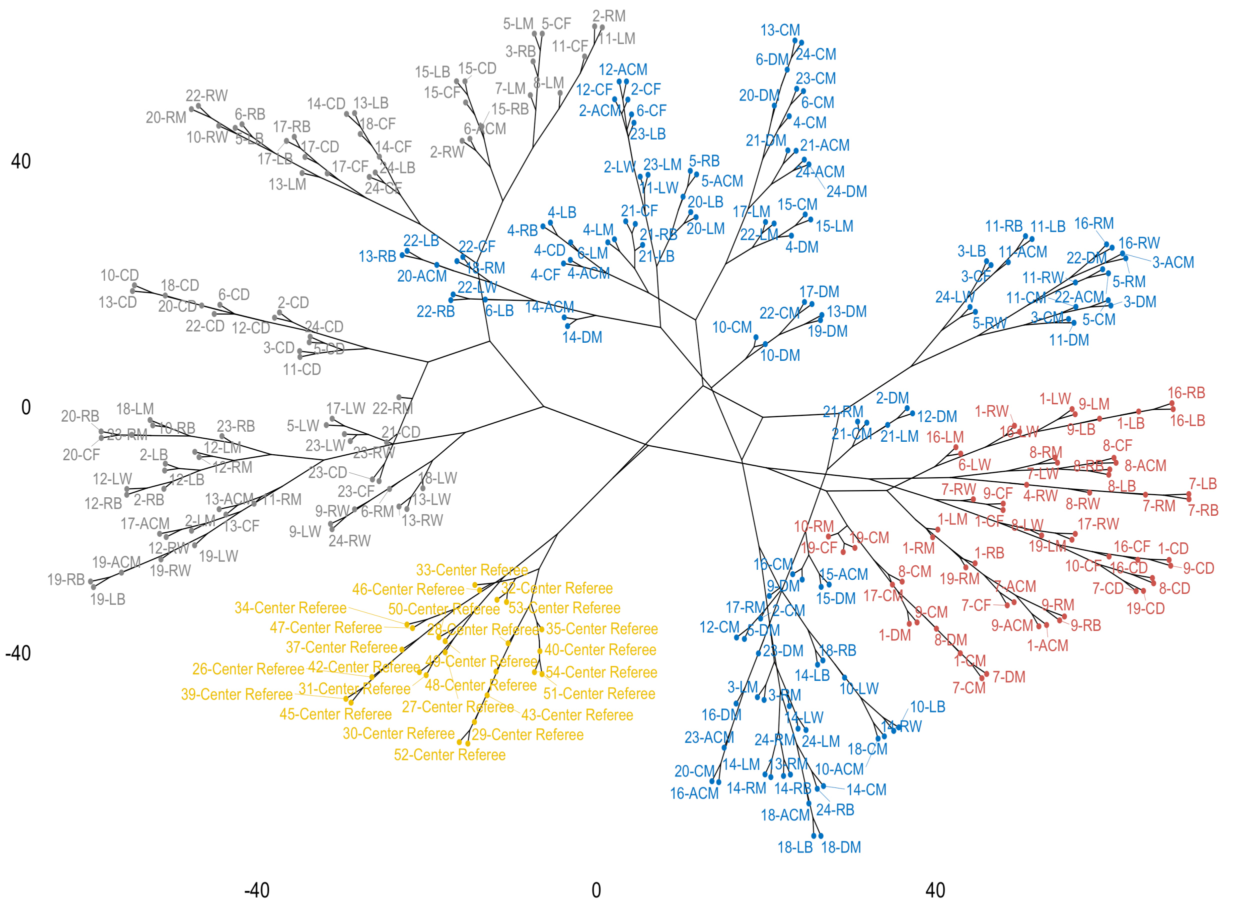


Figure 21. Hierarchical Cluster by Position and Team. *Yellow represents all referees while grey, blue and red are field players.*

6.4.10 Position by Club and Referee – K-Means and Hierarchical

Presenting positional differences by physical trait, demonstrated similar finding between cluster approaches (see Figure 22).

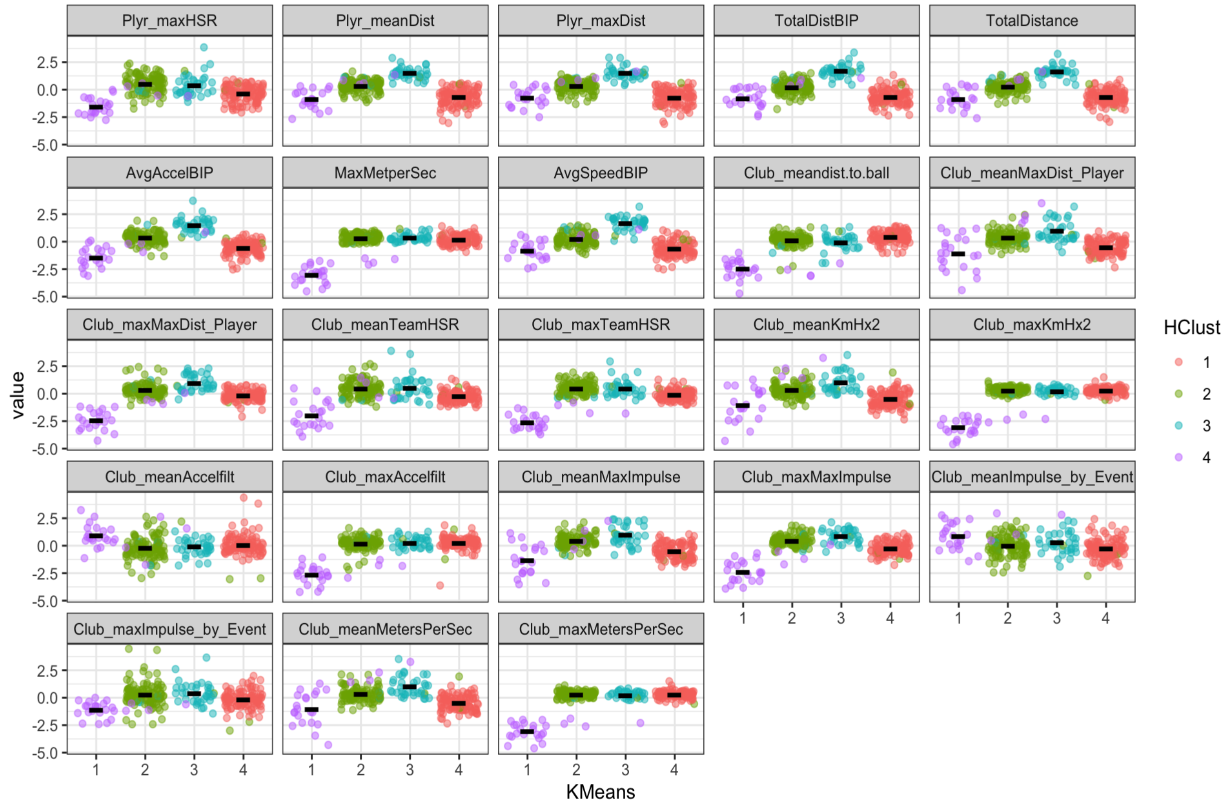
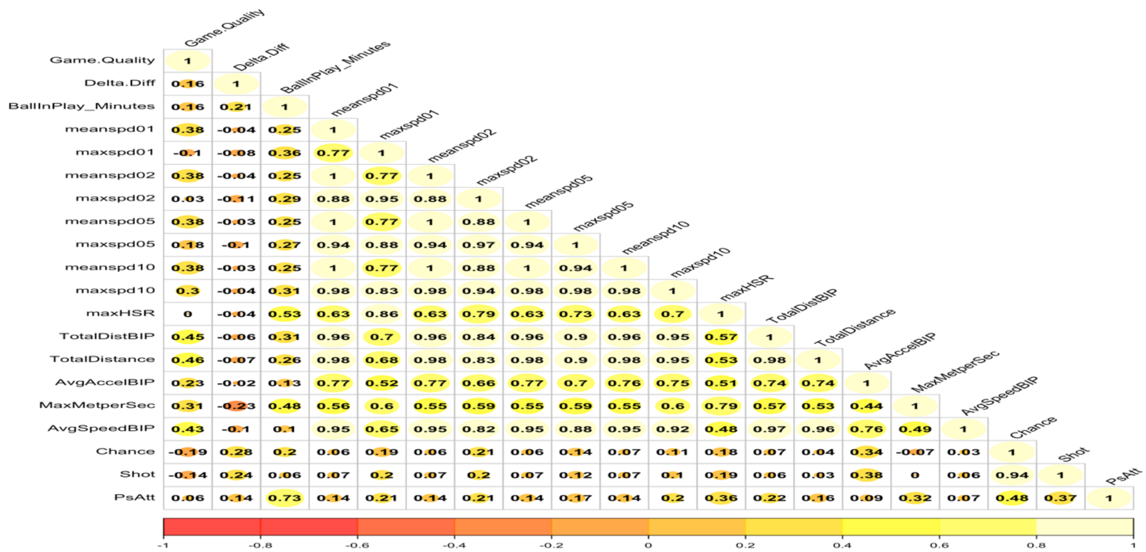
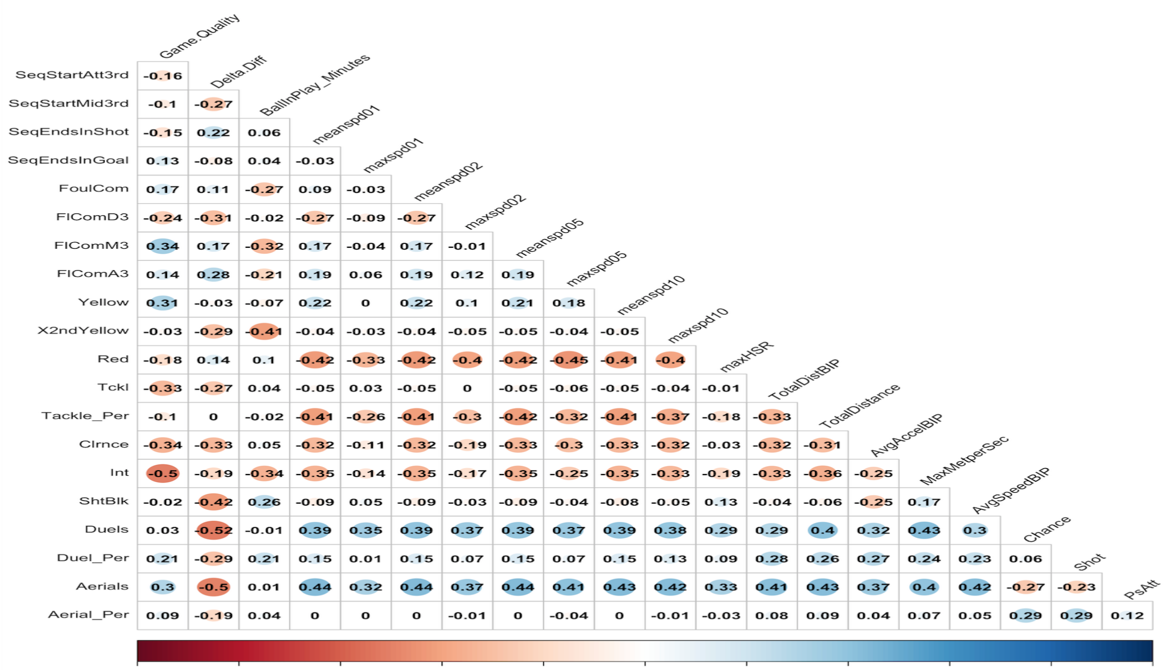


Figure 22. K-Means and Hierarchical Cluster by Position by Physical Trait. *Color represents the hierarchical cluster group while the x-axis compares the K-Means cluster group*

Match official data presented strong correlations between physical metrics and event data. Game quality demonstrated stronger positive relationships to Average Speed BIP, Total Distance, and (.43, .46, .48), and number of ariels (.42, .43, .44) showed positive correlation to Average Speed, Total Distance, and Most Demanding 1-min distance, respectively. Duels showed positive correlations to Total Distance (.4) and Meters/Sec (.43), while red cards, clearances, and interceptions had a negative correlation to all physical metrics (see Figure 23).



Game Quality, Delta difference in rank, BIP minutes, 1-min mean speed, 1-min max speed, 2-min mean speed, 2-min max speed, 5-min, 5 min max speed, 10-min mean speed, 10-min max speed, max HSR, Total Distance while BIP, Total Distance, Mean Acceleration while BIP, Max m/sec, Mean Speed while BIP, Scoring Chances, Shots, and Passing Attempts



Attacking starting in the attacking 3rd, Attack starting in the middle 3rd, Attacking ending in a shot on goal, Fouls commited, Fouls committed in the defending 3rd, Fouls committed in the middle 3rd, Fouls committed in the attacking 3rd, Yellow cards, 2nd yellow cards, Red cards, Tackles, Tackles per min, Clearances, Interceptions, Shots blocked, Duels, Duels per min, Aerial challenges, Ariel challenges per min.

Figure 23, Correlation of Event Data and Only Referee. *Data definitions for abbreviations provided for top and bottom*

6.4.11 Event, Club, and Referee – K-Means

When clustering by event and physical data, 3 of the 4 clusters included referees (see Figure 24).

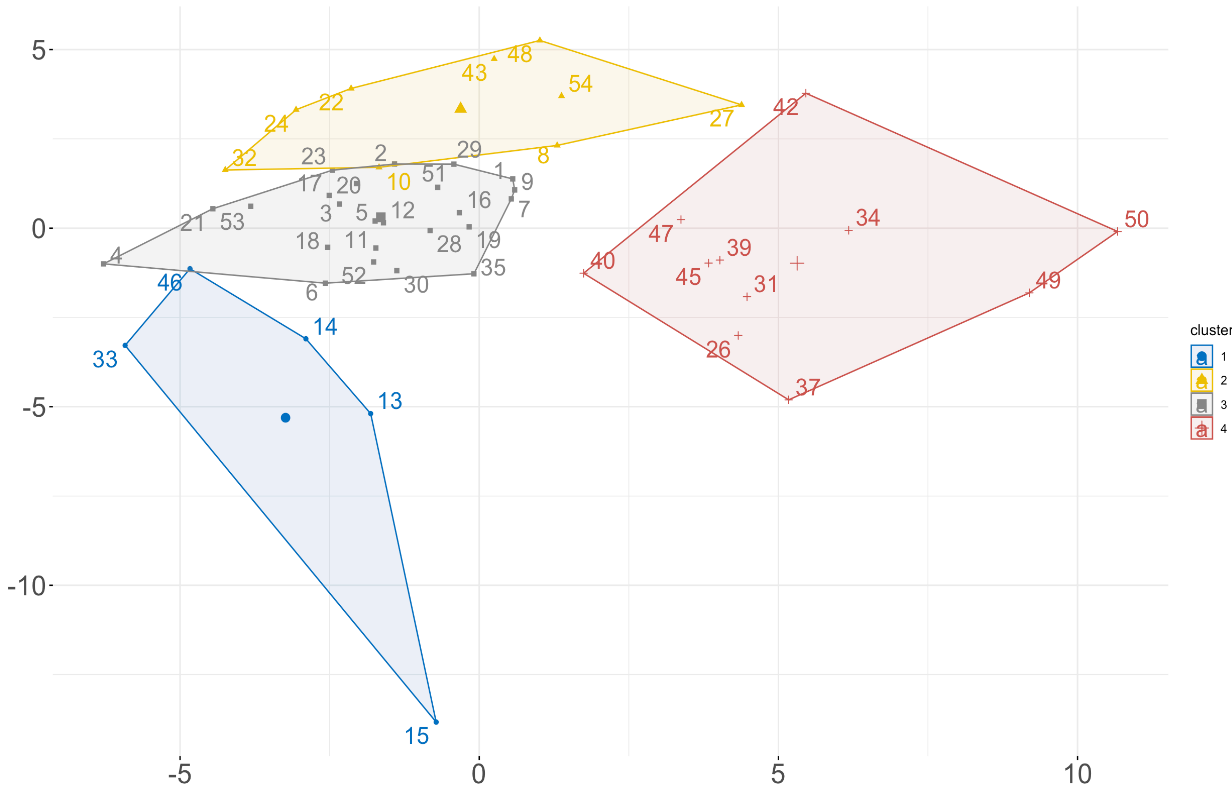


Figure 24. K-Means Cluster of Clubs Including Referees. *Number represents club ID and individual referee ID, color represents cluster group*

6.4.12 Event, Club, and Referee – Hierarchical

Labeling the referee as a club, clustering of clubs using all features presented few teams with features similar to referees (see Figure 25).

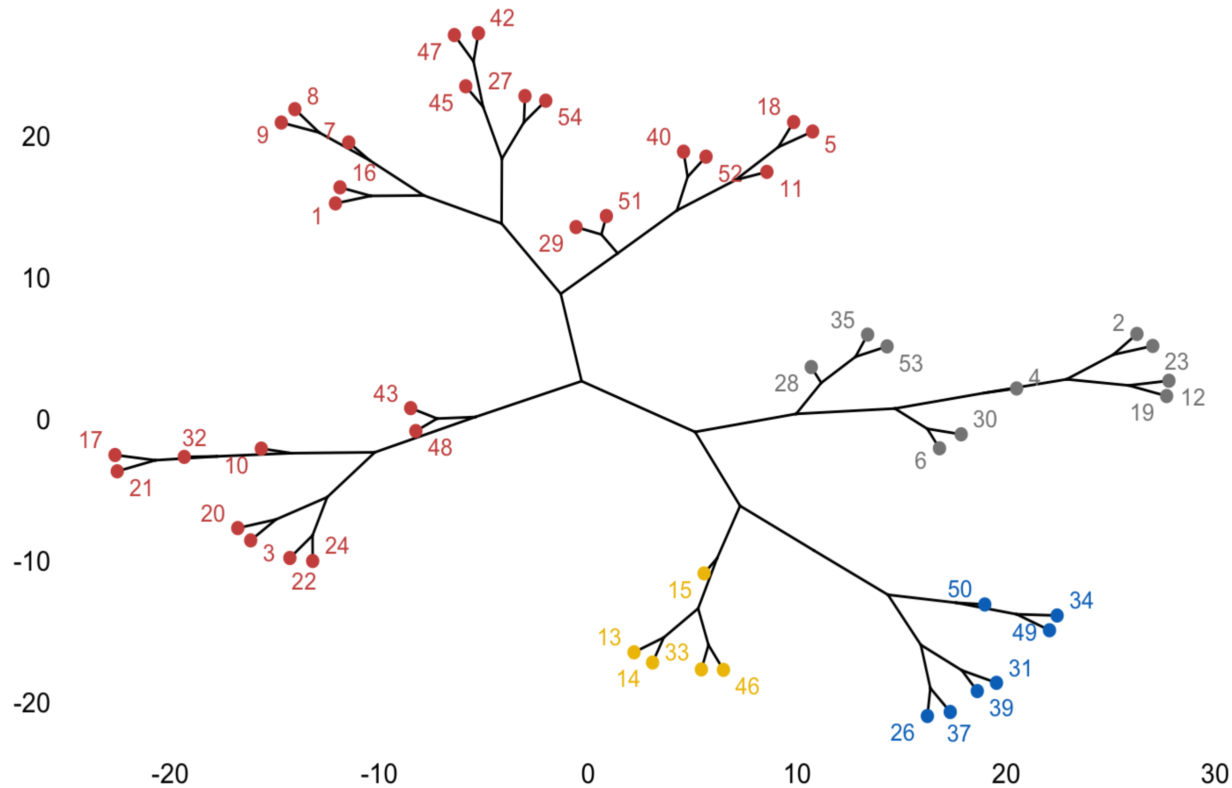


Figure 25. *Hierarchical Cluster of Clubs Including Referees. Colors represent each individual cluster (Red, Yellow, Blue, and Grey)*

6.4.13 Event, Club, and Referee – K-Means and Hierarchical

When comparing clustering techniques for club and referee data and using physical and event data, disagreements appeared across multiple metrics (see Figure 26).

Figure 26. Club Comparison of Unsupervised Approaches. *K-Means and Hierarchical Cluster of Clubs Including Referees*

6.4.14 Event and Referee – K-Means

When clustering by event and physical data for referees only, four clusters were identified (see Figure 27).

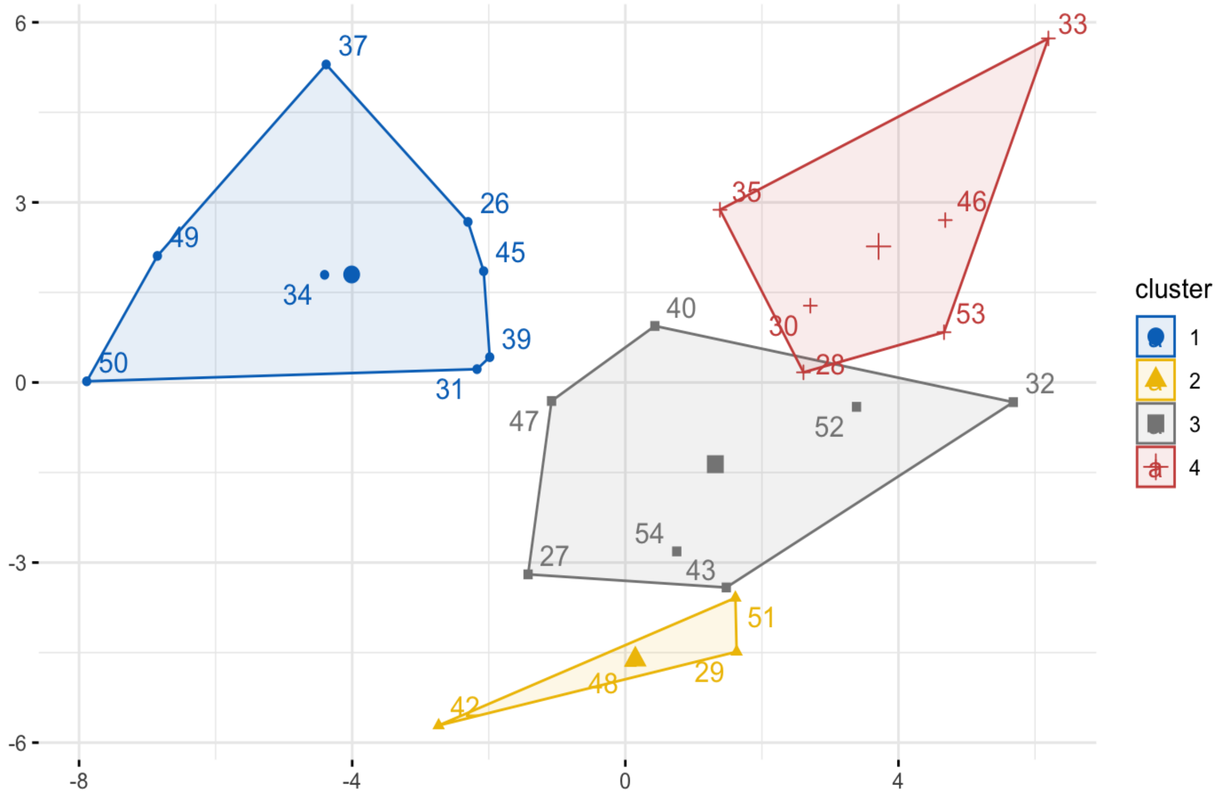


Figure 27. Referee Physical and Match Events Unsupervised Approach. *K-Means Cluster Only Including Match Officials*

6.4.15 Event and Referee – Hierarchical

Figure 28 presents a hierarchical cluster of match officials and those with similar features.

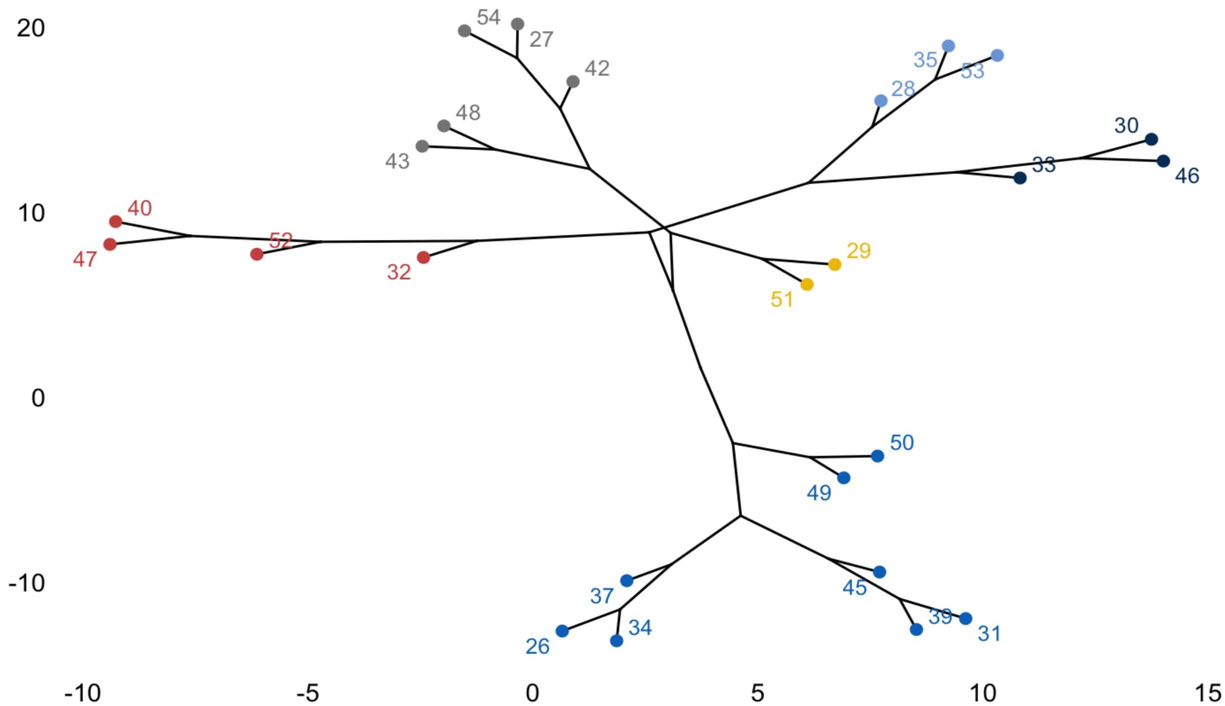


Figure 28. *Hierarchical Cluster Only Including Match Officials*

6.4.16 Event and Referee – K-Means and Hierarchical

Only using match official data presented disagreement between methods while using event and physical data (see Figure 29).

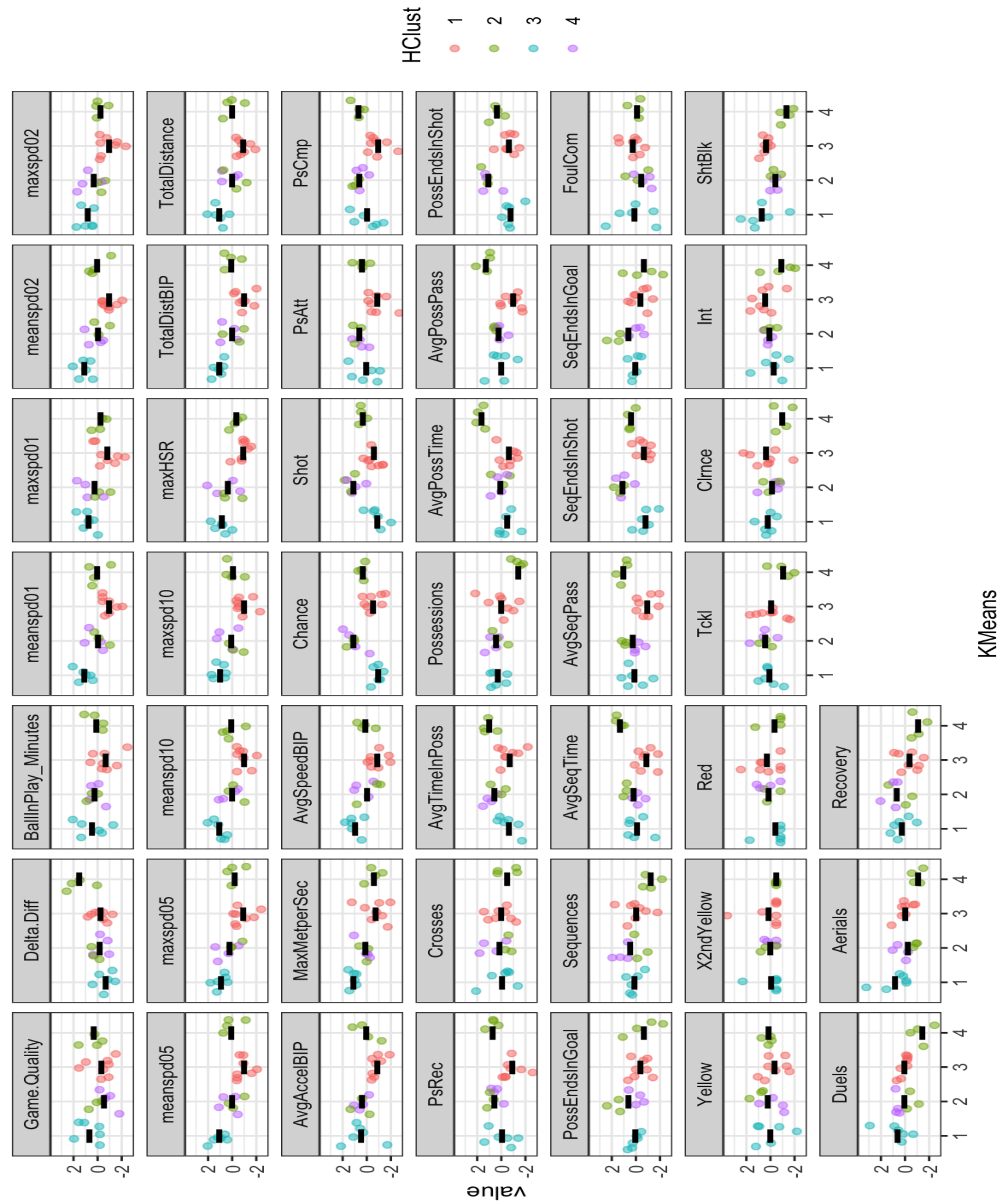


Figure 29. K-Means and Hierarchical Cluster Only Including Match Officials. *Hierachical clusters are represented with colors, while K-Means clustes are indicated on the x-axis.*

6.5 Discussion

This study aimed to implement an integrated approach to explain referee performance by applying two different machine learning techniques on tracking and event data to determine if homogeneous subgroups could be extracted. I explored three different data sets: (a) the first two evaluating physical metrics only by club and position and (b) the final approach pairing both physical and event data. Although only comparing physical data, homogeneous subgroups were found regardless of clustering approach. However, when implementing both K-means and hierarchical clustering using event data and physical output data, despite correlations existing within referee outputs and event data, disagreements occurred between methods. These disagreements were expected considering the differences in calculations between methods. Across sport, the industry has seen an increase in the application of machine learning; however, these disagreements highlight the importance of understanding the differences when using clustering. I found implementing a single clustering approach to evaluating a match official's movement patterns and adjustments within one match or across multiple matches and seasons could be deemed appropriate when trying to compare performances. However, interactions between match officials and teams could inform analysis of a solo referee performance based on differences of selected features within and across referee clusters while providing detail on the influence of the physical performance of players or club and events of a match. Determining similarities in physical and technical components of performance of both match officials and clubs provides an opportunity to further enhance the determination of assigning a match official to a match based on the teams and potential demands required. Pairing a match official to a match based on the cluster grouping would be a novel approach to match assignment. Applying these findings could support organization decision making for match assignment and training topic selection as well as inform analysis when defining performance and finding similarities within referees and between referees and clubs.

Although using different clustering techniques is not new to the sport of soccer (Beernaerts et al., 2020; Park et al., 2019), this is the only study grouping referee and player tracking and event data with the purpose of describing match performance while allowing for practitioners to inform referee match assignment based on similarities among clusters. Furthermore, understanding the "why" behind referee performance is imperative to evaluation, and these findings could be used to better determine a match officials' physical performance with relation to the demands dictated by teams playing the match. Additional application of these findings could support a deliberate focus on specific training topics with the intention to improve future performance.

Isolating analysis on physical metrics provided insight into how match officials, players, and clubs were related. With the inclusion of club and match officials, four clusters were identified but were comprised distinctly as two club-only clusters and two separate match officials' clusters. When only investigating match officials, five separate clusters were presented. These findings would suggest assessing referee performance should reflect the performance metrics from their own cluster rather than the current practice of evaluating based on generalized expectations of subject matter experts or organizations responsible for evaluation. Considering a comparison by position, three separate clusters were identified. Interestingly match officials were clustered with central forwards and central defenders. These data would seem logical when considering the diagonal positioning referenced previously is similar to both positions respectively. Lastly, a variation of position performance by team revealed two separate clusters for players and two separate match official clusters.

Conversely, although both clustering approaches demonstrated similar findings across physical metrics, disagreements occurred when comparing methods on event and tracking data. When considering which approach to use, practitioners should investigate which method most clearly answers relevant questions. Despite disagreements regarding cluster assignment, both methods found 3 of the 4 clusters included match officials mixed with clubs. When considering match official assignment, this key takeaway provides evidence to suggest similarities of match officials and select clubs.

6.5.1 Limitations

This study reflects overall trends among match officials, teams, and players. One current limitation when applying a K-means algorithm is the stability of the results. Coupled with the transient nature of sport where players are in and out of lineups, ladder position changes, and a resultant tactical change, this stability could be compromised further. However, these findings did see similarities in match officials, and this provides reason to explore further contextual factors in future research.

6.6 Conclusion

In conclusion, this study demonstrates the importance of determining how match officials are evaluated. When solely considering physical metrics differences, these findings could help with training decisions; however, when applying these findings to assignment, it appears considering event data and physical metrics would deem more appropriate pairing the match officials who best reflect the demands of the teams in a match. More research should be conducted to implement these event data and tracking data to understand how different machine learning approaches could enhance or inform assignment decision making further.

Chapter 7: Development of a Framework for Match Assignment

7.1 Abstract

Team selection in professional sports is a fundamental component in determining performance. Factors contributing to building a roster and playing time decision making are complex and vast; however, the singular goal is winning. Referee assignment presents a challenge as performance evaluation is rooted in immense subjectivity, and the underpinnings of high performance are yet to be determined. The purpose of this paper was to objectively address problem of referee selection deemed assignment through machine learning. RF, k-NN, and SVM have been used as practical machine learning algorithms for classification and prediction. However, application of supervised learning for prediction and classification in sports is still in its infancy and may be appropriate for assigning soccer match officials. The present study performed a hierarchical cluster technique and then predicted match type and match officials' clusters from a set of physical, technical, and spatial variables trained on a variation of supervised learning algorithms. Data were collected from 24 teams and 25 match officials competing in 559 games across two seasons in Major League Soccer (MLS). Match classification was at acceptable levels of 0.79 or higher for Accuracy, Balanced Accuracy, and F1 score regardless of the machine learning approach. Referee classification had the highest rates for Balanced Accuracy, and both referees by match and the home team level presented balanced Accuracy of 0.68 and above. These findings could be valuable to governing bodies, coaches, and sports managers to recognize combinations of match physical and technical indicators to inform decision making and save time and energy in assigning match officials.

7.2 Introduction

International tournaments, competitions, and worldwide domestic leagues assign soccer match officials weekly. Competition type determines length of time match fixtures will be announced, with challenges arising from the mixture of competitions happening simultaneously. For any such competition, a referee, two assistant referees, a fourth official, and, more recently, a video assistant referee (VAR) require assignment decisions from governing bodies. Similar to a team environment, match official crews are selected by management-level staff from a pool of match officials, with decisions most often subjective, biased, and subject to error. Opportunities exist in automating the selection process based on match officials' performance measures and forecasting alternative officials' performance. While previous work in this project implemented an unsupervised machine learning approach to better understand performance, this chapter implemented both supervised and unsupervised machine learning techniques to identify teams, players, and match officials' trends, and to pair the appropriate match fixture at the right time to promote high performance. Combining these approaches to better predict a referee performance has yet to be explored.

Although scant attention has been given to referee assignments, researchers have presented findings regarding scheduling of matches or fixtures in domestic and international leagues (Alarcón et al., 2014; Durán et al., 2012; Ribeiro, 2012). The emphasis of the limited assignment research has revolved around workload (Alarcón et al., 2009; Durán et al., 2012), fairness in assignment decisions (Yavuz et al., 2008), and travel (Durán et al., 2019; Trick et al., 2012). The premise for which these findings were presented has shown successful implementation regarding solving logistical objectives (Durán et al., 2017).

Using wearable devices or implementing optical tracking technology and the subsequent immense amount of data captured have generated an emphasis on improving evaluation and optimization of player and team performance. Historical evaluation of players has adopted a siloed approach to identifying trends or potential predictors of performance from physical or technical data (Bradley & Ade, 2018). It requires a more holistic view into determining the contextual drivers of good compared to bad performances. Conversely, referee performance presents a challenge when establishing high performers because research on what determines high performance needs to be more conclusive. Three separate investigations (Aragão e Pina et al., 2019; Loghmani et al., 2021; Slack et al., 2013) presented mixed findings when classifying successful match officials with game management as the single consistent characteristic across all three studies. Like evaluating the drivers behind players' performance, match officials have yet to adopt a holistic perspective when explaining performance.

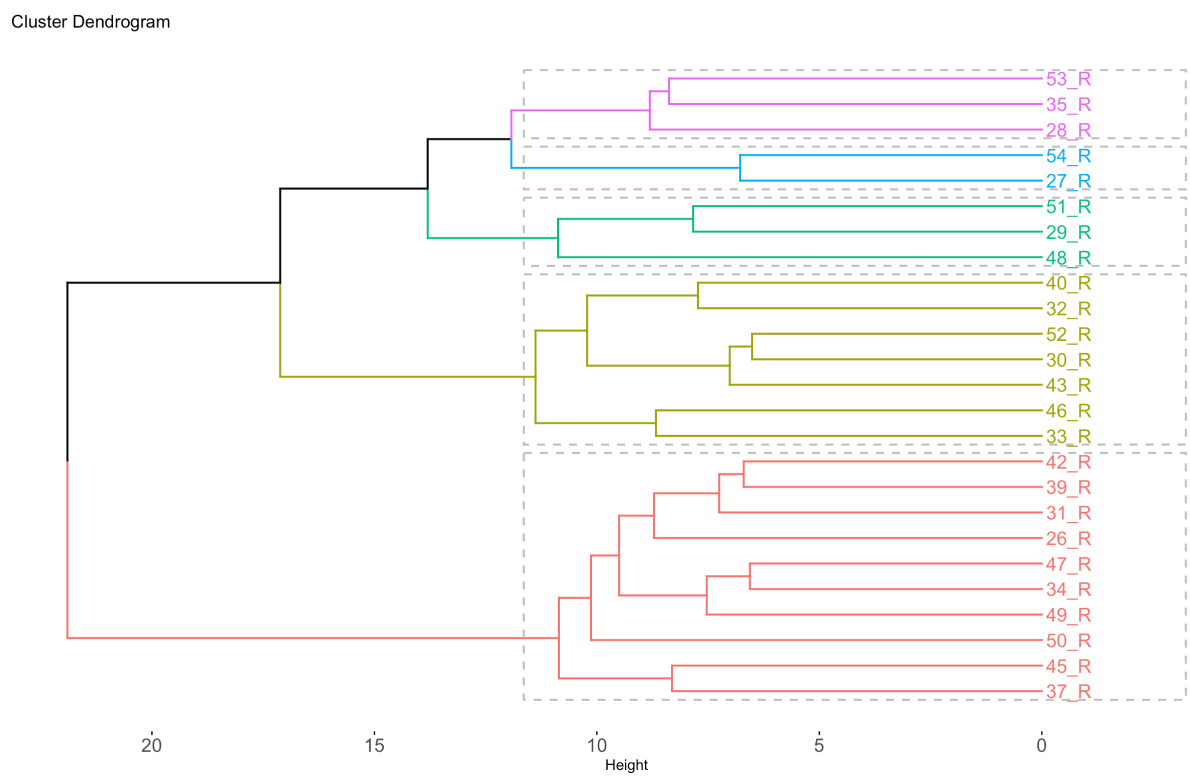
Machine learning is used to identify data patterns for many purposes focused on decision making. Successful application of machine learning from both the prediction and classification side has now ventured into the sports arena. Ashley (2020) presented improvements in the application of machine learning across the spectrum of individual and professional sports. At the same time, future expectations include an increase in the impartiality of the decision-making process in sports (Richter et al., 2019; Rommers et al., 2020). These expectations have not gone unnoticed because various machine learning applications focused on decision making have been presented in basketball (Calvo et al., 2017; Horvat et al., 2020; Leicht et al., 2017a, 2017b) and American football (Fernandes et al., 2020; Mulholland & Jensen, 2014). There are two defined areas of machine learning: supervised and unsupervised. Supervised learning is a typical application of machine learning where the objective is to predict a specified variable, called a class, from a previously unrevealed data set. This process tries to uncover a relationship between an independent variable (i.e., the input) and a dependent variable (i.e., the target). Unlike supervised methods, unsupervised learning is a technique with no outcome or target variable, and patterns are discovered from previously unknown relationships. Determining this relationship through empirical studies using data generated from ever-improving technology are needed to understand the qualities and trends of teams, players, and match officials. The ability to identify the characteristics of teams and match officials will allow governing bodies to pair match officials correctly and inform best practices regarding match assignments. Therefore, the purpose of this study was to label match styles and match officials using a machine learning clustering method and then to predict the clusters of match and match officials using physical and match event data via the machine learning SVM, RF, and k-NN classification methods.

7.3 Methodology

Analysis was conducted using R (version 4.0.2) on a data set that combined optical tracking of x and y coordinates of all players and match officials along with event data (Metrica Play, Amsterdam, NL) and aggregated technical performance data produced by Statsperform (OPTA, product of Statsperform, Chicago, IL). For this study, positional labels included center referee, attacking center midfielder, central defender, central forward, central midfielder, defensive midfielder, left back left midfielder, left winger, right back, right midfielder, and right winger. Data filtering and collection followed the same process as the previous work within this research, where match event and descriptive data included 8,339 observations collected from 559 matches, 24 teams, 711 players, and 25 match officials throughout the 2017–2018 MLS seasons.

7.3.1 Clustering

This study divided matches and referees into clusters according to physical and event data values using hierarchical agglomerative cluster analysis (HACA). The first hierarchical cluster was conducted at the match level, the second and third at the home and away level, the fourth at the referee level by match, and finally at the average data by referee. For all five approaches, mean data were collected for each feature. Without needing other information about the data, the HACA initially puts each sample in a separate cluster and then combines similar groups until they become a single cluster (Alpaydin, 2020). Five clusters were placed as labels on both referee analyses, and three were placed at the respective match levels (see Figure 30).



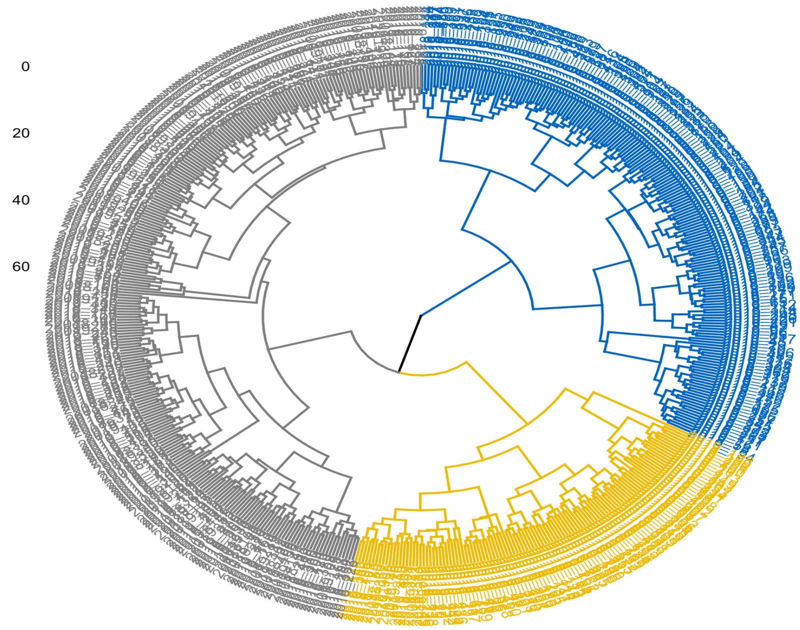


Figure 30. Hierarchical Cluster Tree (dendrogram). *Colors represent the division of five distinct clusters for referees (top) and three at the match level (bottom). Endpoints on the bottom and outside of the circle represent the referee and match number respectively.*

Clustering in this research was appropriate as the intention was to identify any patterns and uncover any relationships and insights of match officials and the fixture that may have been unnoticed previously (Halkidi et al., 2001). An essential step in hierarchical clustering is to select a distance measure. A *standard measure* is a Euclidean distance computed by finding the square of the distance between each variable, summing the squares, and finding the square root of that sum. Ward and Hook's (1963) clustering method was used to cluster the matches and referees. A graphic tree called a dendrogram is usually created to represent the result of the clustering process. Using the parameter R package (Version 0.6.0), the optimal number of clusters was determined from the "easystats," "NbClust," "mclust," "cluster," and "M3C" packages. These were then visualized using the factoextra package (Version 1.0.3).

7.3.2 Feature Extraction and Importance

Feature selection selects a subset of the most relevant features linked to the prediction output. In this study, the output was the cluster grouping, as determined previously. This technique reduces the dimensions of the feature space, making machine learning models easier to interpret and faster to train, with lower risk of multicollinearity issues (James et al., 2021). To select relevant features, I applied the Boruta algorithm (Kursa & Rudnicki, 2010), which also ranks their relevance in addition to selecting relevant features. The Boruta algorithm is a technique using the *z* score of a random forest and has a wide variety and predictability advantages. Recommended by Speiser et al. (2019) for data with many features and subsequent efficient computational time, the Boruta algorithm applies random forest importance measures for feature selection. Selection of features based on forecast results ensures predictability. An extension of the original data set occurs by adding the so-called shadow features whose values are randomly permuted among the training cases to remove their correlations with a decision variable. Next, the shadow features are randomly mixed, and a random forest algorithm is applied to extract an importance measure. The *z* score is computed by dividing the average loss by its standard deviation. The importance measure is used to determine the ranking of features. If the extracted *z* score exceeds the max *z* score, the characteristic feature is selected as the important feature (Speiser et al., 2019). Thirty-three features were extracted at the average referee data: 31 for referee performance by match, 37 at the match level, 28 at the home level, and 37 at the away level (see Figure 31).

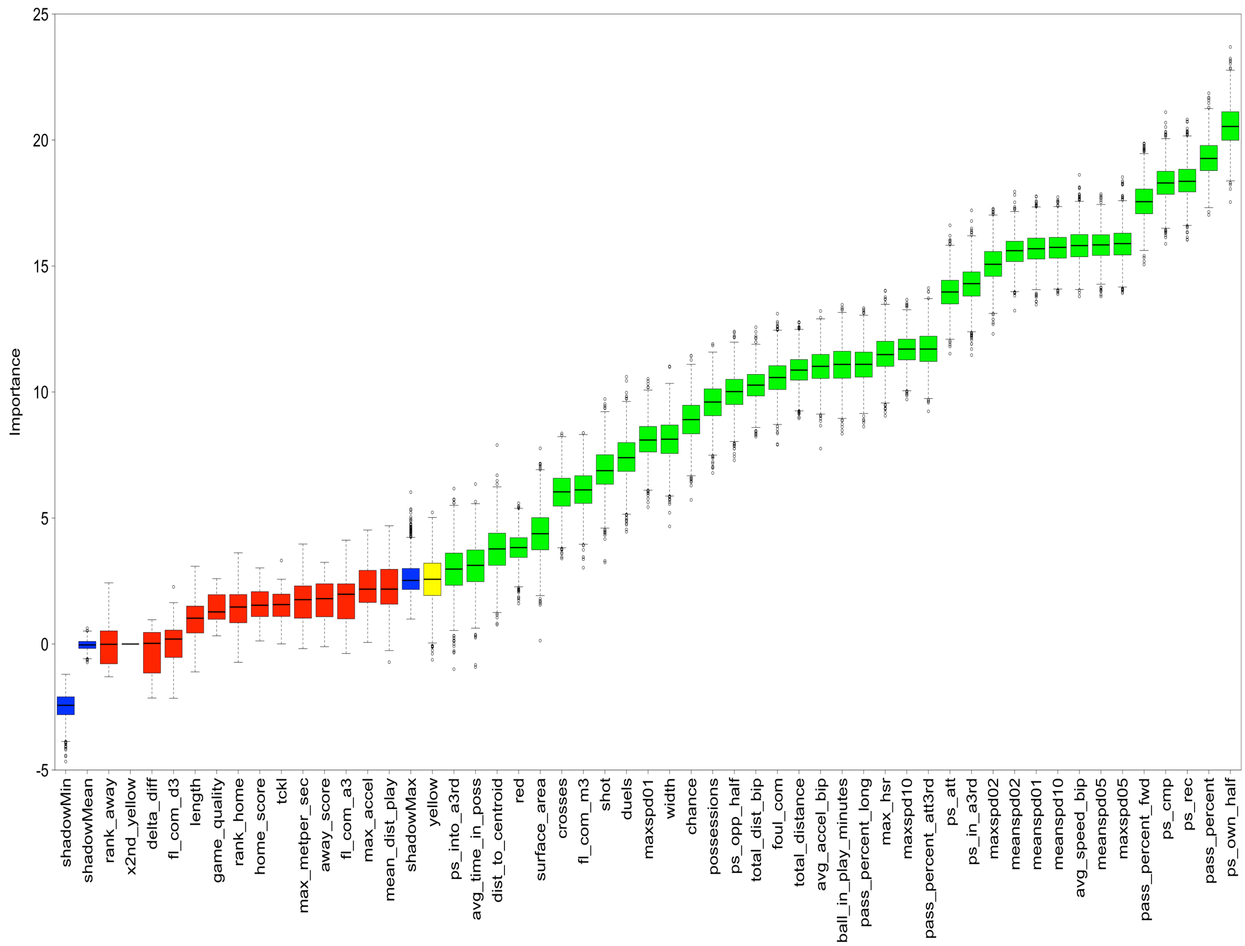


Figure 31. Match Official Variables of Importance. *Blue boxplots correspond to minimal, average, and maximum Z score of a shadow attribute. Red, yellow and green boxplots represent Z scores of rejected, tentative, and confirmed attributes respectively..*

7.3.3 Classification

RF, k-NN, and SVM classification models were applied for analysis. Data were standardized for all supervised learning models and split into a training and test set applying a 70/30 split. Then the training set was divided further into training and validation sets using 10-fold cross validation. This technique was employed to reduce overfitting with the subdividing of the data set into several folds and evaluating the accuracy of each fold (Pavey et al., 2017).

An RF classification method consists of several uncorrelated decision trees. Each tree in the forest may decide on a classification operation, and the class with the high votes decides the final classification. All decision trees have grown under randomization during the learning process. After exploring various randomization methods of decision tree selection (e.g., bagging or boosting), Breiman (2001) coined the term "random forest" (RF) in 1999. Furthermore, RF has been used within various sporting environments (Cust et al., 2021; Woods et al., 2018). Interestingly, Woods et al. (2018) demonstrated classification inaccuracies when focused only on technical skills.

The k-NN is the most straightforward ML algorithm (Brooks et al., 2016). The idea is to memorize the entire data set and classify a point based on the "K" nearest neighbor's class. K-NN is simple and without assumptions; however, it is also challenging to determine the optimal value of K, which is the number of neighbors used. This nonparametric method used for classification relies on a metric distance value. The Euclidian distance was used in line with the other classification methods imposed in this research.

SVM is a supervised ML algorithm mainly used for classification problems and regression. Each data item is depicted as a point (or a vector) in n–dimensional space, where n equals the number of used features. Essentially, the values of each feature generate the coordinate of each point. The challenge is to find the hyperplane that optimizes the differentiation between several classes, which is achieved by maximizing the distance between the hyperplane and the two nearest data points (each one from a different class). The distance is called the margin, and the data points contributing to discovering optimal solutions are called support vectors.

7.3.4 Model Evaluation

Classification performance was evaluated and assessed employing accuracy, balanced accuracy, F1, precision, and recall. The data set obtained from clustering produced imbalanced data. Data sets considered imbalanced for this research would include those classes that appear more often than others (i.e., more matches in Cluster 1). Accuracy is the ratio between the correctly classified observations and the total observations. Balanced accuracy is the arithmetic mean of both sensitivity and specificity. This method was more reliable than the often-seen accuracy with unbalanced data (Chicco et al., 2021). Precision shows the correct positive predictions relative to the sample classified as positive, whereas recall indicates the correct positive predictions relative to the total actual positive predictions. F1 is the harmonic mean of precision and recall combined into one score. If either recall or precision is low, the resulting F1 is also low.

7.4 Results

7.4.1 Clustering

Clustering was used to label match officials, match officials by match demands, match type, and home and away cluster. Table 4 indicates the importance of features across each of the five analyses.

At the referee-only level, the number of passes, passing percentage overall, and the percentage of passes forward comprised the top 3, followed by seven physical metrics. However, when looking at the referee at the match level, despite finding the same top 3, only one physical metric is included in the top 10. A balance of five technical and five physical metrics was deemed most important at the match level, and 7 of the top 10 metrics were physical at the home level compared to only two at the away team measures. Figures 32 and 33 show the distribution of the top 10 features by the cluster at the referee-only and match levels.

Table 4

Top 10 Variables of Importance Determined by the Boruta Algorithm. Each column represents the cluster labels and ordered by rank of importance for each measure.

Ref-only measure	Referee performance measure	Match measure	Home team measure	Away team measure
# Passes in own half	# Passes in own half	Passing %	2-min window max 1-min (m/s)	Passing % forward
Passing %	Passing %	Passing % forward	1-min window max 1-min (m/s)	Passing %
Passing % forward	Passing % forward	Passing % opposing half	5-min window max 1-min (m/s)	# Passing attempts
5-min window max 1-min (m/s)	# passing attempts	10-min window max 1-min (m/s)	Passing %	% of long passes
Mean BIP speed	# passes in attacking 3rd	5-min window max 1-min (m/s)	Total distance BIP	Passing % opposing half
5-min window mean 1-min (m/s)	% of long passes	2-min window max 1-min (m/s)	Possessions	2-min window max 1-min (m/s)
10-min window mean 1-min (m/s)	Passing % opposing half	1-min window max 1-min (m/s)	10-min window max 1-min (m/s)	# passes in opposing half
2-min window mean 1-min (m/s)	Passing % attacking 3rd	Total distance BIP	Passing % forward	1-min window max 1-min (m/s)
1-min window Mean 1-min (m/s)	Crosses	Duels	Mean acceleration BIP	Passing % attacking 3rd
2-min window max 1-min (m/s)	2-min window mean 1-min (m/s)	Passing % attacking 3rd	Total distance	# passes in attacking 3rd

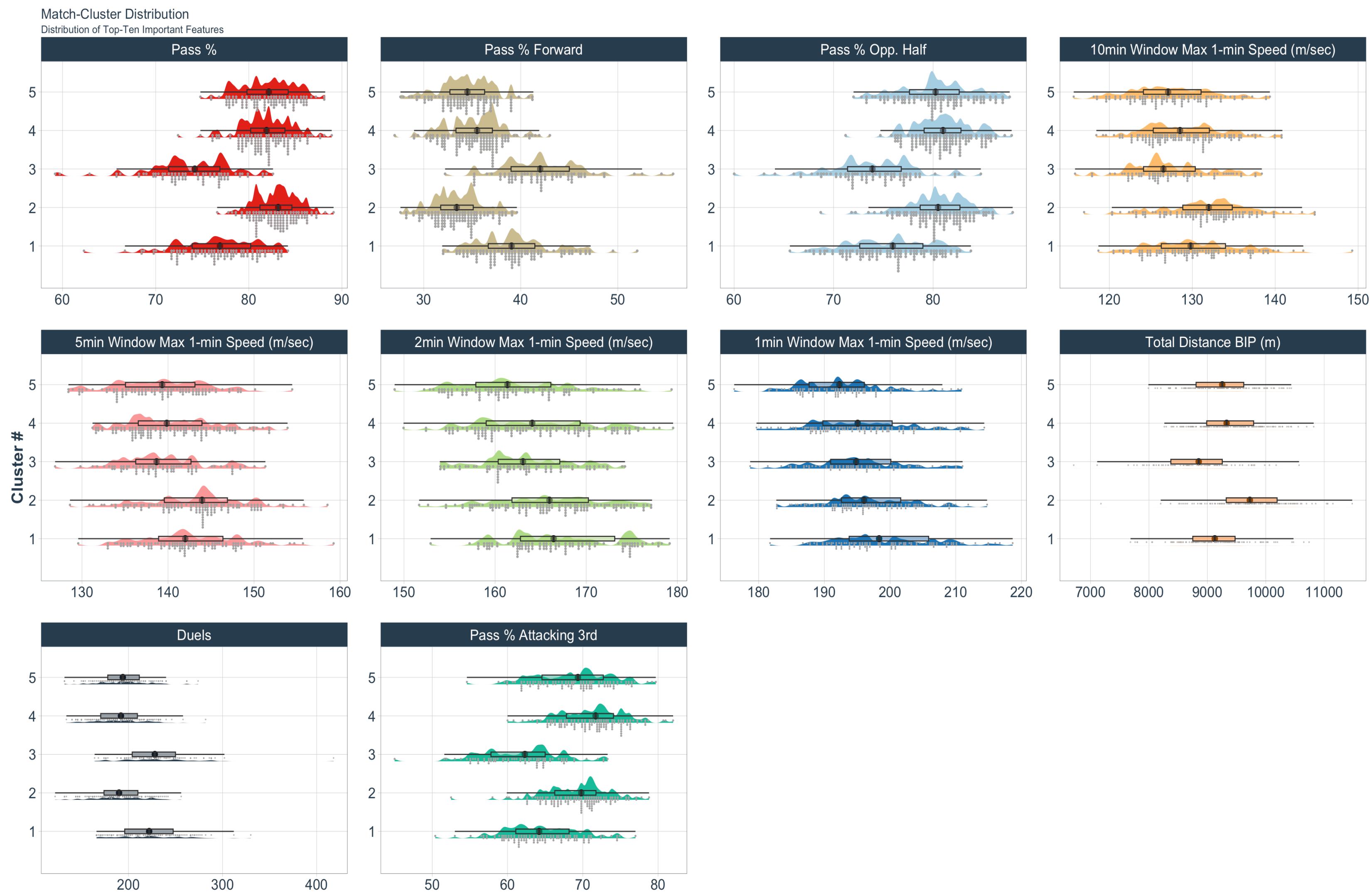


Figure 32. Cluster Distribution of Top 10 Features at Match Level. *Cluster group number is provided along the x-axis and raincloud plot demonstrates the distribution along the y-axis.*

Figure 33. Cluster Distribution of Top 10 Features at Referee Level. *Cluster number is provided along the x-axis and distribution of each measure along the y-axis.*

7.4.2 Classification

Figure 34 shows the evaluation metric for each model across all five clustering approaches. RF models presented higher average results for accuracy and balanced accuracy, and KNN had the highest average Precision, Recall, and F1 score.

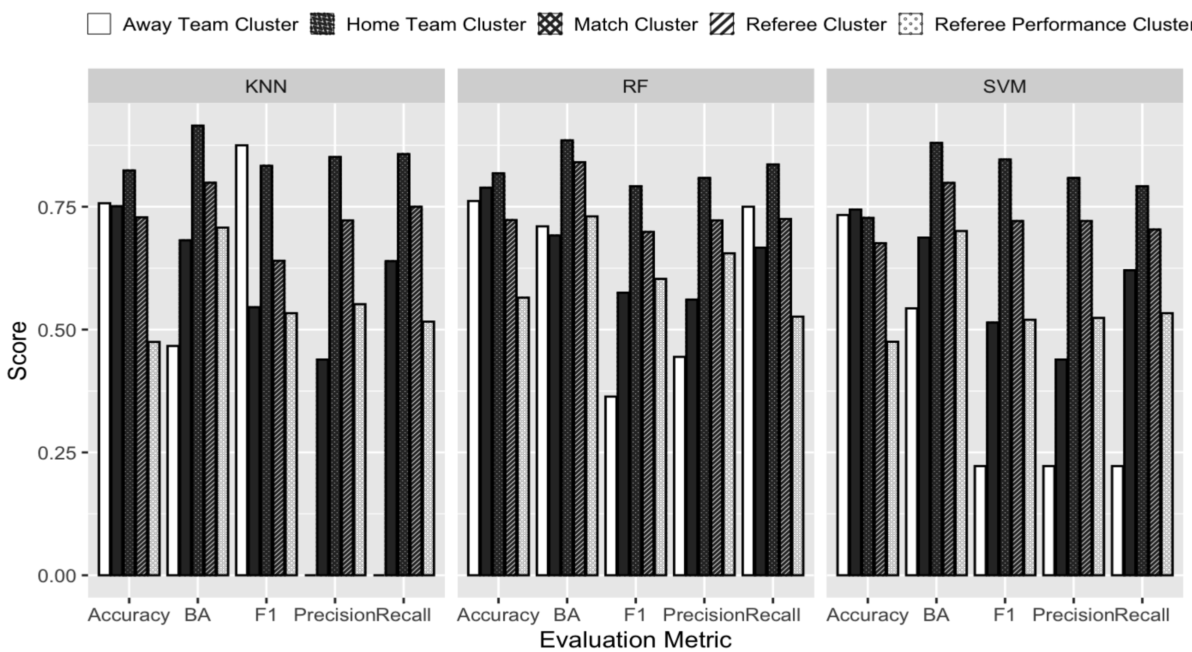


Figure 34. *Comparison of Evaluation Metrics and Performance Predicting the Cluster Label*

As seen in Table 5, at the match level, KNN presented the highest numbers in Accuracy at 0.82, Balanced Accuracy at 0.92, Precision at 0.85, and Recall at 0.86. F1 scores were highest for SVM at 0.85. Home Team level Accuracy scores were notable at 0.79, 0.74, and 0.75 for RF, SVM, and KNN, respectively. Similarly, Accuracy scores were notable at 0.76, 0.73, and 0.76, and RF produced a score of 0.76 for RF. Referee Level scores for Balanced Accuracy were 0.84, 0.80, 0.80, Accuracy 0.72, 0.68, 0.73, and Recall 0.72, 0.70, 0.75 for RF, SVM, and KNN, respectively. Balanced Accuracy showed 0.73, 0.70, and 0.71 for Referees at the Match Level.

Table 5 *Performance Evaluation for Each Algorithm Across All Five Analyses*

Match level					
Type	Accuracy	BA	Precision	Recall	F1
RF	0.82	0.89	0.81	0.84	0.79
SVM	0.73	0.88	0.81	0.79	0.85
KNN	0.82	0.92	0.85	0.88	0.83
Home team cluster					
Type	Accuracy	BA	Precision	Recall	F1
RF	0.79	0.69	0.56	0.67	0.58
SVM	0.74	0.69	0.44	0.62	0.51
KNN	0.75	0.68	0.44	0.64	0.55
Away team cluster					
Type	Accuracy	BA	Precision	Recall	F1
RF	0.76	0.71	0.44	0.75	0.36
SVM	0.73	0.54	0.22	0.22	0.22
KNN	0.76	0.47	0.00	0.00	0.00
Referee only					
Type	Accuracy	BA	Precision	Recall	F1
RF	0.72	0.84	0.72	0.72	0.70
SVM	0.68	0.80	0.72	0.70	0.72
KNN	0.73	0.80	0.72	0.75	0.64
Referee at match level					
Type	Accuracy	BA	Precision	Recall	F1
RF	0.57	0.73	0.66	0.53	0.60
SVM	0.48	0.70	0.52	0.53	0.52
KNN	0.47	0.71	0.55	0.52	0.53

To compare the performance of various models against unseen data, the validation set was explored to investigate the performance of each technique further. Tables 6 and 7 tabulate a performance evaluation from a confusion model for each algorithm employed. The model built using the RF method had the lowest out-of-sample error, and SVM improved accuracy at the referee level. All three models improved, with SVM showing the most significant increase and KNN showing the lowest out-of-sample error.

Table 6 *Ref-Only Decision Tool*

Model type	Accuracy ref only	Misclass error ref only	Validation accuracy ref only	Validation misclass error ref only
KNN	0.73	0.27	0.66	0.34
RF	0.72	0.28	0.71	0.29
SVM	0.68	0.32	0.69	0.31

Table 7 *Match Decision Tool*

Model type	Accuracy match	Misclass error match	Validation accuracy match	Validation misclass error match
KNN	0.82	0.18	0.86	0.14
RF	0.82	0.18	0.82	0.18
SVM	0.73	0.27	0.83	0.17

Figure 35 presents a confusion matrix of the referee only, referee, and match level correctly and incorrectly classified observation.

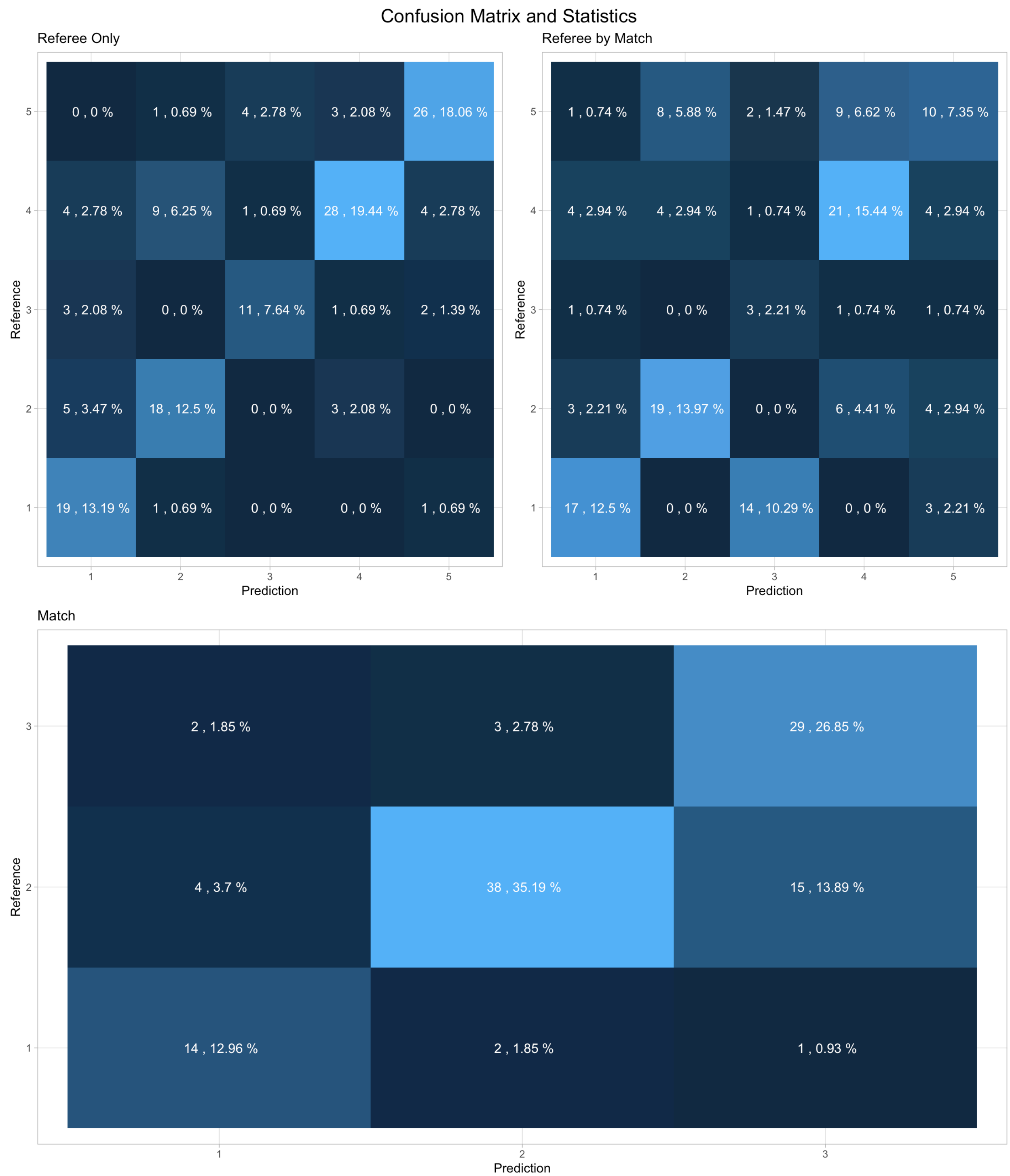


Figure 35. Confusion Matrices Performance for Referee Only, Referee at Match Level, and Match Level. *Utilizing the validation set, correct and incorrect classification labels are identified with frequency of type and percentage of total identified type.*

7.5 Discussion

Intending to provide decision support to governing bodies, the primary aim of this study was to use unsupervised and supervised machine learning approaches to explore implementing a match official assignment tool. This study applied a novel framework of hierarchical clustering, SVM, k-NN, and RF machine learning techniques to deliver findings based on both match official and team physical and technical data within a match. The visual plots provided insights into determining similarities and can be used to explore further how assignments are carried through. The clustering technique provides a picture of determining match type and referee similarities, and the supervised approaches can reasonably predict the cluster. Average values for all metrics across three clusters showcased the notable metrics that would be expected from the match official or match type. When comparing each against each other, the only performance discrepancy occurred on the Away team cluster level, with RF outperforming the others. At the same time, all three supervised machine learning techniques reasonably demonstrated similar performance in all of levels. Further illustrating the advocation of decision support systems (Schelling & Robertson, 2020), these findings throughout two consecutive seasons suggest confidence in the potential for using machine learning techniques to find similar performance metrics, support decision making based on predicted teams, and promote match official performance.

The main findings illuminated how match officials and matches related after applying a feature selection process that trimmed the initial data set of 68 features to a more manageable size. Determining the list of most important features provided insights into establishing what is essential when predicting future performances from each cluster. Although logical from a match or player perspective, it was notable that the top 3 features at the referee level were technical followed by physical metrics. Furthermore, when assessing the referee at match level, the same three technical features were at the top of the list, yet more technical features completed the top 10. This useful finding is highlighted further because similar findings represented the match-level cluster.

Interestingly, Cluster 2 had the lowest duels average, highest passing accuracy, and more passes in the attacking half while also demonstrating the highest physical outputs. As match officials are expected to make decisions accurately, the fewer duels, the fewer foul decisions would be demanded from the game. Average physical data from the match officials displayed a clear difference in physical metrics across clusters; however, the technical features needed to be defined more clearly.

This study investigated the ability of machine learning techniques to identify match officials and match types. As a delineation of match officials' average high and low physical and technical performance, feature selection was a crucial pre-step process for applying supervised learning (Dalton-Barron, 2022; Wundersitz et al., 2015). This global view of the contextual variables, which cluster to aid in understanding match officials' physical performance based on match type, can further lend to more objective performance analysis. This key point also presents a limitation of this study: the reliance on consistent players for each club. Although these findings were presented for two seasons, the tool needed to be investigated to understand a team's acquisition of new players, injuries, new coach or manager, or any other roster change. Further, the turnover over of match officials does not reflect the transient environment of teams. The length of careers as a professional match official is much longer than the average player or coach. This will help in evaluating the findings from this chapter to better determine the utility of this approach with a longer 5,10, or even 20-year analysis of referee performances. Nonetheless, findings from the framework presented would be an essential step in better understanding the features required to support an assignment tool.

7.6 Conclusion

In conclusion, technical and physical metrics can be used to design and apply a decision support system to assign match officials. The current assignment practice includes random subjective opinions, determining match official selection. The findings from this study reflect an approach that allows for a machine learning framework to inform the decisions based on similarities and ultimately implement a supervised approach and predict which official suits the match best. For example, the subjective assignment of an official who may be less experienced and not well-suited for a match where a high number of duels and, subsequently, requires a high number of decisions could potentially be a risk for low performance. Understanding the cluster grouping of this match official would be beneficial to understand better if that official is suited for a predicted match type. Furthermore, automating the selection process based on predicting the future performance of match officials and match fixtures can reduce the human hours required to complete this ongoing task. The performance of supervised learning techniques on determining similarities of match officials and match type in this study provides confidence for further research and application in future applied settings.

CHAPTER 8: DISCUSSION AND CONCLUSION

8.1 Discussion

Referee performance and the subsequent influence on match outcomes, tournament participants, or league standing is debated frequently in most sports. The behavior of referees' decisions, match and player control, and fluency in managing the multitude of factors that could impact viewer perception, both positively and negatively, is substantial. Evaluation from governing bodies has faced increased pressures to elevate match officials' performances to minimize substandard displays and the subsequent negative consequences that arise accordingly. Given the demanding training and competition schedules of professional match officials, optimizing the methodologies that contribute to evaluating substandard, acceptable, or high performances is an obvious requirement at the elite levels of soccer. However, with the increasing amount of data captured across all aspects of referee preparation, performance, and recovery, there needs to be more evidence of using a data support system to evaluate and assign referees to matches that would increase the likelihood of high physical performance. The overarching goal of this thesis was to operationalize data collected from match officials, develop a framework whereby physical performance influencers are understood more clearly, and evaluation is defined, and subsequently forecast future match performances. In this thesis, I presented a novel approach to the performance evaluation of match officials in professional soccer with the primary objective of forecasting performance. The methodologies applied across three separate sections aimed to evaluate the referee's training (Chapter 3); response (Chapter 4); match demands (Chapter 5); and similarities to other referees, players, and matches (Chapter 6). This overarching evaluation of match officials' performance may make it possible to answer the questions posed in Chapter 1:

1. Are training and recovery behaviors of match official's useful metrics for determining the suitability of match officials of specific assignment?
2. What are the physical match demands of match officials and players in Major League Soccer?
3. What are the relationships and similarities between match fixtures, teams, positions, players, and match officials'?
4. Can match type and referee performance be predicted using previous match data?

The first part of this examination focused on match officials' physical training and recovery across a season. Study 1 in Chapter 3 presented training loads and the effect of the type of activity undertaken by referees across a season. As expected, matches presented the highest load; however, the key takeaway from these results indicated evident variability in the RPE_{load} response by training type. Although assignments reflect a typical team fixture schedule, referees must prepare for the matches on their calendars, and known tendencies or patterns for which the referee can prepare are not readily available. In reflection, group analysis on periodization strategies could have been more precise because analyzing anything beyond the individual was deemed unwarranted. This conclusion provides a pathway to analyze further the appropriate training regimen practitioners prescribe that allows for optimal physical performance. Highlighted by a substantial percentage of rest days (38%) across a week, the impact of geography and the subsequent travel days endured by match officials in North America must be accounted for within the analysis. As a match official does not have a "home" game compared to a team, the number of training days is limited because referees must travel to matches and return to their home training city. Nevertheless, this exploration of match officials' training left inconclusive evidence related to the overall objective of forecasting match official performance. .

To understand better if the recovery of match officials would contribute to predicting match officials' performance, I explored the relationship between training, matches, and following day wellness in Chapter 4. Understanding the effects of training and the potential impact on how a match official responds could contribute to match officials' performance in an optimal state. Although theoretically analyzing the short- and long-term impact of matches and the training design and response for match officials appeared to be a good approach, the findings suggested a different takeaway than hypothesized. I implemented general and generalized mixed models to determine the relationship between training and matches, along with the effect of RPE_{load} on following day wellness. Despite similarly reported loads, match officials reported different fluctuations in response to training than players (J. J. Malone et al., 2015; Thorpe et al., 2016a). This key takeaway, although not particularly relevant to the study, could guide future projects and analysis of the important questions posed within morning wellness questionnaires for match officials. Although potentially valuable for practitioners working with match officials, the findings would not suggest a link to the overall objective of contributing to determining future performance.

Chapter 5 presented a multi-season physical, ball-in-play (BIP), and situations where the ball was not in-play activity profile of match officials, players, and positions. Previous exploration of ball in and out of play match variables (Mernagh et al., 2021; Riboli et al., 2021; Wass et al., 2019) did not describe fully the match intensity influence on match officials and specifically on MLS players. Soccer is deemed to have high situational complexities, with vast array of influences contributing to the physical demands on placed on performers. Factors such as playing position, formation, attacking and defending strategies, venues being either home or away, and overall quality of the teams competing contribute to the variation in the physical requirements of games (Barnes et al., 2014; Bradley & Ade, 2018; Bradley & Noakes, 2013; Castellano et al., 2011;

Gregson et al., 2010). These factors and, more recently, others, such as the implementation of Video Assistant Referee, present a challenge when utilizing player and match official profiles. These evaluations of demand from competitive matches that do not account for BIP, and ball-out-of-play situations may not be an applicable approach when examining the efficacy of physical training during match preparation. Therefore, high-performance staff members, medical practitioners, and coaches need to take inventory of what factors should be examined to optimize the planning process and prescription of training. I reported average BIP times and noted 60% of all BIP timeframes were less than 60 s. This helpful information provides a discussion framework for practice design and defines intensities within that session. As stated, understanding typical periods when BIP and ball-out-of-play are imperative; a seemingly large proportion of BIP windows under 60 s would suggest these shorter periods should be emphasized. Our findings are similar to previous research, where intensity dropped as time increased (Delaney et al., 2017). To define peak demands more accurately, identifying intensities while the ball is in play is required. Highlighting the importance of context when defining demand, the distance covered at 5 to 60 s displayed a 53% drop in velocity.

Findings from Chapter 6 provided the platform for the final two chapter, where I examined the relationships between the physical and technical aspects of player and referee performances. The idea behind Chapter 6 was to develop an objective model that crucial stakeholders could use to evaluate in-game performance. In Section 1, I implemented unsupervised learning techniques on three separate analyses to characterize performance and cluster homogeneous subgroups. Four clusters were identified, including club and match officials, but were distinctly comprised of two club-only clusters and two separate match officials' clusters. When considering the findings from Chapter 5, this finding logically groups the teams and referees as separate clusters because intensities deemed higher for outfield players. Performing an analysis solely focused on physical metrics proved valuable in defining the relationships between match officials, players, and clubs.

Furthermore, the insights gathered when only investigating match officials displayed five clusters. This novel approach to defining performance should guide understanding and evaluating match official performance. From a governing body perspective, these findings suggest assessing referee performance should reflect the performance metrics from their cluster. Lastly, considering a comparison by position, three separate clusters were identified. Interestingly, match officials were clustered with central forwards and central defenders. Future supervised approaches could be applied to predict performance based on teams' center backs. Findings from this section could inform decision making with respect to planning training, profiling players and match officials, and, most importantly for this thesis, forecasting performance.

Finally, implementation of both unsupervised and supervised approaches to set the foundation for predicting match physical performance was presented. Applying selected machine learning approaches reasonably predicted the match type and referee from contextual variables while highlighting key vital features. When displaying referees' abilities, the technical performance of the players, interestingly, was top of the list for match officials. Players' technical components are drivers to define match officials' assessment better. When using the work from this thesis, understanding the influence of the fixture is an essential component of match officials' performance. It should be an integral part of the forecasting framework.

The identification and placement of match officials in a game present many challenges and, to date, have been decided upon from factors beyond and without a data-driven support system. By removing biases from key figures in the process, the ability to build upon the structure of this work and contribute to the operational challenge of match assignment may greatly benefit those bodies responsible for these decisions. Furthermore, applying both supervised and unsupervised machine learning in various environments where a selection process is part of the workflow could benefit from the work of this thesis. With the current data landscape and the further potential of artificial intelligence imitating human functions such as learning and problem-solving, or humans applying machine learning techniques, and the application of multilayered neural networks in deep learning becoming a prominent part of the sports landscape, integrating this work in future projects will only enhance match officials' assignment decisions and soccer in the near distant future.

8.1.1 Limitations

The research is limited to analyzing match officials and players across two MLS seasons. Moreover, analysis was based on a domestic league in a certain region of the world and was not a random sample of all matches globally. This fact raises the prospect of these findings as not applying to other parts of the world based on the playing style of players and teams or the referee style conveyed domestically. This could lead to a debate on generalizing these findings and whether they apply to other leagues or competitions or are only specific to MLS and the match officials within the analysis.

Furthermore, despite appropriate analytical approaches employed within the project, there are questions about the stability of these findings when new data are introduced. Investigating the performance of referees can be lifelong project. Specifically in the MLS there are factors such as the geography and the amount travel that is required for players, and match officials, the referee union and the constraints that play into organization decision-making, and the accelerated growth of the league and expansion of new teams and cities. For example, the clusters found in Chapters 5 and 6 relied on the data on hand, potentially leading to a question of how those would change when new data are included in the analysis. The game of soccer is ever-changing as coaches change, education changes, and rules change; consequently, findings would be expected to change, which surfaces the question of how the models would advance with newer data.

Additional limitations would include key performance measures such as decision making and, more specifically, the accuracy of match officials. This is a key performance indicator; however, the subjectivity of the sport needs to provide clear outcome metrics when determining the accuracy of decisions. This project focused on the technical and physical performances of the teams and players, and only the physical performance of the match officials and decision making was not included.

8.1.2 Future Directions

The review of the literature focused on the physical performances of match officials. Research has been conducted predominantly without considering the contextual factors that could describe match officials' behaviors. Within this project, findings suggest the influence of players' technical performance is a key factor in determining the relationships of match official's to matches, players, and teams. Because this is a novel finding and research is in its infancy when describing match official performance, an opportunity is presented to build upon this framework. Anecdotally, the assigning process of match officials lacks a decision framework, and the work performed within this project could be developed further into an assignment tool.

Another area for future research should include a comprehensive data collection approach to factors such as travel, time zones, stadiums, and sleep quality and quantity. As match officials must make decisions fast and accurately, these data could inform and increase forecasting performance further.

Lastly, as stated previously, this research was focused solely on one domestic league. Collaboration across member associations and international competitions is suggested to drive an assigning framework for match officials. The ongoing focus should be on developing a decision support tool that could be implemented within any competition. A high-performing match official's importance is key to the global game's overall progress, and implementing a formalized assignment tool built upon the framework within this project would only enhance the experience of fans, players, and organizations.

8.2 Practical Application

The findings within this work aimed to describe a methodology for informing assignment of match officials. The following applications are proposed base on the findings from each chapter.

8.2.1 Training Loads

- The insights gained form the chapter showed a high variability for match officials' response to training. While logical out the outset, the findings within the chapter further confirm an individual approach to periodization is required. In addition to the findings within the chapter and within this specific environment, match officials are traveling consistently as they do not have home matches. The scheduling for each individual is different and the demands of the travel, most often air travel across multiple time zones, specific to the referee must also be considered.
- With the ever-improving technological advances in tools such as GPS, it would be appropriate to explore more objective ratings where internal and external load could be applied for further and potentially more insightful findings could be utilized in decision making for training. However, this should be approached with caution, as the nature of typical referee training reflects a similar remote environment and could be deemed a challenge for operational reasons. The usage of RPE to understand internal load for this

environment may be potentially the most appropriate method for monitoring. Further exploration of statistical approaches combined with more robust review of the collection tools are warranted.

8.2.2 Training and Response

- An emphasis on the critical selection of appropriate data collection questions and tools can not be stressed enough. There is potential effectiveness gained from utilizing these questionnaires, however as demonstrated the current approach deemed and questions should be conducted to understand best practices for match officials. As noted above and through other concurrent work, the utilization of wellness and subsequent response via wellness can be useful for informing decision making of periodization of training and match selection. However, the results of chapters three and for have left skeptical of applying this data collection method specific to referees and requires further attention.

8.2.3 Physical Match Profiles

- Results showed that the average ball in play is 47 seconds. However, a large number of BIP segments where less time were reported. Referencing these values for training prescription either via an individual training session or a team session should account for both the ability to repeat the maximal physical demands at both the short and long term BIP segments. Furthermore, an analysis of time spent with ball out of play would be appropriate for informing decision making. These are useful starting point in planning for training sessions and coaches should be aware the either too much or too little being carried out within the activity objectives. When designing training content should reflect on the type and the intensity of the training exercise that are placed through the session so that the physical demands reflect the recovery time and work time. Overlooking these key findings in training could diminish the athletic ability to meet the demands placed on match day.

- Maximal speed thresholds could further inform the decision making with regard to flagging or monitoring whether match officials are meeting these demands in training. While these numbers account for any BIP or ball out of play situation, and potentially happen in an isolated portion the match the findings from this analysis are good reference points for all positions on the field. Notably, Match Official guidance to follow a diagonal pathway for movement is lacking across all research. Paired with these findings, the utility of movement could be further explored for optimization of demands. The program design of space, distance, players, and timing could all reflect on the findings from this work to implement within the training environment.

8.2.4 Clustering

- The usefulness of this approach relies on the selection of appropriate methods from the outset of the analysis. Determination of inputs is a key component, but the findings indicated an expected difference in unsupervised learning approaches based on the specific algorithm. This is a key takeaway, and practitioners should be keenly aware of the differences.

- While this project was focused on the similarities within the matches, formation of training decisions and training groups for referees could benefit from the clusters formed.
- Further individual analysis requires attention as it was noted in both the findings and previous research that more experienced referees demonstrated notable differences in high intensity actions compared to younger officials.

8.2.5 Forecasting and a Framework for Match Assignment

- While the data collected throughout this project was plentiful, the ability to trim metrics to a manageable and useful size provides a benefit to practitioners regardless of data infrastructure and organization support.
- Match technical outputs are key to decision making of match officials' assignment. While it has been noted amongst practitioners that the performance of the referee is the ability to keep up with play, these findings could suggest a more technical awareness of referees is more important. Further research of other leagues is needed in this area, but the

implication that the relationship between referees and matches shows that the technical measures within the game cannot be overlooked.

- High intensity running performance at the 1 and 2-minute marks are key training indicators.
- Clustering and then utilizing classification produces interesting results when trying to predict performance. Momentum could be gained from further analysis and comparison by leagues around the world to implement these techniques.

8.3 Conclusion

The conclusions of this thesis are:

1. MLS physical metrics for match officials and outfield players are similar to other domestic leagues worldwide.
2. Evaluation of match BIP is a necessary component of determining match intensities.
3. Unsupervised learning is an appropriate technique application as a contributing factor in determining performance. Determining appropriate technique is imperative when analyzing findings.
4. Match officials clustered into five separate subgroups highlighted by the importance of team technical metrics and compared closely to central defenders' physical performances.
5. Combining machine learning techniques and algorithms presents a novel approach to match assignments.

6. Future work should include further developing an even more in-depth understanding of how a decision support system could aid in the appropriate assignment of match officials.

REFERENCES

Abrantes, C., Maçãs, V., & Sampaio, J. (2004). Variation in football players' sprint test performance across different ages and levels of competition. *Journal of Sports Science & Medicine*, *3*(YISI 1), 44–49.

Akenhead, R., Harley, J. A., & Tweddle, S. P. (2016). Examining the external training load of an English Premier League football team with special reference to acceleration. *The Journal of Strength & Conditioning Research*, *30*(9), 2424–2432. https://doi.org/10.1519/jsc.0000000000001343

Akenhead, R., Hayes, P. R., Thompson, K. G., & French, D. (2013). Diminutions of acceleration and deceleration output during professional football match play. *Journal of Science and Medicine in Sport*, *16*(6), 556–561. https://doi.org/10.1016/j.jsams.2012.12.005

Alarcón, F., Durán, G., & Guajardo, M. (2014). Referee assignment in the Chilean football league using integer programming and patterns. *International Transactions in Operational Research*, *21*(3), 415–438. https://doi.org/10.1111/itor.12049

Alarcón, F., Durán, G., & Guajardo, M. (2009). A referee assignment model for the Chilean soccer tournament and a pattern-based resolution approach. *Systems Engineering Magazine, XXIII*.

Alarcón, F., Durán, G., Guajardo, M., Miranda, J., Muñoz, H., Ramírez, L., Ramírez, M., Suaré, D., Siebert, M., Souyris, S., Weintraub, A., Wolf-Yadlin, R., & Zamorano, G. (2017). Operations research transforms the scheduling of Chilean soccer leagues and South American world cup qualifiers. *Interfaces*, *47*(1), 52–69. https://doi.org/10.1287/inte.2016.0861

Alashwal, H., Halaby, M. E., Crouse, J. J., Abdalla, A., & Moustafa, A. A. (2019). The application of unsupervised clustering methods to Alzheimer's disease. *Frontiers in Computational Neuroscience*, *13*, Article 31. https://doi.org/10.3389/fncom.2019.00031

Alpaydın, E. (2020). *Introduction to machine learning* (4th ed.). The MIT Press.

Altman, D. G., & Bland, J. M. (2011). How to obtain the confidence interval from a P value. *BMJ*, *343*, Article d2090. https://doi.org/10.1136/bmj.d2090

Antal, L. C. (1975). The effects of the changes of the circadian body rhythm on the sportshooter. *British Journal of Sports Medicine*, *9*(1), 9. https://doi.org/10.1136/bjsm.9.1.9

Aquino, R., Martins, G. H. M., Vieira, L. H. P., & Menezes, R. P. (2017). Influence of match location, quality of opponents, and match status on movement patterns in Brazilian professional football players. *Journal of Strength and Conditioning Research*, *31*(8), 2155–2161. https://doi.org/10.1519/jsc.0000000000001674

Aquino, R., Vieira, L. H. P., Carling, C., Martins, G. H. M., Alves, I. S., & Puggina, E. F. (2017). Effects of competitive standard, team formation and playing position on match running performance of Brazilian professional soccer players. *International Journal of Performance Analysis in Sport*, *17*(5), 695–705. https://doi.org/10.1080/24748668.2017.1384976

Aragão e Pina, J., Passos, A. M., Carvalho, H., & Maynard, M. T. (2019). To be or not to be an excellent football referee: Different experts' viewpoints. *Journal of Sports Sciences*, *37*(6), 1–9. https://doi.org/10.1080/02640414.2018.1522940

Araújo, D., Davids, K., & Hristovski, R. (2006). The ecological dynamics of decision making in sport. *Psychology of Sport and Exercise*, *7*(6), 653–676. https://doi.org/10.1016/j.psychsport.2006.07.002

Araújo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology, 19*(1), 69–78. https://doi.org/10.1080/10407410709336951

Araújo, D., Fonseca, C., Davids, K., Garganta, J., Volossovitch, A., Brandao, Re., & Krebs, R. (2010). The role of ecological constraints on expertise development. *Talent Development & Excellence*, *2*(2), 165–179.

Ashley, K. (2020). *Applied machine learning for health and fitness: A practical guide to machine learning with deep vision, sensors and IoT*. Apress. https://doi.org/10.1007/978-1-4842-5772-2

Aughey, R. J. (2011). Applications of GPS technologies to field sports. *International Journal of Sports Physiology and Performance*, *6*(3), 295–310. https://doi.org/10.1123/ijspp.6.3.295

Ayala, V., Martínez-Bebia, M., Latorre, J. A., Gimenez-Blasi, N., Jimenez-Casquet, M. J., Conde-Pipo, J., Bach-Faig, A., & Mariscal-Arcas, M. (2021). Influence of circadian rhythms on sports performance. *Chronobiology International*, *38*(11), 1522–1536. https://doi.org/10.1080/07420528.2021.1933003

Bahill, A. T., Bharathan, K., & Curlee, R. F. (1995). How the testing techniques for a decision support system changed over nine years. *IEEE Transactions on Systems, Man, and Cybernetics*, *25*(12), 1533–1542. https://doi.org/10.1109/21.478443

Balmer, N. J., Nevill, A. M., Lane, A. M., & Ward, P. (2007). Influence of crowd noise on soccer refereeing consistency in soccer. *Journal of Sport Behavior*, *30*(2), 130–145.

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics, 49*(3), 803–821. https://doi.org/10.2307/2532201

Bangsbo, J. (2014). Physiological demands of football. *Sports Science Exchange*, *27*(125), 1–6. https://secure.footprint.net/gatorade/stg/gssiweb/pdf/SSE125_Bangsbo.pdf

Bangsbo, J., Nørregaard, L., & Thorsø, F. (1991). Activity profile of competition soccer. *Canadian Journal of Sport Sciences*, *16*(2), 110–116.

Barnes, C., Archer, D., Hogg, B., Bush, M., & Bradley, P. (2014). The evolution of physical and technical performance parameters in the English Premier League. *International Journal of Sports Medicine*, *35*(13), 1095–1100. https://doi.org/10.1055/s-0034-1375695

Barrett, S. (2017). Monitoring elite soccer players external loads using real-time data. *International Journal of Sports Physiology and Performance*, *12*(10), 1285–1287. https://doi.org/10.1123/ijspp.2016-0516

Barron, D., Ball, G., Robins, M., & Sunderland, C. (2020). Identifying playing talent in professional football using artificial neural networks. *Journal of Sports Sciences*, *38*(11–12), 1211–1220. https://doi.org/10.1080/02640414.2019.1708036

Bartlett, J. D., O'Connor, F., Pitchford, N., Torres-Ronda, L., & Robertson, S. J. (2017). Relationships between internal and external training load in team sport athletes: Evidence for an individualised approach. *International Journal of Sports Physiology and Performance*, *12*(2), 230–234. https://doi.org/10.1123/ijspp.2015-0791

Batterham, A. M., & Hopkins, W. G. (2006). Making meaningful inferences about magnitudes. *International Journal of Sports Physiology and Performance*, *1*(1), 50–57. https://doi.org/10.1123/ijspp.1.1.50

Beernaerts, J., Baets, B. D., Lenoir, M., & Weghe, N. V. de. (2020). Spatial movement pattern recognition in soccer based on relative player movements. *PLoS ONE*, *15*(1), Article e0227746. https://doi.org/10.1371/journal.pone.0227746

Ben-Bassat, M., & Freedy, A. (1982). Knowledge requirements and management in expert decision support systems for (military) situation assessment. *IEEE Transactions on Systems, Man, and Cybernetics*, *12*(4), 479–490. https://doi.org/10.1109/tsmc.1982.4308852

Bland, J. M., & Altman, D. G. (1995). Comparing methods of measurement: Why plotting difference against standard method is misleading. *Lancet*, *346*(8982), 1085–1087. https://doi.org/10.1016/S0140-6736(95)91748-9

Bloomfield, J., Polman, R., & O'Donoghue, P. (2007). Physical demands of different positions in FA Premier League Soccer. *Journal of Sports Science & Medicine*, *6*(1), 63–70.

Bojkowski, Ł., Eider, J., Śliwowski, R., & Wieczorek, A. (2015). Analysis of the longest distances run by the best soccer players at the FIFA World Cup in Brazil in 2014. *Central European Journal of Sport Sciences and Medicine*, *11*(3), 145–151. https://doi.org/10.18276/cej.2015.3-15

Borg, G. A. V. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, *14*(5), 377–381. https://doi.org/10.1249/00005768-198205000-00012

Bradley, P. S., & Ade, J. D. (2018). Are current physical match performance metrics in elite soccer fit for purpose or is the adoption of an integrated approach needed? *International Journal of Sports Physiology and Performance, 13*(5), 656–664. https://doi.org/10.1123/ijspp.2017-0433

Bradley, P. S., Carling, C., Archer, D., Roberts, J., Dodds, A., Di Mascio, M., Paul, D., Diaz, A. G., Peart, D., & Krustrup, P. (2011). The effect of playing formation on high-intensity running and technical profiles in English FA Premier League soccer matches. *Journal of Sports Sciences*, *29*(8), 821–830. https://doi.org/10.1080/02640414.2011.561868

Bradley, P. S., Mascio, M. D., Peart, D., Olsen, P., & Sheldon, B. (2010). High-intensity activity profiles of elite soccer players at different performance levels. *Journal of Strength and Conditioning Research*, *24*(9), 2343–2351. https://doi.org/10.1519/jsc.0b013e3181aeb1b3

Bradley, P. S., & Noakes, T. D. (2013). Match running performance fluctuations in elite soccer: Indicative of fatigue, pacing or situational influences? *Journal of Sports Sciences*, *31*(15), 1627–1638. https://doi.org/10.1080/02640414.2013.796062

Bradley, P. S., Sheldon, W., Wooster, B., Olsen, P., Boanas, P., & Krustrup, P. (2009). High-intensity running in English FA Premier League soccer matches. *Journal of Sports Sciences*, *27*(2), 159–168. https://doi.org/10.1080/02640410802512775

Brooks, J., Kerr, M., & Guttag, J. (2016). Using machine learning to draw inferences from pass location data in soccer. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *9*(5), 338–349. https://doi.org/10.1002/sam.11318

Buchheit, M. (2014). Monitoring training status with HR measures: do all roads lead to Rome? *Frontiers in Physiology*, *5*, Article 73. https://doi.org/10.3389/fphys.2014.00073

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

Buchheit, M., Racinais, S., Bilsborough, J. C., Bourdon, P. C., Voss, S. C., Hocking, J., Cordy, J., Mendez-Villanueva, A., & Coutts, A. J. (2013). Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. *Journal of Science and Medicine in Sport*, *16*(6), 550–555. https://doi.org/10.1016/j.jsams.2012.12.003

Burgess, D. J. (2017). The research doesn't always apply: Practical solutions to evidence-based training load monitoring in elite team sports. *International Journal of Sports Physiology and Performance*, *12*(Suppl 2), 1–19. https://doi.org/10.1123/ijspp.2016-0608

Bush, M. D., Archer, D. T., Hogg, R., & Bradley, P. S. (2015). Factors influencing physical and technical variability in the English Premier League. *International Journal of Sports Physiology and Performance*, *10*(7), 865–872. https://doi.org/10.1123/ijspp.2014-0484

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Calloway, S. P., Hardin, D. M., Crawford, M. D., Hardin, J. M., Lemak, L. J., Giza, E., Forsythe, B., Lu, Y., Patel, B. H., Osbahr, D. C., Gerhardt, M. B., Mandelbaum, B. R., & Baldwin, W. W. (2019). Injury surveillance in Major League Soccer: A 4-year comparison of injury on natural grass versus artificial turf field. *The American Journal of Sports Medicine*, *47*(10), 2279–2286. https://doi.org/10.1177/0363546519860522

Calvo, J. L., García, A. M., & Navandar, A. (2017). Analysis of mismatch after ball screens in Spanish professional basketball. *International Journal of Performance Analysis in Sport*, *17*(4), 555–562. https://doi.org/10.1080/24748668.2017.1367999

Carling, C., Bradley, P., McCall, A., & Dupont, G. (2016). Match-to-match variability in high-speed running activity in a professional soccer team. *Journal of sports sciences*, *34*(24), 2215–2223. https://doi.org/10.1080/02640414.2016.1176228

Carling, C., & Dupont, G. (2011). Are declines in physical performance associated with a reduction in skill-related performance during professional soccer match-play? *Journal of Sports Sciences*, *29*(1), 63–71. https://doi.org/10.1080/02640414.2010.521945

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

Carling, C., Gall, F. L., & Reilly, T. P. (2010). Effects of physical efforts on injury in elite soccer. *International Journal of Sports Medicine*, *31*(3), 180–185. https://doi.org/10.1055/s-0029-1241212

Castagna, C., Abt, G., & D'Ottavio, S. (2007). Physiological aspects of soccer refereeing performance and training. *Sports Medicine*, *37*(7), 625–646. https://doi.org/10.2165/00007256-200737070-00006

Castagna, C., Impellizzeri, F. M., Bizzini, M., Weston, M., & Manzi, V. (2011). Applicability of a change of direction ability field test in soccer assistant referees. *Journal of Strength and Conditioning Research*, *25*(3), 860–866. https://doi.org/10.1519/jsc.0b013e318208ae8e

Castellano, J., Blanco-Villaseñor, A., & Álvarez, D. (2011). Contextual variables and time-motion analysis in soccer. *International Journal of Sports Medicine*, *32*(6), 415–421. https://doi.org/10.1055/s-0031-1271771

Castillo, D., Castagna, C., Cámara, J., Iturricastillo, A., & Yanci, J. (2018). Influence of team's rank on soccer referees' external and internal match loads during official matches. *Journal of Strength and Conditioning Research*, *32*(6), 1715–1722. https://doi.org/10.1519/jsc.0000000000002040

Catterall, C., Reilly, T., Atkinson, G., & Coldwells, A. (1993). Analysis of the work rates and heart rates of association football referees. *British Journal of Sports Medicine*, *27*(3), 193–196. https://doi.org/10.1136/bjsm.27.3.193

Catteeuw, P., Helsen, W., Gilis, B., & Wagemans, J. (2009). Decision-making skills, role specificity, and deliberate practice in association football refereeing. *Journal of Sports Sciences*, *27*(11), 1125–1136. https://doi.org/10.1080/02640410903079179

Cerqueira, M. S., Silva, A. I. da, & Marins, J. C. B. (2011). Analysis of the FIFA's model of physical evaluation applied to the soccer referees. *Revista Brasileira de Medicina Do Esporte*, *17*(6), 425–430. https://doi.org/10.1590/s1517-86922011000600012

Chapman, D. W., Bullock, N., Ross, A., Rosemond, D., & Martin, D. T. (2012). Detrimental effects of West to East trans meridian flight on jump performance. *European Journal of Applied Physiology*, *112*(5), 1663–1669. https://doi.org/10.1007/s00421-011-2134-6

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, *14*(1), Article 13. https://doi.org/10.1186/s13040-021-00244-z

Coutts, A. J., & Reaburn, P. (2008). Monitoring changes in rugby league players' perceived stress and recovery during intensified training. *Perceptual and Motor Skills*, *106*(3), 904–916. https://doi.org/10.2466/pms.106.3.904-916

Cox, R. H., Martens, M. P., & Russell, W. D. (2003). Measuring anxiety in athletics: The Revised Competitive State Anxiety Inventory–2. *Journal of Sport and Exercise Psychology*, *25*(4), 519–533. https://doi.org/10.1123/jsep.25.4.519

Cummins, C., Orr, R., O'Connor, H., & West, C. (2013). Global positioning systems (GPS) and microtechnology sensors in team sports: A systematic review. *Sports Medicine*, *43*(10), 1025–1042. https://doi.org/10.1007/s40279-013-0069-2

Currell, K., & Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports Medicine*, *38*(4), 297–316. https://doi.org/10.2165/00007256-200838040-00003

Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. (2021). Classification of Australian football kick types in-situation via ankle-mounted inertial measurement units. *Journal of Sports Sciences*, *39*(12), 1330–1338. https://doi.org/10.1080/02640414.2020.1868678

Dalton-Barron, N., Palczewska, A., Weaving, D., Rennie, G., Beggs, C., Roe, G., & Jones, B. (2022). Clustering of match running and performance indicators to assess between- and within-playing position similarity in professional rugby league. *Journal of Sports Sciences*, *40*(15), 1712–1721. https://doi.org/10.1080/02640414.2022.2100781

Datta, S., & Datta, S. (2003). Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics*, *19*(4), 459–466. https://doi.org/10.1093/bioinformatics/btg025

Davids, K., & Araújo, D. (2013). An ecological dynamics approach to skill acquisition: Implications for development of talent in sport. *Talent Development & Excellence*, *5*(1), 21–34.

Davids, K., Araújo, D., Hristovski, R., Passos, P., & Chow, J. Y. (2012). Ecological dynamics and motor learning design in sport. In M. Williams & N. Hodges (Eds.), *Skill acquisition in sport: Research, theory and practice* (2nd ed., pp. 112–130). Routledge. https://doi.org/10.13140/RG.2.1.2297.0089

Davids, K., Araújo, D., Seifert, L., & Orth, D. (2015). Expert performance in sport: An ecological dynamics perspective. In J. Baker & D. Farrow (Eds.), *Routledge handbook of sport expertise* (1st ed., pp. 130–144). Routledge.

Davies, M. J., Young, W., Farrow, D., & Bahnert, A. (2013). Comparison of agility demands of small-sided games in elite Australian football. *International Journal of Sports Physiology and Performance*, *8*(2), 139–147. https://doi.org/10.1123/ijspp.8.2.139

Dawson, P., & Dobson, S. (2010). The influence of social pressure and nationality on individual decisions: Evidence from the behaviour of referees. *Journal of Economic Psychology*, *31*(2), 181–191. https://doi.org/10.1016/j.joep.2009.06.001

Delaney, J. A., Thornton, H. R., Rowell, A. E., Dascombe, B. J., Aughey, R. J., & Duthie, G. M. (2017). Modelling the decrement in running intensity within professional soccer players. *Science and Medicine in Football, 2*(2), 86–92. https://doi.org/10.1080/24733938.2017.1383623

Dickinson, R. K., & Hanrahan, S. J. (2009). An investigation of subjective sleep and fatigue measures for use with elite athletes. *Journal of Clinical Sport Psychology*, *3*(3), 244–266. https://doi.org/10.1123/jcsp.3.3.244

Dicks, M., Button, C., & Davids, K. (2010). Examination of gaze behaviors under in situ and video simulation task constraints reveals differences in information pickup for perception and action. *Attention, Perception, & Psychophysics*, *72*(3), 706–720. https://doi.org/10.3758/app.72.3.706

Di Salvo, V., Gregson, W., Atkinson, G., Tordoff, P., & Drust, B. (2009). Analysis of high intensity activity in Premier League soccer. *International Journal of Sports Medicine*, *30*(3), 205–212. https://doi.org/10.1055/s-0028-1105950

Di Salvo, V., Baron, R., González-Haro, C., Gormasz, C., Pigozzi, F., & Bachl, N. (2010). Sprinting analysis of elite soccer players during European Champions League and UEFA Cup matches. *Journal of Sports Sciences*, *28*(14), 1489–1494. https://doi.org/10.1080/02640414.2010.521166

Di Salvo, V., Baron, R., Tschan, H., Montero, F. J. C., Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *International Journal of Sports Medicine*, *28*(3), 222–227. https://doi.org/10.1055/s-2006-924294

Duk, O. S., Min, K. S., Kawczyński, A., Chmura, P., Mroczek, D., & Chmura, J. (2011). Endurance and speed capacity of the Korea republic football national team during the world cup of 2010. *Journal of Human Kinetics*, *30*, 115–121. https://doi.org/10.2478/v10078-011-0079-9

Durán, G. (2021). Sports scheduling and other topics in sports analytics: a survey with special reference to Latin America. *TOP*, *29*(1), 125–155. https://doi.org/10.1007/s11750-020-00576-9

Durán, G., Durán, S., Marenco, J., Mascialino, F., & Rey, P. A. (2019). Scheduling Argentina's professional basketball leagues: A variation on the travelling tournament problem. *European Journal of Operational Research*, *275*(3), 1126–1138. https://doi.org/10.1016/j.ejor.2018.12.018

Durán, G., Guajardo, M., & Gutiérrez, F. (2022). Efficient referee assignment in Argentinean professional basketball leagues using operations research methods. *Annals of Operations Research*, *316*(2), 1121–1139. https://doi.org/10.1007/s10479-020-03897-x

Durán, G., Guajardo, M., & Sauré, D. (2017). Scheduling the South American qualifiers to the 2018 FIFA World Cup by integer programming. *European Journal of Operational Research*, *262*(3), 1109–1115. https://doi.org/10.1016/j.ejor.2017.04.043

Durán, G., Guajardo, M., & Wolf-Yadlin, R. (2012). Operations research techniques for scheduling Chile's Second Division Soccer League. *Interfaces*, *42*(3), 273–285. https://doi.org/10.1287/inte.1110.0608

Elsworthy, N., Burke, D., & Dascombe, B., (2014). Factors relating to the decision performance of Australian football officials. *International Journal of Performance Analysis in Sport*, 14:2, 401-410, https://doi.org/10.1080/24748668.2014.11868730

Erickson, B. J., Harris, J. D., Cvetanovich, G. L., Bach, B. R., Bush-Joseph, C. A., Abrams, G. D., Gupta, A. K., McCormick, F. M., & Cole, B. J. (2013). Performance and return to sport after anterior cruciate ligament reconstruction in male Major League soccer players. *Orthopaedic Journal of Sports Medicine*, *2*(9), 1–10. https://doi.org/10.1177/2325967113497189

Esco, M., Flatt, A., & Nakamura, F. (2016). Initial weekly HRV response is related to the prospective change in VO2max in female soccer players. *International Journal of Sports Medicine*, *37*(6), 436–441. https://doi.org/10.1055/s-0035-1569342

Eston, R. (2012). Use of ratings of perceived exertion in sports. *International Journal of Sports Physiology and Performance*, *7*(2), 175–182. https://doi.org/10.1123/ijspp.7.2.175

Farber, J., Harris, J. D., Kolstad, K., & McCulloch, P. C. (2014). Treatment of anterior cruciate ligament injuries by Major League Soccer team physicians. *Orthopaedic Journal of Sports Medicine*, *2*(11), 1–7. https://doi.org/10.1177/2325967114559892

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, *26*(2), 105–109. https://doi.org/10.3969/j.issn.1002-0829.2014.02.009

Fereday, K., Hills, S. P., Russell, M., Smith, J., Cunningham, D. J., Shearer, D., McNarry, M., & Kilduff, L. P. (2020). A comparison of rolling averages versus discrete time epochs for assessing the worst-case scenario locomotor demands of professional soccer match-play. *Journal of Science and Medicine in Sport*, *23*(8), 764–769. https://doi.org/10.1016/j.jsams.2020.01.002

Fernandes, C. J., Yakubov, R., Li, Y., Prasad, A. K., & Chan, T. C. Y. (2020). Predicting plays in the National Football League. *Journal of Sports Analytics*, *6*(1), 35–43. https://doi.org/10.3233/jsa-190348

Flatt, A. A., & Esco, M. R. (2014). Endurance performance relates to resting heart rate and its variability: A case study of a collegiate male cross-country athlete. *Journal of Australian Strength and Conditioning*, *22*(6), 39–45.

Flatt, A. A., & Esco, M. R. (2015). Smartphone-derived heart-rate variability and training load in a women's soccer team. *International Journal of Sports Physiology and Performance*, *10*(8), 994–1000. https://doi.org/10.1123/ijspp.2014-0556

Forbes-Robertson, S., Dudley, E., Vadgama, P., Cook, C., Drawer, S., & Kilduff, L. (2012). Circadian disruption and remedial interventions. *Sports Medicine*, *42*(3), 185–208. https://doi.org/10.2165/11596850-000000000-00000

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A new approach to monitoring exercise training. *Journal of Strength and Conditioning Research*, *15*(1), 109–115. https://doi.org/10.1519/00124278-200102000-00019

Fowler, P. M., Knez, W., Crowcroft, S., Mendham, A. E., Miller, J., Sargent, C., Halson, S., & Duffield, R. (2017). Greater effect of east versus west travel on jet lag, sleep, and team sport performance. *Medicine & Science in Sports & Exercise*, *49*(12), 2548–2561. https://doi.org/10.1249/mss.0000000000001374

Fry, M. J., & Ohlmann, J. W. (2012). Introduction to the Special Issue on Analytics in Sports, Part II: Sports Scheduling Applications. *Interfaces*, *42*(3), 229–331. http://www.jstor.org/stable/23254859

Gallo, T. F., Cormack, S. J., Gabbett, T. J., William, S., & Lorenzen, C. H. (2015). Characteristics impacting on session rating of perceived exertion training load in Australian footballers. *Journal of Sports Sciences*, *33*(5), 467–475. https://doi.org/10.1080/02640414.2014.947311

Gallo, T. F., Cormack, S. J., Gabbett, T. J., & Lorenzen, C. H. (2016). Pre-training perceived wellness impacts training output in Australian football players. *Journal of Sports Sciences*, *34*(15), 1445–1451. https://doi.org/10.1080/02640414.2015.1119295

Gallo, T. F., Cormack, S. J., Gabbett, T. J., & Lorenzen, C. H. (2017). Self-reported wellness profiles of professional Australian football players during the competition phase of the season. *Journal of Strength and Conditioning Research*, *31*(2), 495–502. https://doi.org/10.1519/jsc.0000000000001515

Gastin, P. B., Fahrner, B., Meyer, D., Robinson, D., & Cook, J. L. (2013). Influence of physical fitness, age, experience, and weekly training load on match performance in elite Australian football. *The Journal of Strength & Conditioning Research*, *27*(5), 1272–1279. https://doi.org/10.1519/jsc.0b013e318267925f

Gastin, P. B., Meyer, D., & Robinson, D. (2013). Perceptions of wellness to monitor adaptive responses to training and competition in elite Australian football. *The Journal of Strength & Conditioning Research*, *27*(9), 2518–2526. https://doi.org/10.1519/jsc.0b013e31827fd600

Gaudino, P., Iaia, F. M., Alberti, G., Hawkins, R. D., Strudwick, A. J., & Gregson, W. (2014). Systematic bias between running speed and metabolic power data in elite soccer players: influence of drill type. *International Journal of Sports Medicine*, *35*(6), 489–493. https://doi.org/10.1055/s-0033-1355418

Gaudino, P., Iaia, F. M., Strudwick, A. J., Hawkins, R. D., Alberti, G., Atkinson, G., & Gregson, W. (2015). Factors influencing perception of effort (session rating of perceived exertion) during elite soccer training. *International Journal of Sports Physiology and Performance*, *10*(7), 860–864. https://doi.org/10.1123/ijspp.2014-0518

Gilis, B., Helsen, W., Catteeuw, P., & Wagemans, J. (2008). Offside decisions by expert assistant referees in association football: Perception and recall of spatial positions in complex dynamic events. *Journal of Experimental Psychology: Applied*, *14*(1), 21–35. https://doi.org/10.1037/1076-898x.14.1.21

Gómez, M.-Á., Mitrotasios, M., Armatas, V., & Lago-Peñas, C. (2018). Analysis of playing styles according to team quality and match location in Greek professional soccer. *International Journal of Performance Analysis in Sport*, *18*(6), 986–997. https://doi.org/10.1080/24748668.2018.1539382

Gomez-Piqueras, P., González-Villora, S., Castellano, J., & Teoldo, I. (2019). Relation between the physical demands and success in professional soccer players. *Journal of Human Sport and Exercise*, *14*(1), 1–11. https://doi.org/10.14198/jhse.2019.141.01

Gong, X. H., & Ling, P. L. (2009). *Discussion on psychological decision support system of table tennis players based on multi-agent system* [Conference session] (pp. 596–601). 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA, United States. https://doi.org/10.1109/csie.2009.507

González-Rodenas, J., Lopez-Bondia, I., Calabuig, F., Pérez-Turpin, J. A., & Aranda, R. (2017a). Association between playing tactics and creating scoring opportunities in counterattacks from United States Major League Soccer games. *International Journal of Performance Analysis in Sport*, *16*(2), 737–752. https://doi.org/10.1080/24748668.2016.11868920

González-Rodenas, J., Lopez-Bondia, I., Calabuig, F., Pérez-Turpin, J. A., & Aranda, R. (2017b). The effects of playing tactics on creating scoring opportunities in random matches from US Major League Soccer. *International Journal of Performance Analysis in Sport*, *15*(3), 851–872. https://doi.org/10.1080/24748668.2015.11868836

Gregory, S., Robertson, S., Aughey, R., & Duthie, G. (2022). The influence of tactical and match context on player movement in football. *Journal of Sports Sciences*, *40*(9), 1063–1077. https://doi.org/10.1080/02640414.2022.2046938

Gregson, W., Drust, B., Atkinson, G., & Salvo, V. (2010). Match-to-match variability of high-speed activities in Premier League Soccer. *International Journal of Sports Medicine*, *31*(4), 237–242. https://doi.org/10.1055/s-0030-1247546

Gyarmati, L., Kwak, H., & Rodriguez, P. (2014). Searching for a unique style in soccer. *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD) Workshop on Large-Scale Sports Analytics.*

Haddad, M., Chaouachi, A., Wong, D. P., Castagna, C., Hambli, M., Hue, O., & Chamari, K. (2013). Influence of fatigue, stress, muscle soreness and sleep on perceived exertion during submaximal effort. *Physiology & Behavior*, *119*, 185–189. https://doi.org/10.1016/j.physbeh.2013.06.016

Harrop, K., & Nevill, A. (2014). Performance indicators that predict success in an English professional League One soccer team. *International Journal of Performance Analysis in Sport*, *14*(3), 907–920. https://doi.org/10.1080/24748668.2014.11868767

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, *17*, 107–145. https://doi.org/10.1023/A:1012801612483

Helsen, W., & Bultynck, J.-B. (2004). Physical and perceptual-cognitive demands of top-class refereeing in association football. *Journal of Sports Sciences*, *22*(2), 179–189. https://doi.org/10.1080/02640410310001641502

Hooper, S. L., Mackinnon, L. T., Howa, A., Goon, R. D., & Chmann, A. W. (1995). Markers for monitoring overtraining and recovery. *Medicine & Science in Sports & Exercise*, *27*(1), 106–112. https://doi.org/10.1249/00005768-199501000-00019

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, *41*(1), 3–12. https://doi.org/10.1249/mss.0b013e31818cb278

Horvat, T., Havaš, L., & Srpak, D. (2020). The impact of selecting a validation method in machine learning on predicting basketball game outcomes. *Symmetry*, *12*(3), Article 431. https://doi.org/10.3390/sym12030431

Hubáček, O., Šourek, G., & Železný, F. (2019). Learning to predict soccer results from relational data with gradient boosted trees. *Machine Learning*, *108*(1), 29–47. https://doi.org/10.1007/s10994-018-5704-6

Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences*, *20*(10), 739–754. https://doi.org/10.1080/026404102320675602

Hüttermann, S., & Memmert, D. (2017). The attention window: A narrative review of limitations and opportunities influencing the focus of attention. *Research Quarterly for Exercise and Sport*, *88*(2), 169–183. https://doi.org/10.1080/02701367.2017.1293228

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

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

International Football Association Board. (n.d.). *Laws of the game*. https://www.theifab.com/

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). Multiple testing. In *An introduction to statistical learning, with applications in R* (pp. 553–595). Springer. https://doi.org/10.1007/978-1-0716-1418-1_13

Janakiraman, V. S., Sarukesi, K., & Surakest, K. (2008). *Decision support systems*. Phi.

Jennings, D., Cormack, S. J., Coutts, A. J., & Aughey, R. J. (2012). GPS analysis of an international field hockey tournament. *International Journal of Sports Physiology and Performance*, *7*(3), 224–231. https://doi.org/10.1123/ijspp.7.3.224

Jiang, J., Ge, H., Du, L., Gomez, M.-A., Gong, B., & Cui, Y. (2022). Impact of match type and match halves on referees' physical performance and decision-making distance in Chinese Football Super League. *Frontiers in Psychology*, *13*, Article 864957. https://doi.org/10.3389/fpsyg.2022.864957

Johansen, B. T., & Erikstad, M. K. (2021). A preliminary analysis of the importance of distance, angle, and insight when soccer referees make penalty decisions. *Frontiers in Sports and Active Living*, *2*, Article 595703. https://doi.org/10.3389/fspor.2020.595703

Johnston, R. D., Thornton, H. R., Wade, J. A., Devlin, P., & Duthie, G. M. (2022). The distribution of match activities relative to the maximal mean intensities in professional rugby league and Australian football. *Journal of Strength and Conditioning Research*, *36*(5), 1360–1366. https://doi.org/10.1519/JSC.0000000000003613

Jones, S. (2003). What is an affordance? *Ecological Psychology, 15*(2), 107–114. https://doi.org/10.1207/S15326969ECO1502

Ju, W., Doran, D., Hawkins, R., Gómez-Díaz, A., Martin-Garcia, A., Ade, J., Laws, A., Evans, M., & Bradley, P. (2022). Contextualised peak periods of play in English Premier League matches. *Biology of Sport*, *39*(4), 973–983. https://doi.org/10.5114/biolsport.2022.112083

Ju, W., Lewis, C. J., Evans, M., Laws, A., & Bradley, P. S. (2022). The validity and reliability of an integrated approach for quantifying match physical-tactical performance. *Biology of Sport*, *39*(2), 253–261. https://doi.org/10.5114/biolsport.2022.104919

Kostuk, K. J., & Willoughby, K. A. (2012). A decision support system for scheduling the Canadian Football League. *Interfaces, 42*(3), 286–295. https://doi.org/10.1287/inte.1110.0561

Krosnick, J. A., & Fabrigar, L. R. (1997). *Designing rating scales for effective measurement in surveys* (Vol. 20). John Wiley & Sons. https://doi.org/10.1002/9781118490013

Krustrup, P., & Bangsbo, J. (2001). Physiological demands of top-class soccer refereeing in relation to physical capacity: Effect of intense intermittent exercise training. *Journal of Sports Sciences*, *19*(11), 881–891. https://doi.org/10.1080/026404101753113831

Krustrup, P., Helsen, W., Randers, M. B., Christensen, J. F., MacDonald, C., Rebelo, A. N., & Bangsbo, J. (2009). Activity profile and physical demands of football referees and assistant referees in international games. *Journal of Sports Sciences*, *27*(11), 1167–1176. https://doi.org/10.1080/02640410903220310

Krustrup, P., Mohr, M., & Bangsbo, J. (2002). Activity profile and physiological demands of top-class soccer assistant refereeing in relation to training status. *Journal of Sports Sciences*, *20*(11), 861–871. https://doi.org/10.1080/026404102320761778

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta Package. *Journal of Statistical Software*, *36*(11), 1–13. https://doi.org/10.18637/jss.v036.i11

Lago, C., Casais, L., Dominguez, E., & Sampaio, J. (2010). The effects of situational variables on distance covered at various speeds in elite soccer. *European Journal of Sport Science*, *10*(2), 103–109. https://doi.org/10.1080/17461390903273994

Lago-Ballesteros, J., & Lago-Peñas, C. (2010). Performance in team sports: Identifying the keys to success in soccer. *Journal of Human Kinetics*, *25*(2010), 85–91. https://doi.org/10.2478/v10078-010-0035-0

Lago-Peñas, C. (2012a). The influence of effective playing time on physical demands of elite soccer players. *The Open Sports Sciences Journal*, *5*(1), 188–192. https://doi.org/10.2174/1875399x01205010188

Lago-Peñas, C. (2012b). The role of situational variables in analysing physical performance in soccer. *Journal of Human Kinetics*, *35*(1), 89–95. https://doi.org/10.2478/v10078-012-0082-9

Lane, A. M., Nevill, A. M., Ahmad, N. S., & Balmer, N. (2006). Soccer referee decision-making: "Shall I blow the whistle?" *Journal of Sports Science & Medicine*, *5*(2), 243–253.

Lastella, M., Roach, G. D., & Sargent, C. (2019). Travel fatigue and sleep/wake behaviors of professional soccer players during international competition. *Sleep Health*, *5*(2), 141–147. https://doi.org/10.1016/j.sleh.2018.10.013

Lazic, J. S., Dekleva, M., Soldatovic, I., Leischik, R., Suzic, S., Radovanovic, D., Djuric, B., Nesic, D., Lazic, M., & Mazic, S. (2017). Heart rate recovery in elite athletes: the impact of age and exercise capacity. *Clinical Physiology and Functional Imaging*, *37*(2), 117–123. https://doi.org/10.1111/cpf.12271

Lee, A., & Galvez, J. C. (2012). Jet lag in athletes. *Sports Health*, *4*(3), 211–216. https://doi.org/10.1177/1941738112442340

Leicht, A. S., Gómez, M. A., & Woods, C. T. (2017a). Explaining match outcome during the men's basketball tournament at the Olympic Games. *Journal of Sports Science & Medicine*, *16*(4), 468–473.

Leicht, A. S., Gomez, M. A., & Woods, C. T. (2017b). Team performance indicators explain outcome during women's basketball matches at the Olympic Games. *Sports*, *5*(4), Article 96. https://doi.org/10.3390/sports5040096

Lemmer, B., Kern, R.-I., Nold, G., & Lohrer, H. (2002). Jet lag in athletes after eastward and westward time-zone transition. *Chronobiology International*, *19*(4), 743–764. https://doi.org/10.1081/cbi-120005391

Leventer, L., Eek, F., Hofstetter, S., & Lames, M. (2016). Injury patterns among elite football players: A media-based analysis over 6 seasons with emphasis on playing position. *International Journal of Sports Medicine*, *37*(11), 898–908. https://doi.org/10.1055/s-0042-108201

Li, H. (2020). Analysis on the construction of sports match prediction model using neural network. *Soft Computing*, *24*(11), 8343–8353. https://doi.org/10.1007/s00500-020-04823-w

Linfati, R., Gatica, G., & Escobar, J. W. (2019). A flexible mathematical model for the planning and designing of a sporting fixture by considering the assignment of referees. *International Journal of Industrial Engineering Computations*, *10*, 281–294. https://doi.org/10.5267/j.ijiec.2018.6.004

Liu, H., Gómez, M. A., Gonçalves, B., & Sampaio, J. (2016). Technical performance and match-to-match variation in elite football teams. *Journal of Sports Sciences*, *34*(6), 509–518. https://doi.org/10.1080/02640414.2015.1117121

Liu, H., Hopkins, W. G., & Gómez, M.-A. (2016). Modelling relationships between match events and match outcome in elite football. *European Journal of Sport Science*, *16*(5), 516–525. https://doi.org/10.1080/17461391.2015.1042527

Loat, C. E. R., & Rhodes, E. C. (1989). Jet-lag and human performance. *Sports Medicine*, *8*(4), 226–238. https://doi.org/10.2165/00007256-198908040-00003

Loghmani, M., Cuskelly, G., & Webb, T. (2021). Examining the career dynamics of elite football referees: a unique identification profile. *Sport Management Review*, *24*(3), 517–542. https://doi.org/10.1080/14413523.2021.1879556

Lovell, T. W. J., Sirotic, A. C., Impellizzeri, F. M., & Coutts, A. J. (2013). Factors affecting perception of effort (session rating of perceived exertion) during rugby league training. *International Journal of Sports Physiology and Performance*, *8*(1), 62–69. https://doi.org/10.1123/ijspp.8.1.62

MacMahon, C., Helsen, W. F., Starkes, J. L., & Weston, M. (2007). Decision-making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences*, *25*(1), 65–78. https://doi.org/10.1080/02640410600718640

Mallo, J., Frutos, P. G., Juárez, D., & Navarro, E. (2012). Effect of positioning on the accuracy of decision making of association football top-class referees and assistant referees during competitive matches. *Journal of Sports Sciences*, *30*(13), 1437–1445. https://doi.org/10.1080/02640414.2012.711485

Mallo, J., Mena, E., Nevado, F., & Paredes, V. (2015). Physical demands of top-class soccer friendly matches in relation to a playing position using global positioning system technology. *Journal of Human Kinetics*, *47*(1), 179–188. https://doi.org/10.1515/hukin-2015-0073

Mallo, J., Navarro, E., García-Aranda, J.-M., & Helsen, W. F. (2009). Activity profile of top-class association football referees in relation to fitness-test performance and match standard. *Journal of Sports Sciences*, *27*(1), 9–17. https://doi.org/10.1080/02640410802298227

Mallo, J., Navarro, E., García-Aranda, J.-M., Gilis, B., & Helsen, W. (2007). Activity profile of top-class association football referees in relation to performance in selected physical tests. *Journal of Sports Sciences*, *25*(7), 805–813. https://doi.org/10.1080/02640410600778602

Malone, J. J., Michele, R. D., Morgans, R., Burgess, D., Morton, J. P., & Drust, B. (2015). Seasonal training-load quantification in elite English Premier League soccer players. *International Journal of Sports Physiology and Performance*, *10*(4), 489–497. https://doi.org/10.1123/ijspp.2014-0352

Malone, S., Owen, A., Newton, M., Mendes, B., Tiernan, L., Hughes, B., & Collins, K. (2017). Wellbeing perception and the impact on external training output among elite soccer players. *Journal of Science and Medicine in Sport*, *21*(1), 29–34. https://doi.org/10.1016/j.jsams.2017.03.019

Mancini, S., & Isabello, A. (2014). Fair referee assignment for the Italian soccer series. *Journal of Quantitative Analysis in Sports*, *10*(2), 153–160. https://doi.org/10.1515/jqas-2013-0108

McLaren, S. J., Weston, M., Smith, A., Cramb, R., & Portas, M. D. (2016). Variability of physical performance and player match loads in professional rugby union. *Journal of Science and Medicine in Sport*, *19*(6), 493–497. https://doi.org/10.1016/j.jsams.2015.05.010

McLean, B. D., Coutts, A. J., Kelly, V., McGuigan, M. R., & Cormack, S. J. (2010). Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. *International Journal of Sports Physiology and Performance*, *5*(3), 367–383. https://doi.org/10.1123/ijspp.5.3.367

Meeusen, R., Duclos, M., Foster, C., Fry, A., Gleeson, M., Nieman, D., Raglin, J., Rietjens, G., Steinacker, J., & Urhausen, A. (2013). Prevention, diagnosis, and treatment of the overtraining syndrome: Joint consensus statement of the European College of Sport Science and the American College of Sports Medicine. *Medicine & Science in Sports & Exercise, 45*(1), 186–205. https://doi.org/10.1249/mss.0b013e318279a10a

Melville, P., Mooney, R. J., & Nagarajan, R. (2002). *Content-boosted collaborative filtering for improved recommendations* [Conference session] (pp. 187–192). Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence, July 28–August 1, 2002, Edmonton, Alberta, Canada. https://cdn.aaai.org/AAAI/2002/AAAI02-029.pdf

Memmert, D., Lemmink, K. A. P. M., & Sampaio, J. (2017). Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, *47*, 1–10. https://doi.org/10.1007/s40279-016-0562-5

Mernagh, D., Weldon, A., Wass, J., Phillips, J., Parmar, N., Waldron, M., & Turner, A. (2021). A comparison of match demands using ball-in-play versus whole match data in professional soccer players of the English Championship. *Sports*, *9*(6), Article 76. https://doi.org/10.3390/sports9060076

Moalla, W., Fessi, M. S., Farhat, F., Nouira, S., Wong, D. P., & Dupont, G. (2016). Relationship between daily training load and psychometric status of professional soccer players. *Research in Sports Medicine*, *24*(4), 387–394. https://doi.org/10.1080/15438627.2016.1239579

Mohr, M., & Krustrup, P., & Bangsbo, J. (2003). Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Sciences*, *21*(7), 519–528. https://doi.org/10.1080/0264041031000071182

Montgomery, A. A., Fahey, T., Peters, T. J., MacIntosh, C., & Sharp, D. J. (2000). Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: Randomised controlled trial. *BMJ*, *320*(7236), 686–690. https://doi.org/10.1136/bmj.320.7236.686

Montgomery, P. G., & Hopkins, W. G. (2013). The effects of game and training loads on perceptual responses of muscle soreness in Australian football. *International Journal of Sports Physiology and Performance*, *8*(3), 312–318. https://doi.org/10.1123/ijspp.8.3.312

Mulholland, J., & Jensen, S. T. (2014). Predicting the draft and career success of tight ends in the National Football League. *Journal of Quantitative Analysis in Sports*, *10*(4), 381–396. https://doi.org/10.1515/jqas-2013-0134

Nevill, A. M., Balmer, N. J., & Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, *3*(4), 261–272. https://doi.org/10.1016/s1469-0292(01)00033-4

Nevill, A. M., Holder, R. L., Bardsley, A., Calvert, H., & Jones, S. (1997). Identifying home advantage in international tennis and golf tournaments. *Journal of Sports Sciences*, *15*(4), 437–443. https://doi.org/10.1080/026404197367227

Nevill, A. M., Newell, S. M., & Gale, S. (1996). Factors associated with home advantage in English and Scottish soccer matches. *Journal of Sports Sciences*, *14*(2), 181–186. https://doi.org/10.1080/02640419608727700

Nicholas, C. W., Nuttall, F. E., & Williams, C. (2000). The Loughborough Intermittent Shuttle Test: A field test that simulates the activity pattern of soccer. *Journal of Sports Sciences*, *18*(2), 97–104. https://doi.org/10.1080/026404100365162

O'Donoghue, P. (2005). Normative profiles of sports performance. *International Journal of Performance Analysis in Sport*, *5*(1), 104–119. https://doi.org/10.1080/24748668.2005.11868319

Oliva-Lozano, J. M., Rojas-Valverde, D., Gómez-Carmona, C. D., Fortes, V., & Pino-Ortega, J. (2020). Worst case scenario match analysis and contextual variables in professional soccer players: a longitudinal study. *Biology of Sport*, *37*(4), 429–436. https://doi.org/10.5114/biolsport.2020.97067

Oliveira, R., Brito, J. P., Martins, A., Mendes, B., Marinho, D. A., Ferraz, R., & Marques, M. C. (2019). In-season internal and external training load quantification of an elite European soccer team. *PLoS ONE*, *14*(4), Article e0209393. https://doi.org/10.1371/journal.pone.0209393

Oudejans, R. R. D., Verheijen, R., Bakker, F. C., Gerrits, J. C., Steinbrückner, M., & Beek, P. J. (2000). Errors in judging 'offside' in football. *Nature*, *404*(6773), 33–33. https://doi.org/10.1038/35003639

Paraskevas, G., Smilios, I., & Hadjicharalambous, M. (2020). Effect of opposition quality and match location on the positional demands of the 4-2-3-1 formation in elite soccer. *Journal of Exercise Science and Fitness*, *18*(1), 40–45. https://doi.org/10.1016/j.jesf.2019.11.001

Park, L. A. F., Scott, D., & Lovell, R. (2019). Velocity zone classification in elite women's football: Where do we draw the lines? *Science and Medicine in Football*, *3*(1), 21–28. https://doi.org/10.1080/24733938.2018.1517947

Paul, D. J., Bradley, P. S., & Nassis, G. P. (2015). Factors affecting match running performance of elite soccer players: Shedding some light on the complexity. *International Journal of Sports Physiology and Performance*, *10*(4), 516–519. https://doi.org/10.1123/IJSPP.2015-0029

Pavey, T. G., Gilson, N. D., Gomersall, S. R., Clark, B., & Trost, S. G. (2017). Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *Journal of Science and Medicine in Sport*, *20*(1), 75–80. https://doi.org/10.1016/j.jsams.2016.06.003

Pion, J., Hohmann, A., Liu, T., Lenoir, M., & Segers, V. (2017). Predictive models reduce talent development costs in female gymnastics. *Journal of Sports Sciences*, *35*(8), 806–811. https://doi.org/10.1080/02640414.2016.1192669

Plews, D. J., Laursen, P. B., Kilding, A. E., & Buchheit, M. (2012). Heart rate variability in elite triathletes, is variation in variability the key to effective training? A case comparison. *European Journal of Applied Physiology*, *112*(11), 3729–3741. https://doi.org/10.1007/s00421-012-2354-4

Plews, D. J., Laursen, P. B., Kilding, A. E., & Buchheit, M. (2013). Evaluating training adaptation with heart-rate measures: A methodological comparison. *International Journal of Sports Physiology and Performance*, *8*(6), 688–691. https://doi.org/10.1123/ijspp.8.6.688

Rampinini, E., Coutts, A., Castagna, C., Sassi, R., & Impellizzeri, F. (2007). Variation in top level soccer match performance. *International Journal of Sports Medicine*, *28*(12), 1018–1024. https://doi.org/10.1055/s-2007-965158

Raghavan, S. R., Ladik, V., & Meyer, K. B. (2005). Developing decision support for dialysis treatment of chronic kidney failure. *IEEE Transactions on Information Technology in Biomedicine*, *9*(2), 229–238. https://doi.org/10.1109/titb.2005.847133

Reilly, T., & Edwards, B. (2007). Altered sleep–wake cycles and physical performance in athletes. *Physiology & Behavior*, *90*(2), 274 284. https://doi.org/10.1016/j.physbeh.2006.09.017

Reilly, T., & Gilbourne, D. (2003). Science and football: A review of applied research in the football codes. *Journal of Sports Sciences*, *21*(9), 693–705. https://doi.org/10.1080/0264041031000102105

Reilly, T., Morris, T., & Whyte, G. (2009). The specificity of training prescription and physiological assessment: A review. *Journal of Sports Sciences*, *27*(6), 575–589. https://doi.org/10.1080/02640410902729741

Renshaw, I., Chow, J. Y., Davids, K., & Hammond, J. (2010). A constraints-led perspective to understanding skill acquisition and game play: A basis for integration of motor learning theory and physical education praxis? *Physical Education and Sport Pedagogy*, *15*(2), 117–137. https://doi.org/10.1080/17408980902791586

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56–58. https://doi.org/10.1145/245108.245121

Ribeiro, C. C. (2012). Sports scheduling: problems and applications. *International Transactions in Operational Research, 19*(1–2), 201–226. https://doi.org/10.1111/j.1475-3995.2011.00819.x

Riboli, A., Semeria, M., Coratella, G., & Esposito, F. (2021). Effect of formation, ball in play and ball possession on peak demands in elite soccer. *Biology of Sport*, *38*(2), 195–205. https://doi.org/10.5114/biolsport.2020.98450

Riedel, S. L., & Pitz, G. F. (1986). Utilization-oriented evaluation of decision support systems. *IEEE Transactions on Systems, Man, and Cybernetics*, *16*(6), 980–996. https://doi.org/10.1109/tsmc.1986.4309016

Riiser, A., Andersen, V., Sæterbakken, A., Ylvisaker, E., & Moe, V. F. (2019). Running performance and position is not related to decision-making accuracy in referees. *Sports Medicine International Open*, *3*(2), E66–E71. https://doi.org/10.1055/a-0958-8608

Ritchie, D., Hopkins, W. G., Buchheit, M., Cordy, J., & Bartlett, J. D. (2016). Quantification of training and competition load across a season in an elite Australian football club. *International Journal of Sports Physiology and Performance*, *11*(4), 474–479. https://doi.org/10.1123/ijspp.2015-0294

Robertson, R., & Noble, B. (1997). Perception of physical exertion: Methods, mediators, and applications. *Exercise and Sport Sciences Reviews*, *25*, 407–452. http://europepmc.org/abstract/MED/9213100

Robertson, S., Bartlett, J. D., & Gastin, P. B. (2016). Red, amber or green? Athlete monitoring in team sport: The need for decision support systems. *International Journal of Sports Physiology and Performance*, *12*(s2), 73–79. https://doi.org/10.1123/ijspp.2016-0541

Rogalski, B., Dawson, B., Heasman, J., & Gabbett, T. J. (2013). Training and game loads and injury risk in elite Australian footballers. *Journal of Science and Medicine in Sport*, *16*(6), 499–503. https://doi.org/10.1016/j.jsams.2012.12.004

Rommers, N., Rössler, R., Verhagen, E., Vandecasteele, F., Verstockt, S., Vaeyens, R., Lenoir, M., D'Hondt, E., & Witvrouw, E. (2020). A machine learning approach to assess injury risk in elite youth football players. *Medicine and Science in Sports and Exercise*, *52*(8), 1745–1751. https://doi.org/10.1249/MSS.0000000000002305

Rowell, A. E., Aughey, R. J., Hopkins, W. G., Stewart, A. M., & Cormack, S. J. (2016). Identification of sensitive measures of recovery following external load from football match play. *International Journal of Sports Physiology and Performance*, *12*(7), 969–976. https://doi.org/10.1123/ijspp.2016-0522

Russell, M., Sparkes, W., Northeast, J., Cook, C. J., Love, T. D., Bracken, R. M., & Kilduff, L. P. (2016). Changes in acceleration and deceleration capacity throughout professional soccer match-play. *The Journal of Strength & Conditioning Research*, *30*(10), 2839–2844. https://doi.org/10.1519/jsc.0000000000000805

Samuels, C. H. (2012). Jet lag and travel fatigue: A comprehensive management plan for sport medicine physicians and high-performance support teams. *Clinical Journal of Sport Medicine*, *22*(3), 268–273. https://doi.org/10.1097/jsm.0b013e31824d2eeb

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), Article 160. https://doi.org/10.1007/s42979-021-00592-x

Saw, A. E., Main, L. C., & Gastin, P. B. (2015a). Monitoring athletes through self-report: Factors influencing implementation. *Journal of Sports Science & Medicine*, *14*(1), 137–146.

Saw, A. E., Main, L. C., & Gastin, P. B. (2015b). Role of a self-report measure in athlete preparation. *The Journal of Strength & Conditioning Research*, *29*(3), 685–691. https://doi.org/10.1519/jsc.0000000000000698

Saw, A. E., Main, L. C., & Gastin, P. B. (2016). Monitoring the athlete training response: Subjective self-reported measures trump commonly used objective measures: a systematic review. *British Journal of Sports Medicine*, *50*(5), 281–291. https://doi.org/10.1136/bjsports-2015-094758

Schelling, X., Fernández, J., Ward, P., Fernández, J., & Robertson, S. (2021). Decision support system applications for scheduling in professional team sport: The team's perspective. *Frontiers in Sports and Active Living*, *3*, Article 678489. https://doi.org/10.3389/fspor.2021.678489

Schelling, X., & Robertson, S. (2020). A development framework for decision support systems in high-performance sport. *International Journal of Computer Science in Sport*, *19*(1), 1–23. https://doi.org/10.2478/ijcss-2020-0001

Schwarz, N., Knäuper, B., Hans-J. Hippler, Noelle-Neumann, E., & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *The Public Opinion Quarterly*, *55*(4), 570–582. http://www.jstor.org/stable/2749407

Scott, A., Malone, J., Morgans, R., Burgess, D., Gregson, W., Morton, J., & Drust, B. (2016). The relationship between physical match performance and 48-h post-game creatine kinase concentrations in English Premier League soccer players. *International Journal of Sports Science & Coaching*, *11*(6), 846–852. https://doi.org/10.1177/1747954116676111

Scott, B. R., Lockie, R. G., Knight, T. J., Clark, A. C., & de Jonge, X. A. K. J. (2013). A comparison of methods to quantify the in-season training load of professional soccer players. *International Journal of Sports Physiology and Performance*, *8*(2), 195–202. https://doi.org/10.1123/ijspp.8.2.195

Shirkhorshidi, A. S., Aghabozorgi, S., & Wah, T. Y. (2015). A comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoS One*, *10*(12), Article e0144059. https://doi.org/10.1371/journal.pone.0144059

Sieghartsleitner, R., Zuber, C., Zibung, M., & Conzelmann, A. (2018). Science or coaches' eye? – both! Beneficial collaboration of multidimensional measurements and coach assessments for efficient talent selection in elite youth football. *Journal of Sports Science & Medicine*, *18*(1), 32–43.

Siegle, M., & Lames, M. (2012). Game interruptions in elite soccer. *Journal of Sports Sciences*, *30*(7), 619–624. https://doi.org/10.1080/02640414.2012.667877

Singh, A., Yadav, A., & Rana, A. (2013). K-means with three different distance metrics. *International Journal of Computer Applications*, *67*(10), 13–17. https://doi.org/10.5120/11430-6785

Slack, L. A., Maynard, I. W., Butt, J., & Olusoga, P. (2013). Factors underpinning football officiating excellence: Perceptions of English Premier League referees. *Journal of Applied Sport Psychology*, *25*(3), 298–315. https://doi.org/10.1080/10413200.2012.726935

Smith, T. B., & Hopkins, W. G. (2011). Variability and predictability of finals times of elite rowers. *Medicine and Science in Sports and Exercise*, *43*(11), 2155–2160. https://doi.org/10.1249/MSS.0b013e31821d3f8e

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems With Applications*, *134*, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028

Stanley, J., Peake, J. M., & Buchheit, M. (2013). Cardiac parasympathetic reactivation following exercise: Implications for training prescription. *Sports Medicine*, *43*(12), 1259–1277. https://doi.org/10.1007/s40279-013-0083-4

Sweeting, A. J., Aughey, R. J., Cormack, S. J., & Morgan, S. (2017). Discovering frequently recurring movement sequences in team-sport athlete spatiotemporal data. *Journal of Sports Sciences*, *35*(24), 2439–2445. https://doi.org/10.1080/02640414.2016.1273536

Taylor, B. J., Mellalieu, D. S., James, N., & Barter, P. (2010). Situation variable effects and tactical performance in professional association football. *International Journal of Performance Analysis in Sport*, *10*(3), 255–269. https://doi.org/10.1080/24748668.2010.11868520

Taylor, J. B., Mellalieu, S. D., James, N., & Shearer, D. A. (2008). The influence of match location, quality of opposition, and match status on technical performance in professional association football. *Journal of Sports Sciences*, *26*(9), 885–895. https://doi.org/10.1080/02640410701836887

Taylor, K.-L. (2012). Fatigue monitoring in high performance sport: A survey of current trends. *Journal of Australian Strength and Conditioning*, *20*(1), 12–23. https://doi.org/10.2307/23601061

Thabtah, F., Zhang, L., & Abdelhamid, N. (2019). NBA game result prediction using feature analysis and machine learning. *Annals of Data Science*, *6*(1), 103–116. https://doi.org/10.1007/s40745-018-00189-x

Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., & Gregson, W. (2015). Monitoring fatigue during the in-season competitive phase in elite soccer players. *International Journal of Sports Physiology and Performance*, *10*(8), 958–964. https://doi.org/10.1123/ijspp.2015-0004

Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., & Gregson, W. (2016a). The influence of changes in acute training load on daily sensitivity of morning-measured fatigue variables in elite soccer players. *International Journal of Sports Physiology and Performance*, *12*(Suppl 2), 1–23. https://doi.org/10.1123/ijspp.2016-0433

Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., & Gregson, W. (2016b). Tracking morning fatigue status across in-season training weeks in elite soccer players. *International Journal of Sports Physiology and Performance*, *11*(7), 947–952. https://doi.org/10.1123/ijspp.2015-0490

Travassos, B., Araújo, D., Davids, K., Vilar, L., Esteves, P., & Vanda, C. (2012). Informational constraints shape emergent functional behaviours during performance of interceptive actions in team sports. *Psychology of Sport and Exercise*, *13*(2), 216–223. https://doi.org/10.1016/j.psychsport.2011.11.009

Trick, M. A., Yildiz, H., & Yunes, T. (2012). Scheduling major league baseball umpires and the traveling umpire problem. *Interfaces*, *42*(3), 232–244. https://www.jstor.org/stable/23254861

Turban, E., Aronson, J. E., & Liang, T.-P. (2004). *Decision support systems and intelligent systems* (7th ed.). Prentice-Hall.

Twist, C., Waldron, M., Highton, J., Burt, D., & Daniels, M. (2012). Neuromuscular, biochemical and perceptual post-match fatigue in professional rugby league forwards and backs. *Journal of Sports Sciences*, *30*(4), 359–367. https://doi.org/10.1080/02640414.2011.640707

Ungar, L., & Foster, D. (2000). *A formal statistical approach to collaborative filtering*. University of Pennsylvania.

Varley, M. C., Gabbett, T., & Aughey, R. J. (2013). Activity profiles of professional soccer, rugby league and Australian football match play. *Journal of Sports Sciences*, *32*(20), 1858–1866. https://doi.org/10.1080/02640414.2013.823227

Vesterinen, V., Nummela, A., Äyrämö, S., Laine, T., Hynynen, E., Mikkola, J., & Häkkinen, K. (2016). Monitoring training adaptation with a submaximal running test under field conditions. *International Journal of Sports Physiology and Performance*, *11*(3), 393–399. https://doi.org/10.1123/ijspp.2015-0366

Wagstaff, K., Cardie, C., Rogers, S., & Schroedl, S. (2001). *Constrained K-means clustering with background knowledge* [Conference session] (pp. 577–584). Proceedings of the Eighteenth International Conference on Machine Learning (ICML `01). San Francisco, CA, United States. http://www.litech.org/~wkiri/Papers/wagstaff-kmeans-01.pdf

Ward, J. H., & Hook, M. E. (1963). Application of an hierarchical grouping procedure to a problem of grouping profiles. *Educational and Psychological Measurement*, *23*(1), 69–81. https://doi.org/10.1177/001316446302300107

Wass, J., Mernagh, D., Pollard, B., Stewart, P., Fox, W., Parmar, N., Jones, B., Kilduff, L., & Turner, A. N. (2019). A comparison of match demands using ball-in-play vs. whole match data in elite male youth soccer players. *Science and Medicine in Football*, *4*(2), 142–147. https://doi.org/10.1080/24733938.2019.1682183

Weaving, D., Dalton, N. E., Black, C., Darrall-Jones, J., Phibbs, P. J., Gray, M., Jones, B., & Roe, G. A. B. (2018). The same story or a unique novel? Within-participant principal-component analysis of measures of training load in Professional Rugby Union skills training. *International Journal of Sports Physiology and Performance*, *13*(9), 1175–1181. https://doi.org/10.1123/ijspp.2017-0565

Webb, T., Dicks, M., Thelwell, R., & Nevill, A. (2016). The impact of referee training: Reflections on the reduction of home advantage in association football. *Soccer & Society*, *19*(7), 1024–1037. https://doi.org/10.1080/14660970.2016.1267626

Weimar, D., & Wicker, P. (2017). Moneyball revisited: Effort and team performance in professional soccer. *Journal of Sports Economics*, *18*(2), 140–161. https://doi.org/10.1177/1527002514561789

Weinberg, M. K., Tronick, E. Z., Beeghly, M., Olson, K. L., Kernan, H., & Riley, J. M. (2001). Subsyndromal depressive symptoms and major depression in postpartum women. *American Journal of Orthopsychiatry*, *71*(1), 87–97. https://doi.org/10.1037/0002-9432.71.1.87

Weston, M. (2014). Match performances of soccer referees: The role of sports science. *Movement & Sport Sciences*, *87*(2015), 113–117. https://doi.org/10.1051/sm/2014011

Weston, M. (2015). Match performances of soccer referees: the role of sport science. *Movement and Sport Sciences*, *1*(87), 113–117. https://doi.org/10.1051/sm/2014011

Weston, M., Castagna, C., Impellizzeri, F. M., Bizzini, M., Williams, A. M., & Gregson, W. (2012). Science and medicine applied to soccer refereeing: An update. *Sports Medicine*, *42*(7), 615–631. https://doi.org/10.2165/11632360-000000000-00000

Weston, M., Castagna, C., Impellizzeri, F. M., Rampinini, E., & Abt, G. (2007). Analysis of physical match performance in English Premier League soccer referees with particular reference to first half and player work rates. *Journal of Science and Medicine in Sport*, *10*(6), 390–397. https://doi.org/10.1016/j.jsams.2006.09.001

Weston, M., Castagna, C., Impellizzeri, F. M., Rampinini, E., & Breivik, S. (2010). Ageing and physical match performance in English Premier League soccer referees. *Journal of Science and Medicine in Sport*, *13*(1), 96–100. https://doi.org/10.1016/j.jsams.2008.07.009

Weston, M., Drust, B., Atkinson, G., & Gregson, W. (2011). Variability of soccer referees' match performances. *International Journal of Sports Medicine*, *32*(3), 190–194. https://doi.org/10.1055/s-0030-1269843

Weston, M., Drust, B., & Gregson, W. (2011). Intensities of exercise during match-play in FA Premier League referees and players. *Journal of Sports Sciences*, *29*(5), 527–532. https://doi.org/10.1080/02640414.2010.543914

Weston, M., Gregson, W., Castagna, C., Breivik, S., Impellizzeri, F. M., & Lovell, R. J. (2011). Changes in a top-level soccer referee's training, match activities, and physiology over an 8-year period: A case study. *International Journal of Sports Physiology and Performance*, *6*(2), 281–286. https://doi.org/10.1123/ijspp.6.2.281

Weston, M., Helsen, W., MacMahon, C., & Kirkendall, D. (2004). The impact of specific high-intensity training sessions on football referees' fitness levels. *The American Journal of Sports Medicine*, *32*(1 Suppl), 54–61. https://doi.org/10.1177/0363546503261421

Wiig, H., Andersen, T. E., Luteberget, L. S., & Spencer, M. (2020). Individual response to external training load in elite football players. *International Journal of Sports Physiology and Performance*, *15*(5), 696–704. https://doi.org/10.1123/ijspp.2019-0453

Winget, C. M., Deroshia, C. W., & Holley, D. C. (1985). Circadian rhythms and athletic performance. *Medicine & Science in Sports & Exercise*, *17*(5), 498–516. https://doi.org/10.1249/00005768-198510000-00002

Winter, W. C., Hammond, W. R., Green, N. H., Zhang, Z., & Bliwise, D. L. (2009). Measuring circadian advantage in Major League Baseball: A 10-year retrospective study. *International Journal of Sports Physiology and Performance*, *4*(3), 394–401. https://doi.org/10.1123/ijspp.4.3.394

Woods, C. T., Veale, J., Fransen, J., Robertson, S., & Collier, N. F. (2018). Classification of playing position in elite junior Australian football using technical skill indicators. *Journal of Sports Sciences*, *36*(1), 97–103. https://doi.org/10.1080/02640414.2017.1282621

Wragg, C. B., Maxwell, N. S., & Doust, J. H. (2000). Evaluation of the reliability and validity of a soccer-specific field test of repeated sprint ability. *European Journal of Applied Physiology*, *83*(1), 77–83. https://doi.org/10.1007/s004210000246

Wundersitz, D. W. T., Josman, C., Gupta, R., Netto, K. J., Gastin, P. B., & Robertson, S. (2015). Classification of team sport activities using a single wearable tracking device. *Journal of Biomechanics*, *48*(15), 3975–3981. https://doi.org/10.1016/j.jbiomech.2015.09.015

Yavuz, M., İnan, U. H., & Fığlalı, A. (2008). Fair referee assignments for professional football leagues. *Computers & Operations Research*, *35*(9), 2937–2951. https://doi.org/10.1016/j.cor.2007.01.004

Zhou, C., Calvo, A. L., Robertson, S., & Gómez, M.-Á. (2020). Long-term influence of technical, physical performance indicators and situational variables on match outcome in male professional Chinese soccer. *Journal of Sports Sciences*, *39*(6), 1–11. https://doi.org/10.1080/02640414.2020.1836793