FACULTY OF MEDICINE
AND HEALTH SCIENCES

Bridging the gap between soccer and science: using technology and data to optimize training and performance

Youri Geurkink

Supervisor(s): Prof. Dr. Jan Bourgois, Prof. Dr. Jan Boone

Ghent University
Faculty of Medicine and Health Sciences

Thesis submitted to fulfill the requirements for the degree of
Doctor in Health Sciences

2021

GHENT
UNIVERSITY

Supervisor

Prof. dr. Jan G. Bourgois (Ghent University)

Co-supervisor

Prof. dr. Jan Boone (Ghent University)

Supervisory Board

dr. Stijn Matthys (KAA Gent)

Examination Board

Prof. dr. Patrick Calders (Ghent University) as chairman

Prof. dr. Frederik Deconinck (Ghent University) as secretary

Prof. dr. Lieven Danneels (Ghent University)

Prof. dr. Steven Verstockt (Ghent University)

dr. Kristof de Mey (Ghent University)

Prof. dr. Johan Pion (University of Applied Sciences Arnhem & Nijmegen)

dr. Arne Jaspers (KU Leuven)

Acknowledgements

The research reported in this thesis was supported by KAA Gent, within the project KAA Gent-UGent Performance Center (supervised by prof. dr. Jan Bourgois).
Graphic design of the front and back cover made by Steven Oonk.

Department of Movement and Sports Sciences, Watersportlaan 2, 9000 Ghent, Belgium.

"Toeval is logisch."

Johan Cruijff

Table of contents

II Original research

III General discussion

IV Addenda

Dankwoord

Het dankwoord is niet alleen een mooie gelegenheid om iedereen te bedanken die in enige mate heeft bijgedragen aan dit doctoraat, maar tevens een goede gelegenheid om terug te kijken naar de afgelopen jaren. Het voltooien van mijn doctoraat, betekent ook het einde van mijn 'Belgisch avontuur', dat begon in 2012, toen ik vanuit Winterswijk in de trein stapte voor de eerste urenlange reis met Gent als bestemming. Na 9 jaar laat ik België en de universiteit achter, maar de ervaringen, kennis en herinneringen neem ik mee.

Een universitaire studie volgen in buitenland, op afstand van familie en vrienden, was niet altijd makkelijk. De opleiding Lichamelijke Opvoeding & Bewegingswetenschappen was echter dermate interessant dat ik deze keuze zo nogmaals zou maken. Vakken zoals inspanningsfysiologie en sportfysiologie, beiden gegevens door **Jan Bourgois**, waren voor mij hoogtepunten binnen de opleiding. Toen ik na het afronden van de opleiding de kans kreeg om onder Jan te doctoreren, heb ik dan ook geen seconde getwijfeld. De ervaringen die ik de afgelopen jaren heb mogen opdoen, in samenwerking met Jan binnen het KAA Gent - UGent Performance Center zijn van goudwaarde. Samenwerken met Jan is niet altijd makkelijk, iedereen die hem een beetje kent weet dat hij enorm veeleisend is. Geen tijd bestaat niet, want een dag heeft 24 uur. Tegelijkertijd heb ik de afgelopen jaren ook altijd blind kunnen varen op de steun van Jan en heb ik enorm veel van hem mogen leren. Mijn dankbaarheid en respect zijn groot en ik hoop dat we, ondanks een grotere afstand in de toekomst, goed contact kunnen blijven onderhouden.

Ook **Jan Boone** ben ik dankbaar voor de hulp van de afgelopen jaren. Eerst en vooral voor de begeleiding van mijn minor onderzoek, die een opmaat heeft gevormd voor dit doctoraat, en vervolgens voor de kans om aan het HILO verbonden te blijven in de periode tussen het afronden van mijn studie en start van het doctoraat. Na het starten van het doctoraat het ik de binnenkant van het HILO nog maar beperkt gezien, maar desondanks kon ik toch altijd terecht bij mijn collega's van het HILO. Kortom, ik wil ook graag **alle collega's op het HILO** bedanken voor de mooie jaren!

Binnen het KAA Gent - UGent Performance Project heb ik op dagelijkse basis mogen werken binnen de eerste ploeg van KAA Gent, en de hoogte- en dieptepunten van de afgelopen jaren van dichtbij meegemaakt. Een professionele voetbalomgeving is een wereld van komen en gaan van spelers en stafleden. Dit maakte de werkomgeving erg dynamisch en leerzaam, omdat je zoveel verschillende personen en karakters ontmoet. In het bijzonder wil ik **Stijn Matthys** bedanken, met wie ik vanaf dag 1 nauw heb mogen samenwerken. Uiteraard ben ik ook alle andere stafleden en medewerkers erkentelijk voor de prettige samenwerking. Tenslotte wil ik KAA Gent bedanken voor de fantastische kans en zal ik de club een warm hart blijven toedragen.

Als sportwetenschapper is het tegenwoordig belangrijk om je weg te kennen achter een computer. Binnen de thesis heb ik leren werken met programmeertaal, een *skill* die mij de afgelopen jaren erg heeft geholpen.

Gilles Vandewiele heeft daarin een belangrijke rol gespeeld. Hij heeft mij kennis laten maken met *Python* en machine learning models, en met vragen kon ik altijd bij hem terecht, en daarvoor wil ik hem bedanken. Ook het data-science team van professor Steven Verstockt wil ik bedanken voor hun ondersteuning.

Ik wil ook alle juryleden hartelijk bedanken om tijd vrij te maken in jullie ongetwijfeld drukke agenda's, om mij van jullie feedback en inzichten te voorzien. Jullie terugkoppeling heeft een meerwaarde hebben gevormd voor deze thesis, waardoor dank!

Familie is erg belangrijk voor mij, en het was daardoor ook niet altijd makkelijk om op afstand te zitten. Met de volgende kans in mijn professionele carrière, als fysiek en wetenschappelijk coördinator binnen de FC Twente/Heracles academie, wordt deze afstand gelukkig wat kleiner. Een wedstrijdje kijken met **pa**, even naar het dorp met **ma**, dollen met **Tycho** en avondjes met de **Hyena's**, ik ben blij dat ik dit in de nabije toekomst weer wat spontaner en frequenter kunnen gaan doen. Even spontaan op bezoek bij **Lynn** en de jongens, die sinds kort in Zweden wonen, zal er niet in zitten, daarvoor zullen we nog gebruik moeten maken van het videobellen. Maar waar ik vooral erg blij om ben, is dat de band met familie en vrienden nog even goed is als 9 jaar terug. Via deze weg ook even een 'shout-out' naar **Steven** voor het ontwerpen van de cover!

De afgelopen jaren heb ik verschillende 'titels' mogen verwerven. Zo heb ik de opleiding Master in de Lichamelijke Opvoeding & Bewegingswetenschappen succesvol afgerond en is dit proefschrift ingediend om '*PhD*' achter mijn naam te kunnen zetten. Echter zijn de allermooiste titels die ik in de afgelopen jaren heb mogen verwerven de titels van *echtgenoot* en *vader* geweest. De lange dagen, onregelmatige planningen en last-minute wijzigingen voor mijn werk binnen KAA Gent, in combinatie met het doctoraat, hebben het niet makkelijk gemaakt om altijd voldoende tijd, aandacht en energie vrij te maken voor het gezin. De afgelopen 2 jaar heb ik **Ferre** mogen zien opgroeien tot een ondeugende peuter, en ik ben elke dag trots dat ik zijn vader ben. Even trots ben ik op jou **Charmaine**. Mijn dankbaarheid voor je begrip en voor alles dat je de afgelopen jaren voor mij hebt gedaan en betekent is groot. Ik kijk uit naar onze toekomst en die van ons gezin!

Summary

Stakeholders within a soccer club, i.e. the players, the coaching staff, people in the management, and every other individual contributing to the club, are part of a never-ending process to help the club be successful. The pursuit to gain a competitive advantage over others competitors incites clubs to invest in players, staff and infrastructure. Advancements in technology allowed its widespread use in soccer, resulting in the availability of a plethora of data, allowing new possibilities in the daily decision-making regarding domains such as performance and training.

In soccer, aiming to improve future performance, past behaviours of players and teams are analysed to shape future decisions. Traditionally, qualitative analysis has been used to analyse historical performances, usually using video footage. Nowadays, especially at the highest levels of professional soccer, an abundance of game-related data is available, which can be used for quantitative game analysis. The insights resulting from both qualitative and quantitative game analysis can subsequently be used to aid the optimization of different processes within the team, of which training is an important aspect. Training is applied to improve all central aspects related to soccer, i.e. technical, tactical, physical and mental abilities of players individually and the team as a whole. Training is monitored using tracking technology, Heart Rate (HR)-monitoring and/or questionnaires, to get insights into the training stimuli being applied and the players' responses to the stimuli. Today, practitioners and academics are exploring how to apply technology in a feasible manner and how data can be used in such way that it contributes to the clubs' goals.

This thesis was conducted within the KAA Gent - UGent Performance Center project, a collaboration between soccer club KAA Gent and Ghent University. Within this project, three studies were conducted, related to performance, training and monitoring of training.

Study I aimed to identify the strongest predictive variables of winning and losing in Belgian professional soccer, using a broad range of performance indicators and contextual variables. The total number of shots on target from the attacking penalty box was identified as the best predictor. It was however shown that not only shot-related variables, but a broad range of variables are amongst the strongest predictors of winning and losing. It seems particularly interesting to look at physical parameters of the second half, given that they are amongst the best predictors of game outcome in soccer. The application of variables such as ELO-ratings, transfer values, match location and Playing Styles can be useful additions to current approaches used to evaluate game performances. Conclusively, the results of this study provided a framework in which practitioners can use quantitative game data to gain more objective insights into performance.

Study II was focused on the structure of training during the preseason, how training intensity is distributed and whether there are differences between the various ways in which the Training Intensity Distribution (TID) can be expressed. During the preseason in a professional soccer team, 57 training sessions were conducted over a period of 36 days, of which 50 sessions were performed outdoor. Players covered, on average, more than 250 kilometres during a total duration of more than 50 hours of outdoor training. Although all TIDs showed that most of the training volume is spent at low intensity, differences were observed for the different methods of quantification. When using distance as a marker of volume, lower proportions of volume were spent at low intensity and higher proportions at moderate and high intensities, as compared to time. Time as a marker of volume, on the other hand, may overestimate the proportion of volume spent at low intensities. Speed as a marker of intensity may better reflect the stochastic nature of soccer in comparison to HR. Speed and distance may be the most relevant and practically feasible combination of respectively intensity and volume markers to describe TID in the context of soccer.

Study III aimed to identify the strongest predictive variables of the session Rate of Perceived Exertion (sRPE) in soccer. A predictive model was constructed, using a total of 70 external load indicators, internal load indicators, individual characteristics, and supplementary variables. The model showed a mean absolute error of 0.67 ($\pm$0.09) Arbitrary Units (AU) and a root-mean-square error of 0.93 ($\pm$0.16) AU, indicating that the sRPE can be predicted quite accurately using only a relatively small number of training observations. External load indicators were identified as the strongest predictors of the sRPE, with total distance as the strongest individual predictor. Including a broad range of variables, other than external load indicators, showed to be useful, as the accumulated importance of these variables accounts for a reasonable component of the total normalized importance. Applications resulting from this study may help a coaching staff to plan, monitor, and evaluate training sessions.

The results of the conducted studies can be situated in the larger context of the continuous endeavour to obtain favourable game outcomes. Since each stakeholder and each process within the club is ideally focused on somehow obtaining a competitive advantage, the domains in this thesis should not be viewed separately and not disregard other aspects relating to game outcome. Conclusively, it is important that each process contributing to the goals of the team is continuously evaluated, in which the correct application of technology and data can provide a proper objective complement for the subjective understanding of this beautiful game.

Samenvatting

Belanghebbenden binnen een club, gaande van de spelers en staf, maar ook alle andere personen die een bijdrage leveren voor de club, maken deel uit van een eindeloos proces om de club te helpen succesvol te zijn. Het streven naar een competitief voordeel ten opzichte van tegenstanders, noopt clubs te investeren in spelers, staf en infrastructuur. Voortschrijdende technische mogelijkheden, ook resulterend in een verhoogde beschikbaarheid van data, biedt nieuwe kansen in de besluitvormingsprocessen met betrekking tot prestatie en training.

Om toekomstige prestaties te verbeteren worden de voorbije verrichtingen van spelers en teams geanalyseerd. Traditioneel wordt kwalitatieve prestatieanalyse toegepast om voorbije prestaties te analyseren, doorgaans door het gebruik van videobeelden. Tegenwoordig, met name op de hoogste niveaus van het professioneel voetbal, is er een overvloed van wedstrijdinformatie beschikbaar dat gebruikt kan worden voor kwantitatieve analyses. De inzichten die voortkomen uit zowel kwalitatieve en kwantitatieve analyses kunnen vervolgens gebruikt worden om verschillende processen binnen het team te optimaliseren, waarvan training een belangrijk aspect is. Training wordt toegepast om alle centrale aspecten, zijnde de technische, tactische, fysieke en mentale vaardigheden van spelers individueel en als team als geheel te gaan verbeteren. Trainingen worden gemonitord met behulp van onder andere *tracking* technologie, hartslagmonitoring en/of vragenlijsten, om inzicht te krijgen in de trainingsprikkels en hoe spelers hierop reageren. Tegenwoordig zoeken zowel mensen die werkzaam zijn binnen het voetbal als academici hoe technologie op een praktisch haalbare manier kan worden gebruikt en hoe data kan worden gebruikt op een manier dat het bijdraagt aan het behalen van de doelen van de club.

Dit proefschrift werd uitgevoerd binnen het KAA Gent – UGent Performance Center project, een samenwerking tussen voetbalclub KAA Gent en de Universiteit Gent. Binnen dit project werden drie studies uitgevoerd, gerelateerd aan prestatie, training en de monitoring van training.

Studie I had het identificeren van de sterkst predictieve variabelen van winst en verlies binnen het Belgisch professioneel voetbal als doel, gebruikmakend van een brede range van prestatie-indicatoren en contextuele variabelen. Het totaal aantal schoten op doel binnen het strafschopgebied werd geïdentificeerd als de belangrijkste voorspeller. Het waren echter niet enkel variabelen gerelateerd aan doelpogingen, maar een ruim scala van variabelen die werden geïdentificeerd als de belangrijkste voorspellers van winst en verlies. Voor de fysieke prestatie-indicatoren lijkt het met name interessant om naar de fysieke parameters van de tweede helft te kijken, omdat deze tot de sterkste voorspellers behoren. Het gebruik van contextuele variabelen, zoals ELO-scores, transferwaarden, wedstrijdlocatie en 'Playing Styles'-variabelen zijn tevens nuttige aanvullingen op de huidige benaderingen om wedstrijdprestaties te evalueren. De resultaten van deze studie bieden een raamwerk waarbinnen kwantitatieve wedstrijddata gebruikt kan worden om meer

objectieve inzichten te krijgen in prestatie.

Studie II focuste op de structuur van de voorbereidingsperiode, de intensiteit van training tijdens deze voorbereidingsperiode en het identificeren van verschillen in de manieren waarop de training intensiteit distributie (TID) kan worden uitgedrukt. Gedurende de voorbereidingsperiode werden 57 activiteiten afgewerkt binnen een periode van 36 dagen, waarvan 50 in de buitenlucht. Gemiddeld legden spelers meer dan 250 kilometer af gedurende meer dan 50 uur aan activiteiten. Volume en intensiteit kunnen respectievelijk in tijd en afstand, en hartslag en snelheid worden uitgedrukt. Voor alle mogelijke combinaties van volume- en intensiteitsparameters werd het grootste deel van het volume aan lage intensiteit afgewerkt. Wanneer afstand als volumeparameter werd gebruikt, dan werden grotere proporties aan matige en hoge intensiteit afgewerkt in vergelijking tot tijd als parameter van volume. Het gebruik van tijd heeft ook als nadeel dat het de proportie van volume aan lage intensiteit mogelijk overschat. Snelheid als intensiteitsparameter geeft een betere expressie van het intermittente karakter van voetbal ten opzichte van hartslag. De combinatie van afstand en tijd om de TID uit te drukken lijkt in de context van voetbal de meest relevante en praktisch uitvoerbare methode.

Studie III had als doel de best voorspellende variabelen van de session Rate of Perceived Exertion (sRPE) te identificeren. Een predictief model werd gevormd, gebruikmakend van totaal 70 factoren, die onderverdeeld konden worden in externe belastingindicatoren, interne belastingindicatoren, individuele karakteristieken, en supplementaire variabelen. Het model had een gemiddelde absolute afwijking van 0.67 ($\pm$0.09) arbitraire eenheden (AU) en een gemiddelde *root-mean-square error* van 0.93 ($\pm$0.16) AU. Dit toont aan dat de session Rate of Perceived Exertion (sRPE) behoorlijk accuraat kon worden voorspeld, gebruikmakend van slechts een relatief klein aantal trainingsobservaties. In het algemeen waren externe belastingindicatoren de belangrijkste voorspellers, met totale afstand als belangrijkste predictor. Het includeren van een wijde range van variabelen naast externe belastingindicatoren is echter nuttig, omdat het geaccumuleerde belang van deze variabelen behoorlijk is. Het predictief model toont aan dat de sRPE behoorlijk accuraat kan worden voorspeld, gebruikmakend van slechts een relatief klein aantal trainingsobservaties. Toepassingen die voortvloeien uit de bevindingen van de studie kunnen trainers helpen bij het plannen, monitoren en evalueren van trainingssessies.

De resultaten van de uitgevoerde studies kunnen binnen het breder perspectief van het voortdurende streven naar goede resultaten worden geplaatst. Aangezien elke belanghebbende en elk proces binnen de club idealiter is gefocust op het creëren van een competitief voordeel, zouden de behandelde onderwerpen binnen dit proefschrift niet afzonderlijk van elkaar moeten worden beschouwd. Ieder proces maakt deel uit van het groter geheel. Het is daarom belangrijk dat de processen die trachten bij te dragen aan het succes van de club steeds weer worden geëvalueerd. Het correct toepassen van technologie en data kan hierbij helpen, door een objectieve aanvulling te bieden voor de subjectieve interpretatie van deze prachtige sport.

Abbrevations

ACQ	Acquisition period
ACWR	Acute/Chronic Workload Ratio
ANN	Artificial Neural Networks
ATP	Adenosine Triphosphate
AU	Arbitrary Units
CMJ	Counter Movement Jump
ELIs	External Load Indicators
FA	Football Association
GNSS	Global Navigation Satellite System
GPS	Global Positioning Systems
HR	Heart Rate
ICs	Individual Characteristics
ILIs	Internal Load Indicators
IMUs	Inertial Measurement Units
KNN	K-Nearest Neighbours
LAS	LASSO Regression
LIR	Linear Regression
LOR	Logistic Regression
LPS	Local Positioning Systems
MD	Match Day
MD+1	First day after the last match
MD-1	Last day prior to the next match
MD+2	Second day after the last match
MD-2	Second last day prior to the next match
NA	Not Available
NAB	Naive Bayes
NBA	National Basketball Association
NLP	Natural Language Processing
NI	Normalized Importance
RFO	Random Forests
RPE	Rate of Perceived Exertion
sRPE	session Rate of Perceived Exertion
SVM	Support Vector Machines
TID	Training Intensity Distribution
TRIMP	Training Impulse
VAR	Video Assistent Referee
VO2max	Maximum Oxygen Uptake
XGB	Extreme Gradient Boosting

Part I

Introduction

Background

Since its official foundation, back in 1863, soccer has spread around the planet, reaching almost every country in the world [240]. The first 14 rules of soccer were described in 1863 by the Football Association (FA) [90], but many rules were added and existing rules have changed over the years. In 1863, the playing duration or even the number of players were not defined but today, a typical soccer game is played by 11 players of each team, subdivided in two halves of 45 minutes, with a 15-minute break in between [242]. The basic rule: *"A goal shall be won when the ball passes over the space between the goal posts"*, was one of the rules defined in 1863 and is a principle in soccer that still stands. Today, regardless of their level, players still adhere to this basic objective during gameplay: to score a goal while preventing the opponent to do so [155]. As the two teams on the pitch pursue the objectives of the game simultaneously, numerous unique interactions arise between players on pitch [155], in which the attacking team attempts to progress the ball towards the opponents' defensive zone to get into a scoring position, while the team in defence tries to deny entry of the opponent with the ball into the defensive zone [151]. The rules and objectives of soccer apparently result in an activity that is enjoyable, as soccer is nowadays considered as the world's most popular sport [75, 145, 190], with hundreds of millions of people competing across the planet [141] and billions of people watching [89]. Soccer is a game that connects people, illustrated by events such as the *Christmas Truce* of 1914, when British and German soldiers ceased fire to play football between the trenches [4]. It can also divide, exampled by the 100-hour war between El Salvador and Honduras after their World Cup game in 1970 [51]. It is a sport over which grown men cry and there is no cultural practice more global than soccer [190]. With its increase in popularity over the years, the interests of stakeholders exponentially increased, and contemporary soccer is a multi-billion euro industry [136]. Gaining a competitive advantage over rival teams can therefore be significantly financially beneficial [195]. Many resources are put into players, staff and infrastructure, to provide the best conditions to be successful. The evolution of technology allowed its widespread use [239], and nowadays, at the highest levels of contemporary soccer, technology is used to collect data from games and training sessions [5], in order to gain insights into domains such as performance, training and monitoring. The domains performance, training and monitoring are all quite broad, but as can be seen in Figure 1, these topics are all interrelated and should therefore be considered accordingly. This thesis is structured in a similar manner, by 1) defining performance and identifying its determinants, 2) describing the goals and structure of training within professional soccer and 3) how training can be monitored.

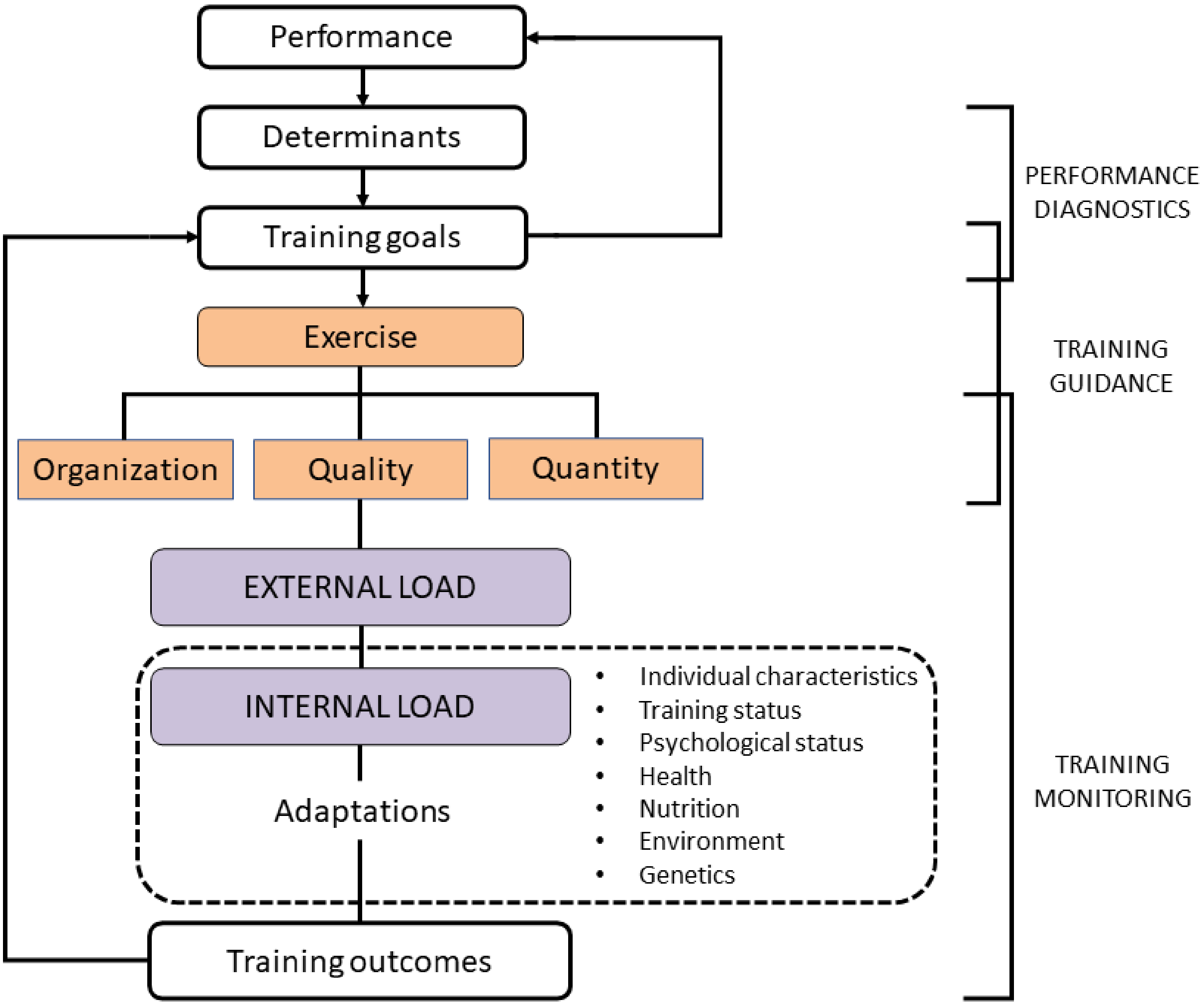


Figure 1: Theoretical framework of the training process by Impellizzeri et al. [130], illustrating the interrelations between performance, training and monitoring in soccer.

1 Performance

Soccer teams play games, of which the result or *game outcome* is determined by the number of goals scored by each of the teams. It is the game outcome that determines the number of points a team earns, or whether the team qualifies for the next stage of a tournament. Given the interests of stakeholders within soccer, *successful* game outcome is important. It can be argued that success in soccer is defined by winning, however, it should be noted that the term *successful* is relative [124]. At the level of a single game, a draw or even a small defeat against a strong opponent can be considered as a success, as not all teams will compete to become champion [107]. This illustrates a dissociation between successful game outcome and successful performance. Therefore, from this point, game outcome is regarded as the *product* of the game, being the ultimate scoreline of a game, while performance is regarded as the *process* of the game [257].

1.1 Performance analysis in soccer

Aiming to improve future performance, the coaching staff tries to shape future decisions and actions of players by providing them with information gathered based on past behaviours [177]. The information on previous performances can be derived from very different sources, such as video footage, tracking data, event data and aggregated statistics [218]. A subdivision can be made between qualitative performance analysis, which is based on subjective, anecdotal information, and quantitative performance analysis, which concerns objective, non-anecdotal data [218]. This chapter will focus on the use of quantitative performance analysis.

1.1.1 Notational data

The foundation of quantitative performance analysis in soccer can be traced back to the early 1950's, in the form of notational analysis [197]. Notational analysis is a technique used to analyse different aspects of performance through a process which involves producing permanent records of game events [134]. Charles Reep was in the 1950's the first to create a comprehensive notational analysis system for football, with the primary focus on the development of an effective playing style [197]. His pioneering work advocated the use of the long-ball strategy [134, 156], as he found that most shots are preceded by a small number of ball possessions, and that regaining ball possession close to the opponent' goal leads to more shots [204]. The logic behind this style of play is that the more times the ball enters the goal-scoring areas of the pitch, the greater the probability of scoring [134]. Although this philosophy has had a big influence on the development of the style of play in England, it is clear that this conclusion in contemporary soccer would be short-sighted, as the playing style of each team should be adapted to skill of the team, the opponent and other situational variables [117]. The use of other styles of play, such as possession play [139], have proven to lead to success as well.

The basic objective of game play has not changed since the time of Charles Reep. Consequently, during game play, the team in possession still attempts to progress the ball towards the opponents' defensive zone to get into a scoring position [151]. Because of the relative scarcity of goals, notational metrics relating to attacking performance often focus on metrics such as shots, crosses or the number of passes into the box [226]. When a team is not in possession, the team will try to deny entry of the opponent with the ball into the defensive zone [151]. The number of metrics describing defensive performances is however limited in comparison to attacking performance. This may be related to the fact that notational statistics are usually frequencies of 'on-the-ball'-events [31]. Metrics such as tackles and interceptions are described in relation to defensive performance but may not accurately describe it, as defending is often more about clever positioning and anticipation [64]. Nowadays, event data is coded by professional sports data companies using specialized computer software based on the videos of the game [31, 192]. This procedure of data collection for event data by professional sports data company *WyScout* (Chiavari, Italy) was recently described by Pappalardo et al. [192]. Each ball touch in a match is added to the software as a new event on the timeline, and supplemented by additional information, such as types and subtypes of actions (e.g. type: shot; subtype: on target). A custom keyboard is used, to ensure that events and data are processed in a streamlined way. The operator subsequently adds the coordinates on the pitch and all the additional attributes for the event. Actions are usually tagged by one operator per team. A third operator acts as the responsible supervisor of the output of the whole match. A fourth operator is optionally used for near-live data delivery, to speed up the collection of complex events which need additional and specific attributes, or a quick review. The tagging procedure is followed by quality control. The first step of this quality control is automatic, in which an algorithm is used to detect the majority of errors made by operators, for example by comparing information on events that are tagged by both operators (e.g. duels), or by searching for impossible combinations of event sequences. The second step of quality control is a manual procedure, through a check of the game events once the game is completed. Over more than 20 years, many attempts and proposals on automated event detection have been put forward by studies [62, 78, 140, 235], illustrating the interest to automate the procedure of event tagging. The reliability of the data delivered by professional sports data companies, is shown to be high [31, 32], which sets high standards for a fully automated system before it can be applied without manual procedures.

1.1.2 Tracking data

During game play, not only the actions on or around the ball are interesting, but the movements and interactions of players on the pitch as well. The first cited studies on motion tracking of soccer players originated from the 1970's [205, 253] and early 1980's [256]. These studies also applied a notational approach [125]. Movement were hand-coded, which made the process of estimating distances very labourintensive, next to methodological problems, as distances were estimated by observing modes of players' movements at pre-determined levels of speed [189]. The first cited study using video-tracking originated from the end of the 1980's [243], but still required a large amount of workload [44]. Today, games in professional soccer are still being tracked using video-tracking systems, however, these systems are currently largely computer automated [44], alleviating the associated workload. Since 2015, it is also allowed to wear Electronic Performance and Tracking Systems (EPTS), approved by the FIFA, during competitive games [88], which allows teams to utilize their own tracking systems. In contemporary football, most games played at the highest professional levels are being tracked using video-tracking systems, consequently, billions of data points are available per season [19].

The evolution of tracking systems provides an alternative approach to notational analytics. The latter applies a reductionist approach to summarize the complex game of soccer into isolated performance indicators [177]. The rationale behind reductionism is to understand the functioning of the whole through an analysis of its individual parts [169]. This approach is however criticized, as it has been stated that games should be analysed as the collective of the interacting parts that comprise the system [177]. In the 1950's, when data had to be hand coded, the notational approach was the only feasible method to obtain a better quantitative understanding of the game, however, as football has evolved, so did technology [215]. The current possibilities that emerge from the use of tracking technologies provide better opportunities to quantify soccer as a complex system, in which player behaviour can be described as an emerging consequence of the interactions between players.

From a tactical perspective, insights into individual and collective actions can be gained using tracking data [43]. Tactical behaviour relates to the positions taken in reaction to the opponent(s) in a game situation and the adaptations of the team to the conditions of play [111]. An increasing number of studies describe tactical behaviour in soccer, as technological advancements allows to collect tracking data during games [206]. In light of achieving the objective of the game, opportunities for scoring arise when offensive players engage in actions that destabilize the defensive response. To destabilize defensive systems, offensive teams displace their players and the ball irregularly to promote a cascade of local instabilities in their opponents' defence [249]. One of the variables that has been described to reflect these instabilities is the centroid position of a team. The team centroid is the geometrical configuration that represents the mean position of the 10 outfield players [101]. The movement of the team centroid during game play demonstrates the dynamic flow between attacking and defending, in longitudinal as well as in lateral direction [99]. It was expected that variability in the inter-team centroid position would reflect a disruption in the balance between two teams and may precede key events [180]. Although no associations were found between high variability in inter-team centroid distance and key events like goal and goal-scoring opportunities [98], the interactions between the teams' centroid do potentially provide interesting information on the collective movements of both teams on the pitch. The position-specific centroid was suggested as a more sensitive marker of instability, describing tactical behaviour more accurately [110]. Memmert et al. [180] showed that eight out of nine critical moments in one studied game showed higher variability in interline distance in the 30 seconds before the critical moment. On an even more individual level, approximate entropy (ApEn) [196], a data processing technique, has been used. ApEn quantifies the amount of regularity and unpredictability of fluctuations over time series [180], and can be used to specify the predictability of players, in relation to their position. For example, central defenders and central attackers were identified as highly predictable [180], while other attackers showed lower predictability [110, 180], which can be related to their specific role on the pitch. Team centroids, position-specific centroids and ApEn are used here as an illustration on how tracking data can be used to respectively map team, position and individual movements on the pitch. Conclusively, tracking data provides opportunities for new approaches regarding performance analysis.

Tracking data can also be used to gain insights into the activity profile of soccer. The activity profile is often expressed in total distance and distances in several speed zones, and has been extensively described in professional football over the last decades [15], including the highest domestic leagues of England [23, 32, 40, 67, 70], Spain [56, 67, 86], Germany [144], Italy [71, 247, 248] and France [66]. Comparisons between studies are difficult, because of the inconsistency in the definitions to demark different activity zones [60, 77]. Furthermore, there are also differences between systems provided by companies because of differences processing methods. For example, even if the speed zones are similar, factors such as the *minimum effort duration* can differ between systems. Company *A* may set the minimum effort duration, which is the minimum duration of an effort above a particular speed or acceleration threshold required to be recorded [172], different compared to company *B*, which will result in differences in the data. There seems however consensus that, in professional soccer, players cover a total distance in the range of 10 to 13 kilometer [15]. In the English Premier League, the total distance has remained relatively stable over the course of several season (2006-2007 until 2012-2013), while the total high intensity activities and running distance have increased, highlighting the increasing demands of top elite soccer [18, 33, 40]. A summary of the general activity profile performed during games is depicted in Figure 2.

1.1.3 Contextual data

Today, practitioners at the highest professional levels often have access to metrics and information from a multitude of sources. Besides notational and tracking metrics, contextual information is available. It has been shown that contextual factors such as match location [87, 144, 148, 150], quality of the opposition [87, 148, 150] and match status [87, 148, 150] influence the game activities of the team and players. Gathering these types of information often does not require highly technological devices or processes. Therefore,

- Between 10-13 kilometers [15]
- Between 18-26% of total distance at speeds exceeding 14.4 km/h [201]
- Between 18-31 actions at speeds exceeding 25.2 km/h [201]
- Over 1400 changes in activities [210]
- Over 600 acc- and 600 decelerations exceeding 0.5 m/s^2 [210]
- Over 6000 impacts exceeding 2 G-forces [210]

Figure 2: A general activity profile of outfield players during game play in soccer, inspired by a figure derived from Dolci et al. [72]. Naturally, there are differences between players playing on different positions and/or systems.

sources offering contextual game information are nowadays abundantly and freely available. The number of relevant sports statistic databases is expanding [122], including websites that offer data on teams and players. An often used website by practitioners, and sometimes even by academics [162], is the website *https://www.transfermarkt.co.uk*, which provides fixtures, results and standings by teams and basic information such as age, position and estimated market values for players. The website *https://www.football-data.co.uk* provides historical results and game odds by several betting companies for 11 nations. An interesting metric, the ELO-rating, is available at *http://clubelo.com*, which can be considered a measure of the team's current strength [128]. Each game, both national and international, results in an exchange in ELO-points. The exchange in ELO-points is higher for a win against a stronger team compared to a victory against an equally strong or weaker team, and vice versa for losses. An illustration on ELO-ratings is depicted in Figure 3. Conclusively, there is a plethora of (open-)sources from which practitioners and academics can gather contextual game data, which can be used to assess game performances and outcome.

1.1.4 Utilization of game data in practice

The use of data is naturally related to its availability. Today, academics tend to use tracking data, while practitioners use a combination of standings, notational metrics and/or tracking data [195], showing a discrepancy between academics and practitioners. It can be argued that the soccer environment is conservationist, or that the lack of academic background of most soccer coaches hinders the coaches' understanding of the methods and terms applied in science [21, 35]. These reasons may explain why many coaches still tend to use more traditional metrics. On the other hand, it should be acknowledged that the use of tracking data can be challenging for reasons such as storing data, accessing data, preprocessing data and handling noisy data [195], and that therefore, the availability of useful metrics derived from tracking data is still limited. Besides the best teams in the world, there are probably not many teams that have the resources to

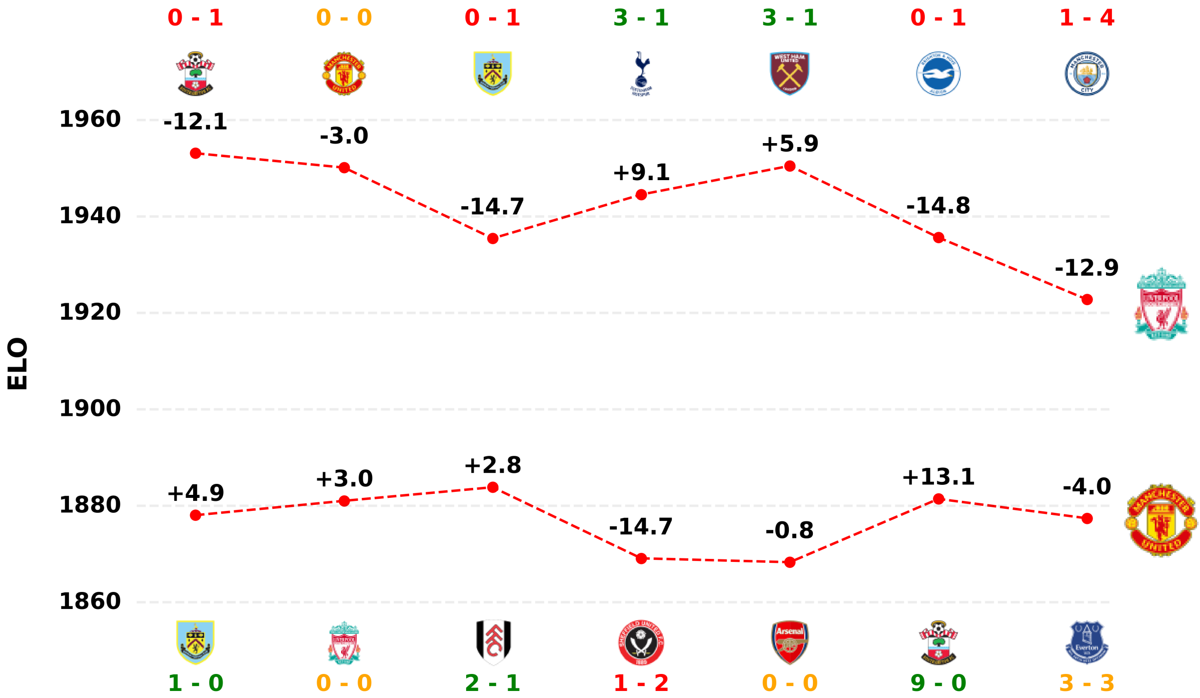


Figure 3: An illustration on the course of the ELO-ratings for two teams, Liverpool and Manchester United, during the first league games of 2021. The opponents and results of Liverpool are depicted on the upper side of the graph, for Manchester United on the lower side of the graph. Data was derived from *ClubELO*.

employ a team of data scientist to process and extract useful information from tracking data. Furthermore, given the costs of purchasing and/or the installation of systems that allow to track players, high-quality tracking data is almost only available at the highest professional levels [195]. Notational analysis systems on the other hand are relatively easy to design, cost-effective in terms of time and resources, and provide coaching staff with valuable and objective information [199]. Interestingly, although the availability of standings and notational metrics is high, academic work combining more traditional metrics and tracking data is scarce [195].

To conclude, notational game data provided the first quantitative insights into game play in soccer. Throughout the years, advancements in technology presented more possibilities with regard to tracking data. Academically, there is a strong interest in the use of tracking data, contrasting the current use of many practitioners, which is still predominantly focused on the more traditional notational game metrics. Therefore, it could be interesting to combine all different types of data into the analysis of performance indicators and game outcome.

1.2 Identifying relations between performance and game outcome

Metrics derived from notational analytics or tracking data are often referred to as performance indicators. Performance indicators are selections or combinations of action variables aiming to define some or all aspects of performance [124]. Performance indicators can be used in a comparative way, to allow a comparison with opponents, other athletes or teams, or in isolation, as a measure of the performance of the individual or team [124]. To be useful, performance indicators should relate to successful performance or outcome [124]. Factors such as ball possession [48, 57, 149, 258], number of passes [36, 117], number of shots [36, 48, 149], number of shots on target [48, 149], entries into the penalty box [209, 258] and successfulness in duels [258] have previously been related to game outcome in soccer. Also tactical indicators such as the centroid-position [180] or physical indicators such as distance in several speed zones [52] have been related to game outcome. It should however be noted that previous studies investigating performance indicators in soccer often suffered from limitations and/or methodological problems, such as small sample sizes and univariate analyses of the observed variables [48, 161]. If presented in isolation, a single set of indicators representing the performance of an individual or a team can give a distorted impression of performance by ignoring other, more or less important, variables [124], which likely explains differences between previous studies [48]. It is therefore important that variables are not considered in isolation when relating performance indicators to game performance. With the advancing number of available metrics, the number of studies to apply a multivariate approach relating performance indicators to game outcome in soccer also increased. Often, regression analyses, discriminant analyses or ANOVA's are applied as statistical methods to identify relations between performance indicators and game outcome [162]. With these kind of statistics, inferences from a sample can be drawn [41]. These types of classical statistical methods are designed for data with a few dozen input variables and small to medium sample sizes [41]. However, more complex statistical approaches are needed to explain the underlying mechanisms of performance determinants in soccer [162].

1.2.1 Machine Learning

In recent years, new ways of statistical analyses were introduced [162], more specifically in the field of machine learning. Machine learning is the practice of utilising statistical and computational methods for classification, pattern recognition and prediction [120]. Machine learning is particularly helpful when dealing with many input variables in relation to the number of samples [41]. When the number of input variables and the possible relationships between the input variables increase, the model that captures these relationships becomes more complex [41]. Consequently, statistical inferences become less precise and the boundary between statistical and machine learning approaches becomes hazier. [41]. Furthermore, machine learning can also be used when data is gathered without a carefully controlled experimental design and in the presence of complicated non-linear interactions [41]. Machine learning is typically subdivided into two areas: supervised and unsupervised learning [120]. In supervised learning, the aim is to optimise a model on a set of labelled training data to fit to a given response [120]. Finding a model with the appropriate complexity for a data set requires finding a balance between bias and variance [163]. Bias is the error introduced using a predictive model that is incapable of capturing the underlying data, and variance is the error due to sensitivity to noise in the data [163]. A model that is too simple is likely to have a high bias and low variance (underfit), while an overly complex model typically has low bias and high variance (overfitting) [163], illustrated in Figure 4. Overfitting and underfitting can both be problematic when applying the model to unseen data. To assess the performance of a machine learning model on unseen data, especially for relatively small datasets, the use of K-fold cross-validation is recommended [163]. When using K-fold cross-validation, the dataset is split in K-folds. For example, when K is equal to 5, the dataset is divided into 5 parts. From those 5 parts, 4 parts are used to fit the model and 1 part to test the model. This process is conducted 5 times, so every part of the dataset is used to train the model 4 times and once to test it. This will result in 5 difference predictions, which can be used for the evaluation of the model [163].

Several machine learning algorithms have been used to predict game outcome [122]. Initially, the most often used algorithm for sports prediction were neural networks [122], but the downside of this algorithm is that it is a 'Black Box', indicating that it is not possible to determine the most important factors of the model [120]. 'White Box'-machine learning algorithms, on the other hand, allow understanding on how the input features are used to make predictions [167]. Nowadays, tree-based machine learning algorithms, which can be considered as 'White Box'-algorithms, are the most popular models in use [168]. Examples of tree-based machine learning algorithms are Random Forests [34] and Extreme Gradient Boosting [50]. Both algorithms can be defined as ensemble methods, which rely on a large number of relatively weak and simple models, in this case decision-trees, to obtain a stronger ensemble prediction [186]. This technique can be somewhat compared to the 'wisdom-of-the-crowds'-phenomenon, in which the aggregation of different sources results in decisions that are superior than those that are made by single individuals [208]. In

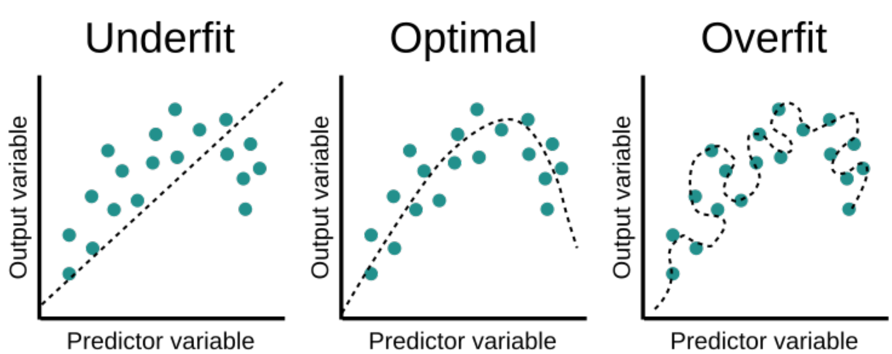


Figure 4: An illustration on under- and overfitting (source: *Edspresso; https://www.educative.io/edpresso/overfitting-and-underfitting.*)

Figure 5, an illustration is provided on Tree Emsemble Models.

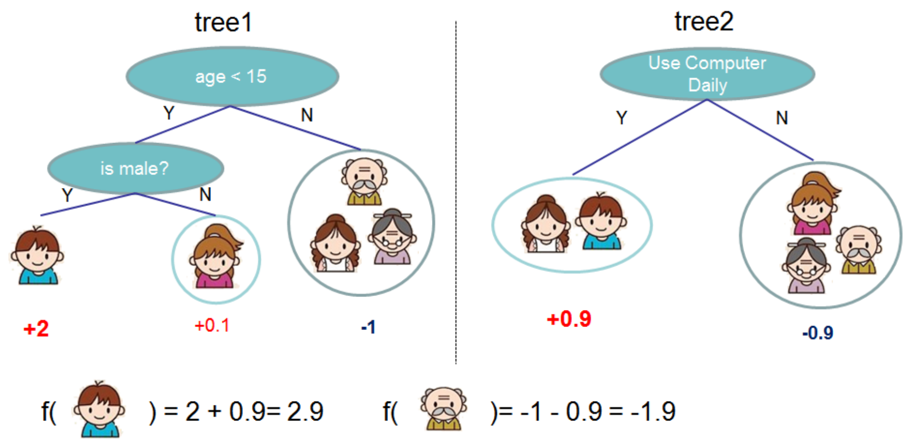


Figure 5: An illustration on a Tree Ensemble Model, which shows that the final prediction for a given sample is the sum of predictions from each tree, as depicted by Chen and Guestrin [50].

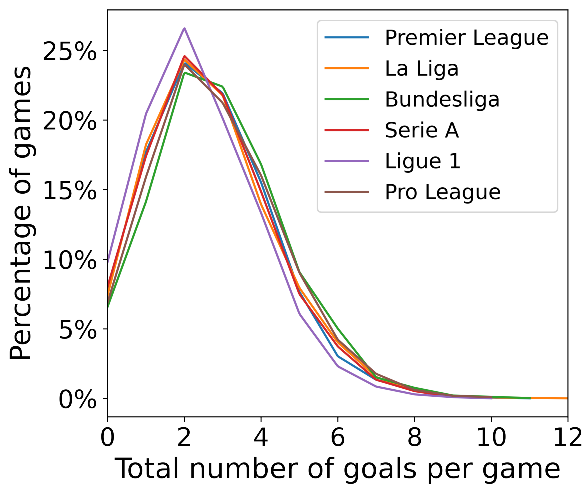
Machine learning algorithms are used in very different branches [168], with very divergent applications using the same algorithms. When applying a machine learning algorithm, there are a number of default settings which describe how the model is constructed. These settings are called hyperparameters. For tree-based algorithms, examples of hyperparameters are the depth of the tree or the maximum number of leaves. It is however possible, that the default hyperparameter settings make the model too simple or complex for the data, which respectively results in under- or overfitting. To improve performance of the model, the hyperparameters can be optimized. Model performance can also be improved by removing redundant features [122]. The larger the number of variables, the greater the number of 'dimensions' [147], which increases the computational cost of the algorithm exponentially. Therefore, feature selection is an important step in machine learning and its main purpose is to achieve better performance by reducing the redundant features [122].

1.2.2 Predicting game outcome in soccer

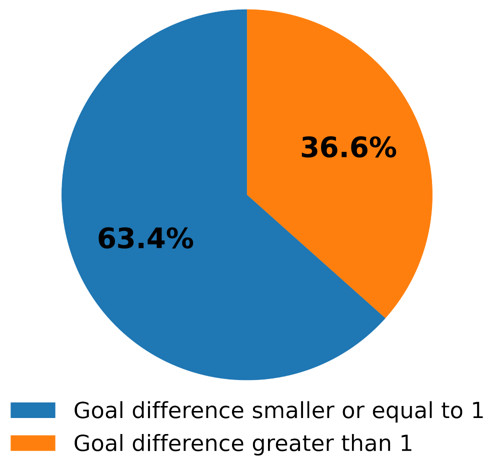
The predictive accuracy of soccer games is, as demonstrated by the accuracies in Table 1, usually relatively low. This can partly be explained by the low-scoring nature of soccer (Figure 6), resulting in a lowered predictability of games [154], and thus lower predictive accuracy. Furthermore, in contrast to many other sports, soccer games can result in three (win-draw-lose) instead of two separate outcomes (win-lose). It should also be noted that most previous studies relating to the prediction of game outcome in soccer focused on betting [75] and these studies often used open-source data, instead of detailed game statistics based on notational or tracking data. It may be that 'better' input features will improve model performance. It could be very interesting to construct predictive machine learning models using a combination of detailed notational and/or tracking data combined with contextual game data, as there are techniques, such as TreeExplainer [168], that can be used to identify the most important predictors of a predictive model.

Table 1: An overview on previous studies predicting game outcome in soccer. Abbreviations: Artificial Neural Networks (ANN), Random Forests (RFO), Logit Boost (LGB), K-Nearest Neighbours (KNN), Naive Bayes (NAB), Bayesian Networks (BYN), Logistic Regression (LOR), LASSO Regression (LAS), Extreme Gradient Boosting (XGB), Support Vector Machines (SVM), Linear Regression (LIR)

Study	**Description**	**Accuracy**
Hucaljuk and Rakipović [123]	Prediction of win/draw/lose based on contextual variables, such as form, head-to-head results and current ranking from a single Champions League group stage.	ANN: 68.8% RFO: 65.6% LGB: 62.5% KNN: 62.5% NAB: 56.3% BYN: 56.3%
Tax and Joustra [234]	Prediction of win/draw/lose based on 65 open-source attributes, of which the majority can be described as contextual, such as the number of days since the last game, travel distance and form using data from 13 season of the highest division of the Netherlands	LGB: 56.1%
Prasetio and Harlili [198]	Prediction of win/draw/lose using player characteristics derived from the video-game FIFA, using data of 5 season in the highest division of England.	LOR: 69.5%
Groll et al. [112]	Prediction game outcome of World Cup 2018 games using contextual variables such as ELO-rating, FIFA Rank and average age, combined with World Cup 2002-2014 game data and 8 years of historical game data preceding the 2018 World Cup	RFO: 60.9% LAS: 59.4
Knoll and Stübinger [142]	Prediction of win/draw/lose using basic metrics on shots, passes, possession, duels, set-pieces, fouls and cards of over 8000 games in the highest division of England, Spain, Germany, Italy and France.	RFO: 75.6% XGB: 71.0% SVM: 66.1% LIR: 63.6%
Stübinger et al. [230]	Prediction of win/draw/lose using player characteristics derived from the video-game FIFA, using over 20000 games in the highest and second highest divisions England, Spain, Germany, Italy and France.	RFO: 81.3% XGB: 79.1% SVM: 69.7% LIR: 72.9%



(A) Distribution of total goals per game, demonstrating the (generally and relatively) low-scoring nature of soccer.



(B) Goal differences in game outcome, illustrating that in over 63% of all games, a single goal could have changed the game outcome.

Figure 6: An illustration on the low-scoring nature of soccer, using all game outcomes from the last 20 seasons in the highest domestic leagues of England, Spain, Germany, Italy, France and Belgium. Data was derived from *https://www.football-data.co.uk*.

To conclude, nowadays there are new approaches, in the form of machine learning, to relate performance indicators to game outcome. Predictions by machine learning models can be used to identify the best predictors of game outcome on a global and local level. Generally, the accuracy of machine learning models applied to game outcome has been relatively low, however a broad range of detailed performance indicators may provide better modelling accuracies.

2 Training & Monitoring

Training is an important piece of the puzzle with regard to improving the central aspects of performance in soccer [229], being the technical, tactical, physiological and mental aspects, and consists of systematically performed exercises to improve sport specific skills and physical abilities [250]. These central aspects should all be considered when planning, monitoring and evaluating the training process, for players individually and the team as a whole. Training in soccer is a comprehensive concept, as for a single exercise, coaches should consider factors such as the number of players, duration of the exercise, space size, technical-tactical predominant actions, complexity and physical load [65]. As this would be too all-encompassing to comprehensively discuss, this chapter will mainly focus on the general physical aspect of training, although without disregarding the other aspects of soccer. Each activity will impact the physical conditioning of soccer players differently, therefore, it is important to monitor the workload imposed on players [82]. Monitoring of workload - *workload* is also often referred to as *training load* - is done with the aim of making evidence-based decisions on appropriate loading schemes to enhance team performance and avoid negative training effects [5]. There is a close interaction between training and monitoring (Figure 7), therefore, these topics will be jointly discussed within the this chapter.

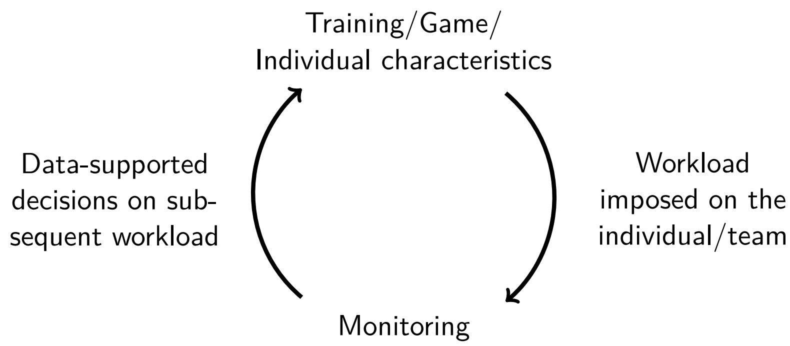


Figure 7: Graphical illustration on the interaction between training and monitoring in soccer, based on Impellizzeri et al. [130].

2.1 Training

The aim of training is to get an athlete ready for any game demands within reason [102]. The intermittent activity profile of soccer necessitates contributions from both the aerobic and anaerobic energy systems, while sufficient strength, speed, endurance, coordination and flexibility are required to perform soccer-specific actions [183], demonstrating that soccer players require a multitude of fitness components at a high level [119]. To develop and maintain these fitness components, as well as working on the other central aspects of soccer, different types of training sessions are performed [251]. It should be noted that technical/tactical sessions are frequently the prioritized aspect of the training plan and will therefore often take precedent over the other training activities [183]. Physical conditioning sessions on and off the pitch are nevertheless also usually a consistent part of the training schedule. To optimally structure training in soccer, it is first important to understand the determinants and influencing factors of training, as depicted in Figure 8.

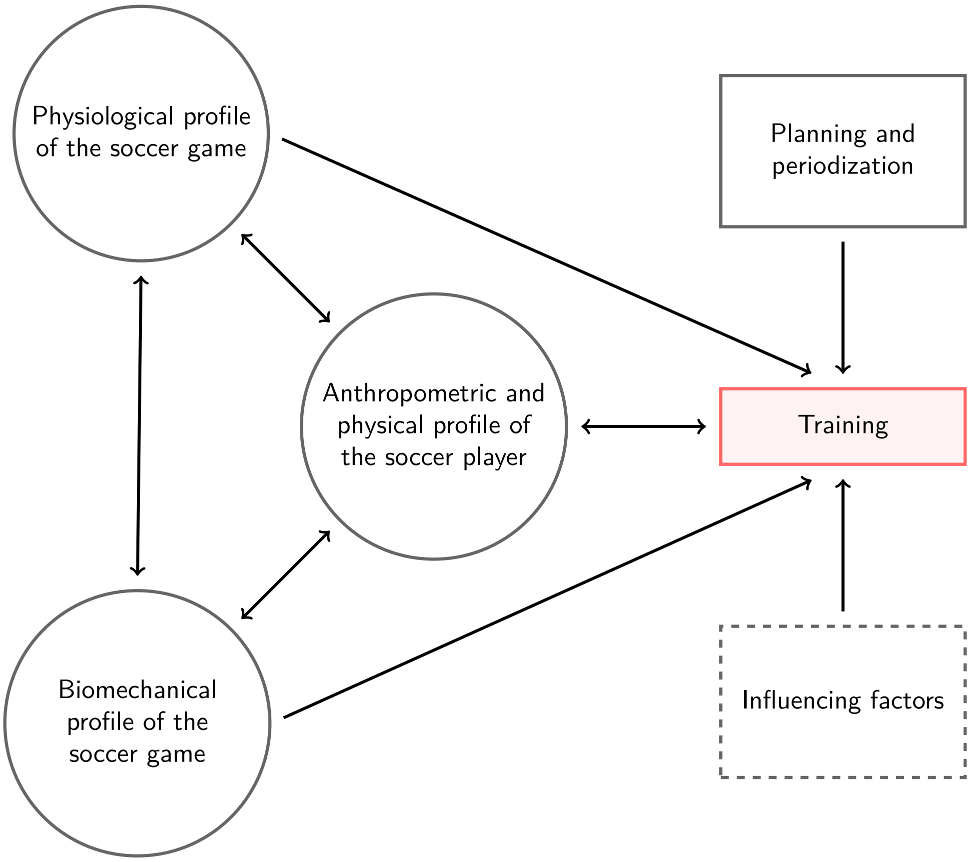


Figure 8: Adaptation of the model by Bourgois et al. [30], showing the interactions between the determinants and influencing factors of training in soccer.

2.1.1 Determinants of training

2.1.1.1 Physiological profile The physiological demands in soccer are multifactorial because of its complex nature [14], and include factors such as the cardiorespiratory and -vascular load, metabolic load, muscle fiber recruitment and the contribution of different energy systems [30]. In terms of energy systems, soccer mainly depends on the aerobic metabolism [14, 229], using aerobic pathways for glycolysis and fat metabolism by the mitochondria [8]. It is estimated that about 98% of the total energy is derived from the aerobic metabolism [121], using muscle glycogen as the most important substrate for energy production [16]. The anaerobic system is however a predominant energy source for short, intense actions, indicated by high rates of creatine-phosphate utilization and glycolysis [16], to resynthesize adenosine triphosphate (ATP).

The intermittent pattern of exercise and recovery in team sports can best be described as random, as it is imposed by tactical factors and the player's ability to self-select the intensity and nature of their efforts [108]. There are parts of the game where the recovery duration is sufficient to allow full recovery of sprint performance [13], but there are also parts in which recovery duration will be too short to allow full recovery. The ability to repeatedly perform high intense actions is therefore an important requirement for soccer players [22]. Even small impairments in the ability to perform repeated high intense actions could be detrimental for performance on the pitch, as an estimated ≈0.8% decrease in sprint speed would already have a substantial effect on the likelihood of a player losing possession of the ball, when sprinting against an opponent for the ball [193]. Phosphocreatine is particularly important for repeated high-intense actions, since better maintenance of muscle power output has been attributed to a faster rate of phosphocreatine resynthesis during the recovery between sprints [108].

2.1.1.2 Biomechanical profile The biomechanical load of a sport in general is determined by different characteristics such as generated forces by muscles and joints, movement frequency, changes in direction, technique, and impacts with the environment [30]. Playing soccer will lead to mechanical stress on different body tissues, such as cartilage, bones, muscles and tendons [245]. These body tissues will be largely stressed by propulsive and breaking forces against the ground [245], but also by actions such as kicking, tackling, falling and jumping [160]. Over the course of a full game, the body also repeatedly needs to absorb the high forces from impacts with other players and the ground [245]. The biomechanical load of soccer is however challenging to quantify, because of the difficulties of measuring biomechanical loads in vivo, which is a major issue limiting the progress of understanding biomechanical load-response pathways [246]. It is however clear that the biomechanical impact of soccer on the body is a significant factor to consider.

2.1.1.3 Anthropometric and physical profile of soccer players

Anthropometric profile of soccer players Anthropometry is the study of the human body measurements, and for soccer players often expressed in body height and body weight (Table 2). It is clear that there is considerable variation in weight and height between and within playing position. This is in contrast with many other sports, in which there is a necessity for a distinct anthropometric profile to perform at a high level in that particular sport. For sports such as gymnastics [214] and basketball [259], body height is considered as a determinant of performance. To illustrate the anthropometric profile of different sports, body height is considered for elite soccer players, gymnasts and basketball players in comparison to the average male body height. For elite gymnasts, the average body height is 167 cm [10], while for professional basketball players in the National Basketball Association (NBA), the average body height lies around 206 cm [194]. For soccer players playing in the English Premier League, the average height is about 182 cm [231], which is closer to average male body height (178 cm [187]), as compared to the average body height of elite gymnasts and professional basketball players (Figure 9).

Table 2: General physiological and physical profile of professional soccer players. Only studies that included a few hundred players from different teams within the same league were used, as to include sufficient numbers of players for each position. Abbreviation: Maximum Oxygen Uptake (VO_2max), Counter Movement Jump (CMJ), Not Available (NA).

Playing Position	Country	Body Height (cm)	Body Weight (kg)	Maximum Oxygen Uptake (VO_2max) (ml/kg/min)	10m sprint (s)	Counter Movement Jump (CMJ) (cm)
Goalkeeper	Belgium [24]	188.2 ± 4.5	84.2 ± 4.5	52.1 ± 5.0	NA	45.6 ± 2.6
	Croatia [227]	185.0 ± 3.1	81.0 ± 2.3	50.5 ± 2.7	2.27 ± 0.04	45.1 ± 1.7
	Iceland [9]	185.2 ± 4.7	81.4 ± 7.7	57.3 ± 4.7	NA	38.0 ± 5.6
Defender	Croatia [227]	177.2 ± 4.5	74.5 ± 5.6	59.2 ± 1.5	2.14 ± 0.07	44.2 ± 1.9
	Iceland [9]	181.1 ± 5.4	76.9 ± 6.1	62.8 ± 4.4	NA	39.3 ± 5.5
Central Defender	Belgium [24]	186.4 ± 4.3	82.5 ± 5.0	55.6 ± 3.5	NA	46.0 ± 4.1
Full Back	Belgium [24]	179.3 ± 4.8	73.4 ± 6.4	61.2 ± 2.7	NA	41.0 ± 3.8
Midfielder	Belgium [24]	181.3 ± 4.1	76.7 ± 5.1	60.4 ± 2.8	NA	41.4 ± 3.7
	Croatia [227]	169.4 ± 5.6	64.4 ± 3.2	62.3 ± 3.1	2.23 ± 0.05	44.3 ± 2.1
	Iceland [9]	179.3 ± 5.2	75.9 ± 7.0	63.0 ± 4.3	NA	39.3 ± 4.9
Attacker	Belgium [24]	183.5 ± 6.7	78.6 ± 4.8	56.8 ± 3.1	NA	44.2 ± 4.2
	Croatia [227]	180.7 ± 3.4	78.4 ± 5.2	58.9 ± 2.1	2.03 ± 0.09	48.5 ± 1.5
	Iceland [9]	180.2 ± 5.3	75.3 ± 5.9	62.9 ± 5.5	NA	39.4 ± 4.2
ALL	Croatia [227]	181.4 ± 2.5	78.4 ± 3.1	60.1 ± 2.3	2.27 ± 0.04	45.1 ± 1.7
	Iceland [9]	180.6 ± 5.4	76.5 ± 6.6	62.5 ± 4.8	NA	39.2 ± 5.0

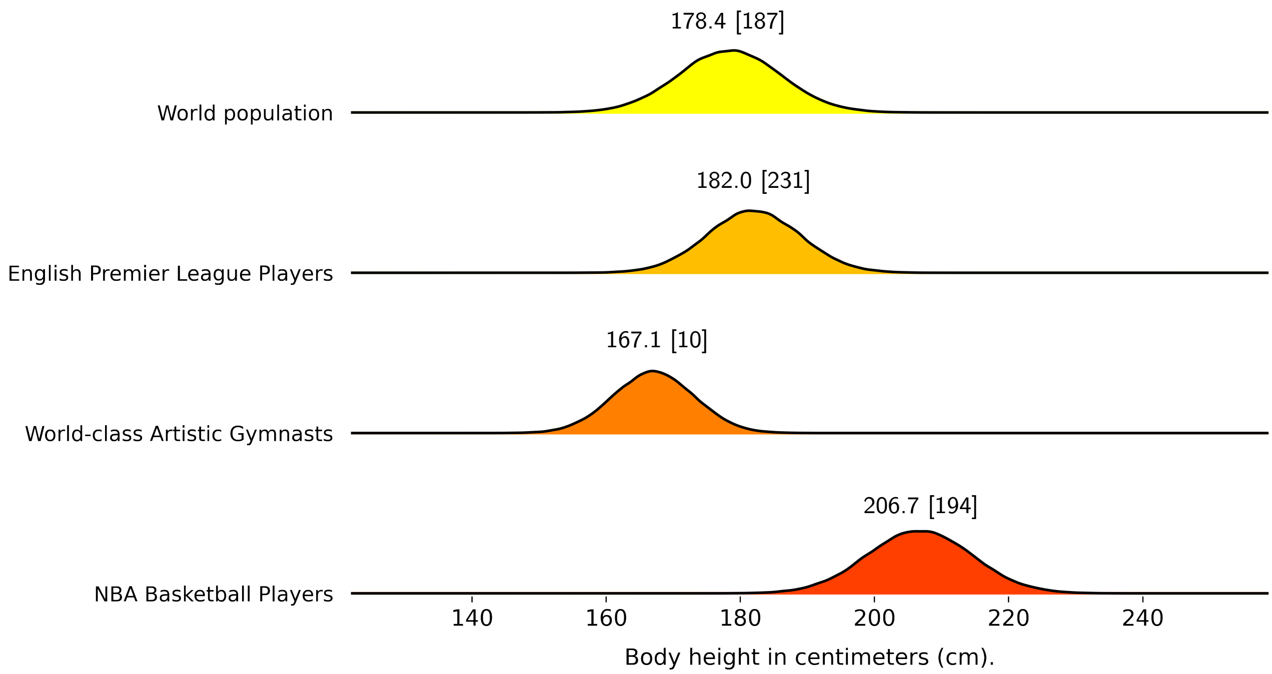


Figure 9: Graphical illustration on the distribution of body height in elite athletes in soccer, gymnastics and basketball, compared to the distribution of body height of the world population. It is known that the distribution of body height is normally distributed within the general population [143]. For the sample of soccer players, gymnasts and basketball players data on distribution was not available, therefore, it was assumed that their body height was normally distributed as well. The majority of the world population has an adult body height within the yellow area of the normal distribution of the male world population, illustrating that only a small sample of the world population has an adult body height which conforms to the conditions for body height in artistic gymnastics or basketball, in contrast to soccer. Above the distributions, the average height for the concerning population is given.

Physical profile of soccer players The physical profile concerns all traits connected to the basic fitness components, such as endurance, speed, strength, flexibility and coordination. As for anthropometry, there is a reasonable variation in the performance on several physical tests (Maximum Oxygen Uptake (VO_2max), 10-meter sprint and CMJ) between and within playing position (Table 2), which may be because of the versatility in physical traits that a soccer player needs to possess. A more physiological example relates to the muscle fibre typology of soccer players, in comparison to athletes in two athletic disciplines, namely sprinting and marathon running. The proportion of fast-twitch fibres in professional soccer players is reported to be between 59.8% $\pm$ 10.8% [133], while the proportions of fast-twitch fibers in high-level sprinters and marathon runners are reported to be on average respectively >70% [58, 241] and <31% [58]. The ratio between slow- and fast-twitch fibres lies around 50%-50% in the general population [213], which indicates that soccer players generally do not possess an extraordinary muscle fibre typology. Moreover, extraordinary low proportions of fast-twitch fibres may even be detrimental with regard to speed-related capacities [108], which are considered as important for soccer players [118], while extraordinary high proportions may be unfavourable in terms of recovery between repeated sprints [224] or the recovery of muscle glycogen content between games [113].

To conclude on the anthropometric and physical profile, one of the reasons that soccer may be so popular worldwide [229], is because the anthropometric and physical characteristics do not *a priori* minimize the chances of competing at a high level. It should nevertheless be highlighted that over the years, the physical demands of top elite soccer have increased [18, 33, 40], and it can be argued that, nowadays, possessing a mixture of physical qualities at a high level is important to play at a top level, besides qualities on the technical, tactical and mental aspects of soccer. For example, it is not only important to be fast [118], but it is necessary to be capable to repeatedly perform high intensity actions. High levels of aerobic fitness are therefore also important, as it may aid the ability to resist fatigue during repeated high intensity actions [109, 238], given that the restoration of creatine-phosphate is limited by the availability of oxygen [211]. With regard to the anthropometry, a more average profile compared to the general population seems advantageous, while for the physical profile, versatility is beneficial, with many physical qualities at a high level [119].

2.1.1.4 Planning and periodization A soccer season is commonly planned in three distinct periods: transition periods, preparatory periods and competition periods [221] (Figure 10). The duration of each period is influenced by intrinsic (e.g., environmental conditions) and extrinsic factors (e.g., international competitions). Each of these phases have specific goals and associated challenges with regard to periodization of training [251]. Periodization can be defined as the process of systematically planning a short- and long-term training program by varying workloads and incorporating adequate rest and recovery [153], and should be viewed as a framework in which a more detailed training plan can be scheduled [152]. Typically, the preparatory season in the summer, usually mentioned as the *preseason*, is considered as the start of the season. The competition period will however be first discussed, as this concerns the vast majority of the season. First discussing the competition period also allows to emphasize the importance of both the preparatory and transition periods as paramount phases during the season.

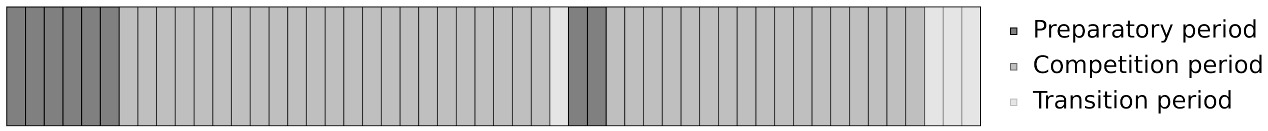


Figure 10: Example on an overall overview on the competition phases during a season, inspired by Malone et al. [170]. Each separate rectangle represents a single week.

Competition period(s) At the highest levels of contemporary soccer, some teams play over 60 competitive games per season [45], over the course of about 10 months [221]. High frequencies of games forces teams to work from game to game [251], in which the planning and periodization strategy will be strongly determined by the game schedule during the competition period [30]. The number of days between games influences the training schedule, as such that the training schedule of a week with one game differs from a week with two games [65, 203, 251]. Next to an often congested fixture schedule, practitioners have to cope with other challenges, such as (inter)national travel, which impacts sleep, recovery and stress [188]. Players also need to recover from games, which makes providing the appropriate training stimuli and preparing players for the following game(s) more challenging. Although the specific periodization of training depends on the philosophy of the coaching staff [7], several studies describe a similar pattern [65, 251], illustrated in Figure 11.

As previously mentioned, during the in-season period, teams usually work from game to game [251]. It can be argued that the preparation for a new game starts at the end of the previous game. The first day after the last match (MD+1) and second day after the last match (MD+2) are usually focused on recovery from the game. A soccer game is a taxing activity for in-field players, as it results in inflammation, muscle damage and metabolic disturbances [222]. These effects are more pronounced compared to sports such as basketball, handball and volleyball, possibly because of, amongst others, a high amount of eccentric work [225]. A game also results in depleted glycogen stores [133], which is linked to reduced exercise performance at the end of a soccer game [15]. Muscle glycogen stores are reported to be restored after 48 hours [146]. In light of recovery, the workload during the first two days after a game is therefore usually low.

When the next game is approaching, there usually is a reduction in workload towards the game on the

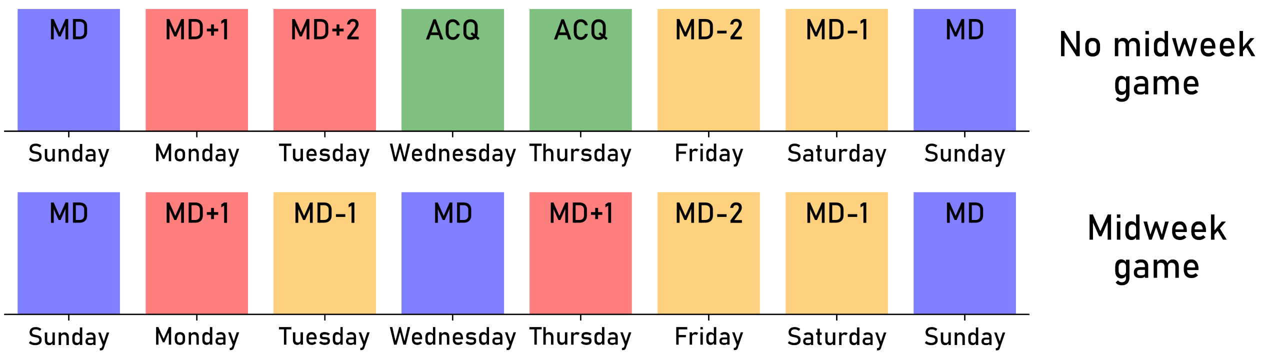


Figure 11: An example on the weekly training structure of a training week with no midweek game and a training week with a midweek game, based on examples from professional teams [65, 203, 251]. Red bar: focus on recovery from the previous game; yellow bar: focus on preparation for the next game; green bar: opportunities for acquisition session(s); blue bar: game. Abbreviations: Match Day (MD), first day after the last match (MD+1), second day after the last match (MD+2), second last day prior to the next match (MD-2), last day prior to the next match (MD-1), Acquisition Period (ACQ).

second last day prior to the next match (MD-2) and the last day prior to the next match (MD-1). An example on the content of the sessions of a Spanish first division team, also competing in the Champions League, on MD-2 and MD-1 was provided by Martín-García et al. [174]. On MD-2, the focus is often on technical-tactical elements, with control and passing sequences, a positional game with a low number of players per team and tactical exercises. On MD-1, activation drills replicating the tactical competition scenarios are conducted and the session is concluded with set pieces.

The remaining days between two games can be used for acquisition session(s). These sessions are, amongst others, usually focussed on applying a physical overload [251]. An example of the training content of these days are a gym session followed by positional and small sided games and/or a more tactical session with positional games and an 11-vs-11 [174]. It should however be noted that at the highest levels of international soccer, often two games are played during a single week. In these cases, factors such as the quality of the opposition, number of training days between matches and travel activities also influence the between match-periodization [138]. The second game during the week will usually have a big impact on the training schedule, as this would eliminate the feasibility of acquisition session(s), because all training sessions are conducted on the days directly prior or after a game [251]. A congested fixture schedule will also highlight the need for supplemental training for the benched and non-selected players to maintain fitness [251].

Transition & preparatory period(s) The last game of the competitive season usually marks the end of the competition period [221]. As previously explained, during the competition period, professional soccer players repeatedly have to perform at a high level [24, 229], resulting in an accumulation of physiological and psychological stress [220]. The off-season period, often between 3 to 6 weeks [221, 251], allows players to recover from this accumulated stress, and this period is normally characterized by a complete or substantial reduction in training [220]. This reduction in training will result in detraining effects, being a partial or complete loss of training-induced adaptations [184], impacting cardiorespiratory and -vascular, metabolic and muscular functions. The physical training status at the start of the preseason of the players is however often unknown, as activities are usually not comprehensively monitored during the off-season period. It is therefore important that the coaching staff considers both recovery from the accumulated loads during off-season [220], as well as educating players on the importance of keeping a minimal chronic load to minimize injury risk in the subsequent preseason [103].

The contemporary sequences of congested in-season fixture periods in soccer place a greater importance on the preparatory periods, and the preseason in particular, as this is for some teams a unique window during the season where physical overloads on players can be placed during training sessions [185]. The duration of the preseason is influenced by the last game of the previous season and the first game of the new season [97], as well as by international tournaments such as the FIFA World Cup or UEFA European Championship [97, 251]. Preseason protocols with a length of 4 to 8 weeks are commonly reported [97, 173]. The preseason generally consists of a variety of training types, including match play, running sessions and fitness training [80]. Weekly training hours of 14-18 hours are reported [81], accumulated over 9 to 11 separate activities [81, 137, 173], including both team training sessions and friendly games. These activities are usually spread over 6 days, as players usually have 1 day off [81, 137]. Preseason training has previously been associated with an increase in VO2max [81, 173] and several lactate thresholds [47, 173]. To elicit these positive training effects, training should be carefully planned, especially given the importance of the preseason [80]. The importance of the preseason is confirmed by a study of Ekstrand et al. [80], which showed that a greater number of preseason team training sessions was associated with a lower injury burden, fewer severe injuries, higher training attendance and higher player availability during the season. It should also be noted that the preseason is not only an important period of the season from a physical, but also from a technical and tactical point of view. Acquisition sessions must therefore be planned in the context of the other training sessions that players are required to perform [105].

Most professional soccer teams from the best European leagues have a second transition and preparatory period during the winter break, between the two competition periods [79]. The length of this transition period, defined by the number of days between the last team training or match before the break and the first training or match after the break, usually ranges between 6 to 14 days [79]. After the transition period in the winter, there is often a short preparatory period to prepare the players for the second part of the competitive season. The length of both the winter transition period as well as the preparatory period depend on the game schedule from the league association and the teams' training schedules [12, 79].

2.1.2 Influencing factors

Based on the determinants of training, the general framework of training can be made. There are however other factors that may influence the implementation of training, such as recovery, environmental conditions and other influencing factors. These topics will be briefly discussed below.

2.1.2.1 Recovery, environment and other influencing factors

Recovery Various recovery interventions are used to facilitate physiological and psychological recovery after exercise, to help individuals perform better during subsequent training sessions or competition, and lower injury risk [244]. Elite athletes often train or compete more than once a day, so recovery interventions between training sessions or events may help to restore exercise performance [244]. Appropriate (or inappropriate) recovery strategies, being the basic principles such as sleep [115] and nutrition [203], combined with supplemental recovery strategies such as a cooling down and/or stretching [244], may influence subsequent training.

Environmental conditions Soccer is a sport that is played all around the globe, leading to soccer to be played in differing environmental conditions. From temperatures far below zero and snowy conditions, to temperatures well above 30°C, and games and training sessions at high humidity, or altitudes higher than 3000 meters [76]. Each of these conditions, or even combinations of these conditions, stress the human physiology. Cardiorespiratory and -vascular, metabolic, neuromuscular and thermoregulatory alterations have been associated with exercise in these conditions [73, 116, 175, 176]. These environmental factors may induce alterations of the training schedule.

Other influencing factors There are a plethora of factors that may affect training. These factors may be personal and consequently affect a single individual or several individuals within a team, but can also be factors such as unforeseen circumstances when travelling, which could affect training of the whole team. A very topical example of a factor impacting training was the first lockdown of the COVID-pandemic, which confined players to train at their homes and compelled creative solutions to be sought by staff members to help their players maintain their physical conditioning and lifestyle [63]. Religion may also be a reoccurring factor influencing training, with the Ramadan as a good example. During the month of Ramadan, training approaches may need to be changed for some individuals or even the whole squad, as Muslims fast throughout the daylight hours, resulting in altered dietary plans and sleep schedules [179].

All conditions considered, both the determinants of training, as well as the influencing factors, programming appropriate training schedules for the team as a whole and the players individually is challenging. Efficient communication within the coaching staff is important, as in many cases it is the head coach who dictates the training program and therefore determines a large part of the workload [5, 39]. It is a constant communication and management of on-field and off-field training stimuli that is required to successfully integrate the different physical conditioning components within the overall training plan [251].

2.2 Monitoring

To accurately prescribe and evaluate training, in order to elicit subsequent maintenance or improvements of physical fitness, quantification of workload is considered to be crucial [94]. In the context of athletic training, the external load, represents the activities performed by the athlete [250]. The external load will invoke an internal load, which incorporates all the psychophysiological responses occurring during the execution of the exercise (single or sequence) prescribed by the coach [130]. Acute and chronic adaptations of the physiological and musculoskeletal system take place in reaction to the internal load [130, 245]. Monitoring both the external and internal load is important to properly adjust training plans and organize training load [55]. Inappropriate training could elicit negative training effects such as injuries, decreased fitness and decreased performance [104].

Developments in technology and analytical methods have led to new possibilities in the applied environment, and practitioners now have the ability to monitor workload using Global Navigation Satellite Systems (GNSS), HR-sensors and other microtechnology [92]. Contemporary technology produces a plethora of variables enabling practitioners to quantify workload in greater detail than ever before [5] and technology is nowadays considered as an essential aspect in the monitoring of players [92]. Workload is commonly described by a combination of volume and intensity parameters [217], which are either external or internal [129]. The adopted technological tools to quantify workload, provide coaches with means to measure progress, helping them to make day-to-day or week-to-week decisions [92]. Concurrently, these tools should also be minimally disruptive to training comfort [92]. In the remainder of this chapter, the monitoring of training in soccer will be introduced according to the structural representation of training monitoring in soccer depicted in Figure 12, followed by the current monitoring approaches in practice.

2.2.1 Expression of workload

2.2.1.1 GNSS- and IMU-technology The first documented form of monitoring in sport was conducted in the early 1900's by endurance track athletes [95], using stopwatches to track running times. As the distances on a running track were also known, the running pace could be deduced. A similar approach to monitor soccer is unfeasible, as soccer activities are usually associated with unstructured movement patterns [170]. Hand-coding the unstructured movements patterns of players, as it was done in the 1970's and 1980's [125], is an unsuitable method to monitor the daily workload as well. Therefore, the feasibility to monitor soccer specific activities was strongly related to the developments of technology.

Nowadays, utilization of technology to track the movement of players is used on a daily basis, with

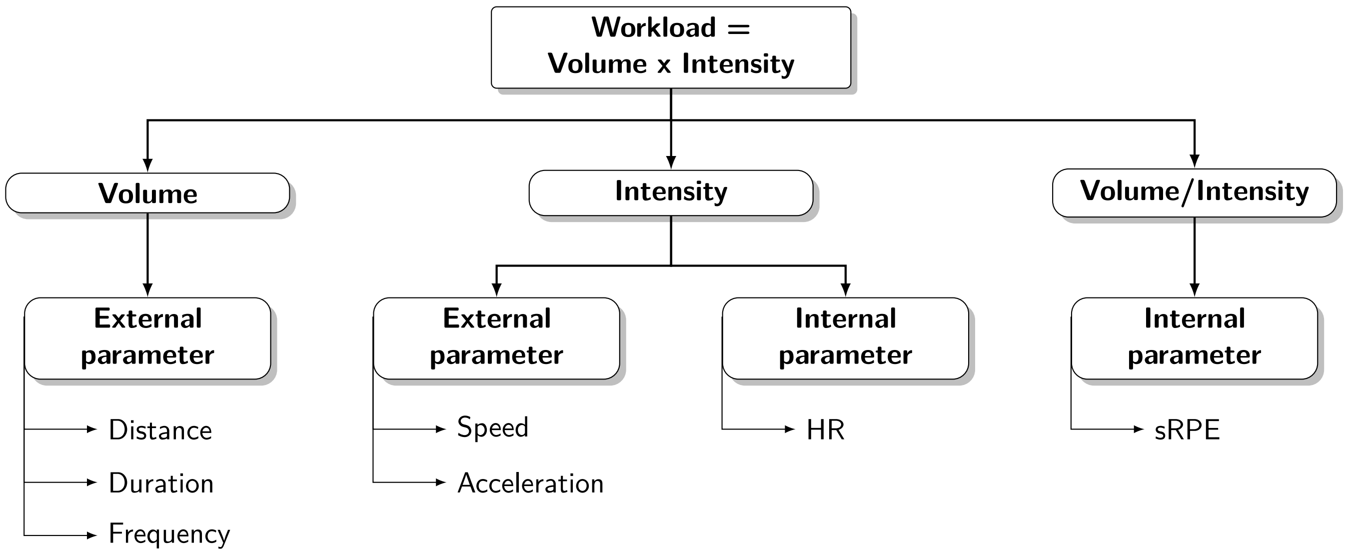


Figure 12: An overview on the basic parameters which are nowadays used to quantify workload in soccer. Abbrevations: Heart Rate (HR); session Rate of Perceived Exertion (sRPE).

three main technologies i.e. GNSS, Local Positioning Systems (LPS) and video technology. GNSS is the most widely used technology for the day-to-day follow-up [5]. GNSS entails the different types of satellite systems, such as the Global Position System (GPS) developed by the Americans, Russia's GLONASS, China's BeiDou and Europe's GALILEO. GNSS, in the form of GPS-technology, was initially developed for military use in the 1970, but was made available for civilian use in the 1980's [119]. Around the year 2000, the first studies were conducted to study the use of GNSS-technology for locomotor activities [132, 216]. GNSS is a satellite-based navigation system, which uses orbiting satellites to allow accurate determination location [119, 216]. Each satellite transmits the exact time and its location [216], which are subsequently used by GNSS-devices to calculate the distance to the satellites by the amount of time it takes to receive a transmitted signal [119]. As the GNSS receives signals from several satellites simultaneously, the GNSS-device can be accurately located [119]. LPS uses 'local satellites' to locate the position of the receiver [207]. The local satellites are base stations surrounding the pitch [100] of which the location is known [207]. The last often used method to track players is video technology, which is mainly used for games. Quantification of game workloads are often part of a league-wide deal (e.g. partnership of the Pro League (Belgium) with STATS Perform (Chicago, USA) [200], partnership of the Deutsche Fußball Liga (Germany) with ChyronHego (New York, USA) [165]). As a result of these partnership, all teams are monitored by the same company, using a standardized methodology to obtain game data. Video technology to track games is currently largely computer automated [44].

The use of tracking technology has become an integral part of physical performance analysis, allowing coaches and support staff to understand the physical demands induced on team sport athletes [49]. The choice of technology should be based on the strengths and limitations of each system. The cost of GNSS-technology is (relatively) low [11] and GNSS-technology generally has a high portability [11, 119], but is bound to the connectivity to satellites and can thus also only be used outdoors [11]. Data quality from GNSS-technology is lowered when dealing with an insufficient number of satellites [119], objects that block communication between satellites and the GNSS [60], or an insufficient spread between satellites [207]. LPS can be used to counter these problems, which showed to have a higher validity for measuring athlete position compared to GNSS and video technology [164]. Disadvantages of the LPS-technology are however the costs and the lowered portability. A main advantage of video technology is that the players do not need to wear a sensor, in contrast to GNSS and LPS [164]. On the other hand, the error of measurement of instantaneous speed is higher for video technology than for both GNSS and LPS [164]. It should be noted that in the field of video technology, current developments show promising results with regard to accuracy [165], possibly already outperforming LPS. What should be considered for all types of technology, are higher measurement errors as the speed of the tracking object increases [164]. It is also suggested that technologies with higher frequency provide greater validity for measurements of distance [11, 60, 202]. Conclusively, when interpreting physical data, it is thus important to know which technology was used and its sampling frequency. Data regarding high-intensity movements should be interpreted with caution, especially when utilizing technology with a low sampling frequency. Because of differences between data from various types of tracking technology, integration of data from different systems should also be cautiously done [37].

Tracking systems basically express the location of a player in X-Y-coordinates [212], and this location is repeatedly measured. The earth is spherical in shape, which makes the calculation to measure distance more complex than depicted in Equation 1, however, when applying a Cartesian coordinate system [212], the distance could theoretically be calculated using the Pythagoras theorem. When considering two time points, T_1 and T_2, the associated X-Y-coordinates are (X_1,Y_1) and (X_2,Y_2). The distance between the locations at T_1 and T_2 can be calculated using the formula depicted in Equation 1. The displacement of a player in a certain time interval allows to calculate speed (Equation 2); the change in speed over time allows to calculate acceleration (Equation 3) [191].

$$Distance = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \tag{1}$$

$$Speed = \frac{\Delta Distance}{\Delta Time} \tag{2}$$

$$Acceleration = \frac{\Delta Distance}{\Delta Time^2} = \frac{\Delta Speed}{\Delta Time} \tag{3}$$

This indicates that acceleration can be determined using the X-Y-coordinates measured by tracking systems. However, as the acceleration is the second derivative of distance, small measurement errors in the location of a player may be magnified, leading to artefacts in the acceleration data [6, 191]. There has however been significant work on the integration of microtechnology [59], including the integration of Inertial

Measurement Units (IMUs). [119]. Inertial Measurement Units (IMUs), electronic devices usually containing an accelerometer, gyroscope and magnetometer to measure respectively linear acceleration, angular velocity and orientation [233], are used to measure accelerations in three axes [49]. Today, most commercially available GNSS-devices contain IMUs [49], allowing more accurate determination of accelerations in soccer.

Monitoring accelerations in soccer is relevant, as players do not move at a fixed pace, but increase their speed (acceleration) or decrease their speed (negative acceleration or, as it is usually referred to, deceleration) [61]. Accelerations and decelerations during soccer-specific activities contribute to the workload [210]. Although high-intensity movements, such as jumping, tackling, collisions, passing, shooting, accelerations and decelerations may induce high levels of physiological and/or mechanical stress, these movements may be classified in the low-speed locomotor category, causing an underestimation of the workload during training and matches [61]. Various workload terminologies with regard to accelerations have been developed, such as *PlayerLoad* (Capapult Sports, Melbourne, VIC, Australia) and *Body Load* (GPSports Systems, Canberra, Australian Capital Territory, Australia), which both use the accelerations in different planes (anterior-posterior, medial-lateral, vertical) to calculate a general measure of workload resulting from acceleration [49]. A different popular concept resulting from acceleration-related data is the metabolic power concept, based on the findings of Di Prampero et al. [69] regarding estimations on the energy cost of sprinting. It has been suggested that high-intensity demands, in terms of energy cost, are underestimated by traditional measurements of running speed alone [106], and the use of data regarding accelerations may thus be useful. The study of Osgnach et al. [191] was the first to propose the use of metabolic power in soccer, which uses both the speed and acceleration to estimate the metabolic cost of soccer. The metabolic power concept is currently widely used to quantify workload [5]. There are however also considerations for the use of metabolic power calculated using GNSS. As it is stated by Buchheit et al. [38], the metabolic power concept is still an underestimation of the energy demands of soccer-specific drills. The concept behind metabolic power is nevertheless interesting, as it tries to estimate the energy cost of soccer, which can be considered as an internal load parameter, using speed and acceleration as external load parameters.

2.2.1.2 Internal load The ability to quantify internal load is important as it allows practitioners and coaches to quantify the implications of the external load and training prescriptions on various physiological systems [42]. The concept of internal load incorporates all the psychophysiological responses occurring during the execution an external load [130]. Examples of internal load measures are HR, the Rate of Perceived Exertion (RPE), and biochemical, hormonal and immunological assessments [74]. It is also important to note that internal load parameters are generally more difficult to quantify in a feasible manner [42]. Assessing muscle oxygenation during training using Near-infrared Spectroscopy (NIRS), for example, or examining humoral and neuromuscular parameters in a feasible manner for daily use requires technology that has not yet been developed [42]. Practicality is an important factor to consider when monitoring internal load, as the implementation of too many devices or measurements may interfere with the athlete's training activities [42]. For instance, owing to the continuous competitive characteristics of soccer, it is impossible to collect blood samples during matches [68]. In practice, HR and Rate of Perceived Exertion (RPE) are the most common internal load measures in soccer [5].

HR In the late 1930's, the use of HR was first documented in track athletes, when heart-rate responses were monitored by palpating the HR between interval sets [95]. The first wireless heart rate monitors were introduced in 1983, consisting of a transmitter and receiver [157]. This development resulted in an increased use of HR-monitoring in athletes [3]. HR-monitoring represents a noninvasive method that is universally used to monitor the physiological response to external load in team sports [68], and has the advantage that it is relatively easy to measure and HR-data can immediately be used to adjust the intensity [3]. It is well known that HR shows a quasi-linear relationship with VO_2 up to nearly maximal intensities, although for precise estimations of exercise intensities, the relationship between HR and VO_2 should be determined individually [3]. There are however some considerations for the use of HR to monitor internal load. Intra-individual variation of 2-4 beats/min are likely to occur when individuals are measured on different days [3]. Furthermore, physiological factors such as cardiovascular drift and hydration status, environmental factors such as temperature and altitude [3], and psychological factors such as stress [68] influence HR. Furthermore, due to the inertia of HR, the measurements of HR may not reflect the instantaneous internal load [3, 228]. However, more importantly in a soccer context, the HR does not accurately reflect the anaerobic energy production during training [15]. Therefore, HR cannot be considered as the most reliable indicator of exercise intensity during all exercises including directional changes, small-sided games, high-intensity intermittent exercises, exercises with much anaerobic work and power training [68].

(s)RPE The current best alternative method to monitor internal load in soccer is the RPE. The RPE was introduced by Borg in the 1960's [25], and additional studies on this concept were published in 1970's [27] and 1980's [26]. Perceived exertion has been described as a psycho-biological process [182], and the rating of perceived exertion (RPE) involves the collective integration of afferent feedback from cardiorespiratory, metabolic, thermal stimuli and feed-forward mechanisms to enable an individual to evaluate how easy or hard an exercise task feels at any point in time [84]. The RPE is moderated by psychological factors (eg, cognition, memory, previous experience, understanding of the task) and situational factors (eg, knowledge of the end point, duration, temporal characteristics of the task). The use of the RPE in sport, independent of age, is founded on its strong relationships with exercise intensity (eg, work, speed, power) and physiological factors (eg, heart rate, ventilation, oxygen uptake, blood lactate) [84]. Initially, a 21-grade scale was used to rate the perceived exertion [26], later, a scale from 6-20 was introduced, based on the range between the resting HR (60 = **6** X 10) and the 'maximum' HR (200 = **20** x 10) [28]. Today scales to assess RPE range from 6-20 or from 0-10 [26, 94] (Figure 13).

Borg RPE-scale [28].		**CR10-scale [94].**	
6	No Exertion At All	0	Rest
7	Extremely Light	1	Very, Very Easy
8	-	2	Easy
9	Very Light	3	Moderate
10	-	4	Somewhat Hard
11	Light	5	Hard
12	-	6	-
13	Somewhat Hard	7	Very Hard
14	-	8	-
15	Hard (Heavy)	9	-
16	-	10	Maximal
17	Very Hard		
18	-		
19	Extremely Hard		
20	Maximal Exertion		

Figure 13: Illustration on the most used sRPE-scales.

The RPE is usually assessed as the global perceived exertion for a complete session/training, therefore, it is referred to as the sRPE [93]. The sRPE is the most common method to assess subjective rating of perceived exertion in soccer. The sRPE-method to monitor internal load has been shown to be valid, reliable and useful on the field [114]. The sRPE is usually assessed 30 minutes after training. The use of this 30-minute latency period when assessing the sRPE is based on the assumption that a latency effect exists and that a timeframe of 30 minutes is sufficient to neutralize the influence of the intensity at the final stages of a training session [85]. It was however shown that the exercise intensity distribution of a soccer session does not influence the sRPE directly of 30 minutes after a training session [85]. Therefore, it appears that a latency period is not required [53, 85]. In practice, the sRPE is commonly used [5], although is not always assessed after matches [5, 85], as a consequence of the sensitivity of the postmatch environment and the psychological states of players and coaches.

An advantage of the sRPE compared to HR as a measure of internal training, is that the sRPE appears to work well with a variety of types of exercise, from resistance training to game activities, within a large variety of sporting activities [96]. An additional advantage of the sRPE is that it does not require additional technology. Furthermore, according to Vanrenterghem et al. [245], the sRPE can be used to assess both the physiological and biomechanical internal load. The use of sRPE is time-dependent [29, 84], which indicates that this metric cannot be strictly subdivided between a volume or intensity parameters, as it integrates both factors.

2.2.2 Monitoring workload in practice

Nowadays, almost all professional soccer teams at the highest standards of international soccer monitor workload [5]. Most of these clubs track each field-training session and every game [5], which is important to understand the total load of a longer period [15, 104]. The total load is usually expressed as a combination of an intensity and a volume parameter [217], for example the time (volume) in a specific HR-zone (intensity), or the distance (volume) in a specific speed zone [5]. It is common practice to collect multiple workload variables concurrently [252], with teams recording between 4-10 variables for training and 0-7 variables for games [5]. Important to note is that for both external and internal load, there is no single or gold standard measure. The external and internal load are rather expressed by a myriad of variables describing the activity [130]. Consequently, different monitoring approaches are applied by clubs [5].

When monitoring workload, the players' individual characteristics should be considered, besides the external and internal load [129, 130]. Examples of individual characteristics are genetic factors and previous training experience [130], or aspects of physical fitness such as aerobic and anaerobic power, muscle strength, flexibility and agility [232]. To aid the interpretation of workload, practitioners may employ discrete physiological, physical, or psychological assessments to detect acute and chronic differences in training responses [5]. Assessments may include blood or saliva analysis, monitoring of the autonomic nervous system function using heart-rate indices, various maximal and submaximal performance measures, and subjective athlete self-report measures [5]. The latter is often monitored in practice, using psychometric questionnaires, usually comprised of 3 to 5 questions on a 5-7 point scale [171, 178, 236, 237]. Differences in the general individual characteristics, combined with the day-to-day variations in the general status of the players, could uncover inter-individual differences in the players' physiological and biomechanical responses to external workload [130]. The process of continuous monitoring of external and internal load, with consideration of the inter-individual differences between players [130], could lead to better insights into the actual workload and potentially result in changes to training prescription in subsequent sessions (i.e., increase or decrease workload) [135].

It is apparent that workload should be viewed in relation to previous and future (expected) workloads. To avoid negative training effects, sufficient variation in the magnitude of workload is important, which can be expressed as *training monotony* (Equation 4) [91]. The product of workload and training monotony expresses the *training strain* (Equation 5), which can also be used to gain more insight into the variation of workloads, as high workloads combined with high monotony are associated with negative training effects [91].

$$\text{Training Monotony} = \text{Mean daily workload} \,/\, \text{Standard Deviation of weekly workload} \tag{4}$$

$$\text{Training Strain} = \text{Weekly workload} * \text{Training Monotony} \tag{5}$$

A topical method on the interrelation between 'acute' workloads and 'chronic' workload is expressed by the Acute/Chronic Workload Ratio (ACWR). One of the first studies on the Acute/Chronic Workload Ratio (ACWR) was published in 2015 by Hulin et al. [126]. The ACWR was applied as the ratio between the acute workload, the total load accumulated over the last 7 days, and the chronic workload, the total load accumulated over the last 28 days. The ACWR is based on the fitness-fatigue concept by Banister et al. [17], in which 'fatigue' is represented by the acute workload and 'fitness' by the chronic workload [104]. It has been suggested that spikes in workload (a ratio greater than 1.5) are associated with injury [127] and that practitioners should aim to maintain a ACWR between approximately 0.8 and 1.3 [104]. ACWR has however been criticized and/or limitations/challenges have been stated. It has been suggested that use of rolling averages is problematic, as it assigns the same *weight* to a session carried out the day before the analysis and a session occurring 4 weeks prior [181]. The use of exponentially weighted moving averages has been proposed as an alternative to rolling averages, assigning decreasing weightings for each older workload value [254]. It is nevertheless suggested that this method does not compensates for a different problem with the ACWR, namely the lack of a physiological and biomechanical rationale behind the acute and chronic load time windows [131]. Furthermore, the mathematical coupling of the numerator and denominator has been stated as a limitation [166]. Also, in practice, challenges exists, such as the integration of data from different systems, missing data and a lack of data during the off-season [37]. There is however interest in the use of the ACWR by academics, given the number of studies (>130) produced on this topic [131], and the use of acute and chronic workloads by practitioners [5]. Despite its limitations, it could still be useful to monitor how recent workload relate to historic workloads [255], whether this is done using the ACWR or using different methods.

There is consensus over the fact that soccer is an activity that alternates low- and high-intensity efforts, but not on how to define low- or high-intensity efforts or how to demarcate different intensity zones, although this is an essential aspect of the analysis of workload. For locomotor activities, there are no consistent definitions for speed thresholds [77]. Often, different absolute speed zones, based on meters per seconds or kilometres an hour, are used to subdivide the total training volume [5]. These thresholds are usually however not physiologically justified [2]. Furthermore, the demarcation points in which exercise transitions between the low(er)- and high(er)-intensity domains differs between individuals [2]. There have been proposals to individualize speed thresholds [1, 2], based on peak speed or speed at the VO2max [1]. The physiologically best justified demarcations points may however be the first and second lactate/ventilatory threshold [30]. The first and second lactate/ventilatory threshold, also often referred to as respectively the aerobic and anaerobic threshold [223], subdivide the exercise domain into three separate windows. Exercise at an intensity below the first lactate/ventilatory threshold (*low intensity*) does not lead to a significant increase of the blood lactate concentration, and the oxygen consumption is linearly related to the CO_2-output and ventilation. At an intensity between the first and second lactate/ventilatory threshold (*moderate intensity*), the lactate production is higher than the metabolizing capacity in the muscle cells, which results in an increase in the blood lactate concentration. Both the production of CO_2 and the ventilation increases. The total oxidative capacity of the system, including non-working muscles, liver and ventricular muscle mass, is still sufficiently high to cope with the incoming lactate. The second lactate/ventilatory threshold is attained at the equilibrium of lactate production and elimination. An exercise intensity above the second lactate/ventilatory threshold threshold (*high intensity*) results in a muscular lactate production which exceeds the total lactate elimination rate. The associated increase in H^+ leads to a further increase in the ventilation [20]. The specific physiological effects of surpassing an exercise intensity provides a solid physiological justification for its use to demarcate exercise volume.

In endurance sports, the use of the first and second lactate/ventilatory threshold to demarcate the exercise volume in the low-, moderate- and high-intensity domain is common and has been abundantly described [228]. The distribution of volume over the different intensity domains can be defined as the Training Intensity Distribution (TID). In endurance sports, HR is often used as a marker of exercise intensity, while time is used as a marker of volume [228]. Three major TID-models have been described in the scientific literature with regard to TID in elite endurance athletes [30] (graphically illustrated in Figure 14):

1. The polarized model (POL), emphasizing a large volume of low intensity training (>=75%), and sequentially increasing proportions of volume at moderate and high intensity.
2. The pyramidal model (PYR), with a large volume of training at low intensity (>=70%), combined with sequentially decreasing proportions of volume at moderate and high intensity.
3. The threshold model (THR), with a large of volume spent at moderate intensity (>=40%).

The TID of elite endurance athletes, competing in different endurance sports, can often be categorized as POL or PYR [228]. Thus, elite endurance athletes generally spent a large proportion of the total volume (>=70%) in the low intensity zone, regardless of the seasonal period. This may be considered as paradoxical, as Olympic endurance events are performed at or above the lactate threshold [219]. It should however be noted that large volumes of training at moderate and high intensity can result in negative training effects [83]. For endurance athletes, it is speculated that a large volume of training (>=70%) should be performed at low intensity, combined with small to medium volumes of moderate and high intensity training [30]. The effectiveness of a TID with a large proportion of volume spent at low intensity

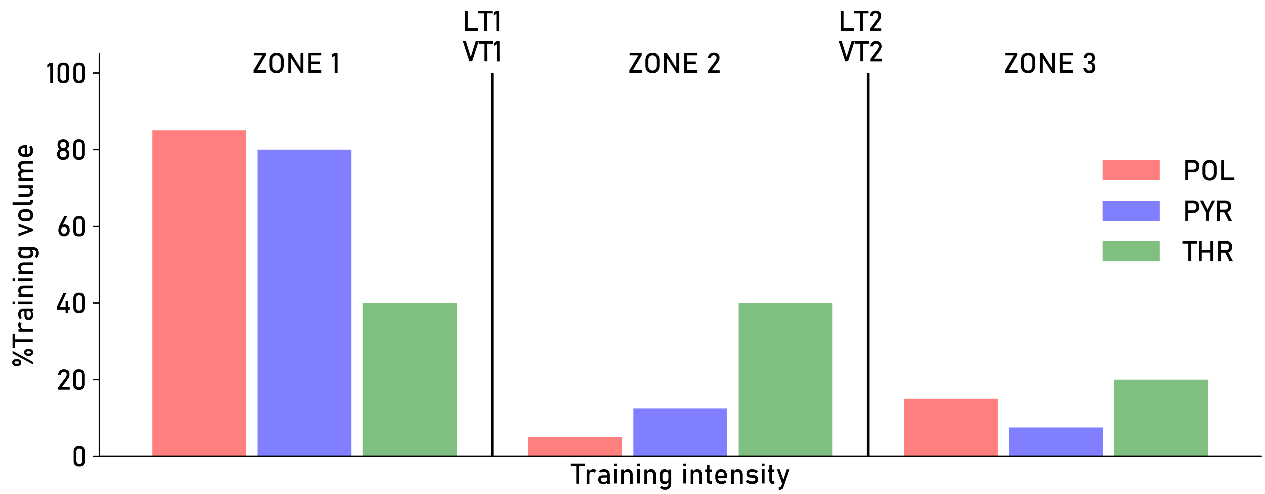


Figure 14: An illustration on the 3 major TID-models, based on a figure by Bourgois et al. [30].

can be explained from an evolutionary and physiological perspective, as described by Bourgois et al. [30]. The genetic make-up of the Homo Sapiens has not changed much over the last 10000 years. From a genetic point of view, humans fit an exercise pattern that combines prolonged low-intensity activities combined with short bursts of high-intensity activities. This has had implications for the physiological responses to training. It has been stated that the dose-response characteristics of high intensity training saturates at fairly low levels of volume in highly trained athletes, and specifically induces more functional cell- and organ-specific adaptions. Moreover, exceeding the 'optimum' volume of moderate- and high-intensity exercise might even generate negative training effects [54, 83]. Training at lower intensities, on the other hand, induces and synchronizes structural adaptations, and counteracts possible negative training effects induced by training at moderate and high intensities [30].

In contrast to endurance sports, the use of the *traditional* TID in soccer has received limited scientific interest, despite its physiological justification and the different possible methods to describe TID in soccer. The first and second lactate/ventilatory thresholds are attained at a specific exercise intensity, which can be expressed at a specific sRPE-rating [7], HR [20] or speed [2]. The first studies on TID in soccer used HR and training duration as markers of respectively exercise intensity and volume, in line with studies in endurance sports. Algrøy et al. [7] also used the sRPE as a marker of intensity. sRPE-ratings from 0-4, 5-7 and 8-10 were respectively categorized as low-, moderate- and high-intensity. The use of the sRPE may however be problematic, as sessions may contain mixed amounts of intensity [158], as well as twice accounting for the duration of training [29, 84]. A recent study by Lee and Mukherjee [159] used speed and distance as a marker of training intensity and volume. Practitioners in soccer often use combinations of speed and distance to monitor workload, which can be considered as alternative markers of respectively training intensity and volume [5]. An overview on the previous studies describing TID in soccer is shown in Figure 15. It is demonstrated that a large proportion of the workload in soccer is performed at lower intensity.

Conclusively, contemporary soccer subjects players to congested schedules of demanding games, often accompanied with fatiguing (inter)national travel activities. To maximise their performance and thus the probability of favourable game outcomes, players require a multitude of fitness components at high level. Between games, training sessions are performed to improve tactical, technical, physical and mental abilities, but these sessions need to be considered in relation to the upcoming games. This poses significant challenges for practitioners. From a physical viewpoint, both inadequate and excessive workloads can lead to negative training effects, such as injuries, decreased fitness and performance [104]. Monitoring of workload is done with the aim of making evidence-based decisions on appropriate loading schemes to enhance team performance and prevent negative training effects [5].

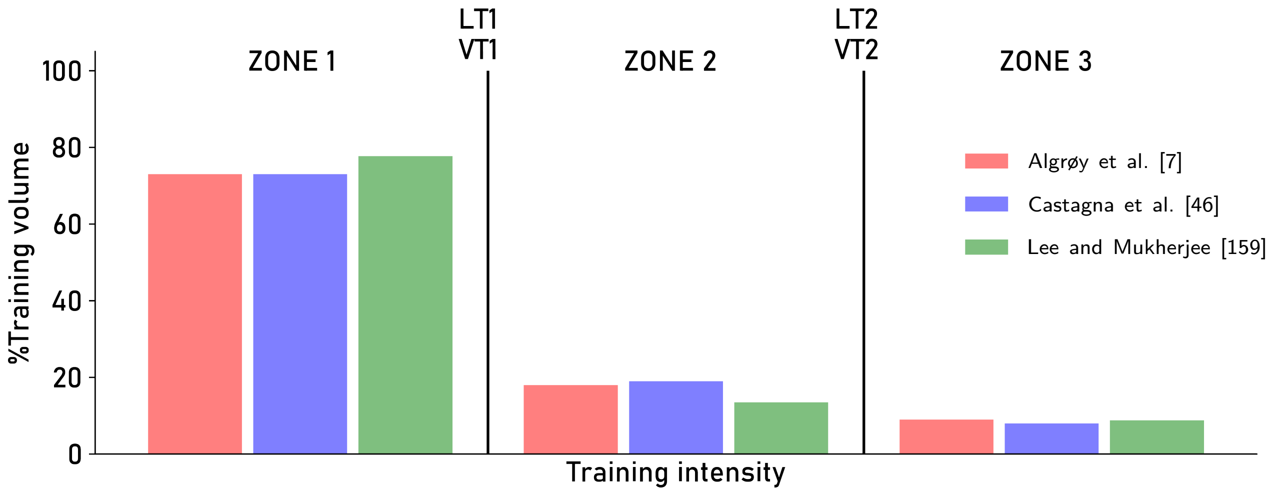


Figure 15: Overview on previous studies describing TID in soccer. To express respectively exercise intensity and volume, Algrøy et al. [7] and Castagna et al. [46] both used HR and training duration; Lee and Mukherjee [159] used a combination of speed and distance.

Objectives and outline of the thesis

To be successful, clubs attempt to provide the best conditions for good performances. Games are analysed to find ways to optimally perform during games, and training sessions are extensively planned and monitored to prepare players for any game demands within reason. So although the topics within this thesis, i.e. performance, training and monitoring, are all quite broad, they serve the same purpose, namely improving team performance, depicted in Figure 16. Another similarity is that the amounts of data on these topics, as a result of advancements in technology, are ever increasing. This wealth of data may however be overwhelming for the stakeholders within the club, as they cannot see *the forest for the trees*. This thesis aims to provide tools and context for practitioners to condense the available data to useful information.

Study I - Performance - What are the strongest predictive variables of winning and losing in Belgian professional soccer? The purpose of the study was to identify the strongest predictive variables of winning and losing in soccer using a broad range of variables, in order to reduce the number of metrics into a manageable quantity. During games, both teams attempt to score, so it was hypothesized that shot-related variables are amongst the strongest predictors of soccer.

Study II - Training - How is training structured in a professional soccer team? And how is training intensity distributed and can it best be quantified? The purpose of this study was twofold. First, this study aimed to extensively describe the preseason period in a professional soccer team, as there is surprisingly little detailed information on how teams structure this important period of the season. The second purpose concerns the TID during this phase of the season using different markers of intensity and volume. It was hypothesized that there are differences between the combinations of intensity (HR and speed) and volume markers (time and distance), showing different proportions of volume spent at low, moderate and high intensities.

Study III - Monitoring - What are the strongest predictive variables of the session Rate of Perceived Exertion in soccer? The purpose of this study was to construct an accurate model predicting the session Rate of Perceived Exertion, based on the external load, HR-data, players' individual characteristics and contextual training information within a first division Belgian soccer team, and subsequently, determine the best predictors. It was hypothesized that external load variables are the most important predictors of the sRPE.

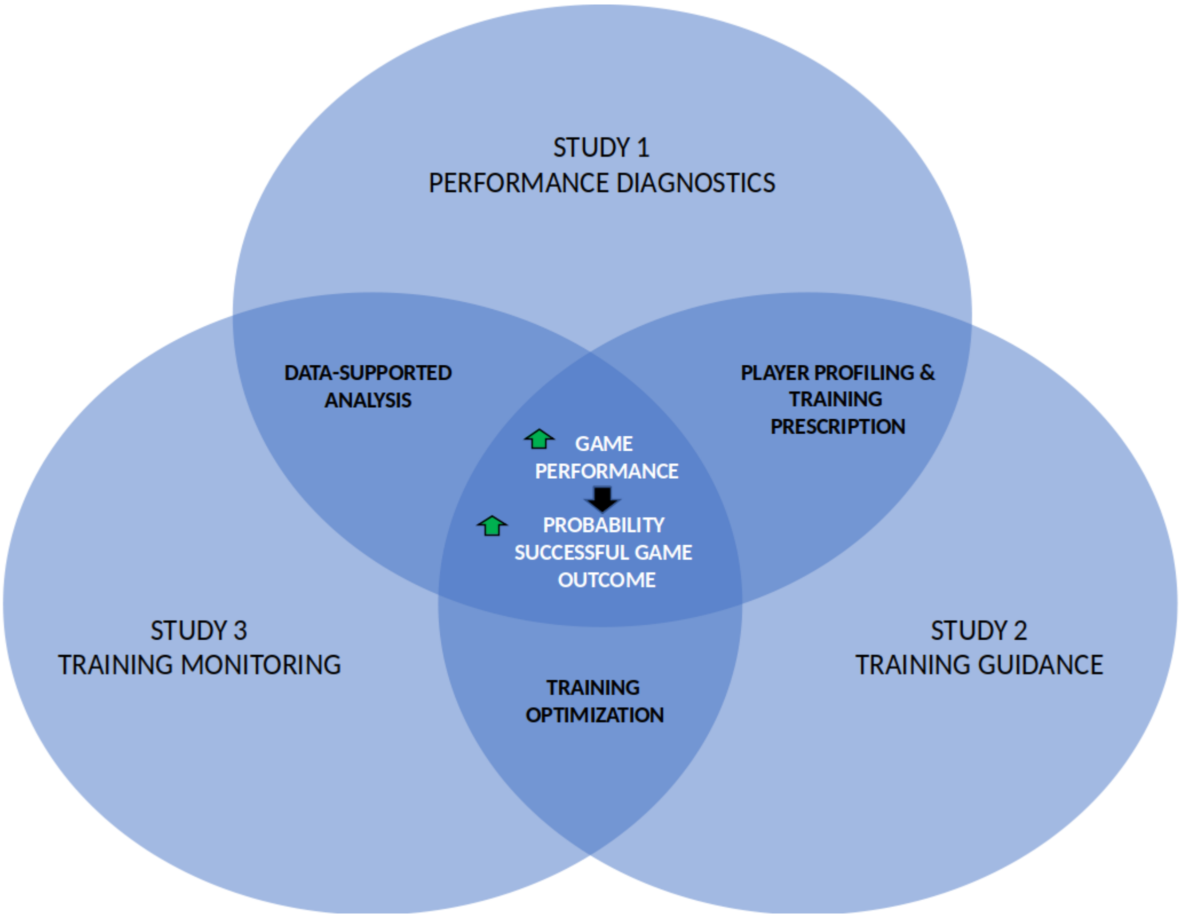


Figure 16: Graphical illustration on the overlap between the studies conducted within this thesis .

References

[1] Abbott W, Brickley G, and Smeeton NJ. An individual approach to monitoring locomotive training load in English Premier League academy soccer players. *International Journal of Sports Science and Coaching*, 13(3):421–428, 2018.

[2] Abt G and Lovell R. The use of individualized speed and intensity thresholds for determining the distance run at high-intensity in professional soccer. *Journal of Sports Sciences*, 27(9):893–898, 2009.

[3] Achten J and Jeukendrup AE. Heart rate monitoring: Applications and limitations. *Sports Medicine*, 33(7):517–538, 2003.

[4] Adams I. A Game for Christmas? The Argylls, Saxons and Football on the Western Front, December 1914. *International Journal of the History of Sport*, 32(11-12):1395–1415, 2015.

[5] Akenhead R and Nassis GP. Training load and player monitoring in high-level football: Current practice and perceptions. *International Journal of Sports Physiology and Performance*, 11(5):587–593, 2016.

[6] Akenhead R, French D, Thompson KG, and Hayes PR. The acceleration dependent validity and reliability of 10Hz GPS. *Journal of Science and Medicine in Sport*, 17(5):562–566, 2014.

[7] Algrøy EA, Hetlelid KJ, Seiler S, and Pedersen JI. Quantifying training intensity distribution in a group of norwegian professional soccer players. *International Journal of Sports Physiology and Performance*, 6(1):70–81, 2011.

[8] Ament W and Verkerke GJ. Exercise and Fatigue. *Educational Psychology*, 39(5):273–282, 2020.

[9] Arnason A, Sigurdsson SB, Gudmundsson A, Holme I, Engebretsen L, and Bahr R. Physical Fitness, Injuries, and Team Performance in Soccer. *Medicine and Science in Sports and Exercise*, 36(2):278–285, 2004.

[10] Atikovic A. Anthropometric characteristics of olympic female and male artistic gymnasts from 1996 to 2016. *International Journal of Morphology*, 38(4):990–996, 2020.

[11] Aughey RJ. Applications of GPS technologies to field sports. *International Journal of Sports Physiology and Performance*, 6(3):295–310, 2011.

[12] Aus Der Fünten K, Faude O, Lensch J, and Meyer T. Injury characteristics in the German professional male soccer leagues after a shortened winter break. *Journal of Athletic Training*, 49(6):786–793, 2014.

[13] Balsom PD, Seger JY, Sjodin B, and Ekblom B. Maximal-Intensity Intermittent Exercise - Effect of Recovery Duration. *International Journal of Sports Medicine*, 13(7):528–533, 1992.

[14] Bangsbo J. Energy demands in competitive soccer. *Journal of Sports Sciences*, 12(1):5–12, 1994.

[15] Bangsbo J, Mohr M, and Krustrup P. Physical and metabolic demands of training and match-play in the elite football player. *Nutrition and Football: The FIFA/FMARC Consensus on Sports Nutrition*, 0414:1–18, 2006.

[16] Bangsbo J, Iaia M, and Krustrup P. Metabolic response and fatigue in soccer. *International journal of sports physiology and performance*, 2(2):111–127, 2007.

[17] Banister E, Calvert T, Savage M, and Bach T. A systems model of training for athletic performance. *Aust J Sports Med*, SMC-6(2):94–102, 1975.

[18] Barnes C, Archer DT, Hogg B, Bush M, and Bradley PS. The evolution of physical and technical performance parameters in the english premier league. *International Journal of Sports Medicine*, 35(13):1095–1100, 2014.

[19] Bialkowski A, Lucey P, Carr P, Yue Y, Sridharan S, and Matthews I. Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 725–730, Washington DC, United States, 2014. IEEE.

[20] Binder RK, Wonisch M, Corra U, Cohen-Solal A, Vanhees L, Saner H, and Schmid JP. Methodological approach to the first and second lactate threshold in incremental cardiopulmonary exercise testing. *European Journal of Preventive Cardiology*, 15(6):726–734, 2008.

[21] Bishop DJ. An applied research model for the sport sciences. *Sports Medicine*, 38(3):253–263, 2008.

[22] Bishop DJ, Girard O, and Mendez-Villanueva A. Repeated-sprint ability part II: Recommendations for training. *Sports Medicine*, 41(9):741–756, 2011.

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

[24] Boone J, Vaeyens R, Steyaert A, Vanden Bossche L, and Bourgois JG. Physical Fitness of Elite Belgian Soccer Players by Player Position. *Journal of Strength and Conditioning Research*, 26(8):2051–2057, 2012.

[25] Borg GA. A simple rating scale for use in physical work test. *Kungliga fysiografiska sallskapet i Lund*, 32:7–15, 1962.

[26] Borg GA. Psychophysical bases of perceived exertion, 1982.

[27] Borg GAV. Perceived Exertion as an Indicator of Somatic Stress. *Scandinavian Journal of Rehabilitation Medicine*, 2 (3):92–98, 1970.

[28] Borg GAV. Borg perceived exertion and pain scales., 1998.

[29] Bourdon PC, Cardinale M, Murray A, Gastin PB, Kellmann M, Varley MC, Gabbett TJ, Coutts AJ, Burgess DJ, Gregson W, and Cable NT. Monitoring athlete training loads: Consensus statement. *International Journal of Sports Physiology and Performance*, 12(2):161–170, 2017.

[30] Bourgois JG, Bourgois G, and Boone J. Perspectives and Determinants for Training-Intensity Distribution in Elite Endurance Athletes. *International Journal of Sports Physiology and Performance*, 14(8):1151–1156, 2019.

[31] Bradley PS, O'Donoghue P, Wooster B, Tordoff P, O'Donoghue P, Wooster B, and Tordoff P. The reliability of ProZone MatchViewer: a videobased technical performance analysis system. *International Journal of Performance Analysis in Sport*, 7(3):117–129, 2007.

[32] Bradley PS, Di Mascio M, Peart D, Olsen P, and Sheldon B. High-Intensity Activity Profiles of Elite Soccer Players at Different Performance Levels. *Journal of Strength and Conditioning Research*, 24(9):2343–2351, 2010.

[33] Bradley PS, Archer DT, Hogg B, Schuth G, Bush M, Carling C, and Barnes C. Tier-specific evolution of match performance characteristics in the English Premier League: it's getting tougher at the top. *Journal of Sports Sciences*, 34(10):980–987, 2016.

[34] Breiman L. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[35] Brink MS, Kuyvenhoven JP, Toering T, Jordet G, and Frencken WGP. What do football coaches want from sport science? *Kinesiology*, 50:150–154, 2018.

[36] Broich H, Mester J, Seifriz F, and Yue Z. Statistical Analysis for the First Bundesliga in the Current Soccer Season. *Progress in Applied Mathematics*, 7(2):1–8, 2011.

[37] Buchheit M. Applying the acute: chronic workload ratio in elite football: worth the effort? *British Journal of Sports Medicine*, 51(18):bjsports–2016–097017, 2016.

[38] Buchheit M, Manouvrier C, Cassirame J, and Morin JB. Monitoring locomotor load in soccer: Is metabolic power, powerful? *International Journal of Sports Medicine*, 36(14):1149–1155, 2015.

[39] Burgess DJ. 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(2):136–141, 2017.

[40] Bush M, Barnes C, Archer DT, Hogg B, and Bradley PS. Evolution of match performance parameters for various playing positions in the English Premier League. *Human movement science*, 39:1–11, 2015.

[41] Bzdok D, Altman N, and Krzywinski M. Points of Significance: Statistics versus machine learning. *Nature Methods*, 15(4):233–234, 2018.

[42] Cardinale M and Varley MC. Wearable Training-Monitoring Technology : Applications , Challenges , and Opportunities. *International Journal of Sports Physiology and Performance*, 12(2):55–62, 2017.

[43] Carling C, Williams AM, and Reilly T. *Handbook of soccer match analysis: A systematic approach to improving performance*. Psychology Press, 2005.

[44] Carling C, Bloomfield J, Nelson LJ, and Reilly T. The role of motion analysis in elite soccer: Contemporary performance measurement techniques and work rate data. *Sports Medicine*, 38(10):839–862, 2008.

[45] Carling C, Le Gall F, and Dupont G. Are physical performance and injury risk in a professional soccer team in match-play affected over a prolonged period of fixture congestion? *International journal of sports medicine*, 33(1):36–42, 2012.

[46] Castagna C, Impellizzeri FM, Chaouachi A, Bordon C, and Manzi V. Effect of Training Intensity Distribution on Aerobic Fitness Variables in Elite Soccer Players: A Case Study. *Journal of Strength and Conditioning Research*, 25(1):66–71, 2011.

[47] Castagna C, Impellizzeri FM, Chaouachi A, and Manzi V. Preseason variations in aerobic fitness and performance in elite-standard soccer players: A team study. *Journal of Strength and Conditioning Research*, 27(11):2959–2965, 2013.

[48] Castellano J, Casamichana D, and Lago C. The Use of Match Statistics that Discriminate Between Successful and Unsuccessful Soccer Teams. *Journal of Human Kinetics*, 31(1):139–147, 2012.

[49] Chambers R, Gabbett TJ, Cole MH, and Beard A. The Use of Wearable Microsensors to Quantify Sport-Specific Movements. *Sports Medicine*, 45(7):1065–1081, 2015.

[50] Chen T and Guestrin C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[51] Chirinos EA. National Identity and Sports in Latin America. *Mapping Politics*, 9:19–27, 2018.

[52] Chmura P, Konefał M, Chmura J, Kowalczuk E, Zajac T, Rokita A, Andrzejewski M, Chmura Pawełand Konefał M, Chmura J, Kowalczuk E, Zajac T, Rokita A, and Andrzejewski M. Match outcome and running performance in different intensity ranges among elite soccer players. *Biology of Sport*, 35(2):197–203, 2018.

[53] Christen J, Foster C, Porcari JP, and Mikat RP. Temporal Robustness of the Session Rating of Perceived Exertion. *International Journal of Sports Physiology and Performance*, 11(8):1088–1093, 2016.

[54] Chwalbińska-Moneta J, Kaciuba-Uściłko H, Krysztofiak H, Ziemba A, Krzemiński K, Kruk B, and Nazar K. Relationship between EMG, blood lactate, and plasma catecholamine thresholds during graded exercise in men. *Journal of Physiology and Pharmacology*, 49(3):433–441, 1998.

[55] Clemente FM, Seerden G, and van der Linden CMI. Quantifying the physical loading of five weeks of pre-season training in professional soccer teams from Dutch and Portuguese leagues. *Physiology and Behavior*, 209(June), 2019.

[56] Clemente JA, Requena, Jukic I, Nayler, Hernández, and Carling C. Is Physical Performance a Differentiating Element between More or Less Successful Football Teams? *Sports*, 7(10):216, 2019.

[57] Collet C. The possession game? A comparative analysis of ball retention and team success in European and international football, 2007-2010. *Journal of Sports Sciences*, 31(2):123–136, 2013.

[58] Costill DL, Daniels J, Evans W, Fink W, Krahenbuhl G, and Saltin B. Skeletal-muscle enzymes and fiber composition in male and female track athletes. *Journal of Applied Physiology*, 40(2):149–154, 1976.

[59] Coutts AJ. In the Age of Technology, Occam's Razor Still Applies. *International Journal of Sports Physiology and Performance*, 9(6):741, 2014.

[60] Cummins C, Orr R, O'Connor H, and West C. Global Positioning Systems (GPS) and Microtechnology Sensors in Team Sports: A Systematic Review. *Sports Medicine*, 43(10):1025–1042, 2013.

[61] Dalen T, Jørgen I, Gertjan E, Havard HG, and Ulrik W. Player load, acceleration, and deceleration during forty-five competitive matches of elite soccer. *Journal of Strength and Conditioning Research*, 30(2):351–359, 2016.

[62] De Sousa Júnior SF, De A. Araújo A, and Menotti D. An overview of automatic event detection in soccer matches. In *2011 IEEE Workshop on Applications of Computer Vision, WACV 2011*, number December 2013, pages 31–38, 2011.

[63] De Souza DB, González-García J, Campo RLD, Resta R, Buldú JM, Wilk M, and Coso JD. Players' physical performance in LaLiga across the season: Insights for competition continuation after COVID-19. *Biology of Sport*, 38(1):3–7, 2021.

[64] Decroos T, Van Haaren J, Bransen L, and Davis JJ. Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, volume 25, pages 1851–1861, Anchorage, AK, USA, 2019.

[65] Delgado-Bordonau J and Mendez-Villanueva A. Tactical Periodization. *Fitness in Soccer: The Science and Practical Application*, 2012.

[66] Dellal A, Wong DP, Moalla W, and Chamari K. Physical and technical activity of soccer players in the French first league- with special reference to their playing position. *International SportMed Journal*, 11(2), 2010.

[67] Dellal A, Chamari K, Wong DP, Ahmaidi S, Keller D, Barros RML, Bisciotti GN, and Carling C. Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European Journal of Sport Science*, 11(1):51–59, 2011.

[68] Dellal A, Silva CDD, Hill-Haas S, Wong DP, Natali AJ, De Lima JRP, Filho MG, Marins JJCB, Garcia ES, and Karim C. Heart ratemonitoring in soccer: Interest and limits during competitive match play and training, practical application, 2012.

[69] Di Prampero PE, Fusi S, Sepulcri L, Morin JB, Belli A, and Antonutto G. Sprint running: A new energetic approach. *Journal of Experimental Biology*, 208(14):2809–2816, 2005.

[70] Di Salvo V, Gregson W, Atkinson G, Tordoff P, and Drust B. Analysis of high intensity activity in premier league soccer. *International Journal of Sports Medicine*, 30(3):205–212, 2009.

[71] Di Salvo V, Baron R, Tschan H, Calderon Montero FJ, Bachl N, and Pigozzi F. Performance characteristics according to playing position in elite soccer. *International Journal of Sports Medicine*, 28(3):222–227, 2007.

[72] Dolci F, Hart NH, Kilding AE, Chivers P, Piggott B, and Spiteri T. Physical and Energetic Demand of Soccer. *Strength & Conditioning Journal*, 42(3):70–77, 2020.

[73] Doubt TJ. Physiology of Exercise in the Cold. *Sports Medicine*, 11(6):367–381, 1991.

[74] Drew MK and Finch CF. The Relationship Between Training Load and Injury, Illness and Soreness: A Systematic and Literature Review. *Sports Medicine*, 46:861–883, 2016.

[75] Dubitzky W, Lopes P, Davis JJ, and Berrar D. The Open International Soccer Database for machine learning. *Machine Learning*, 108(1):9–28, 2019.

[76] Dvorak J and Racinais S. Training and playing football in hot environments. *Scandinavian Journal of Medicine and Science in Sports*, 20(suppl. 3):4–5, 2010.

[77] Dwyer DB and Gabbett TJ. Global positioning system data analysis: Velocity ranges and a new definition of sprinting for field sport athletes. *Journal of Strength and Conditioning Research*, 26(3):818–824, 2012.

[78] Ekin A, Tekalp AM, and Mehrotra R. Automatic soccer video analysis and summarization. *IEEE Transactions on Image Processing*, 12(7):796–807, 2003.

[79] Ekstrand J, Spreco A, and Davison M. Elite football teams that do not have a winter break lose on average 303 player-days more per season to injuries than those teams that do: A comparison among 35 professional European teams. *British Journal of Sports Medicine*, pages 1231–1235, 2018.

[80] Ekstrand J, Spreco A, Windt J, and Khan KM. Are Elite Soccer Teams ' Preseason Training Sessions Associated With Fewer In-Season Injuries ? A 15-Year Analysis From the Union of European Football Associations (UEFA) Elite Club Injury Study. *The American Journal of Sports Medicine*, 48(3):723–729, 2020.

[81] Eliakim E, Doron O, Meckel Y, Nemet D, and Eliakim A. Pre-season Fitness Level and Injury Rate in Professional Soccer – A Prospective Study. *Sports Medicine International Open*, 02(03):E84–E90, 2018.

[82] Eniseler N, Nursing P, and Source AH. Heart Rate and Blood Lactate Concentrations As Predictors of Physiological Load on Elite Soccer Players during Various Soccer Training Activities. *Journal of strength and conditioning research / National Strength & Conditioning Association*, 19(4):799–804, 2005.

[83] Esteve-Lanao J, Foster C, Seiler S, and Lucia A. Impact of training intensity distribution on performance in endurance athletes. *Journal of Strength and Conditioning Research*, 21(3):943–949, 2007.

[84] Eston R. Use of Ratings of Perceived Exertion in Sports. *International Journal of Sports Physiology and Performance*, 7:175–182, 2012.

[85] Fanchini M, Ghielmetti R, Coutts AJ, Schena F, and Impellizzeri FM. Effect of training-session intensity distribution on session rating of perceived exertion in soccer players. *International Journal of Sports Physiology and Performance*, 10 (4):426–430, 2015.

[86] Felipe JL, Garcia-Unanue J, Viejo-Romero D, Navandar A, and Sánchez-Sánchez J. Validation of a video-based performance analysis system (Mediacoach®) to analyze the physical demands during matches in LaLiga. *Sensors (Switzerland)*, 19(19):4–13, 2019.

[87] Fernandez-Navarro J, Fradua L, Zubillaga A, and McRobert AP. Influence of contextual variables on styles of play in soccer. *International Journal of Performance Analysis in Sport*, 18(3):423–436, 2018.

[88] FIFA. FIFA and IFAB to develop global standard for electronic performance and tracking systems, 2015. URL `https://www.fifa.com/who-we-are/news/fifa-and-ifab-to-develop-global-standard-for-electronic-performance-an-2709918`.

[89] FIFA. More than half the world watched record-breaking 2018 World Cup, 2018. URL `https://www.fifa.com/worldcup/news/more-than-half-the-world-watched-record-breaking-2018-world-cup`.

[90] Football Association. Laws of the Game, 1863. URL `https://the18.com/soccer-entertainment/5-fun-rules-1863-fifa-needs-consider-bringing-back`.

[91] Foster C. Monitoring training in athletes with reference to overtraining syndrome. *Medicine and Science in Sports and Exercise*, 30(7):1164–1168, 1998.

[92] Foster C. Back to the future. *International Journal of Sports Physiology and Performance*, 11(1):281–282, 2016.

[93] Foster C, Hector LL, Welsh R, Schrager M, Green MA, and Snyder AC. Effects of specific versus cross-training on running performance. *European Journal of Applied Physiology and Occupational Physiology*, 70(4):367–372, 1995.

[94] Foster C, FLorhaug JA, Franklin J, Gottschall L, Hrovatin LA, Parker S, Doleshal P, and Dodge C. A New Approach to Monitoring Exercise Training. *Journal of Strength and Conditioning Research*, 15(1):109–115, 2001.

[95] Foster C, Rodríguez-Marroyo JA, and De Koning JJ. Monitoring training loads: The past, the present, and the future. *International Journal of Sports Physiology and Performance*, 12(May):2–8, 2017.

[96] Foster C, Boullosa DA, McGuigan MR, Fusco A, Cortis C, Arney BE, Orton B, Dodge C, Jaime S, Radtke K, van Erp T, de Koning JJ, Bok D, Rodríguez-Marroyo JA, and Porcari JP. 25 Years of Session Rating of Perceived Exertion: Historical Perspective and Development. *International Journal of Sports Physiology and Performance*, pages 1–10, 2021.

[97] Francioni FM, Capranica L, Tessitore A, Conte D, Figueiredo AJ, and Lupo C. Preseason Strategies of Italian First League Soccer Clubs in Relation to their Championship Ranking: A Five-Year Analysis. *Journal of Human Kinetics*, 50 (1):145–155, 2016.

[98] Frencken WG, de Poel H, Lemmink KA, Delleman NJ, de Poel H, and Lemmink KA. Analysis of game dynamics and related game events in 11 vs 11 soccer. In *7th World Congress on Science & football*, page 102. 7th World Congress on Science & Football, 2011.

[99] Frencken WG, Lemmink KA, Delleman NJ, and Visscher C. Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science*, 11(4):215–223, 2011.

[100] Frencken WGP, Lemmink KAPM, and Delleman NJ. Soccer-specific accuracy and validity of the local position measurement (LPM) system. *Journal of Science and Medicine in Sport*, 13(6):641–645, 2010.

[101] Frencken WGP, de Poel H, Visscher C, Lemmink KAPM, de Poel H, Visscher C, Lemmink KAPM, de Poel H, Visscher C, and Lemmink KAPM. Variability of inter-team distances associated with match events in elite-standard soccer. *Journal of Sports Sciences*, 30(12):1207–1213, 2012.

[102] Gabbett TJ and Whiteley R. Two training-load paradoxes: Can we work harder and smarter, can physical preparation and medical be teammates?, 2017.

[103] Gabbett TJ, Ekstrand J, Whiteley R, King E, Milsom J, Sheehan J, Hawkins RD, and Kennelly S. If overuse injury is a 'training load error', should undertraining be viewed the same way? *British Journal of Sports Medicine*, 50(17): 1017–1018, 2016.

[104] Gabbett TJ, Hulin BT, Blanch P, and Whiteley R. High training workloads alone do not cause sports injuries: How you get there is the real issue. *British Journal of Sports Medicine*, 50(8):444–445, 2016.

[105] Gamble P. Periodization of Training for Team Sports Athletes. *Strength and Conditioning Journal*, 28(5):56–66, 2006.

[106] Gaudino P, Iaia M, Alberti G, Strudwick AJ, Atkinson G, and Gregson W. Monitoring training in elite soccer players: Systematic bias between running speed and metabolic power data. *International Journal of Sports Medicine*, 34(11): 963–968, 2013.

[107] Georgievski B, Labadze L, and Aboelsoud ME. Comparative advantage as a success factor in football clubs: Evidence from the English Premier League (EPL). *Journal of Human Sport and Exercise*, 14(2):292–314, 2019.

[108] Girard O, Mendez-Villanueva A, and Bishop DJ. Repeated-Sprint Ability – Part I. *Sports Medicine*, 41(8):673–694, 2011.

[109] Glaister M. Multiple Sprint Work. *Sports Medicine*, 35(9):757–777, 2005.

[110] Gonçalves BV, Figueira BE, Maçãs V, and Sampaio JE. Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. *Journal of Sports Sciences*, 32(2):191–199, 2014.

[111] Gréhaigne JF and Godbout P. Tactical knowledge in team sports from a constructivist and cognitivist perspective. *Quest*, 47(4):490–505, 1995.

[112] Groll A, Ley C, Schauberger G, and Van Eetvelde H. A hybrid random forest to predict soccer matches in international tournaments. *Journal of Quantitative Analysis in Sports*, 2019.

[113] Gunnarsson TP, Bendiksen M, Bischoff R, Christensen PM, Lesivig B, Madsen K, Stephens F, Greenhaff P, Krustrup P, and Bangsbo J. Effect of whey protein- and carbohydrate-enriched diet on glycogen resynthesis during the first 48h after a soccer game. *Scandinavian Journal of Medicine and Science in Sports*, 23(4):508–515, 2013.

[114] Haddad M, Stylianides G, Djaoui L, Dellal A, and Chamari K. Session-RPE Method for Training Load Monitoring: Validity, Ecological Usefulness, and Influencing Factors. *Frontiers in Neuroscience*, 11(November), 2017.

[115] Halson SL. Nutrition, sleep and recovery. *European Journal of Sport Science*, 8(2):119–126, 2008.

[116] Hargreaves M. Physiological limits to exercise performance in the heat. *Journal of Science and Medicine in Sport*, 11 (1):66–71, 2008.

[117] Harrop K and Nevill A. Performance indicators that predict success in an English Professional League One Soccer Team. *International Journal of Performance Analysis in Sport*, 14(3):907–920, 2014.

[118] Haugen T, Tønnessen E, Hisdal J, and Seiler S. The Role and Development of Sprinting Speed in Soccer. *International Journal of Sports Physiology and Performance*, 9(1):432–441, 2014.

[119] Hennessy L and Jeffreys I. The current use of GPS, its potential, and limitations in soccer. *Strength and Conditioning Journal*, 40(3):83–94, 2018.

[120] Herold M, Goes FR, Nopp S, Bauer P, Thompson C, and Meyer T. Machine learning in men's professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science and Coaching*, 14(6):798–817, 2019.

[121] Hoff J and Helgerud J. Endurance and Strength Training for Soccer Players. *Sports Medicine*, 34(3):165–180, 2004.

[122] Horvat T and Job J. The use of machine learning in sport outcome prediction: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5):1–28, 2020.

[123] Hucaljuk J and Rakipović A. Predicting football scores using machine learning techniques. *MIPRO 2011 - 34th International Convention on Information and Communication Technology, Electronics and Microelectronics - Proceedings*, 48: 1623–1627, 2011.

[124] Hughes MD and Bartlett RM. The use of performance indicators in performance analysis. *Journal of Sports Sciences*, 20:739–754, 2002.

[125] Hughes MD and James N. Notational analysis of soccer. In *Proceedings of the 5th International Scientific Conference on Kinesiology*, pages 644–660, 2008.

[126] Hulin BT, Gabbett TJ, Blanch P, Chapman P, Bailey D, and Orchard JW. Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *British Journal of Sports Medicine*, 48(8):708–712, 2014.

[127] Hulin BT, Gabbett TJ, Caputi P, Lawson DW, and Sampson JA. Low chronic workload and the acute:Chronic workload ratio are more predictive of injury than between-match recovery time: A two-season prospective cohort study in elite rugby league players. *British Journal of Sports Medicine*, 50(16):1008–1012, 2016.

[128] Hvattum LM and Arntzen H. Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, 26(3):460–470, 2010.

[129] Impellizzeri FM, Rampinini E, and Marcora SM. Physiological assessment of aerobic training in soccer. *Journal of Sports Sciences*, 23(6):583–592, 2005.

[130] Impellizzeri FM, Marcora SM, and Coutts AJ. Internal and external training load: 15 years on. *International Journal of Sports Physiology and Performance*, 14(2):270–273, 2019.

[131] Impellizzeri FM, Tenan MS, Kempton T, Novak A, and Coutts AJ. Acute: Chronic workload ratio: Conceptual issues and fundamental pitfalls. *International Journal of Sports Physiology and Performance*, 15(6):907–913, 2020.

[132] Ishii T, Masuda T, Kurogi H, Yabuno S, and Yamamoto H. The distance covered of soccer and rugby referees during the match using a mobile "GPS". In *International Symposium of Biomechanics in Sports*, pages 322–325, 2002.

[133] Jacobs I, Westlin N, Karlsson J, Rasmusson M, and Houghton B. Muscle glycogen and diet in elite soccer players. *European Journal of Applied Physiology and Occupational Physiology*, 48(3):297–302, 1982.

[134] James N. Notational analysis in soccer: past, present and future. *International Journal of Performance Analysis in Sport*, 6(2):67–81, 2006.

[135] Jaspers A, Brink MS, Probst SG, Frencken WG, and Helsen WF. Relationships Between Training Load Indicators and Training Outcomes in Professional Soccer. *Sports Medicine*, 47(3):533–544, 2017.

[136] Jaworska M. European football at the cross-roads. *Journal of Physical Education and Sport*, 20(2):1188–1194, 2020.

[137] Jeong TS, Reilly T, Morton JP, Bae SW, and Drust B. Quantification of the physiological loading of one week of "pre-season" and one week of "in-season" training in professional soccer players. *Journal of Sports Sciences*, 29(11): 1161–1166, 2011.

[138] Kelly VG and Coutts AJ. Planning and monitoring training loads during the competition phase in team sportd. *Strength and Conditioning Journal*, 29(4):32–37, 2007.

[139] Kempe M, Vogelbein M, Memmert D, and Nopp S. Possession vs. Direct Play: Evaluating Tactical Behavior in Elite Soccer. *International Journal of Sports Science*, 4(6A):35–41, 2014.

[140] Khan A, Lazzerini B, Calabrese G, and Serafini L. Soccer Event Detection. In *Khan, Abdullah, et al. "Soccer event detection." 4th international conference on image processing and pattern recognition (IPPR 2018)*, pages 119–129, 2018.

[141] Khan KM, Thompson AM, Blair SN, Sallis JF, Powell KE, Bull FC, and Bauman AE. Sport and exercise as contributors to the health of nations. *The Lancet*, 380(9836):59–64, 2012.

[142] Knoll J and Stübinger J. Machine-Learning-Based Statistical Arbitrage Football Betting. *KI - Kunstliche Intelligenz*, 34 (1):69–80, 2020.

[143] Komlos J and Kim JH. Estimating trends in historical heights. *Historical Methods*, 23(3):116–120, 1990.

[144] Konefał M, Chmura Paweland Tessitore A, Melcer T, Kowalczuk E, Chmura J, Andrzejewski M, Chmura P, Tessitore A, Melcer T, Kowalczuk E, Chmura J, and Andrzejewski M. The Impact of Match Location and Players' Physical and Technical Activities on Winning in the German Bundesliga. *Frontiers in Psychology*, 11(July):1–9, 2020.

[145] Krustrup P and Krustrup BR. Football is medicine: It is time for patients to play! *British Journal of Sports Medicine*, 52(22):1412–1413, 2018.

[146] Krustrup P, Ortenblad N, Nielsen J, Nybo L, Gunnarsson TP, Iaia M, Madsen K, Stephens F, Greenhaff P, and Bangsbo J. Maximal voluntary contraction force, SR function and glycogen resynthesis during the Wrst 72 h after a high-level competitive soccer game. *European Journal of Applied Physiology*, 111(12):2987–2995, 2011.

[147] Kuo FY and Sloan IH. Lifting the curse of Dimensionality. *Notices of the AMS*, 52(11):1320–1328, 2005.

[148] Lago C. The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *Journal of Sports Sciences*, 27(13):1463–1469, 2009.

[149] Lago-ballesteros J and Lago-Peñas C. Performance in Team Sports: Identifying the Keys to Success in Soccer. *Journal of Human Kinetics*, 25(1):85–91, 2010.

[150] Lago-Peñas C. The role of situational variables in analysing physical performance in soccer. *Journal of Human Kinetics*, 35(1):89–95, 2012.

[151] Lamas L, Barrera J, Otranto G, and Ugrinowitsch C. Invasion team sports: Strategy and match modeling. *International Journal of Performance Analysis in Sport*, 14(1):307–329, 2014.

[152] Lambert MI and Borresen J. A Theoretical Basis of Monitoring Fatigue: A Practical Approach for Coaches. *International Journal of Sports Science & Coaching*, 1(4):371–388, 2006.

[153] Lambert MI, Viljoen W, Bosch A, Pearce AJ, and Sayers M. General principles of trauma. In Schwellnus M, editor, *The Olympic Textbook of Medicine in Sport*, pages 1–48. Blackwell Publishing, 2008.

[154] Lames M. Chance involvement in goal scoring in football – an empirical approach. *German Journal of Exercise and Sport Research*, 48(2):278–286, 2018.

[155] Lames M and McGarry T. On the search for reliable performance indicators in game sports. *International Journal of Performance Analysis in Sport*, 7(1):62–79, 2017.

[156] Larson O. Charles Reep: A Major Influence on British and Norwegian Football. *Soccer & Society*, 2(3):58–78, 2001.

[157] Laukkanen R and Virtanen P. Heart Rate Monitors: State of the Art. *Journal of Sports Sciences*, 16(1):3–7, 1998.

[158] Laursen PB. Training for intense exercise performance: High-intensity or high-volume training? *Scandinavian Journal of Medicine and Science in Sports*, 20(SUPPL. 2):1–10, 2010.

[159] Lee M and Mukherjee S. Relationship of Training Load with High-intensity Running in Professional Soccer Players. *International Journal of Sports Medicine*, 40(5):336–343, 2019.

[160] Lees A and Nolan L. The biomechanics of soccer : A review The biomechanics of soccer : A review. *Journal of Sports Sciences*, 16(3):211–234, 2011.

[161] Lepschy H, Wäsche H, and Woll A. How to be Successful in Football: A Systematic Review. *The Open Sports Sciences Journal*, 11(1):3–23, 2018.

[162] Lepschy H, Wäsche H, and Woll A. Success factors in football: an analysis of the German Bundesliga. *International Journal of Performance Analysis in Sport*, 20(2):150–164, 2020.

[163] Lever J, Krzywinski M, and Altman N. Points of Significance: Model selection and overfitting. *Nature Methods*, 13(9): 703–704, 2016.

[164] Linke D, Link D, and Lames M. Validation of electronic performance and tracking systems EPTS under field conditions. *PLoS ONE*, 13(7):1–19, 2018.

[165] Linke D, Link D, and Lames M. Football-specific validity of TRACAB's optical video tracking systems. *PLoS ONE*, 15 (3):1–17, 2020.

[166] Lolli L, Batterham AM, Hawkins RD, Kelly DM, Strudwick AJ, Thorpe RT, Gregson W, and Atkinson G. Mathematical coupling causes spurious correlation within the conventional acute-to-chronic workload ratio calculations. *British journal of sports medicine*, 53(15):921–922, 2019.

[167] Lundberg SM and Lee SI. A Unified Approach to Interpreting Model Predictions Scott. In *Neural Information Processing Systems*, number 31, pages 4765–4774, Long Beach, CA, USA, 2017.

[168] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, and Lee SI. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67, 2020.

[169] MacKenzie R and Cushion C. Performance analysis in football: A critical review and implications for future research'. *Journal of Sports Sciences*, 32(1):2–7, 2013.

[170] Malone JJ, Di Michele R, Morgans R, Burgess DJ, Morton JP, and Drust B. Seasonal Training-Load Quantification in Elite English Premier League Soccer Players. *International Journal of Sports Physiology and Performance*, 10(4): 489–497, may 2015.

[171] Malone JJ, Frencken WGP, Jaspers A, Brink MS, Helsen WF, and Merks B. Seasonal Training Load and Wellness Monitoring in a Professional Soccer Goalkeeper. *International Journal of Sports Physiology and Performance*, 13(5): 672–675, 2017.

[172] Malone JJ, Lovell R, Varley MC, and Coutts AJ. Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *International Journal of Sports Physiology and Performance*, 12(S2):18–26, 2017.

[173] Manzi V, Bovenzi A, Impellizzeri FM, Carminati I, and Castagna C. Individual training-load and aerobic-fitness variables in premiership soccer players during the precompetitive season. *Journal of Strength and Conditioning Research*, 27(3): 631–636, 2013.

[174] Martín-García A, Gómez Díaz A, Bradley PS, Morera F, and Casamichana D. Quantification of a professional football team's external load using a microcycle structure. *Journal of strength and conditioning research*, 32(12):3511–3518, 2018.

[175] Maughan RJ, Otani H, and Watson P. Influence of relative humidity on prolonged exercise capacity in a warm environment. *European Journal of Applied Physiology*, 112(6):2313–2321, 2012.

[176] Mazzeo RS. Physiological Responses to Exercise at Altitude. *Sports Medicine*, 38(1):1–8, 2008.

[177] McGarry T. Applied and theoretical perspectives of performance analysis in sport: Scientific issues and challenges. *International Journal of Performance Analysis in Sport*, 9(1):128–140, 2009.

[178] Mclean BD, Coutts AJ, Kelly V, McGuigan MR, and Cormack SJ. 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, 2010.

[179] Meckel Y, Ismaeel A, and Eliakim A. The effect of the Ramadan fast on physical performance and dietary habits in adolescent soccer players. *European Journal of Applied Physiology*, 102(6):651–657, 2008.

[180] Memmert D, Lemmink KA, and Sampaio JE. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine*, 47(1):1–10, 2017.

[181] Menaspà P. Are rolling averages a good way to assess training load for injury prevention? *British Journal of Sports Medicine*, 2017.

[182] Morgan WP. Psychological factors influencing perceived exertion. *Medicine and Science in Sports and Exercise*, 5(2): 97–103, 1973.

[183] Morgans R, Orme P, Anderson L, and Drust B. Principles and practices of training for soccer. *Journal of Sport and Health Science*, 3(4):251–257, 2014.

[184] Mujika I and Padilla S. Detraining: Loss of Training-Induced Physiological and Performance Adaptations. Part I. *Sports Medicine*, 30(3):79–87, 2000.

[185] Mujika I, Halson S, Burke LM, Balagué G, and Farrow D. An Integrated, Multifactorial Approach to Periodization for Optimal Performance in Individual and Team Sports. *International Journal of Sports Physiology and Performance*, 13 (1):538–561, 2018.

[186] Natekin A and Knoll A. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7(DEC), 2013.

[187] NCD Risk Factor Collaboration. A century of trends in adult human height. *ELife*, 5(e13410), 2016.

[188] Nédélec M, Halson S, Abaidia AE, Ahmaidi S, and Dupont G. Stress, Sleep and Recovery in Elite Soccer: A Critical Review of the Literature. *Sports Medicine*, 45(10):1387–1400, 2015.

[189] Ohashi J, Togari H, Isokawa M, and Suzuki S. Measuring movement speeds and distances covered during soccer match-play. In Reilly T, Lees A, Davids K, and Murphy W, editors, *Science and Football*, pages 329–333, 1988.

[190] Orejan J. *Football/Soccer: history and tactics*. McFarland, 2011.

[191] Osgnach C, Poser S, Bernardini R, Rinaldo R, and Di Prampero PE. Energy cost and metabolic power in elite soccer: A new match analysis approach. *Medicine and Science in Sports and Exercise*, 42(1):170–178, 2010.

[192] Pappalardo L, Cintia P, Rossi A, Massucco E, Ferragina P, Pedreschi D, and Giannotti F. A public data set of spatio-temporal match events in soccer competitions. *Scientific data*, 6(1):236, 2019.

[193] Paton CD, Hopkins WG, and Vollebregt L. Sprints in Team-Sport Athletes. *Medicine & Science in Sports & Exercise*, 33(5):822–825, 2001.

[194] Paulauskas R, Masiulis N, Vaquera A, Figueira BE, and Sampaio JE. Basketball game-related statistics that discriminate between european players competing in the nba and in the euroleague. *Journal of Human Kinetics*, 65(1):225–233, 2018.

[195] Perin C, Vuillemot R, Stolper CD, Stasko JT, Wood J, and Carpendale S. State of the Art of Sports Data Visualization. *Computer Graphics Forum*, 37(3):663–686, 2018.

[196] Pincus SM, Gladstone IM, and Ehrenkranz RA. A regularity statistic for medical data analysis. *Journal of Clinical Monitoring*, 7(4):335–345, 1991.

[197] Pollard R. Charles reep (1904 – 2002): Pioneer of notational and performance analysis in football. *Journal of Sports Sciences*, 20(10):853–855, 2002.

[198] Prasetio D and Harlili. Predicting football match results with logistic regression. *4th IGNITE Conference and 2016 International Conference on Advanced Informatics: Concepts, Theory and Application, ICAICTA 2016*, pages 2–6, 2016.

[199] Pratas JM, Volossovitch A, and Carita AI. Goal scoring in elite male football: A systematic review. *Journal of Human Sport and Exercise*, 13(1):218–230, 2018.

[200] Pro League. Pro League & Stats Perform breiden partnership uit, 2020. URL `https://www.proleague.be/nl/pl/news/pro-league-and-stats-perform-breiden-partnership-uit`.

[201] Rampinini E, Coutts AJ, Castagna C, Sassi R, and Impellizzeri FM. Variation in top level soccer match performance. *International Journal of Sports Medicine*, 28(12):1018–1024, 2007.

[202] Rampinini E, Alberti G, Fiorenza M, Riggio M, Sassi R, Borges TO, and Coutts AJ. Accuracy of GPS Devices for Measuring High-intensity Running in Field-based Team Sports. *International Journal of Sports Medicine*, 36(1):49–53, 2015.

[203] Ranchordas MK, Dawson JT, and Russell M. Practical nutritional recovery strategies for elite soccer players when limited time separates repeated matches. *Journal of the International Society of Sports Nutrition*, 14(1), 2017.

[204] Reep C and Benjamin B. Skill and Chance in Association Football. *Journal of the Royal Statistical Society*, 131(4): 581–585, 1968.

[205] Reilly T and Thomas V. A Motion Analysis of Work-Rate in Different Positional Roles in Professional Football Match-Play. *Journal of Human Movement Studies*, 2:87–97, 1976.

[206] Rein R and Memmert D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, 5(1):13, 2016.

[207] Rico-González M, Pino-Ortega J, Nakamura FY, Arruda Moura F, Rojas-Valverde D, and Los Arcos A. Past, present, and future of the technological tracking methods to assess tactical variables in team sports: A systematic review. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 234(4): 281–290, 2020.

[208] Rokach L. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2):1–39, 2010.

[209] Ruiz-Ruiz C, Fradua L, Fernández-García Á, and Zubillaga A. Analysis of entries into the penalty area as a performance indicator in soccer. *European Journal of Sport Science*, 13(3):241–248, 2013.

[210] Russell M, Sparkes W, Northeast J, Cook CJ, Love TD, Bracken RM, and Kilduff LP. Changes in Acceleration and Deceleration Capacity Throughout Professional Soccer Match-Play. *Journal of Strength and Conditioning Research*, 30 (10):2839–2844, 2016.

[211] Sahlin K, Harris RC, and Hultman E. Resynthesis of creatine phosphate in human muscle after exercise in relation to intramuscular ph and availability of oxygen. *Scandinavian Journal of Clinical and Laboratory Investigation*, 39(6): 551–557, 1979.

[212] Sajeevan G. Latitude and longitude - A misunderstanding. *Current Science*, 94(5):568–569, 2008.

[213] Saltin B, Henriksson J, Nygaard E, Andersen P, and Jansson E. Fiber types and metabolic potentials of skeletal muscles in sedentary man and endurance runners*. *Annals of the New York Academy of Sciences*, 301(1):3–29, 1977.

[214] Sands WA, Slater C, McNeal JR, Murray SR, and Stone MH. Historical Trends in the Size of US Olympic Female Artistic Gymnasts. *International Journal of Sports Physiology and Performance*, 7(4):350–356, 2012.

[215] Sarmento H, Marcelino R, Anguera MT, CampaniÇo J, Matos N, and Leitão J. Match analysis in football: a systematic review. *Journal of Sports Sciences*, 32(20):1831–1843, 2014.

[216] Schutz Y and Chambaz A. Could a satellite-based navigation system (GPS) be used to assess the physical activity of individuals on earth? *European Journal of Clinical Nutrition*, 51(5):338–339, 1997.

[217] Scott BR, Lockie RG, Knight TJ, Clark AC, and De Jonge XA. 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, 2013.

[218] Seidenschwarz P, Rumo M, Jonsson A, Probst L, Rauschenbach F, and Schuldt H. Combining qualitative and quantitative analysis in football with sportsense. *MMSports 2019 - Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports, co-located with MM 2019*, pages 34–41, 2019.

[219] Seiler S. What is Best Practice for Training Intensity and Duration Distribution in Endurance Athletes. *International Journal of Sports Physiology and Performance*, 5:276–291, 2010.

[220] Silva JR, Rebelo A, Marques F, Pereira L, Seabra A, Ascensão A, and Magalhães J. Biochemical impact of soccer: an analysis of hormonal, muscle damage, and redox markers during the season. *Applied Physiology, Nutrition, and Metabolism*, 39(4):432–438, apr 2014.

[221] Silva JR, Brito J, Akenhead R, and Nassis GP. The Transition Period in Soccer: A Window of Opportunity. *Sports Medicine*, 46(3):305–313, 2016.

[222] Silva JR, Rumpf MC, Hertzog M, Castagna C, Farooq A, Girard O, and Hader K. Acute and Residual Soccer Match-Related Fatigue: A Systematic Review and Meta-analysis. *Sports Medicine*, 48(3):539–583, 2018.

[223] Skinner JS and McLellan TH. The Transition from Aerobic to Anaerobic Metabolism. *Research Quarterly for Exercise and Sport*, 51(1):234–248, 1980.

[224] Soderlund K and Hultman E. ATP and phosphocreatine changes in single human muscle fibers after intense electrical stimulation. *American Journal of Physiology - Endocrinology and Metabolism*, 261(6 24-6), 1991.

[225] Souglis A, Bogdanis GC, Giannopoulou I, Papadopoulos C, and Apostolidis N. Comparison of Inflammatory Responses and Muscle Damage Indices Following a Soccer, Basketball, Volleyball and Handball Game at an Elite Competitive Level. *Research in Sports Medicine*, 23(1):59–72, 2015.

[226] Spearman W. Beyond Expected Goals. In *12th Annual MIT Sloan Sports Analytics Conference*, number August, pages 1–17, 2018.

[227] Sporis G, Jukic I, Ostojic SM, and Milanovic D. Fitness Profiling in Soccer: Physical and Physiologic Characteristics of Elite Players. *Journal of Strength and Conditioning Research*, 23(7):1947–1953, 2009.

[228] Stöggl TL and Sperlich B. The training intensity distribution among well-trained and elite endurance athletes. *Frontiers in Physiology*, 6:295, oct 2015.

[229] Stolen T, Chamari K, Castagna C, and Wisloff U. Physiology of Soccer. *Sports Medicine*, 35(6):501–536, 2005.

[230] Stübinger J, Mangold B, and Knoll J. Machine learning in football betting: Prediction of match results based on player characteristics. *Applied Sciences*, 10(1), 2020.

[231] Sutton L, Scott M, Wallace J, and Reilly T. Body composition of English Premier League soccer players: Influence of playing position, international status, and ethnicity. *Journal of Sports Sciences*, 27(10):1019–1026, 2009.

[232] Svensson M and Drust B. Testing soccer players. *Journal of Sports Sciences*, 23(6):601–618, 2005.

[233] Taborri J, Keogh J, Kos A, Santuz A, Umek A, Urbanczyk C, van der Kruk E, and Rossi S. Sport biomechanics applications using inertial, force, and EMG sensors: A literature overview. *Applied Bionics and Biomechanics*, 2020, 2020.

[234] Tax N and Joustra Y. Predicting The Dutch Football Competition Using Public Data: A Machine Learning Approach. *Transactions on knowledge and data engineering*, 10(10):1–13, 2015.

[235] Thirumurugan P and Hasan Hussain S. Event Detection in Videos Using Data Mining Techniques. *International Journal of Computer Science and Information Technologies*, 3(2):3473–3475, 2012.

[236] Thorpe RT, Strudwick AJ, Buchheit M, Atkinson G, Drust B, and Gregson W. Monitoring fatigue during the in-season competitive phase in elite soccer players. *International Journal of Sports Physiology and Performance*, 10(8):958–964, 2015.

[237] Thorpe RT, Strudwick AJ, Buchheit M, Atkinson G, Drust B, and Gregson W. Tracking Morning Fatigue Status Across In-Season Training Weeks in Elite Soccer Players. *International Journal of Sports Physiology and Performance*, 11(7): 947–952, 2016.

[238] Tomlin DL and Wenger HA. Relationship Between Aerobic Fitness and Recovery from High Intensity Intermittent Exercise. *Sports Medicine*, 31(1):1–11, 2001.

[239] Torres-Ronda L and Schelling X. Critical process for the implementation of technology in sport organizations. *Strength and Conditioning Journal*, 39(6):54–59, 2017.

[240] Tovar J. Soccer, World War II and coronavirus: a comparative analysis of how the sport shut down. *Soccer and Society*, 22(1-2):66–74, 2021.

[241] Trappe S, Luden N, Minchev K, Raue U, Jemiolo B, and Trappe TA. Skeletal muscle signature of a champion sprint runner. *Journal of Applied Physiology*, 118(12):1460–1466, 2015.

[242] Tumilty D. Physiological Characteristics of Elite Soccer Players. *Sports Medicine: Evaluations of Research in Exercise Science and Sports Medicine*, 16(2):80–96, 1993.

[243] Van Gool D, Van Gervan D, and Boutmans J. *The physiological load imposed on soccer players during real match-play*, volume 1. E. & F.N. Spon, London, 1988.

[244] Van Hooren B and Peak JM. Do We Need a Cool-Down After Exercise A Narrative Review of the Psychophysiological Effects and the Effects on Performance, Injuries and the Long-Term Adaptive Response. *Sports Medicine*, 48(7): 1575–1595, 2018.

[245] Vanrenterghem J, Nedergaard NJ, Robinson MA, and Drust B. Training Load Monitoring in Team Sports: A Novel Framework Separating Physiological and Biomechanical Load-Adaptation Pathways. *Sports Medicine*, 47(11):2135–2142, 2017.

[246] Verheul J, Nedergaard NJ, Vanrenterghem J, and Robinson MA. Measuring biomechanical loads in team sports–from lab to field. *Science and Medicine in Football*, 4(3):246–252, 2020.

[247] Vigne G, Gaudino C, Rogowski I, Alloatti G, and Hautier C. Activity profile in elite Italian soccer team. *International journal of sports medicine*, 31(5):304–310, 2010.

[248] Vigne G, Dellal A, Gaudino C, Chamari K, Rogowski I, Alloatti G, Del Wong P, Owen AL, and Hautier C. Physical outcome in a successful italian serie a soccer team over three consecutive seasons. *Journal of Strength and Conditioning Research*, 27(5):1400–1406, 2013.

[249] Vilar L, Araújo D, Davids K, and Bar-Yam Y. Science of winning soccer: Emergent pattern-forming dynamics in association football. *Journal of Systems Science and Complexity*, 26(1):73–84, 2013.

[250] Viru A and Viru M. Nature of training effects. *Exerc Sport Sci*, 6(7):67–95, 2000.

[251] Walker GJ and Hawkins RD. Structuring a Program in Elite Professional Soccer. *Strength and Conditioning Journal*, 40(3):72–82, 2018.

[252] Weaving D, Jones B, Till K, Abt G, and Beggs C. The case for adopting a multivariate approach to optimize training load quantification in team sports. *Frontiers in Physiology*, 8(DEC):1–3, 2017.

[253] Whitehead N. *Conditioning for Sport*. E.P. Publishing Co. Ltd., Yorkshire, 1975.

[254] Williams S, West S, Cross MJ, and Stokes KA. Better way to determine the acute: chronic workload ratio? *Br J Sports Med*, 51(3):209–210, 2017.

[255] Windt J and Gabbett TJ. Is it all for naught? What does mathematical coupling mean for acute:chronic workload ratios? *British Journal of Sports Medicine*, 53(16):988–990, 2019.

[256] Withers R, Maricic Z, Wasilewski S, and Kelly L. Match Analyses of Australian Professional Soccer Players. *Journal of Human Movement Studies*, 8:159–176, 1982.

[257] Woods CT, McKeown I, Shuttleworth RJ, Davids K, and Robertson SJ. Training programme designs in professional team sport: An ecological dynamics exemplar. *Human Movement Science*, 66(March):318–326, 2019.

[258] Yang G, Leicht AS, Lago C, and Gómez MáÁ. Key team physical and technical performance indicators indicative of team quality in the soccer Chinese super league. *Research in Sports Medicine*, 26(2):158–167, 2018.

[259] Zarić I, Kukić F, Jovićević N, Zarić M, Marković M, Toskić L, and Dopsaj M. Body height of elite basketball players: Do taller basketball teams rank better at the FIBA world cup? *International Journal of Environmental Research and Public Health*, 17(9), 2020.

Part II

Original research

Study I

Machine Learning-based Identification of the Strongest Predictive Variables of Winning and Losing in Belgian Professional Soccer

Geurkink, Y. , Boone, J., Verstockt, S., & Bourgois, J. G. (2021). Machine Learning-Based Identification of the Strongest Predictive Variables of Winning and Losing in Belgian Professional Soccer. *Applied Sciences*, 11(5), 2378.

Youri Geurkink [1], Jan Boone [1,2], Steven Verstockt [3 *], and Jan G. Bourgois [1,2 *]

[1] Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

[2] Center of Sports Medicine, Ghent University Hospital, Ghent, Belgium

[3] Department of Electronics and Information Systems, Research group IDLab, Ghent University—IMEC, Ghent, Belgium

* These authors share last authorship

Abstract

This study aimed to identify the strongest predictive variables of winning and losing in the highest Belgian soccer division. A predictive model based on a broad range of variables (n=100) was constructed, using a dataset consisting of 576 games. To avoid multicollinearity and reduce dimensionality, respectively variance inflation factor (threshold of 5) and BorutaShap were applied. A total of 13 variables remained and were used to predict winning or losing using Extreme Gradient Boosting. TreeExplainer was applied to determine feature importance on a global and local level. The model showed an accuracy of 89.6% $\pm$ 3.1% (precision: 88.9%; recall: 90.1%, f1-score: 89.5%), correctly classifying 516 out of 576 games. Shots on target from the attacking penalty box showed to be the best predictor. Several physical indicators are amongst the best predictors, as well as contextual variables such as ELO-ratings, added transfers value of the benched players and match location. The results show the added value of the inclusion of a broad spectrum of variables when predicting and evaluating game outcomes. Similar modelling approaches can be used by clubs to identify the strongest predictive variables for their leagues, and evaluate and improve their current quantitative analyses.

Keywords

Association Football; Performance; Performance Analysis; KPI; Game Result

Introduction

The numerous unique chains of dynamic interactions between players during soccer games can be reduced to different performance indicators, to allow more insight into game performance [34]. A performance indicator is a selection, or combination, of action variables aiming to define some or all aspects of performance, and can be used to assess the performances of an individual, a team or the elements of a team [27]. Generally, sports performance judgments are prone to bias [44], as good outcomes are attributed to internal causes and bad outcomes to external causes [39]. Better insights into performance indicators could therefore not only aid the decision-making process of coaches and players with regard to training and game preparation, but also help other stakeholders related to the team, such as scouting and management, to correctly evaluate team and player performances [6].

Previously, several performance indicators have already been linked to performance in soccer, such as ball possession [10, 12, 32, 51], number of passes [7, 23], number of shots [7, 10, 32, 35], number of shots on target [10, 32, 35], entries into the penalty box [46, 51] and successfulness in duels [51]. Previous studies investigating performance indicators in soccer however often suffer from limitations and/or methodological problems, such as small sample sizes and univariate analyses of the observed variables [10]. If presented in isolation, a single set of indicators representing the performance of an individual or a team can give a distorted impression of a performance by ignoring other, more or less important, variables [27], which likely explains differences between previous studies [10]. It is therefore important that variables are not considered in isolation when relating performance indicators to game performance.

Technical innovations have led to an increasing availability of different performance indicators. Advanced metrics can currently be calculated in soccer using tracking data [21]. As tracking data results in millions of data points per season [2], calculating these metrics challenges data management and analytical methods of analysts [22]. Advanced metrics based on tracking data are also often provided by professional sports data companies, but obtaining these metrics is usually not free. So although metrics based on tracking data may better capture the complex nature of soccer [21], obtaining these metrics may at the moment not be feasible because of pratical and/or financial reasons for many teams. Furthermore, although there has been an increased interest into the utilization of tracking data by academics, practitioners still tend to use more traditional performance indicators [43].

In contemporary soccer, there is a wealth of different performance indicators, describing technical, tactical and physical performances, as well as contextual information. In order to provide to-the-point information to the coaching staff, a selection of the most important performance indicators can be helpful. Therefore, this study aimed to identify the strongest predictive variables of winning and losing in soccer using a broad range of variables. Machine learning is used instead of inferential statistics, as machine learning is better at handling large amounts of input variables [8]. Based on previous research, it can be hypothesized that shot-related variables such as shots and shots on target are closely related to performance in soccer, however, it is also expected that other indicators will be amongst the most important predictors.

Materials and Methods

Sample

All data was collected in the highest Belgian soccer division, over the course of 3 seasons (2017-2018, 2018-2019, 2019-2020), totalling 771 games. It should be noted that, because of the COVID19 pandemic, the last part of the 2019-2020 season was not played. Games with missing values because of measurements errors, such as missing physical data and/or technical data, and games that resulted in a draw, were excluded from the analysis. This resulted in the availability of 576 games for the analysis. All games were tracked using the SportVU system (Stats Perform, Chicago, IL, USA), an optical tracking system using three high-definition camera's [36]. Consent was given by STATS Perform and the Belgian Pro League for the use of the data for scientific purposes. The reliability of the data delivered by professional sports data companies, is shown to be high [4, 5]. The study was conducted in accordance with the Declaration of Helsinki.

Variables

A total of 100 variables were included into the analysis (Table 1). All variables, with the exception of contextual game information, were derived from three sources, STATS Viewer, STATS Dynamix and STATS Edge, all offered by STATS Perform. Data regarding the teams' playing styles and Expected Goals were derived from STATS Edge. Definitions on Playing Styles and the calculation of Expected Goals can be found on the website of STATS Perform (*https://www.statsperform.com*). Physical game data was available at STATS Dynamix. The default speed, acceleration and deceleration thresholds for physical data were used, including the minimum effort duration of 0.5 seconds set for variables related to speed, accelerations and deceleration. All other variables, with the exception of contextual game information, were derived from STATS Viewer. Inside the STATS Viewer software, the user can derive its own metrics, based on criteria such as destination and direction. Direction could be set in 3 different directions, all relative to the opponent's goal. A ball played in an angle of -45° to +45° was defined as forward, backwards was defined if the ball was played in an angle from -135° to +135° relative to the opponent's goal. Balls played in an angle of +45° to +135°, and -45° to -135° were defined as sideways.

Contextual variables were obtained from other sources. ELO-ratings were included as a measure of team strength, provided by the API of *http://www.clubelo.com*, which is freely available. This source provides ELO-ratings for each team of over 50 (inter)national leagues, including the UEFA Europa League and Champions League, and has previously been used for scientific purposes [14, 17]. Each game, both national and international, results in an exchange in ELO-points (Equation 1), where *dr* is the difference in ELO-rating between two clubs and *R* is the result (1 for win, 0.5 for draw and 0 for loss). The exchange in ELO-points is thus higher for a win against a stronger team compared to a victory against an equally strong or weaker team, and vice versa for losses. This equation was derived from *http://www.clubelo.com*.

$$\Delta ELO = (R - \frac{1}{10^{(-dr/400)} + 1}) * 20 \tag{1}$$

Form was defined as the difference in a clubs' current ELO-rating compared to the ELO-rating before their previous game. Market values, ages and nationalities were obtained from *https://www.transfermarkt.co.uk*. These variables were previously used by [35] for scientific purposes. Lastly, the number of days between games was used as a variable, also including other competitive games such as cup and European games, to consider the possible effect of additional games for one team compared to the other on winning or losing.

Differences between the two teams competing during each game were calculated and used as input variable in this study, with the exception of match location. During games, there are interactions between the two competing teams [34], and these interactions will result in different performance statistics for both teams. The difference of each performance statistic between the two teams may provide insights into the difference in performance by the two teams on the pitch, and therefore be more informative to the model than the separate performance statistics of both the teams. Moreover, by not including the performance

Table 1: An overview on the 100 included variables.
[A] Playing styles: Maintenaince, Build-Up, Sustained Threat, Fast Tempo, Direct Play, Counter Attack, Crossing, High Press

Category	Expression	Parameter	Part Game
Shot-related	Number (n)	- Shots, shots on target, shots not on target - Shots/shots on target/shots not on target inside attacking penalty box - Expected Goals - Shots resulting from 8 playing styles [A]	Full Game
Defense	Number (n)	- Duels won, possession won/loss	Full Game
Technical	Number (n)	- Total/forward/sideways/backward passes, successful passes - Total/forward/sideways/backward passes to attacking half/third - Dribbles, successful dribbles - Ball touches in attacking penalty box - Passes <10 meter, Passes <25 meter, Passes >25 meter - Possession in 8 playing styles [A]	Full Game
	Percentage (%)	- Total ball possession - Ball possession on defensive half/attacking half/attacking third/attacking penalty box	
Physical	Distance (m)	- Total distance - Distance between 0-6 km/h, 6-15 km/h, 15-20 km/h, 20-25 km/h, >25 km/h - Distance at accelerations/decelerations >2ms^2/>3ms^2	First/Second Half
	Number (n)	- Actions at speed >15 km/h/>25 km/h - Accelerations/decelerations >2ms^2/>3ms^2	
Disciplinal	Number (n)	- Fouls, fouls at attacking half, yellow cards, red cards, offside	Full Game
Set-Pieces	Number (n)	- Corners, penalties, free kick/throw-in on attacking third	Full Game
Contextual	Currency (€)	- Line-up/bench current estimated total transfer value - Line-up/bench estimated paid total transfer value	Full Game
	Age (years)	- Line-up/bench average age	
	Arbitrary Units (AU)	- ClubELO, Form	
	Number (n)	- Days between games	
	Match location	- Home/Away	

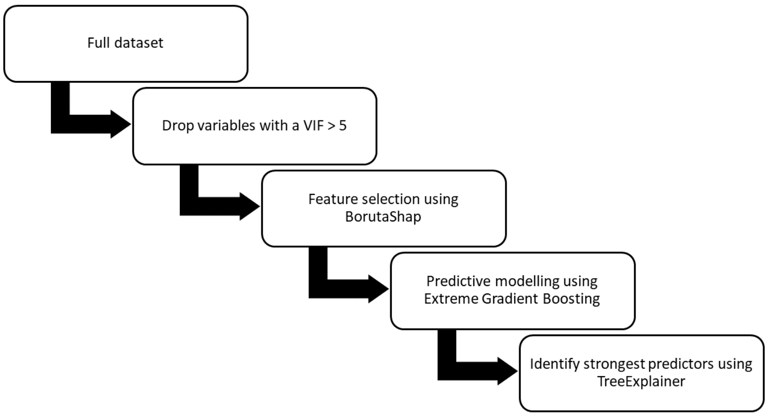


Figure 1: Graphical illustration on the workflow used for this study.

statistics of both teams into the model, the dimensionality of the model can be limited. An appendix with the full description of each variable was added on pages 128-133.

Procedures

Data from all sources was loaded into Python (version 3.7.1). Variables included in the model were selected based on a combination of expert knowledge and availability. To avoid multicollinearity, a Variance Inflation Factor (VIF) analysis was conducted, with a threshold of 5, using the *statsmodels* package. BorutaShap was applied as a feature selection technique, using the *BorutaShap* package. Extreme Gradient Boosting, a tree-based machine learning technique, was applied for both BorutaShap and predicting game outcome (win or lose), using the *xgboost* package (distributed by *SkLearn*). A 5-fold stratified cross-validation was used to validate the results. Cross-validation uses a large part of the data to fit the model, in this study 80%, and a small part of the data to test the model, in this study 20% [25]. Each part was thus used 4 times to fit the model and once for validation. Stratified K-fold was used to preserve balance between the frequency of each class of the dependent variable. The *StratifiedKFold* and *cross_val_score* packages, distributed by *SkLearn*, were used for the cross-validation. The average classification accuracy, precision, recall and F1-score of each cross-validation fold is reported, as well as the standard deviation.

The aim of this study was to identify the best predictive variables, therefore, it was opted to apply a tree-based machine learning, which has the advantage of high interpretability and the possibility to apply a theoretically well grounded method such as TreeExplainer to identify the strongest predictors [38]. TreeExplainer uses Shapley values to explain the global model structure, by combining local explanations of each prediction [38]. Using TreeExplainer, it is possible to determine the importance of each feature [41], on a global and local level. TreeExplainer was applied using the *shap* package. A graphical illustration on the workflow from the dataset to the identification of the best predictors is depicted in Figure 1.

Results

After the removal of variables using VIF and the feature selection procedure using BorutaShap, a total of 13 variables were used during the modelling procedure. The model showed a predictive accuracy of 89.6% ± 3.1%, correctly classifying 516 out of 576 games that resulted in a win or loss (precision: 88.9%; recall: 90.1%, f1-score: 89.5%; Figure 2). In Table 2, the misclassifications are further specified in relation to total goal difference.

The most important predictors of the model are presented in Figure 3. As an illustration of the possibilities of local explanations using TreeExplainer, two individual game predictions are presented in Figure 4A and Figure 4B.

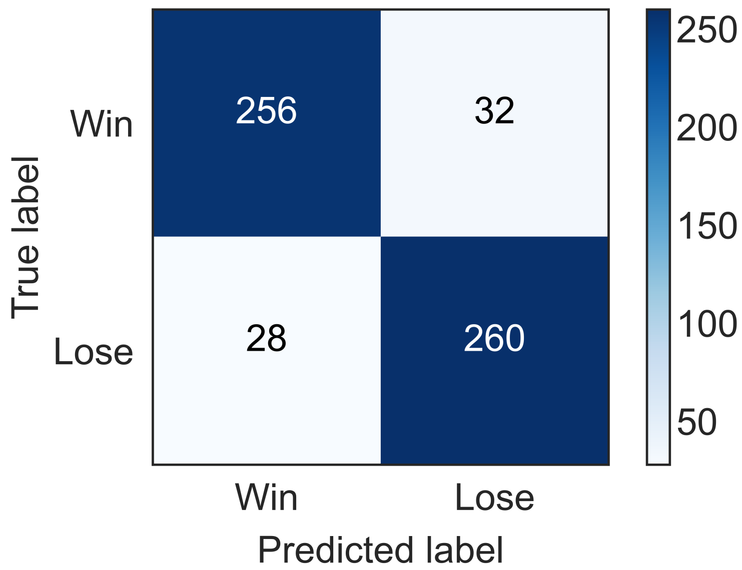


Figure 2: Confusion matrix showing the true and predicted results.

Table 2: An overview on the total and relative number of games and misclassifications for each goal difference.

Goal Difference	Games (n=576)		Misclassifications (n=60)	
	Total	%	Total	%
1	302	52.4%	43	71.7%
2	160	27.8%	16	26.7%
3	69	12.0%	1	1.7%
4	33	5.7%	0	0%
5	8	1.4%	0	0%
6	3	0.5%	0	0%
7	1	0.2%	0	0%

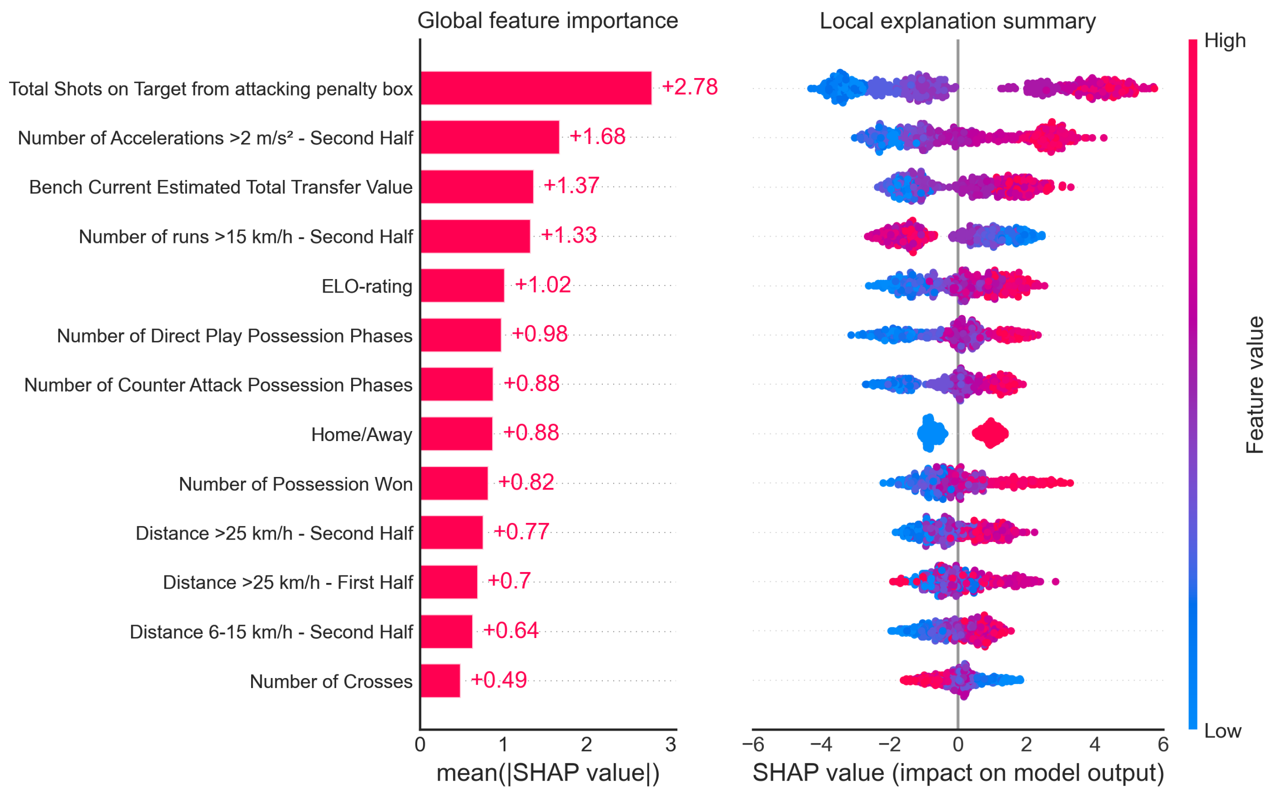
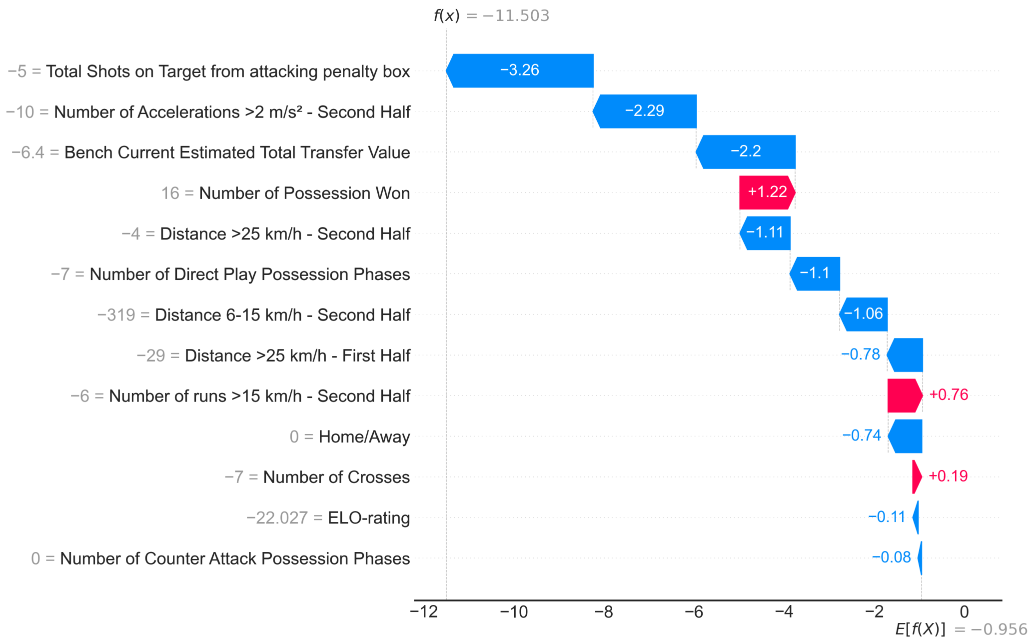


Figure 3: Feature importance based on SHAP-values. On the left side, the mean absolute SHAP-values are depicted, to illustrate global feature importance. On the right side, the local explanation summary shows the direction of the relationship between a variable and game outcome. Positive SHAP-values are indicative of winning, while negative SHAP-values are indicative of losing. As demonstrated by the colorbar, higher values are shown in red, while lower values are shown in blue. As an example, this shows that positive differences between total shots on target from the attacking penalty box between teams are associated with winning, while negative differences are associated with losing.

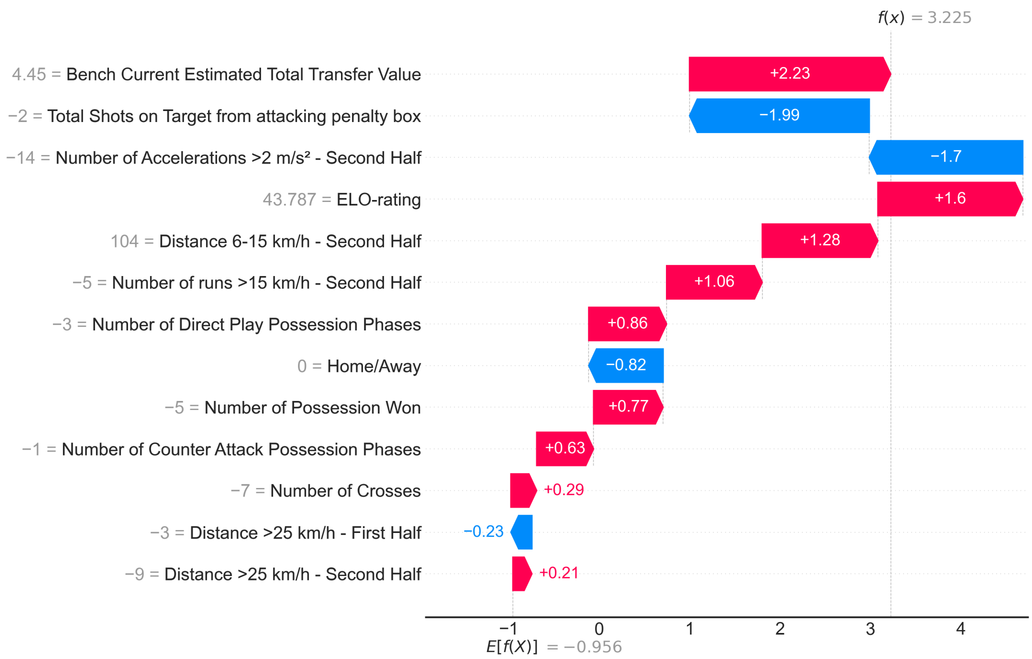
Discussion

This study aimed to identify the strongest predictive variables of winning and losing in Belgian professional soccer. A broad spectrum of variables was used to build a predictive model. The results showed that more than 89% of the games resulting in a win or loss could be correctly classified. Total shots on target from the attacking penalty box showed to be the strongest predictor. Interestingly, a broad range of variables, including physical indicators such as the distance in several speed zones, the number of accelerations (>2 m/s^2) and number of actions $>$ 15 km/h showed to be among the most important variables related to game results, as well as contextual variables such as ELO-rating, the total added transfer value of the benched players and match location.

There are different purposes for predicting game outcome. Previous studies relating to the prediction of game outcome in soccer often focused on betting [16]. These studies used historic games to construct a model, which was then evaluated by predicting future soccer games. In our study however, the aim was to



(A) Game "Team A" - "Team B" (2-1), from the perspective of Team B (loss predicted).



(B) Game "Team C" - "Team D" (0-1), from the perspective of Team D (win predicted).

Figure 4: Two waterfall plots exampling local game predictions, showing the contribution of each variable to the prediction. The variable name is preceded by the value of the particular variable. Below the x-axis, the baseline value (E[f(X)]) is displayed, indicative of the expected value of the model evaluated on the background dataset. The SHAP-values of each variable are summed to match the model output with all variables included. Positive SHAP-values push the model to predict a win, while negative SHAP-values push the model to predict a loss.

identify the strongest predictive variables, so instead of a division between historic and future games, a cross-validation approach was used to evaluate model performance. The prediction accuracy was considerably higher compared to a similar study conducted by Hassan et al. [24], who reported classification accuracies of 72.7% and 83.3% when predicting winning and losing in professional soccer using artificial neural networks. A classification accuracy of amply 89% can be considered as high, as it has been shown that chance plays a major role in goal-scoring [33]. The results show that the majority of the misclassifications occur when the final goal difference between two teams is small. Given these small goal differences between teams, and the impact that each goal has on game outcome [33], the occurrence of "lucky winners" or "unlucky losers" may be frequent, and classification accuracies of close to 100% seem improbable.

It was hypothesized that shot-related variables were among the strongest predictors, as previous studies [7, 10, 32, 35] already showed that shot-related variables closely relate to game outcome in soccer. In this study, the total shots on target from the attacking penalty box showed to be the best predictor of winning and losing. Of all shot-related variables, total shots on target from the attacking penalty box was the only variable which was not rejected by VIF or BorutaShap, showing that shot-related variables are closely interrelated. It should be noted that this does not directly indicate that other shot-related variables can be deemed as unimportant, but that the information entailed by those measures is already captured, or better captured, by other metrics in relation to game outcome. The number of shots on or off target are often reported by sports data providers, as opposed to location, which is often not reported. It may therefore be useful to either use both total shots on target from the penalty box, or use a metric such as Expected Goals, which also takes shot location into account. Different often reported metrics such as the total number of passes and the number of successful passes are not among the best important predictors of game outcome, which may also be explained by the informative value to the model in comparison to other variables. This may also explain why Playing Styles-related possessions, such as Direct Play and Counter Attack, are amongst the best predictors, as they may provide more information to the model than total ball possession.

Physical fitness is deemed as an important factor relating to performance in soccer, however, the role of physical game output in relation to other performance indicators remained to be elucidated [9]. In our study, several physical indicators are shown to be among the best predictors of game outcome in soccer. All variables relating to physical variables were subdivided into the first and second half, as it was previously shown that physical game performances, such as high-speed running [5], the number of acc- and decelerations and the distance in several acc- and deceleration zones decrease at the end of the game [47]. Interestingly, most of the selected predictive physical variables relate to the second half, with the exception of the distance $>$25 km/h, of which both halves were included into the modelling procedure. This study also shows that total difference in the number of medium accelerations (>2m/s^2) is associated with winning or losing, further confirming the importance of high-intensity efforts in soccer [18]. As indicated by the inclusion of the variable distance between 6-15 km/h in the second half, the ability to maintain the physical capabilities to not only perform high intensity actions, but also low to medium intensity efforts throughout the game seems to be important. With regard to the interpretation of physical performance indicators, it is however important to note that 'more' does not always indicate 'better', as shown by the inverse relationship of the number of actions >15 km/h and game outcome. This finding is also partly confirmed by a study of Chmura et al. [11], showing that players covered less distance in the zone between 17-21 km/h during games that were won. As physical game output depends on a myriad of factors, such as ball possession [15], pacing strategy [42], match location and match status [31], physical game output, regardless of its expression, should be viewed in relation to other performance indicators [3, 9] and contextual information [31].

Some variables that can be related to attacking play are negatively associated to game outcome. In accordance with previous research [19, 37], higher frequencies of crosses are negatively associated with game outcome. Crossing, which can be defined as an airborne delivery of the ball into the opponent's penalty area [50], may therefore be labelled as an inefficient method to create good scoring opportunities [37, 50]. It should however be noted that playing style depend on the qualities and characteristics of the team. Therefore, the coaching staff may decide to apply a playing style that can generally be characterized as inefficient, because it matches the teams' qualities and characteristics.

Technical innovations have led to an increasing availability of different performance indicators. Advanced metrics such as Packing Rate and Impect [21] can currently be calculated in soccer using tracking data. As tracking data results in millions of data points per season [2], calculating these metrics challenges data management and analytical methods of analysts [22]. Advanced metrics based on tracking data are also often provided by professional sports data companies (STATS Perform, OPTA), but obtaining these metrics is usually not free. So although metrics based on tracking data may better capture the complex nature of soccer [21], obtaining these metrics may at the moment not be feasible because of pratical and/or financial reasons for many teams. Moreover, in a low-scoring game such as soccer, rare events are often those that lead to success [21]. These events should be captured to accurately predict game outcome, however, these events cannot always be properly quantified, even with more advanced metrics based on tracking data. Actions that are currently difficult to quantify, such as good positioning or the ability to give defense-splitting passes, are however often recognized by clubs, media and/or fans, resulting in higher transfer values reported by sources such as Transfermarkt.co.uk. These actions also help teams to get better game outcomes, resulting in improved ELO-ratings. Including transfer values and ELO-ratings may thus be useful, possibly by partly filling the gap of what cannot be (currently) quantified, also considering the feasibility of obtaining these variables in terms of pracitical and financial reasons.

Machine learning was used in this study to identify the strongest predictive variables. It has also previously been used in a soccer context not only in relation to game outcome [13, 26, 49] and tactics [40], but also in relation to training load [20, 29] and injuries [45], showing the broad window of applications of machine learning in soccer. Given that game performance [48], training load [28] and injury [1] are all multidimensional, with many relating factors, the application of machine learning can be useful since it is particularly helpful when dealing with many input variables [8]. Developments in the area of machine learning, such as TreeExplainer [38], which is not only a strong theoretically grounded method to calculate feature importance [41], but also allows to build visualisations that indicate the direction of the relation of a variable with performance, can be helpful in the translation from science to practice. Illustrations such as those displayed in Figure 3 and Figure 4 can be useful for analysts to show how features impact game outcome.

Future studies should attempt to add more detailed information for several features, for example, total distance in and out of possession or detailed information on the position of crosses. This information can be informative to the model and aid the explanation of results. As data is often provided by sports data companies, these companies should be encouraged to add more detail to the provided data to allow more thorough analyses. The use of 'new' features should also be encouraged, to test their added value in relation to other, more established features. It should also be noted that the results from this study cannot fully distinguish whether a variable is the cause of a (un)favourable scoreline, or the effect. To illustrate, losing teams attempt to turn the game and therefore may engage in more high-intensity efforts [11]. The winning side on the other hand, may fall back, allowing less space for losing side to play and perform sprints, while the losing team may apply a more risk-taking strategy, which could result in more counter-attacking play of the winning side [19, 30]. Therefore, more research is necessary to gain more insight into the cause-effect relation between performance indicators and game outcome.

The results from our study show which variables can be considered as the best predictors of an accurate model predicting winning and losing in professional Belgian soccer. It also provides the direction of the relationship of these variables with winning and losing, also in relation to the other predictors. It was shown that not only shot-related variables, but a broad range of variables are amongst the strongest predictors of winning and losing. As the workflow from dataset to predictive modelling was also described in detail, similar approaches can be used to evaluate the current performance indicators provided to the coaching staff and other stakeholders connected to the team. It seems particularly interesting to look at physical parameters of the second half, given that they are amongst the best predictors of game outcome in soccer. Using variables such as ELO-ratings, transfer values, match location and Playing Styles can be useful additions to current approaches used to evaluate game performances.

Acknowledgement(s)

The authors would like to thank the team of Steven Verstockt for their help with the analyses.

Disclosure statement

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Bahr R and Krosshaug T. Understanding injury mechanisms: A key component of preventing injuries in sport. *British Journal of Sports Medicine*, 39(6):324–329, 2005.

[2] Bialkowski A, Lucey P, Carr P, Yue Y, Sridharan S, and Matthews I. Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 725–730, Washington DC, United States, 2014. IEEE.

[3] Bradley PS and Ade JD. 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, 2018.

[4] Bradley PS, O'Donoghue P, Wooster B, Tordoff P, O'Donoghue P, Wooster B, and Tordoff P. The reliability of ProZone MatchViewer: a videobased technical performance analysis system. *International Journal of Performance Analysis in Sport*, 7(3):117–129, 2007.

[5] Bradley PS, Di Mascio M, Peart D, Olsen P, and Sheldon B. High-Intensity Activity Profiles of Elite Soccer Players at Different Performance Levels. *Journal of Strength and Conditioning Research*, 24(9):2343–2351, 2010.

[6] Brechot M and Flepp R. Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals. *Journal of Sports Economics*, 21(4):335–362, 2020.

[7] Broich H, Mester J, Seifriz F, and Yue Z. Statistical Analysis for the First Bundesliga in the Current Soccer Season. *Progress in Applied Mathematics*, 7(2):1–8, 2011.

[8] Bzdok D, Altman N, and Krzywinski M. Points of Significance: Statistics versus machine learning. *Nature Methods*, 15 (4):233–234, 2018.

[9] Carling C. Interpreting physical performance in professional soccer match-play: Should we be more pragmatic in our approach? *Sports Medicine*, 43(8):655–663, 2013.

[10] Castellano J, Casamichana D, and Lago C. The Use of Match Statistics that Discriminate Between Successful and Unsuccessful Soccer Teams. *Journal of Human Kinetics*, 31(1):139–147, 2012.

[11] Chmura P, Konefał M, Chmura J, Kowalczuk E, Zajac T, Rokita A, Andrzejewski M, Chmura Pawełand Konefał M, Chmura J, Kowalczuk E, Zajac T, Rokita A, and Andrzejewski M. Match outcome and running performance in different intensity ranges among elite soccer players. *Biology of Sport*, 35(2):197–203, 2018.

[12] Collet C. The possession game? A comparative analysis of ball retention and team success in European and international football, 2007-2010. *Journal of Sports Sciences*, 31(2):123–136, 2013.

[13] Constantinou AC. Dolores: a model that predicts football match outcomes from all over the world. *Machine Learning*, 108(1):49–75, 2019.

[14] Csató L. The UEFA Champions League seeding is not strategy-proof since the 2015/16 season. *Annals of Operations Research*, 292(1):161–169, 2020.

[15] Dellal A, Chamari K, Wong DP, Ahmaidi S, Keller D, Barros RML, Bisciotti GN, and Carling C. Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European Journal of Sport Science*, 11(1):51–59, 2011.

[16] Dubitzky W, Lopes P, Davis JJ, and Berrar D. The Open International Soccer Database for machine learning. *Machine Learning*, 108(1):9–28, 2019.

[17] Engist O, Merkus E, and Schafmeister F. The Effect of Seeding on Tournament Outcomes: Evidence From a Regression-Discontinuity Design. *Journal of Sports Economics*, 22(1):115–136, 2021.

[18] Faude O, Koch T, and Meyer T. Straight sprinting is the most frequent action in goal situations in professional football. *Journal of Sports Sciences*, 30(7):625–631, 2012.

[19] Fernandez-Navarro J, Fradua L, Zubillaga A, and McRobert AP. Influence of contextual variables on styles of play in soccer. *International Journal of Performance Analysis in Sport*, 18(3):423–436, 2018.

[20] Geurkink Y, Lievens M, de Turck F, Ongenae F, Matthys SPJ, Boone J, Bourgois JG, Vandewiele G, Lievens M, de Turck F, Ongenae F, Matthys SPJ, Boone J, Bourgois JG, Vandewiele G, Lievens M, de Turck F, Ongenae F, Matthys SPJ, Boone J, and Bourgois JG. Modeling the Prediction of the Session Rating of Perceived Exertion in Soccer: Unraveling the Puzzle of Predictive Indicators. *International Journal of Sports Physiology and Performance*, 14(6):1–6, 2018.

[21] Goes FR, Kempe M, and Lemmink KAPM. Predicting match outcome in professional Dutch football using tactical performance metrics computed from position tracking data. In *MathSport International*, number June, pages 105–115, 2019.

[22] Goes FR, Meerhoff RLA, Bueno MJO, Rodrigues DM, Moura FA, Brink MS, Elferink-Gemser MT, Knobbe AJ, Cunha SA, and Lemmink KAPM. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science*, pages 1–16, 2020.

[23] Harrop K and Nevill A. Performance indicators that predict success in an English Professional League One Soccer Team. *International Journal of Performance Analysis in Sport*, 14(3):907–920, 2014.

[24] Hassan A, Akl AR, Hassan I, and Sunderland C. Predicting wins, losses and attributes' sensitivities in the soccer world cup 2018 using neural network analysis. *Sensors*, 20(11):1–13, 2020.

[25] Hastie T, Tibshirani R, and Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Number 2. Springer Science & Business Media, 2nd edition, 2009.

[26] Hucaljuk J and Rakipović A. Predicting football scores using machine learning techniques. *MIPRO 2011 - 34th International Convention on Information and Communication Technology, Electronics and Microelectronics - Proceedings*, 48: 1623–1627, 2011.

[27] Hughes MD and Bartlett RM. The use of performance indicators in performance analysis. *Journal of Sports Sciences*, 20:739–754, 2002.

[28] Impellizzeri FM, Rampinini E, and Marcora SM. Physiological assessment of aerobic training in soccer. *Journal of Sports Sciences*, 23(6):583–592, 2005.

[29] Jaspers A, Brink MS, Probst SG, Frencken WG, and Helsen WF. Relationships Between Training Load Indicators and Training Outcomes in Professional Soccer. *Sports Medicine*, 47(3):533–544, 2017.

[30] Lago C. The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *Journal of Sports Sciences*, 27(13):1463–1469, 2009.

[31] Lago C, Casais L, Dominguez E, and Sampaio JE. The effects of situational variables on distance covered at various speeds in elite soccer. *European Journal of Sport Science*, 10(2):103–109, 2010.

[32] Lago-ballesteros J and Lago-Peñas C. Performance in Team Sports: Identifying the Keys to Success in Soccer. *Journal of Human Kinetics*, 25(1):85–91, 2010.

[33] Lames M. Chance involvement in goal scoring in football – an empirical approach. *German Journal of Exercise and Sport Research*, 48(2):278–286, 2018.

[34] Lames M and McGarry T. On the search for reliable performance indicators in game sports. *International Journal of Performance Analysis in Sport*, 7(1):62–79, 2017.

[35] Lepschy H, Wäsche H, and Woll A. Success factors in football: an analysis of the German Bundesliga. *International Journal of Performance Analysis in Sport*, 20(2):150–164, 2020.

[36] Linke D, Link D, and Lames M. Validation of electronic performance and tracking systems EPTS under field conditions. *PLoS ONE*, 13(7):1–19, 2018.

[37] Liu H, Gomez MÁ, Lago-Peñas C, and Sampaio JE. Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. *Journal of Sports Sciences*, 33(12):1205–1213, 2015.

[38] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, and Lee SI. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67, 2020.

[39] Mark MM, Mutrie N, Brooks DR, and Harris DV. Causal Attributions of Winners and Losers in Individual Competitive Sports: Toward a Reformulation of the Self-Serving Bias. *Journal of Sport Psychology*, 6:184–196, 1984.

[40] Memmert D, Lemmink KA, and Sampaio JE. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine*, 47(1):1–10, 2017.

[41] Molnar C. *Interpretable Machine Learning*. Lulu.com, 2020.

[42] Paul DJ, Bradley PS, and Nassis GP. 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, 2015.

[43] Perin C, Vuillemot R, Stolper CD, Stasko JT, Wood J, and Carpendale S. State of the Art of Sports Data Visualization. *Computer Graphics Forum*, 37(3):663–686, 2018.

[44] Plessner H and Haar T. Sports performance judgments from a social cognitive perspective. *Psychology of Sport and Exercise*, 7(6):555–575, 2006.

[45] Rommers N, Rössler R, Verhagen E, Vandecasteele F, Verstockt S, Vaeyens R, Lenoir M, D'Hondt E, and Witvrouw E. A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players. *Medicine and Science in Sports and Exercise*, 52(8):1745–1751, 2020.

[46] Ruiz-Ruiz C, Fradua L, Fernández-García Á, and Zubillaga A. Analysis of entries into the penalty area as a performance indicator in soccer. *European Journal of Sport Science*, 13(3):241–248, 2013.

[47] Russell M, Sparkes W, Northeast J, Cook CJ, Love TD, Bracken RM, and Kilduff LP. Changes in Acceleration and Deceleration Capacity Throughout Professional Soccer Match-Play. *Journal of Strength and Conditioning Research*, 30 (10):2839–2844, 2016.

[48] Stolen T, Chamari K, Castagna C, and Wisloff U. Physiology of Soccer. *Sports Medicine*, 35(6):501–536, 2005.

[49] Stübinger J, Mangold B, and Knoll J. Machine learning in football betting: Prediction of match results based on player characteristics. *Applied Sciences*, 10(1), 2020.

[50] Vecer J. Crossing in Soccer has a Strong Negative Impact on Scoring: Evidence from the English Premier League the German Bundesliga and the World Cup 2014. Technical report, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic, 2014.

[51] Yang G, Leicht AS, Lago C, and Gómez MáÁ. Key team physical and technical performance indicators indicative of team quality in the soccer Chinese super league. *Research in Sports Medicine*, 26(2):158–167, 2018.

Study II

Training Characteristics and Training Intensity Distribution of the Preseason in a Professional Soccer Team

Submitted to Journal of Exercise Science and Fitness

Youri Geurkink [1,2], Jan Boone [1], Stijn P.J. Matthys [1,2], and Jan G. Bourgois [1,2]

[1] Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

[2] KAA Gent – UGent Performance Center, Ghent, Belgium

Abstract

Objectives: To describe the training characteristics of the preseason in a professional soccer team, and describe and compare different training intensity distribution (TID) methods, using 4 combinations of intensity (heart rate (HR) and speed) and volume (time, distance) parameters. Materials and Methods: Twenty professional soccer players participated in this study. HR and speed at the first and second lactate threshold were determined during an incremental running test on a treadmill. All outdoor training sessions (n=50) were performed with a GPS/HR-sensor. Results: Players covered on average 254.0 $\pm$ 47.3 km during 53.8 $\pm$ 10.4 hours of outdoor preseason training activities, over a preseason duration of 36 days. The proportion of volume spent at low intensity was higher for Speed-Time (87.9%) as compared to HR-Time (76.6%), HR-Distance (66.5%) and Speed-Distance (65.6%) and higher for HR-Time than for HR-Distance and Speed-Distance. At moderate intensity, the proportion of volume spent was higher for HR-Distance 28.9%) compared to HR-Time (20.4%), Speed-Time (7.4%) and Speed-Distance (17.8%), and for Speed-Time lower than for HR-Time and Speed-Distance. At high intensity, the proportion of volume spent was higher for Speed-Distance (16.5%) compared to HR-Time (3.1%), HR-Distance (4.8%) and Speed-Time (4.7%). Conclusions: All TIDs showed that most of the training volume is spent at low intensity. The proportions of volume spent at low intensity are lower when using distance as a marker of volume and higher at moderate and high intensities, as compared to time. The combination of speed and distance may be the most relevant and practically feasible method to describe TID in soccer.

Keywords:

TID; Association Football; Training Load; Preparation; Off-season; Transition Period

Introduction

A competitive season in most professional European soccer leagues generally lasts over 10 months [25]. Over the course of a season, professional soccer players repeatedly have to perform at a high level, playing games once or twice a week, emphasizing the need to develop and maintain strong physical capacities [5, 29]. Next to the competitive season, two other periods can be distinguished: the transition period and preseason [25]. The transition period is characterized by a complete or substantial reduction in training, to allow recovery from the physiological and psychological stress of the competitive season [24]. The preseason that follows the transition period is used to prepare players for a new competitive season. Preseason schedules often consist of one to three training sessions a day, being not only soccer sessions or friendly games, but also physical fitness sessions such as running and strength sessions [15]. The combination of a decline in physical fitness because of reduced training during the transition period [21] and a high number of training sessions directly after the transition period [15] emphasizes the need for a well-planned preseason.

To maximize player fitness and availability, intensity and volume parameters are closely monitored during outdoor training-related activities [3]. Total training volume is often subdivided based on intensity thresholds [11], which are often heart rate (HR) [6] or speed based [1], allowing insight into the training intensity distribution (TID). TID can be defined as the intensity of exercise and its distribution over time [28]. Generally, a conceptual 3-zone intensity distribution model has been used in endurance sports [23], based on physiological and/or perceptual intensity thresholds, demarking the training intensity into a low (below the first ventilatory/lactate threshold), moderate (between the first and second ventilatory/lactate threshold) and high intensity zone (above the second ventilatory/lactate threshold). Previous studies in endurance sports [28] and soccer [4, 9] predominantly used HR as a marker of training intensity and time as a marker of training volume. However, practitioners in soccer often use the parameters speed and distance to monitor training load, which can be considered as alternative markers of respectively training intensity and volume [3]. The previously mentioned definition of TID, as proposed by Stöggl and Sperlich [28], using time as the sole marker of volume, should in the context of soccer therefore be broadened to include distance as an alternative marker of volume. The combination of speed with distance as markers of respectively intensity and volume to describe training load and TID was previously used by Lee and Mukherjee [17]. It is thus clear that there are more methods to describe TID in soccer, but unclear if there are differences in the TID when using different combinations of intensity (HR, speed) and volume parameters (time, distance) for the same external training load. Comparisons between studies using different methods to describe TID is problematic, since periodization of training can be markedly different depending on the philosophy of the coaching staff [9], thus not allowing an accurate comparison between studies using different methods to describe TID.

The aims of this study are to 1) extensively describe the training characteristics of the preseason in a professional soccer team and 2) describe and compare the TID using different combinations of intensity (HR, speed) and volume (time, distance) parameters. It is hypothesized that there are differences between the combinations of intensity and volume markers, showing different proportions of the volume spent at low, moderate and high intensities.

Materials and Methods

Participants

Twenty professional male soccer players (age 23.6 ± 3.7; range 18.3 - 31.9) participated in the preseason in a Belgian first division soccer team. Data from goalkeepers was excluded from the analysis. This study was approved by the ethical committee of Ghent University Hospital (approval number 2019-1198). Subjects were informed of the benefits and risks of the study prior to signing an institutionally approved informed consent document. This study was performed in accordance with the ethical standards of the Helsinki Declaration.

Design

Physical testing. Before the first day of the preseason, players performed several physical and medical tests. Anthropometric measurements and maximal incremental running tests were conducted following the conditions and protocols described by Boone et al. [5]. Players' body height (± 0.1 cm) and body weight (± 0.1 kg) was determined (Seca Balance). Body fat percentage was calculated by measurements of thickness of 8 skinfolds, using a Harpenden skinfold caliper [31]. The incremental running tests on a treadmill until volitional exhaustion were conducted to determine peak oxygen uptake (VO2peak), and HR, speed and VO2 at the first and second lactate threshold. The gradient of the treadmill was set at 1.5%, which was based on unpublished results from our laboratory. The gradient was used to compensate for the lower energy cost of indoor running [16], and as such, the speeds of incremental exercise test could be translated to outdoor running speed. Players' VO2peak was determined as the highest 30-second average of the breath-by-breath values obtained from the metabolic measurement system (Jaeger Oxycon Pro, Hochenhausen, Germany). The first lactate threshold was determined as the first increase in blood lactate concentrations above resting levels. The second lactate threshold was determined using the modified Dmax method [10].

Preseason. The preseason in this observational study spanned over a period of 36 days, ranging from the first day of training after the transition period until the last day of training before the first official game. The players were not only prepared for the club's national league, but also to play in the Europa League, an official tournament of the Union of European Football Associations (UEFA). During preseason, no activities were planned on Sunday, which resulted in 31 days with group-based activities. In total, 57 training session were performed, of which 50 were performed outdoor. Outdoor sessions were subdivided into soccer sessions (n=29), friendly games (n=11) and conditioning sessions (n=10). Soccer sessions were group-based sessions with the objective of improving all factors related to performance in soccer, including technical and tactical qualities. Friendly games were defined as games played against an opponent, outside the team's official (inter)national leagues or cup, using official soccer rules. Conditioning sessions consisted of all sessions aiming to improve players' physical qualities. These conditioning sessions were subdivided into running and complex sessions. Running sessions were defined as group-based sessions on the pitch or in the forest with the purpose to run at a pre-determined speed. Complex sessions were defined as group-based sessions, designed by the team's physical department, with the aim of improving a range of physical qualities using a combination of running, sprinting, jumping, flexibility, coordinative and functional strength drills. Indoor sessions comprised of strength sessions and injury prevention sessions. Strength sessions (n=7) were performed in the team's gym, aiming to improve the players' strength and power qualities. Players performed injury preventive exercises individually before most soccer sessions. All training sessions were prescribed and guided by the coaching staff.

GPS and TID. All team-based outdoor activities were undertaken with players wearing a sensor with a 10-Hz portable HR-monitor and 10-Hz GPS-unit (Polar Team Pro, Kempele, Finland). During strength and injury prevention sessions players did not wear a GPS/HR-sensor, because of the inability to establish a satellite connection by the GPS-units when an activity is performed indoor [20]. After each session, GPS- and HR-data were downloaded to a PC. The volume in each intensity zone was calculated based on the players' intensity thresholds (Visual Basic for Applications, Microsoft Office 365, Redmond, Washington, USA). TID was described using the following combinations between respectively intensity and volume parameters: HR distribution over time (HR-Time), HR distribution over distance (HR-Distance), speed distribution over time (Speed-Time), and speed distribution over distance (Speed-Distance).

Statistical analysis

Data was analysed using SPSS (Statistical Package for the Social Science, version 26.0). Descriptive analytics are reported as mean ± standard deviation. The means of the 4 combinations to describe TID were compared using the Kruskal-Wallis test, because the normality assumption was not met. Dunn's pairwise tests were carried out to detect differences between the combinations of intensity and volume parameters. Statistical significance was set at $P<0.05$. Because of the exploratory nature of the study, no alpha level adjustments were applied to correct for multiple testing, as it may lead to potentially missing meaningful findings in an exploratory and non-clinical setting [26].

Results

Players' anthropometric and physiological characteristics at the time of the maximal incremental running test are presented in Table 1.

Table 1: Overview on the average physical and physiological characteristics..

Antropometric characteristics (n=20)	
Length (cm)	179.1 ± 7.4
Weight (kg)	74.6 ± 7.7
Fat percentage (%)	9.5 ± 0.9
Values at volitional exhaustion	
VO_2peak (L/min^{-1})	4.2 ± 0.4
VO_2peak (ml/kg^{-1}/min^{-1})	57.0 ± 2.1
HR (bpm)	192.2 ± 10.7
[LA] (mmol/L)	11.1 ± 1.8
Ve (L/min)	143.1 ± 15.0
Values at the first lactate threshold	
VO_2 (L/min^{-1})	3.0 ± 0.3
VO_2 (ml/kg^{-1}/min^{-1})	40.5 ± 3.0
Speed (km/h)	10.6 ± 0.5
HR (bpm)	154.5 ± 8.5
VO_2 (%VO_2max)	70.6 ± 4.3
Values at the second lactate threshold	
VO_2 (L/min^{-1})	3.6 ± 0.4
VO_2 (ml/kg^{-1}/min^{-1})	48.8 ± 1.0
Speed (km/h)	13.7 ± 0.6
HR (bpm)	178.1 ± 9.2
VO_2 (%VO_2max)	85.6 ± 3.1

Abbreviations: VO_2 = oxygen consumption; VO_2peak = peak oxygen uptake; VO_2max = maximum oxygen uptake; bpm = beats per minute; [LA] = blood lactate concentration; Ve = pulmonary ventilation; %VO_2max = percentage of maximum oxygen uptake.

A total of 817 individual sessions were recorded with the Polar sensors, averaging 40.9 ± 5.8 sessions per player. A complete overview of all activities can be found in Figure 1. During the preseason, the average total distance covered and session duration per player during all outdoor team-based activities were respectively 254.0 ± 47.3 km and 53.8 ± 10.4 hours. Figure 2 shows the daily average total distance covered, duration of all outdoor team-based activities and the average distance covered per minute.

Differences in TID between the different combinations of intensity and volume markers are shown in Figure 3. The proportion of volume spent at low intensity was higher for Speed-Time than for HR-Time ($p<0.001$), HR-Distance ($p<0.001$) and Speed-Distance ($p<0.001$). The proportion of volume spent at low intensity was higher for HR-Time than for HR-Distance ($p=0.014$) and Speed-Distance ($p=0.002$). At moderate intensity, the proportion of volume spent was higher for HR-Distance than for HR-Time

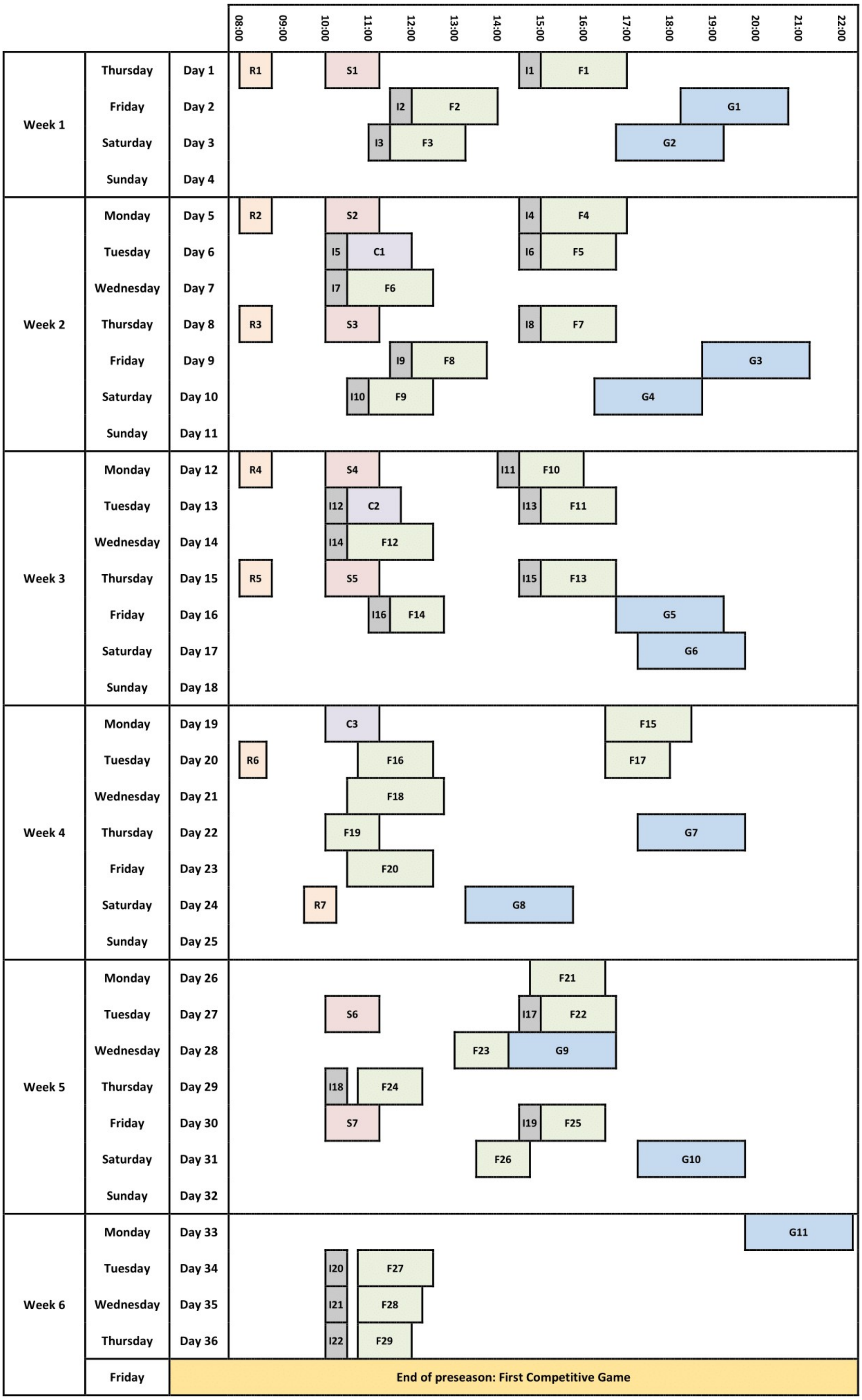


Figure 1: A complete overview of all group-based activities and individual injury prevention sessions during the preseason. Each letter represents a specific type of sessions: R = Running; S = Strength; I = Individual; C = Complex; F = Soccer; G = Game .

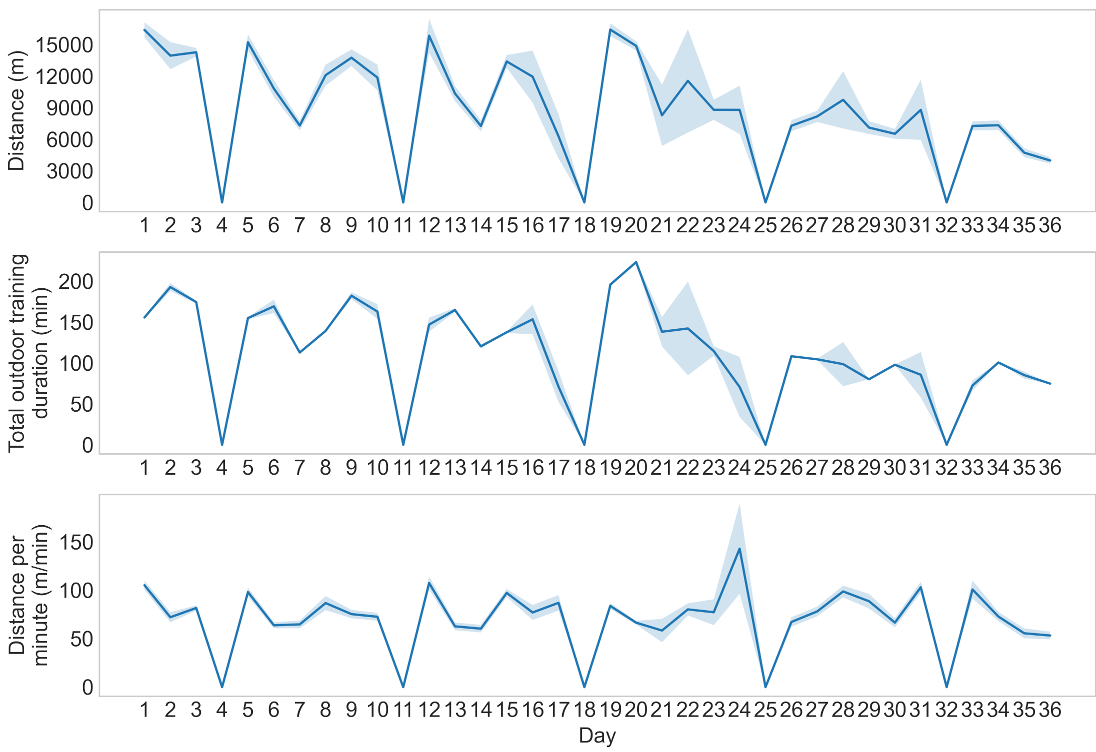


Figure 2: Overview on the average total distance covered, total outdoor training duration and distance per minute per day of the preseason. The error band shows the standard deviation.

(p=0.024), Speed-Time (p<0.001) and Speed-Distance (p=0.002), and for Speed-Time lower than for HR-Time (p<0.001) and Speed-Distance (p<0.001). At high intensity, the proportion of volume spent was higher for Speed-Distance than for HR-Time (p<0.001), HR-Distance (p<0.001) and Speed-Time (p<0.001). In Figure 4, an overview of the TID for each combination of the intensity and volume markers, provided for each individual player.

Discussion

The present study aimed to describe the training characteristics of the preseason in a first division Belgian soccer team, also participating in the Europa League, an official tournament of the Union of European Football Associations (UEFA). Secondly, this study aimed to describe and compare the TID using different combinations of intensity (HR and speed) and volume (time and distance) markers. The results showed that there are differences in TID depending on the used combination of intensity and volume markers. This study is, to the best of our knowledge, the first to compare different combinations of volume and intensity markers to describe TID in soccer.

The players' endurance capacities, such as VO2peak, and first and second lactate threshold, were assessed

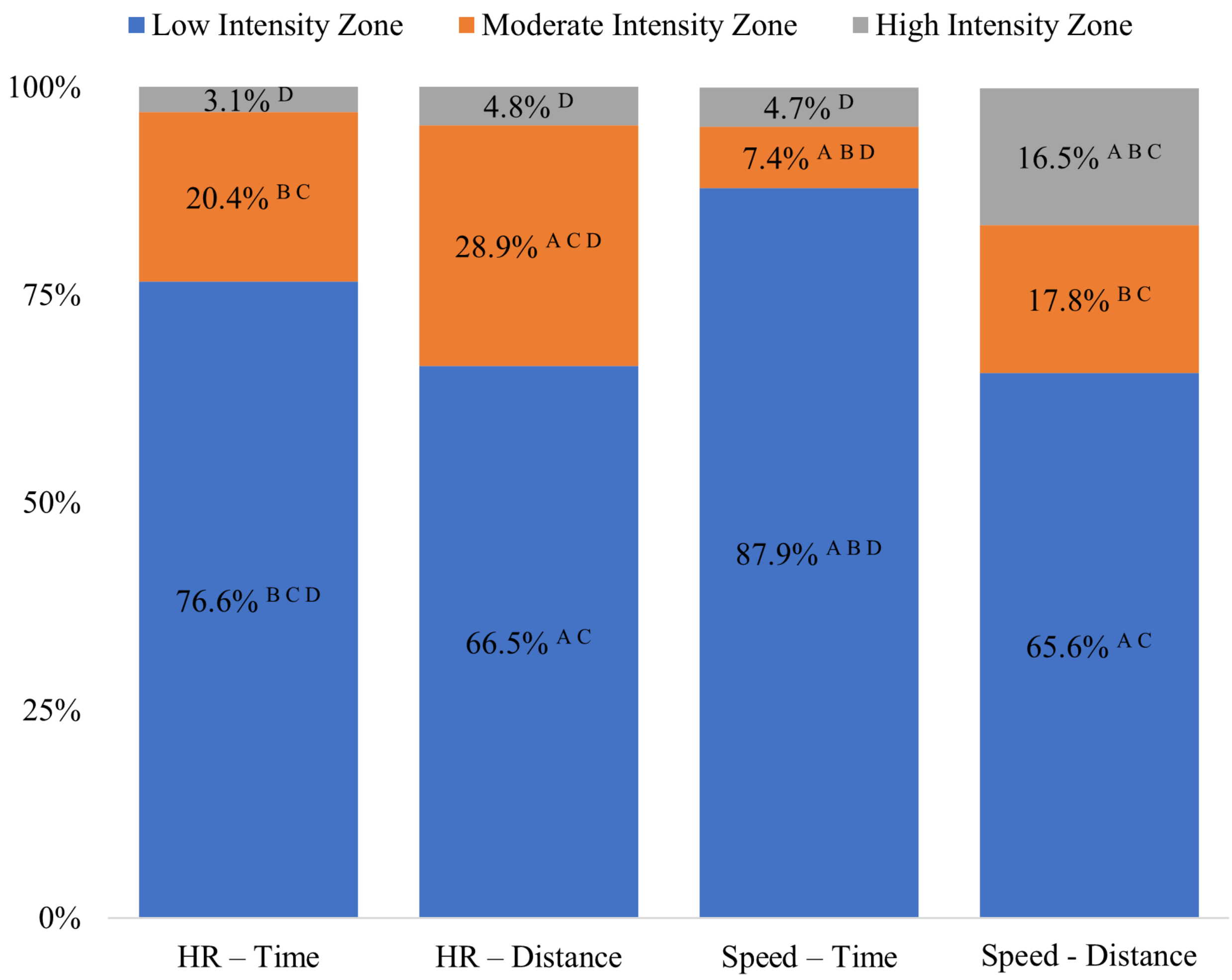


Figure 3: Differences between the combinations of intensity and volume parameters to describe TID. 'A' = different from HR-Time; 'B' = different from HR - Distance; 'C' = different from Speed - Time; 'D' = different from Speed - Distance.

before the start of the preseason. The endurance capacities showed to be comparable to that of players in several other international leagues [4, 9]. The duration of the complete preseason was 36 days, which is similar to the preseason duration of professional teams within other international leagues [9, 19]. It should however be noted that the duration of both the transition period and preseason depends on the duration of the competitive season [25], which typically lasts between August and May in the best leagues in Europe. The duration of the preseason should therefore be considered in context of the transition period, also allowing players sufficient time to recover from the physiological and psychological stress that accumulates during a season [24]. Although the in-season training characteristics are not described, it can be expected that the total number of weekly soccer-specific and conditioning sessions are higher during the preseason than during a regular week in-season, in line with previous research. Jeong et al. [15] showed that both training duration and intensity are generally higher during preseason than in-season. Since there are more training sessions and those sessions may be longer and at a higher intensity during preseason compared

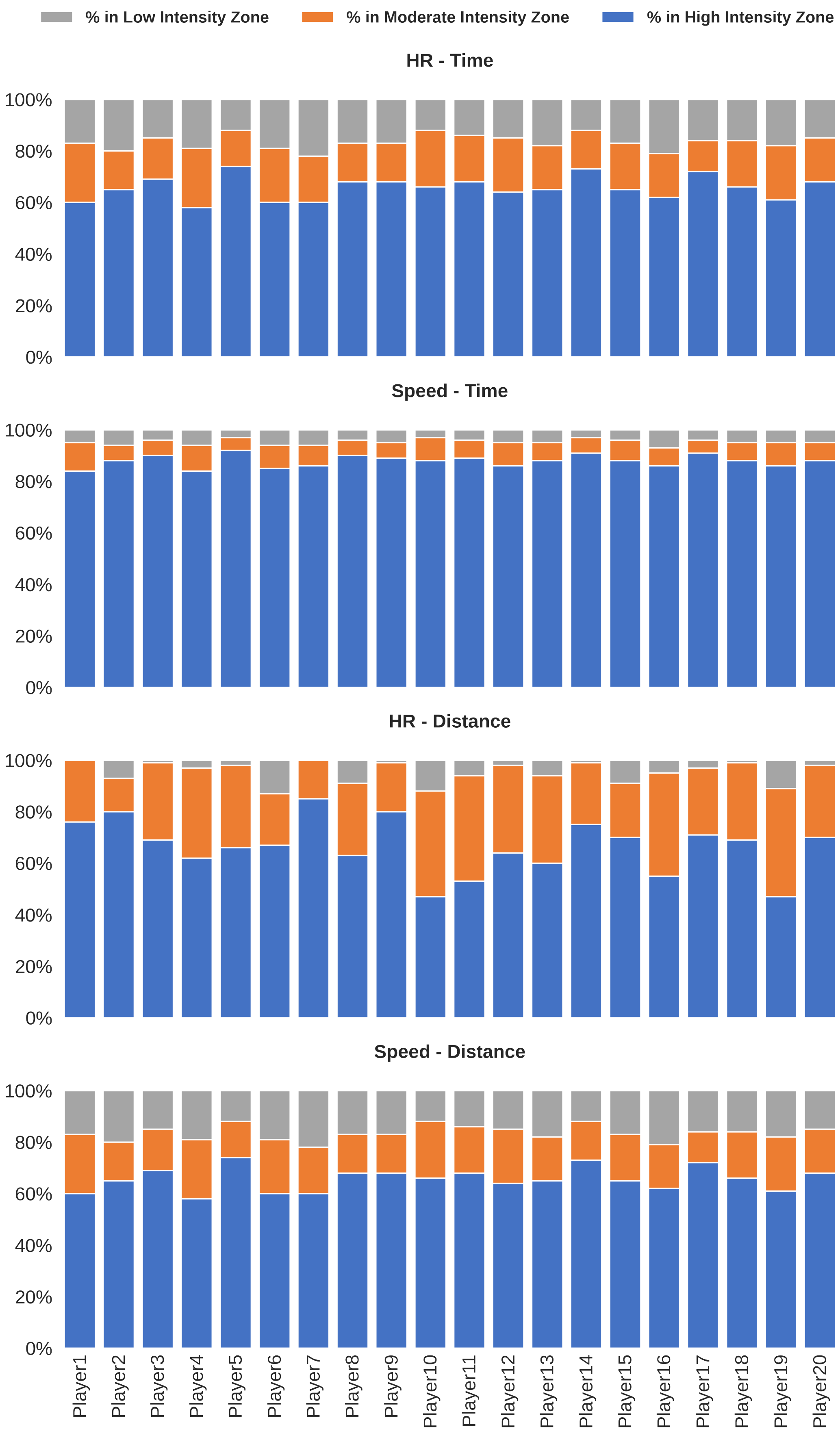


Figure 4: Overview of the TID for each method, provided for each individual player. The volume in each intensity zone is shown as a percentage.

to the in-season weekly training characteristics, it is important to carefully schedule the preseason, also bearing in mind that players' fitness is reduced as a consequence of a reduction in training during the transition period [21]. The biomechanical response rate, which is generally slower than the physiological response rate [30], should be considered during the scheduling process, allowing sufficient recovery between sessions. When the time and/or circumstances to recover between consecutive sessions are insufficient, the biomechanical adaptations may still be incomplete, which could lead to injury [30]. It is therefore important that conditioning sessions are complementary to the soccer sessions [14], to avoid large amounts of stress on similar physical components.

The total weekly training volume, expressed in distance, is lower for soccer players as compared to elite weight-bearing endurance athletes, such as marathon runners [12]. This may be partly explained by the fact that soccer players have to endure a high number of changes in speed, direction and impacts with other players [22], resulting in high biomechanical loads. These high biomechanical loads constrain the total training volume that can be endured [7]. Since the total training volume is constraint, coaching staff may opt to maximise the volume at moderate and high intensities, because high intensity efforts are generally perceived as important in soccer [8]. However, the dose-response characteristics of high intensity training, which specifically induces more functional cell- and organ-specific adaptions [7], saturates at fairly low levels of volume in highly trained athletes, while large volumes of training at moderate and high intensity can result in negative training effects [13]. This may explain why the results in this study and the findings of previous studies [4, 9, 17] (Figure 5), showed that the majority of the total training volume in professional soccer is spent at low intensity instead of moderate and/or high intensities. This should also be regarded as an important principle, since training volume spent at low intensity induces and synchronizes structural adaptations [7]. Differences in the proportions of volume in the three intensity zones between this study and previous studies describing TID in soccer [4, 9, 17], may be explained by factors such as differences in training philosophy [9], the level of competition and methodological differences. When scheduling the preseason or other mesocycles in the season, the distribution of volume in all intensity zones should therefore be carefully considered to optimize training effects. It is also apparent that variation in between the TID between players exists within the preseason. This could be explained by factors such as the number and type of training sessions that players conducted, position on the pitch or the physiology of the player. Monitoring the individual TID could therefore be interesting, in order to appropriately adjust the volume in certain intensity zones.

The results also showed that there are differences in the magnitude of the proportions spent at low, moderate, and high intensities between the combinations of intensity and volume parameters to describe TID. The proportion of total volume spent at low intensity is lower when using distance as parameter of volume, as compared to time. Consequently, the proportions of volume at moderate and/or high intensities

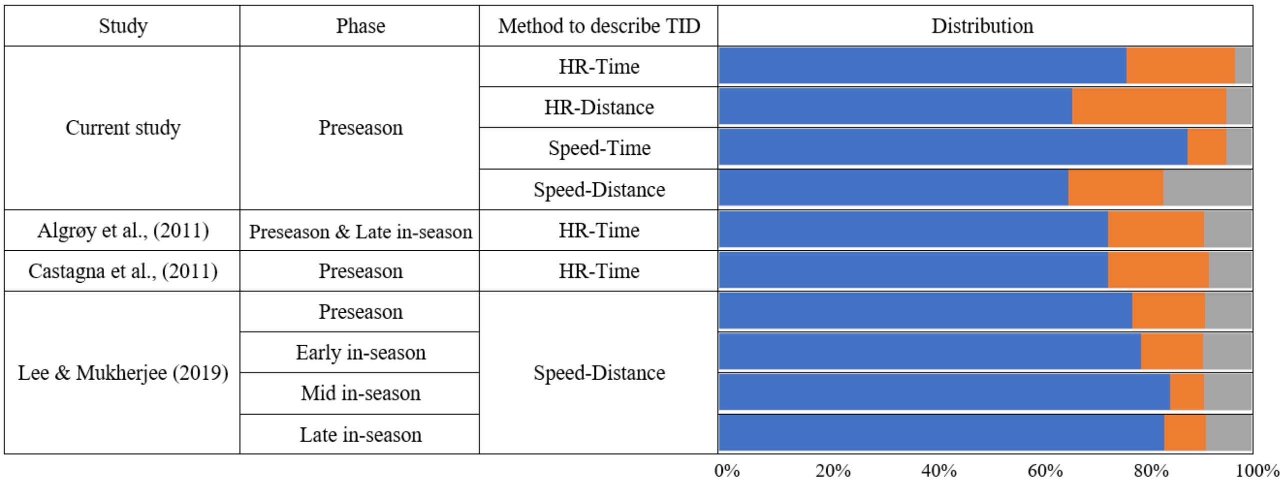


Figure 5: The TIDs reported by the current study and previous studies in soccer.

are higher. The use of time as parameter of volume in soccer may however be problematic because there are moments during a training session where players are not or less active, for example when the coaching staff provides instructions, during drinking breaks or when players must wait for their turn to re-participate in an exercise. During these moments, players are likely to stand still or move at low intensity, and as such this contributes to a higher proportion of time spent at low intensity. It can be argued that these moments are planned by the coaching staff and are therefore part of the training philosophy. However, in case of unplanned breaks or a complex organization of a training session, accurately filtering those moments may be practically difficult. When using distance as a parameter of volume, those breaks have a relatively small impact, since players are likely to stand still or only cover limited distances during these breaks. As a parameter of intensity, HR is commonly used when describing the TID in sports. The use of HR to describe intensity in soccer has however been a topic of discussion since HR can under- or overestimate intensity in soccer [2, 18]. During intermittent activities such as soccer, the use of speed as a measure of intensity may be more suitable, since it may better reflect the stochastic nature of soccer than HR. Also, there seems to be a greater interest of professionals working in elite soccer teams in expressing the intensity based on speed thresholds [3]. Given the advantages of speed and distance as respectively parameters of intensity and volume, this combination may be the most relevant and practically feasible method to describe TID in soccer.

Future research could monitor the TID over a prolonged period, thus including data from competitive games and in-season load. This may provide more information about possible differences in TID between different phases of the season, in line with the study of Lee and Mukherjee [17]. Secondly, the impact of not only strength and individual sessions, but also activities outside the training context could not be accurately taken into account when describing the TID, although also these activities provoke psycho-biological responses [27]. Researchers should however feel encouraged to find applicable methods to account for off-pitch and daily life activities when describing TID [27].

Conclusions

During the preseason in a professional soccer team, 57 training session were conducted over a period of 36 days, of which 50 session were performed outdoor. Players covered, on average, more than 250 kilometres during a total duration of more than 50 hours of outdoor training sessions. A large proportion of the total TID is spent at low intensity, regardless of the combination of intensity (HR, speed) and volume (time, distance) parameters. Speed as a marker of intensity may better reflect the stochastic nature of soccer in comparison to HR. Time as a marker of volume may overestimate the proportion of volume spent at low intensities. Speed and distance can be used as markers of respectively intensity and volume and may be in the context of soccer be the most relevant and practically feasible method to describe TID.

Acknowledgements

The authors would like to thank the staff and the players for their corporation.

Disclosure of Interest

The authors report no conflict of interest.

References

[1] Abt G and Lovell R. The use of individualized speed and intensity thresholds for determining the distance run at high-intensity in professional soccer. *Journal of Sports Sciences*, 27(9):893–898, 2009.

[2] Achten J and Jeukendrup AE. Heart rate monitoring: Applications and limitations. *Sports Medicine*, 33(7):517–538, 2003.

[3] Akenhead R and Nassis GP. Training load and player monitoring in high-level football: Current practice and perceptions. *International Journal of Sports Physiology and Performance*, 11(5):587–593, 2016.

[4] Algrøy EA, Hetlelid KJ, Seiler S, and Pedersen JI. Quantifying training intensity distribution in a group of norwegian professional soccer players. *International Journal of Sports Physiology and Performance*, 6(1):70–81, 2011.

[5] Boone J, Vaeyens R, Steyaert A, Vanden Bossche L, and Bourgois JG. Physical Fitness of Elite Belgian Soccer Players by Player Position. *Journal of Strength and Conditioning Research*, 26(8):2051–2057, 2012.

[6] Borresen J and Lambert MI. Quantifying Training Load: A Comparison of Subjective and Objective Methods. *International Journal of Sports Physiology and Performance*, 3(1):16–30, 2008.

[7] Bourgois JG, Bourgois G, and Boone J. Perspectives and Determinants for Training-Intensity Distribution in Elite Endurance Athletes. *International Journal of Sports Physiology and Performance*, 14(8):1151–1156, 2019.

[8] Carling C, Bloomfield J, Nelson LJ, and Reilly T. The role of motion analysis in elite soccer: Contemporary performance measurement techniques and work rate data. *Sports Medicine*, 38(10):839–862, 2008.

[9] Castagna C, Impellizzeri FM, Chaouachi A, Bordon C, and Manzi V. Effect of Training Intensity Distribution on Aerobic Fitness Variables in Elite Soccer Players: A Case Study. *Journal of Strength and Conditioning Research*, 25(1):66–71, 2011.

[10] Chalmers S, Esterman A, Eston R, and Norton K. Standardisation of the Dmax Method for Calculating the Second Lactate. *International Journal of Sports Physiology and Performance*, 10:921–926, 2011.

[11] Cummins C, Orr R, O'Connor H, and West C. Global Positioning Systems (GPS) and Microtechnology Sensors in Team Sports: A Systematic Review. *Sports Medicine*, 43(10):1025–1042, 2013.

[12] Enoksen E, Tjelta AR, and Tjelta LI. Distribution of training volume and intensity of elite male and female track and marathon runners. *International Journal of Sports Science and Coaching*, 6(2):273–293, 2011.

[13] Esteve-Lanao J, Foster C, Seiler S, and Lucia A. Impact of training intensity distribution on performance in endurance athletes. *Journal of Strength and Conditioning Research*, 21(3):943–949, 2007.

[14] Gamble P. Periodization of Training for Team Sports Athletes. *Strength and Conditioning Journal*, 28(5):56–66, 2006.

[15] Jeong TS, Reilly T, Morton JP, Bae SW, and Drust B. Quantification of the physiological loading of one week of "pre-season" and one week of "in-season" training in professional soccer players. *Journal of Sports Sciences*, 29(11): 1161–1166, 2011.

[16] Jones AM and Doust JH. A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *Journal of Sports Sciences*, 14(4):321–327, 1996.

[17] Lee M and Mukherjee S. Relationship of Training Load with High-intensity Running in Professional Soccer Players. *International Journal of Sports Medicine*, 40(5):336–343, 2019.

[18] Little T and Williams AG. Measures of Exercise Intensity During Soccer Training Drills With Professional Soccer Players. *Journal of Strength and Conditioning Research*, 21(2):367–371, may 2007.

[19] Malone JJ, Di Michele R, Morgans R, Burgess DJ, Morton JP, and Drust B. Seasonal Training-Load Quantification in Elite English Premier League Soccer Players. *International Journal of Sports Physiology and Performance*, 10(4):489–497, may 2015.

[20] Malone JJ, Lovell R, Varley MC, and Coutts AJ. Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *International Journal of Sports Physiology and Performance*, 12(S2):18–26, 2017.

[21] Mujika I and Padilla S. Detraining: Loss of Training-Induced Physiological and Performance Adaptations. Part I. *Sports Medicine*, 30(2):79–87, 2000.

[22] Russell M, Sparkes W, Northeast J, Cook CJ, Love TD, Bracken RM, and Kilduff LP. Changes in Acceleration and Deceleration Capacity Throughout Professional Soccer Match-Play. *Journal of Strength and Conditioning Research*, 30 (10):2839–2844, 2016.

[23] Seiler S, Haugen O, and Kuffel E. Autonomic recovery after exercise in trained athletes: Intensity and duration effects. *Medicine and Science in Sports and Exercise*, 39(8):1366–1373, 2007.

[24] Silva JR, Rebelo A, Marques F, Pereira L, Seabra A, Ascensão A, and Magalhães J. Biochemical impact of soccer: an analysis of hormonal, muscle damage, and redox markers during the season. *Applied Physiology, Nutrition, and Metabolism*, 39(4):432–438, apr 2014.

[25] Silva JR, Brito J, Akenhead R, and Nassis GP. The Transition Period in Soccer: A Window of Opportunity. *Sports Medicine*, 46(3):305–313, 2016.

[26] Sinclair J, Taylor PJ, and Hobbs SJ. Alpha Level Adjustments for Multiple Dependent Variable Analyses and Their Applicability – A Review. *International Journal of Sports Science and Engineering*, 7(1):17–20, 2013.

[27] Sperlich B and Holmberg HC. The Responses of Elite Athletes to Exercise: An All-Day, 24-h Integrative View Is Required! *International Journal of Sports Physiology and Performance*, 8(1):4–18, jan 2013.

[28] Stöggl TL and Sperlich B. The training intensity distribution among well-trained and elite endurance athletes. *Frontiers in Physiology*, 6:295, oct 2015.

[29] Stolen T, Chamari K, Castagna C, and Wisloff U. Physiology of Soccer. *Sports Medicine*, 35(6):501–536, 2005.

[30] Vanrenterghem J, Nedergaard NJ, Robinson MA, and Drust B. Training Load Monitoring in Team Sports: A Novel Framework Separating Physiological and Biomechanical Load-Adaptation Pathways. *Sports Medicine*, 47(11):2135–2142, 2017.

[31] Wallace J, Marfell-Jones M, George K, and Reilly T. A comparison of skinfold thickness measurements and dual-energy x-ray absorptiometry analysis of percent body fat in football players. In *VIth World Congress on Science and Football*, volume 1, pages 66–70, Antalya, Turkey, 2007.

Study III

Modeling the Prediction of the Session Rate of Perceived Exertion in Soccer: Unravelling the Puzzle of Predictive Indicators

Geurkink, Y., Vandewiele, G., Lievens, M., De Turck, F., Ongenae, F., Matthys, S.P.J., Boone, J., & Bourgois, J. G. (2019). Modeling the prediction of the session rating of perceived exertion in soccer: Unraveling the puzzle of predictive indicators. *International journal of sports physiology and performance*, 14(6), 841-846.

Youri Geurkink [1,2], Gilles Vandewiele [3], Maarten Lievens [1], Filip de Turck [3], Femke Ongenae [3], Stijn P.J. Matthys [1,2], Jan Boone [1 *] and Jan G. Bourgois [1,2 *]

[1] Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

[2] KAA Ghent (Football Club),Ghent, Belgium

[3] Department of of Information Technology, Ghent University, Ghent, Belgium

* These authors share last authorship

Abstract

Purpose: To predict the session Rate of Perceived Exertion (sRPE) in soccer and determine its main predictive indicators. Methods: A total of 70 external load indicators (ELIs), internal load indicators (ILIs), individual characteristics, and supplementary variables were used to build a predictive model. Results: The analysis using gradient-boosting machines showed a mean absolute error of 0.67 (±0.09) Arbitrary Units (AU) and a root-mean-square error of 0.93 (±0.16) AU. ELIs were found to be the strongest predictors of the sRPE, accounting for 61.5% of the total Normalized Importance (NI), with total distance as the strongest predictor. The included internal-load indicators and individual characteristics accounted only for 1.0% and 4.5%, respectively, of the total NI. Predictive accuracy improved when including supplementary variables such as group-based sRPE predictions (10.5% of Normalized Importance (NI)), individual deviation variables (5.8% of NI), and individual player markers (17.0% of NI). Conclusions: The results showed that the sRPE can be predicted quite accurately using only a relatively limited number of training observations. ELIs are the strongest predictors of the sRPE. However, it is useful to include a broad range of variables other than ELIs, because the accumulated importance of these variables accounts for a reasonable component of the total NI. Applications resulting from predictive modeling of the sRPE can help coaching staff plan, monitor, and evaluate both the external and internal training load.

Keywords:

sRPE, training load, machine learning, soccer, team sports

Introduction

Soccer training consists of structurally and systematically performed general and specific exercises to improve physical abilities and acquire skills. The adaptations imposed by a training stimulus take place on an anatomical, physiological, biochemical, molecular, and functional level [28]. The content of a training session is usually prescribed by the coach and can be defined as the external training load. The total external training load comprises all of the players' actions during a training session [19] and is generally quantified using tracking technology [10]. The external training load elicits an internal physiological stress, or the internal training load. However, the internal training load is not only dependent on the imposed external training load, but also on the players' Individual Characteristics (ICs) [19].

It is possible to assess the internal training load through quantification of a training session's duration and intensity [19]. Duration is quantifiable in time and relatively easy to measure. On the other hand, intensity can be quantified using different methods, such as HR monitoring, blood lactate concentrations, and the session rating of perceived exertion (sRPE) [19]. Heart rate (HR) monitoring is widely used in soccer, but it has been suggested that HR monitoring underestimates or overestimates the intensity during intermittent activities [2, 21]. Furthermore, HR monitoring requires both technical and physiological expertise to make an appropriate analysis [1]. The blood lactate concentration is not often monitored during soccer training for practical reasons [19]. On the other hand, the sRPE is a simple and practical tool that represents the players' own perceptions of training stress, which include both physiological and psychological stress [18]. The sRPE has shown to be a valid indicator of intensity in soccer [6, 14, 15, 18] and a more valid marker of exercise intensity in soccer over a broad range of activities than HR monitoring [21].

The external training load is the first factor influencing the internal training load [19] and has previously been investigated by Gaudino et al. [14]. They confirmed that the sRPE is related to external load indicators (ELIs), such as high-speed distance, impacts, and accelerations. Second, the players' ICs influences the internal training load. Because of the variability in adaptations to a training stimulus [26], coaches should consider the players' Individual Characteristics (ICs) when prescribing the external training load. It has been suggested that fitter athletes within a team may not receive the optimal stimulus for physiological adaptations through extensive use of group training exercises [16]. On the other hand, players with inferior fitness may be overstressed through group training sessions [19]. To provide an appropriate training stimulus, it is important to establish a level of agreement between the athlete's sRPE and the coach's sRPE. However, previous research has shown that, in several sports, coaches tend to underestimate or overestimate the athlete's sRPE [7, 11, 22].

Based on the previous statement, it seems that coaches can experience difficulties controlling the external training load and may put athletes at risk of maladaptive responses to training [29], such as fatigue, injury, and a reduction in performance [19]. To further understand the underlying indicators of external training load, internal training load, and ICs contributing to the sRPE, this study aims to predict the sRPE and identify the main predictors of the sRPE.

Methods

Subjects

A total of 46 players (age = 25.6 [4.2] y) from an elite Belgian soccer team participated in this study. Prior to the start of the study, players were informed about the study protocol and the criteria set to assess the sRPE. Data were obtained between June 2015 and March 2017. To ensure a higher level of homogeneity concerning the experimental data, goalkeepers were excluded from this study, given the differences in external training load of goalkeepers compared with the other playing positions. This study was approved by the ethical committee of the Ghent University Hospital. Before the start of the study, all participants signed an informed consent.

Design

During 61 training sessions, a total of 913 individual training observations were obtained. The analyzed training sessions were prescribed and guided by the coaching staff, without interference of the research staff. Strength and recuperation sessions were excluded to ensure a greater similarity between training sessions. During the training sessions, players' HR was assessed through 20-Hz portable HR-monitors and global positioning systems (GPS; Polar Team Pro, Kempele, Finland). GPS-units with a frequency of 10 Hz were used; the accelerometers inside the Global Positioning System (GPS)-units had a frequency of 100 Hz. No control subject participated in this observational study.

Methodology

Session Rating of Perceived Exertion. The sRPE was assessed 15 minutes after each training session. Players were asked to rate the training session between 1 and 10 using the sRPE scale [13]. All players were familiarized with the protocol.

External-Load Indicators. Using the GPS-units, the following variables were measured during the training sessions: total distance (m), training duration (s), distance (m) in 5 speed zones (3.00–6.99 km/h, 7.00–10.99 km/h, 11.00–14.99 km/h, 15.00–18.99 km/h, and >19.00 km/h), the number of accelerations (m/s^2) (0.50–0.99, 1.00–1.99, 2.00–2.99, and 3.00–50.00), the number of decelerations (m/s^2) (0.50–0.99, 1.00–1.99, 2.00–2.99, and 3.00–50.00), and the number of sprints (>25 km/h). Average speed (m/s) was derived using distance and time. The proportion of workload in every acceleration, deceleration, and speed zone relative to the total workload was also included into the predictive model. Finally, individual deviations of several External-Load Indicators (ELIs) (total distance, total time, and number of sprints) compared with the group mean, based on historical training data, were derived. This way, the model could take differences in external workload for players compared with the group mean into account.

Internal-Load Indicators. Data concerning HR was subdivided into 5 different HR-zones relative to their maximum HR (50–60%, 60–70%, 70–80%, 80–90%, and 90–100%) (s). The maximum HR was determined during an incremental test protocol on a treadmill. The proportion of the time spent in each HR-zone relative to the total time spent in the 5 HR-zones and the Edwards' training impulse (TRIMP) were also included into the model. Edwards' TRIMP is expressed as the product of the accumulated training duration in each HR-zone, with a coefficient relative to each HR-zone (50–60%= 1, 60–70%= 2, 70–80%= 3, 80–90%= 4, and 90–100% = 5) [12].

Individual Characteristics. The ICs can be subdivided into physiological and personal characteristics. The physiological characteristics consisted of assessment of conditional parameters, sprinting speed, acceleration, anaerobic power, and muscle fiber composition. An assessment of conditional parameters was conducted on a treadmill using an incremental test protocol, starting at 8 km/h, with an increase in speed of 2 km/h every 3 minutes. The sprinting tests consisted of 2 different exercises, namely, a 10-m sprint and a 5-times-10-m shuttle run. The 10-m sprint was used as a measure of starting speed and acceleration, whereas the shuttle run provided information about the players' agility in combination with speed. The jumping test was used as a measure of anaerobic power and consisted of a countermovement jump with arms. The complete test battery was performed on the same day. The participants did not perform strenuous exercise 24 hours before testing. The tests were conducted every 6 months, and changes in test performances were registered. Because of injuries or decreased fitness status, 5 out of the 46 players were unable to perform the sprinting tests. For the complete description of the tests, the experimental protocol from Boone et al. [5] can be consulted.

Muscle fiber composition was estimated in a noninvasive way, through measurements of the carnosine content in the gastrocnemius and soleus by proton magnetic resonance spectroscopy. For a complete description of the protocol, Baguet et al [3] can be consulted. The muscle fiber distribution was only measured once, as muscle fiber is largely dependent on genetics [24], and substantial changes in the distribution of muscle fibers because of training are, in our case, unlikely. Because of unavailability of the magnetic resonance imaging scanner, muscle fiber type was not determined in 21 out of the 46 players.

The personal characteristics consisted of age, playing position, and nationality. Players' age was determined during each training observation. Playing position was subdivided into 5 categories (central defender, full back or wide midfielder, central midfielder, winger, and central attacker). Nationality was defined using the latitude and longitude of each player's country of birth.

Supplementary Variables. Several variables and techniques were used in an attempt to improve predictive accuracy and were categorized as "supplementary variables." First, group-based sRPE prediction variables were included. Based on the average external and internal workload, several predictive models were used to generate a group-based prediction of the sRPE. The following models were used to predict the group-based sRPE: generalized additive models, multivariate adaptive regression splines, decision tree, random forest, linear regression, and support vector regression. The predicted group-based sRPE values were then included into the model. Each player's individual deviance from the group mean (mean deviation, SD, maximum deviation, and minimum deviation) was also included, as individual differences in the interpretation of the sRPE scale can influence the sRPE value provided by the players. To further account for differences between players, the effect of every player on the predictive accuracy was quantified using 1-hot encoding. One-hot encoding is defined as mapping of a variable to a binary vector of length equal to the number of categories, in our case, the number of players (n = 46). All elements in the vector are converted to zero except at the index corresponding to the category of that sample. This technique created 46 features for the model but, in this study, they are regarded as 1 variable. Finally, 4 variables regarding weather (temperature, humidity, visibility, and wind speed) were included to account for the effect of weather on the perceived training load.

Statistical Analyses

Gradient boosting machines were used to identify the main predictive indicators of the sRPE and to predict the sRPE. This machine-learning technique creates a large number of different decision trees, resulting in an integrated model predicting the outcome. Missing values for muscle fiber type, sprinting times, and jumping performance were replaced with the group mean. Data were analyzed using Python 3.5 (Python Software Foundation, Wilmington, DE).

To evaluate predictive accuracy, 2 different standard statistical metrics were used, namely, the mean absolute error and the root mean squared error. It was opted to also express model accuracy in the number of correct classifications. The predicted value was a rounded number, which could be evaluated if the model predicted correctly. To provide more information about the deviation from the mean, a "loose accuracy" approach was used. This approach classified a difference within a range of 1 as correct, when the predicted value was compared with the observed sRPE. For example, if the observed value was 4, all predicted values between 3 and 5 were classified as correct.

Measurements were taken using 5-fold cross-validation, in which the data were partitioned according to training identifiers. In other words, on average, 50 and 13 training sessions were used, respectively, in the training and test set in each fold. This was done to avoid contamination between the training and test set. The influence of each included indicator was expressed as a normalized importance (NI) value. The importance of each variable is a measure of the magnitude by which the model-predicted sRPE value is altered for various values of a variable. In our case, the NI was determined through the number of expressions of a variable in the created decision trees.

Results

The average training sRPE was 4.34 (±1.06) AU; the distribution of sRPE values can be found in (Figure 1). The median number of records per player was 18 (ranging from 2 to 46). Other descriptive values of regularly reported ELIs were: duration 90 (±20) minutes and total distance covered 5977 (±1893) m.

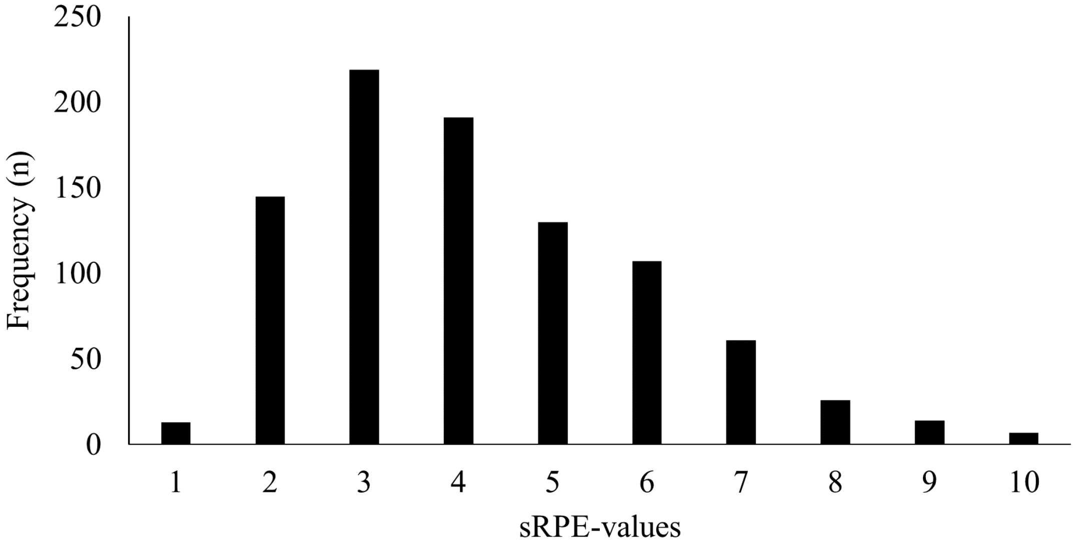


Figure 1: Distribution of observed sRPE-values from 46 players, resulting in 913 training observations.

Predicting the sRPE The predictive model showed a mean absolute error of 0.67 (±0.09) AU and a root-mean-square error of 0.93 (±0.16) AU. In total, 47.6% (±7.17%) of the cases were correctly classified. The "loose accuracy" approach, classifying all predictions within a range of ±1 relative to the observed sRPE score as correct, resulted in 91.7% (±3.45%) correctly classified cases. The model did not predict sRPE values of 8, 9, or 10 because of the limited number of observed high sRPE values. In (Figure 2), a confusion matrix is depicted showing the observed versus the predicted sRPE values.

Predictive Indicators of the sRPE Most of the strongest individual predictors of the sRPE can be regarded as ELIs, as total distance, total time, and number of sprints are among the strongest predictors. The NI of all ELIs together accounted for 61.5%. The sRPE showed a positive relationship with most measures, presenting an increased sRPE when external workload increased. Players' ICs accounted only for 4.5% of the total NI. HR showed to be a poor predictor of the sRPE, as internal load indicators (Internal-Load Indicators (ILIs)) only accounted for 1.0% of the total NI.

The supplementary variables were subdivided in machine learning variables (10.5% of NI), individual deviation variables (5.8% of NI), individual player markers (17.0% of NI), and weather variables (0.2% of NI), which accounted for 33% of the total NI. An overview of the NI of all indicators can be found in Table 1.

Table 1: Overview of the Normalized Importance (NI) for all indicators

Category	Indicator	NI	Description
ELIs	Total distance	0.250	Total distance (m) covered during a training session.
	Total duration	0.067	Total time (s) during a training session.
	Number of sprints	0.072	Total number of sprints (N) during a training session.
	Deviation of total distance	0.061	Differences in total distance between players during a training session.
	Deviation of total sprints	0.109	Differences in the total number of sprints between players during a training session.
	Other ELIs (n = 28)	0.058	Average speed, distance covered in 5 speed zones (km/h) (3.00-6.99, 7.00-10.99, 11.00-14.99, 15.00-18.99, and >19.00); normalized proportion of distance covered in each speed zone, number of accelerations in each zone (m/s2) (0.50–0.99, 1.00–1.99, 2.00–2.99, and 3.00–50.00); number of decelerations (m/s2) (0.50–0.99, 1.00–1.99, 2.00–2.99, and 3.00–50.00); normalized proportion of number of accelerations and decelerations in each zone, differences in training duration between players during a training session.
	Sum of all ELI variables (n = 33)	**0.615**	
ICs	Physiological characteristics (n = 8)	0.013	VO_2max, speed at aerobic and anaerobic threshold, muscle fiber composition, sprint time (time at 5 and 10 m), 5-times-10-m shuttle run, and CMJ with arms.
	Personal characteristics (n = 3)	0.032	Age, nationality, and playing position (central defender, full back, central midfielder, winger, and striker).
	Sum of all IC variables (n = 11)	**0.045**	
ILIs	HR variables (n = 11)	0.010	Time spent in 5 HR-zones relative to the maximum HR (50–60%, 60–70%, 70–80%, 80–90%, and 90–100%), proportion of time spent in each HR-zone, training load based on Edwards' TRIMP.
	Sum of all ILI variables (n = 11)	**0.010**	
SVs	Group-based RPE-predictions (n = 6)	0.105	The sRPE of every training session was predicted by the following models: generalized additive model, multivariate adaptive regression spline, decision tree, random forest, linear regression, and support vector regression.
	Individual deviation variables (n = 4)	0.058	Individual sRPE deviances from group mean (mean, maximum, minimum, and SD).
	Individual player marker (n = 1)	0.170	Each player's individual influence on predictive accuracy was quantified by 1-hot encoding.
	Weather variables (n = 4)	0.002	Weather circumstances (temperature, humidity, visibility, and wind speed).
	Sum of all SV variables (n = 15)	**0.330**	

Abbreviations: ELIs, external-load indicators; HR, heart rate; ICs, individual characteristics; Internal-Load Indicators (ILIs), internal-load indicators; sRPE, session rating of perceived exertion; SVs, supplementary variables; VO_2max, maximal oxygen uptake.

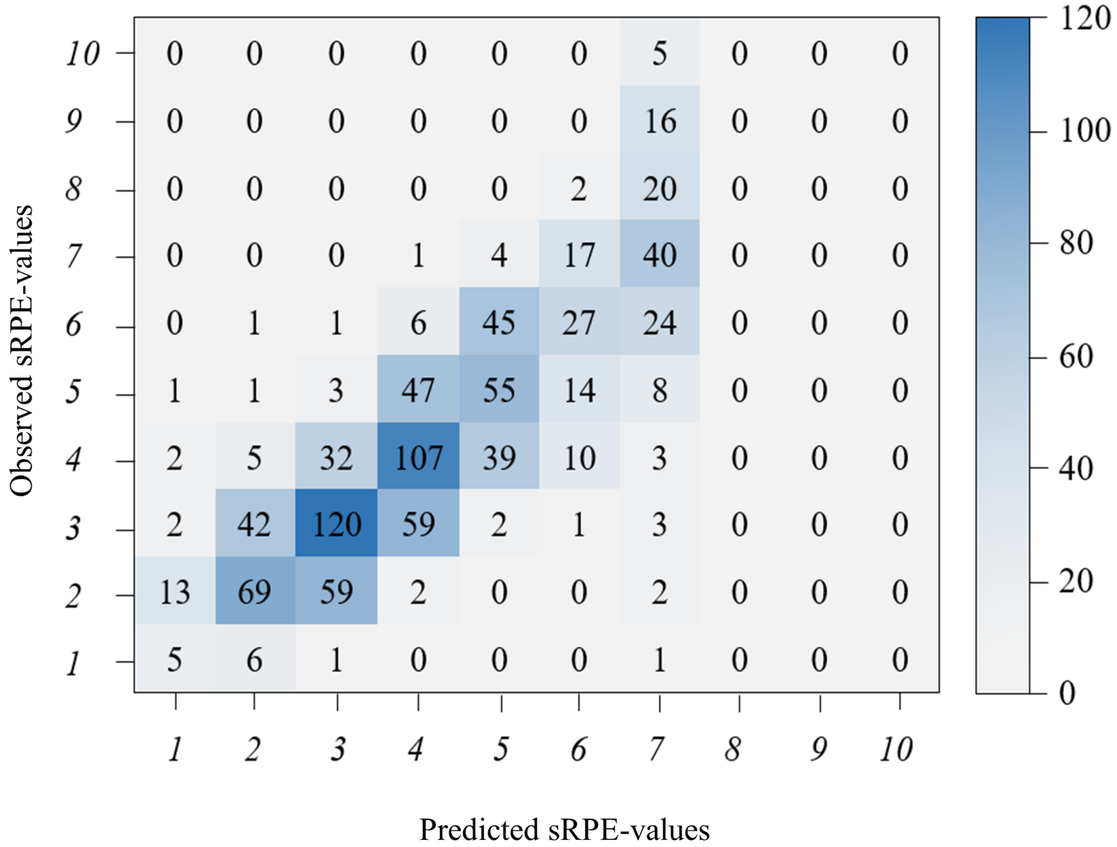


Figure 2: Confusion matrix showing the observed versus the predicted sRPE-values. The matrix provides insight into the predictive model by showing the number of correct and incorrect prediction for each sRPE-value.

Discussion

The aim of this study was to predict the sRPE and determine the main predictive indicators. The sRPE was predicted quite accurately, using a broad range of variables. To our knowledge, this study in soccer is the first to incorporate a large set of predictive indicators other than ELIs, namely, ICs, ILIs, and supplementary variables. The findings demonstrate that ELIs are the strongest predictors of the sRPE. It is also useful to improve predictive accuracy by including machine-learning variables, individual deviation variables, and individual players' markers. The included ILIs and ICs showed poor predictive value.

The findings from this study show higher predictive accuracy than previous studies conducted in soccer [20, 23]. Despite a smaller data set, our results showed a lower mean absolute error and root mean squared error. This may be explained using different machine-learning techniques and/or other predictive indicators. Our findings showed that even on a relatively small data set, the sRPE can be predicted quite accurately. This may offer solutions to (sub)elite teams, especially when teams consist of large numbers of players. The sRPE is regularly monitored in team sports; however, it can take considerable time and effort [8]. The model used in this research could help to predict players' sRPE, which might make daily assessment of the sRPE unnecessary [9]. The coaching staff may choose to closely monitor the external and internal training load for a limited period to build a model that is appropriately predictive. After this period, the coaching staff can choose to monitor the predicted sRPE, instead of daily collection of the sRPE. The constructed model may also form a basis for live monitoring of the sRPE during a training session [9, 27]. This would allow coaches to adapt players' training content, possibly preventing negative training effects. However, live monitoring of the sRPE may not only provide live information but could also provide information regarding the process of accumulating external training load on the sRPE. Currently, the sRPE only provides a global indication about the "product" of the internal training load, rather than provide information about the dynamical "process" in which a certain internal training load is achieved. As different combinations of external training load can elicit a similar sRPE, more insight into temporal changes of the sRPE during training sessions could help the coaching staff to prescribe and adapt the external training load.

Previous research already showed that ELIs can be used to predict the sRPE in soccer [20, 23]. Our results showed that total distance was the strongest predictor of the sRPE, while total duration and the number of sprints were also strong predictors. The importance of these factors in relation to perceived training load can be explained by the fact that these measures capture information about both session volume and intensity [9]. However, it should be noted that findings from our research show some differences in regard to previous research [20, 23], as the importance assigned to each ELI varies between studies. Differences between studies may be explained by differences in training content, the use of different devices to register players' activity, or different modeling approaches. Although several findings with regard to variable importance can be generalized, the differences in variable importance between studies may imply limited transferability of predictive models. However, previous studies did not include variables other than ELIs. As these variables account for 38.5% of the NI, it is useful to include a broad range of variables with a possible impact on the perceived exertion. However, the ICs, ILIs, and weather variables showed only small predictive value. A set of variables that showed promise are the individual deviation variables and the individual players' markers (5.8% and 17.0% of the total NI, respectively). The NI values assigned to these variables show that there are differences in interpretation of the sRPE-scale, as some players consistently rate a training session higher or lower than the group mean. This corresponds with a previous statement by Bartlett et al [4], endorsing the importance of personal interpretation of the sRPE scale and an individualized interpretation of the perceived exertion.

There are some limitations and opportunities to be stated. A limitation of this study is the small number of extreme sRPE values, especially on the higher end of the continuum. For that reason, the model was adapted to predict sRPE scores between 1 and 7, which explains the error for sRPE values of 8, 9, or

10. However, extreme sRPE values could provide very useful information to both the model as well as the coaching staff. Furthermore, the error was larger when predicting lower sRPE values. Future studies should include more extreme sRPE values. The results also showed that it is useful to include a broad range of variables into a predictive model. Including more variables may improve model accuracy, offering opportunities for future research. In this study, the external training load was not registered over a prolonged series of training sessions, so changes in external training load over time, which can be quantified using measures such as the acute/chronic workload ratio [17] and exponentially moving weighted averages [30], are not considered. Finally, information about the training modalities could be quantified to provide useful information to the predictive model. When including more factors in a predictive model, researchers should consider the methodological quality, but also the practicality, of the predictive factor. Baseline feeling or lifestyle involving, for example, diet or rest, may likely have an influence on the sRPE. However, these factors are practically more difficult to quantify. Subjective ratings may, at the moment, be a suitable alternative when attempting to get quantitative insight into lifestyle, as previous research showed that morning-measured ratings of fatigue, sleep quality, and delayed onset of muscle soreness are sensitive markers to daily fluctuations in the sRPE in elite soccer players [25]. Although this study may already be regarded as fairly comprehensive, given the number of ICs, variables such as the aforementioned can also be incorporated to further improve model accuracy and explore the role of the different predictive indicators.

Practical Implications

Predictive modeling of the sRPE may aid the process of planning and adjusting training load. In our model, regularly monitored ELIs, such as distance, training duration, the number of sprints, and derivatives of these ELIs were found to be strong predictors, showing that it is important to carefully consider these ELIs when planning and monitoring training load. Furthermore, differences between individuals on a physiological and personal basis, but also differences when interpreting the sRPE scale, can influence the sRPE values. This study provides further confirmation for an individualized approach when planning and monitoring training load.

Conclusions

Our predictive model showed that the sRPE can be predicted quite accurately using only a relatively small number of training observations. ELIs are the strongest predictors of the sRPE, with total distance as the strongest predictor. Including a broad range of variables, other than ELIs, is useful, as the accumulated importance of these variables accounts for a reasonable component of the total NI. Applications resulting from predictive models may help a coaching staff to plan, monitor, and evaluate training sessions. Future research could focus on including more variables into the model, such as factors of daily status and changes in external training load over time, to create an even more comprehensive model predicting the sRPE.

Acknowlegdements

The authors would like to thank the coaches and players for their cooperation.

J.B. and J.G.B. share last authorship.

References

[1] Alexiou H and Coutts AJ. A comparison of methods used for quantifying internal training load in women soccer players. *International Journal of Sports Physiology and Performance*, 3(3):320–330, 2008.

[2] Algrøy EA, Hetlelid KJ, Seiler S, and Pedersen JI. Quantifying training intensity distribution in a group of norwegian professional soccer players. *International Journal of Sports Physiology and Performance*, 6(1):70–81, 2011.

[3] Baguet A, Everaert I, Hespel P, Petrovic M, Achten E, and Derave W. A New Method for Non-Invasive Estimation of Human Muscle Fiber Type Composition. *Plos One*, 6(7):6, 2011.

[4] Bartlett JD, O'Connor F, Pitchford N, Torres-Ronda L, and Robertson SJ. Relationships between internal and external training load in team-sport athletes: Evidence for an individualized approach. *International Journal of Sports Physiology and Performance*, 2017.

[5] Boone J, Vaeyens R, Steyaert A, Vanden Bossche L, and Bourgois JG. Physical Fitness of Elite Belgian Soccer Players by Player Position. *Journal of Strength and Conditioning Research*, 26(8):2051–2057, 2012.

[6] Borresen J and Lambert MI. Quantifying Training Load: A Comparison of Subjective and Objective Methods. *International Journal of Sports Physiology and Performance*, 3(1):16–30, 2008.

[7] Brink MS, Frencken WG, Jordet G, and Lemmink KA. Coaches' and players' perceptions of training dose: Not a perfect match. *International Journal of Sports Physiology and Performance*, 2014.

[8] Burgess DJ. 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*, pages 1–19, 2016.

[9] Carey DL, Ong KL, Whiteley R, Crossley KM, Crow J, and Morris ME. Predictive modelling of training loads and injury in Australian football. *International Journal of Computer Science in Sport*, 17(1):49–66, 2017.

[10] Cummins C, Orr R, O'Connor H, and West C. Global Positioning Systems (GPS) and Microtechnology Sensors in Team Sports: A Systematic Review. *Sports Medicine*, 43(10):1025–1042, 2013.

[11] Doeven SH, Brink MS, Frencken WG, and Lemmink KA. Impaired Player-Coach Perceptions of Exertion and Recovery During Match Congestion. *International journal of sports physiology and performance*, pages 1–20, 2017.

[12] Edwards S. The heart rate monitor book. *Medicine & Science in Sports & Exercise*, 26(5):647, 1994.

[13] Foster C, FLorhaug JA, Franklin J, Gottschall L, Hrovatin LA, Parker S, Doleshal P, and Dodge C. A New Approach to Monitoring Exercise Training. *Journal of Strength and Conditioning Research*, 15(1):109–115, 2001.

[14] Gaudino P, Iaia FM, Strudwick AJ, Hawkins RD, Alberti G, Atkinson G, and Gregson W. 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, 2015.

[15] Haddad M, Chaouachi A, Wong DP, Castagna C, Hambli M, Hue O, and Chamari K. Influence of fatigue, stress, muscle soreness and sleep on perceived exertion during submaximal effort. *Physiology and Behavior*, 119:185–189, 2013.

[16] Hoff J. Soccer specific aerobic endurance training. *British Journal of Sports Medicine*, 2002.

[17] Hulin BT, Gabbett TJ, Lawson DW, Caputi P, and Sampson JA. The acute:chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *British Journal of Sports Medicine*, 50(4): 231–U123, 2016.

[18] Impellizzeri FM, Rampinini E, Coutts AJ, Sassi A, and Marcora SM. Use of RPE-based training load in soccer. *Medicine and Science in Sports and Exercise*, 2004.

[19] Impellizzeri FM, Rampinini E, and Marcora SM. Physiological assessment of aerobic training in soccer. *Journal of Sports Sciences*, 23(6):583–592, 2005.

[20] Jaspers A, Op De Beeck T, Brink MS, Frencken WG, Staes F, Davis JJ, and Helsen WF. Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn From Machine Learning. *International Journal of Sports Physiology and Performance*, 13(5):625–630, 2017.

[21] Little T and Williams AG. Measures of Exercise Intensity During Soccer Training Drills With Professional Soccer Players. *Journal of Strength and Conditioning Research*, 21(2):367–371, may 2007.

[22] Murphy AP, Duffield R, Kellett A, and Reid M. Comparison of Athlete-Coach Perceptions of Internal and External Load Markers for Elite Junior Tennis Training. *International Journal of Sports Physiology and Performance*, 9(5):751–756, 2014.

[23] Rossi A, Perri E, Trecroci A, Savino M, Alberti G, and Iaia M. GPS data reflect players' internal load in soccer. In *IEEE International Conference on Data Mining Workshops, ICDMW*, pages 890–893, New Orleans, LA, USA, 2017. IEEE.

[24] Simoneau JA and Bouchard C. Genetic determinism of fiber type proportion in human skeletal muscle. *The FASEB Journal*, 9(11):1091–1095, 1995.

[25] Thorpe RT, Strudwick AJ, Buchheit M, Atkinson G, Drust B, and Gregson W. Tracking Morning Fatigue Status Across In-Season Training Weeks in Elite Soccer Players. *International Journal of Sports Physiology and Performance*, 11(7): 947–952, 2016.

[26] Timmons JA. Variability in training-induced skeletal muscle adaptation. *Journal of Applied Physiology*, 2011.

[27] Vandewiele G, Geurkink Y, Lievens M, Ongenae F, De Turck F, and Boone J. Enabling Training Personalization by Predicting the Session Rate of Perceived Exertion (sRPE). In *Machine Learning and Data Mining for Sports Analytics ECML/PKDD*, number September, Skopje, Macedonia, 2017.

[28] Viru A and Viru M. Nature of training effects. *Exerc Sport Sci*, 6(7):67–95, 2000.

[29] Wallace LK, Coutts AJ, Bell J, Simpson N, and Slattery KM. Using Session-RPE to Monitor Training Load in Swimmers. *Strength and Conditioning Journal*, 30(6):72–76, 2008.

[30] Williams S, West S, Cross MJ, and Stokes KA. Better way to determine the acute: Chronic workload ratio? *British Journal of Sports Medicine*, 51(3):209–210, 2017.

Part III

General discussion

1 Discussion

The interest in soccer by millions of people across the globe, and the interests of stakeholders within soccer teams to perform well, remains undiminished. At the highest levels of soccer, specialists in several domains work together to provide the best conditions for good performances. Technology is implemented to gather new data, aiming to acquire data which can help to further improve these conditions. This thesis touches on three major topics within soccer that can (potentially) benefit from a correct implementation of technology and use of data, i.e. performance, training and monitoring. In this chapter, the most important results of the three studies are summarized, and subsequently, discussed and reflected on. Conclusively, the insights from the studies, scientific literature and applied setting are synthesized.

1.1 Overall results

Study I aimed to identify the strongest predictive variables of winning and losing in Belgian professional soccer. A broad range of performance indicators and contextual variables were used. As hypothesized, a shot-related variable, i.e. the total number of shots on target from the attacking penalty box, was identified as the best predictor. It should nevertheless be noted that a broad range of divergent variables were amongst the strongest predictors of winning and losing. Physical parameters, such as the number of accelerations and distances in several speed thresholds were identified as strong predictors. More specifically, it seems particularly interesting to look at physical parameters of the second half. Contextual variables such as ELO-ratings, transfer values, match location and Playing Styles can be useful additions to current approaches used to evaluate game performances, as they were identified as strong predictors too. This study also provided some useful insights for practical use. The application of *TreeExplainer* allowed to determine the best predictors of each game, which could be potentially useful when analysing games. As the workflow from dataset to predictive modelling was described in detail, similar approaches can be used to evaluate the current performance indicators provided to the coaching staff and other stakeholders connected to the team. Conclusively, the results of this study provided a framework in which practitioners can use quantitative game data to gain more objective insights into performance.

Study II focused on the structure of training during the preseason, how training intensity is distributed and whether there are differences between the various ways in which the TID can be expressed. The preseason started with a 2-day testing schedule in which several physical and medical tests were conducted, including a maximal incremental running test. Subsequently, the players were exposed to a 36-days preseason training program that included a mix of soccer sessions, running sessions, strength sessions, complex sessions and friendly games. The total average distance covered by players over the complete preseason averaged 254.0 $\pm$ 47.3 km, during 53.8 $\pm$ 10.4 hours of outdoor preseason training activities. Regardless of the method that was used to quantify TID, most of the training volume is spent at low intensity. Differences in the proportion of volume spent at low, moderate and high intensity were however observed for the different methods of quantification. When using distance as a marker of volume, lower proportions of volume were spent at low intensity and consequently higher proportion of volume were spent at moderate and high intensities, as compared to time. Time as a marker of volume may overestimate the proportion of volume spent at low intensities. Speed as a marker of intensity may better reflect the stochastic nature of soccer in comparison to HR. In the context of soccer, speed and distance can thus be used as markers of respectively intensity and volume and may be the most relevant and practically feasible method to describe TID.

Study III aimed to identify the strongest predictive variables of the session Rate of Perceived Exertion in soccer. A total of 70 external load indicators, internal load indicators, individual characteristics, and supplementary variables were used to build a predictive model, which showed a mean absolute error of 0.67 ($\pm$0.09) Arbitrary Units (AU) and a root-mean-square error of 0.93 ($\pm$0.16) AU. The results indicated that the sRPE can be predicted quite accurately using only a relatively small number of training observations, with little over 900 training observations. External load indicators were found to be the strongest predictors of the sRPE, accounting for 61.5% of the total importance, with total distance as the strongest individual predictor. The included internal-load indicators and individual characteristics accounted only for respectively 1.0% and 4.5% of the total importance. Predictive accuracy improved when supplementary variables such as group-based sRPE predictions (10.5% of the total importance), individual deviation variables (5.8% of the total importance), and individual player markers (17.0% of the total importance) were included. The results showed that including a broad range of variables, other than external load indicators, is useful, as the accumulated importance of these variables accounts for a reasonable component of the total importance. Applications resulting from this study may help a coaching staff to plan, monitor, and evaluate training sessions.

1.2 Discussion and reflection on the research findings

In this section, the topics of the studies are reflected on, and should provide relevant information on further progression of performance, training and monitoring in soccer. Throughout this reflection, text boxes are added with research opportunities.

The topic of **Study I** closely relates to the current technological possibilities to quantify soccer-related game-activities, and consequently an increase in the availability of game data, providing new opportunities and insights in the field of quantitative game analysis. It is however also known that the responsibilities of soccer coaches have increased over the years and that providing relevant and to-the point information is important [48]. In order to cope with the first challenge, namely an increased amount of data, machine learning was applied. Machine learning works better with large numbers of input variables as compared to inferential statistics [20]. The use of machine learning consequently allows to use many input variables, which in the case of **Study I** resulted in the inclusion of a divergent set of variables. One of the main findings of this study, that the total number of shots on target from the attacking penalty box, a shot-related variable, is the best predictor of winning of losing, may in its essence not be a very interesting finding. After all, for coaches the importance of shots for scoring goals is obvious [51]. It is however interesting that the magnitude of importance of each of the predictors can be quantified, using a strongly theoretically grounded method such as *TreeExplainer* [54], which allows to rank the variables. That way, the 'share' of the total importance of individual variables can be quantified. The direction of the relationships between the predictors and game outcome was also established. It was found that for several variables, a negative relationship with the game outcome was found, which indicates that higher values lower the probability of winning. Establishing the direction of the relationship between variables and game outcome is therefore useful, as it aids a correct interpretation.

Opportunities

At the highest levels of professional soccer, a plethora of game information is available, from different sources. To effectively gather and process data, the use of *nifty* techniques such as *web scraping* and Natural Language Processing (NLP) can be explored. Web scraping is a technique to extract data from the internet and save it to a file system or database for later retrieval or analysis [82]. NLP employs computational techniques for the purpose of learning, understanding, and producing human language content [39], which could potentially be used to extract useful data from text. These *new* techniques exemplify the possibilities that may emerge with regard to the use of contextual information for analyses purposes.

Most of the used performance indicators in **Study I** can be best described as notational metrics. The field of quantitative performance analysis is currently in a transition phase, where new possibilities arise because of the increased availability of (tracking) data but the more traditional metrics derived from notational performance analysis are still widely used [68]. There has been critique on the use of several more traditional notational performance metrics, as they have been used as a result of availability rather than to develop a deeper understanding of performance [56]. Furthermore, an exclusive focus on notational metrics will probably not be sufficient to explain the underlying mechanisms of performance and game outcome in soccer [51]. The possibilities that tracking data provide may in this context be thus an opportunity. It should however also be noted that the use of tracking technology yields millions of data points per game [7]. Consequently, the use of tracking data will challenge data management and analytical skills of those involved with the data [37]. Also, while notational metrics are regularly quite broadly available in professional soccer, the availability of technical/tactical metrics resulting from tracking data is often still scarce. The last few years nevertheless, an increasing number of studies utilized tracking data, providing new and interesting insights. The exploration and use of tracking data should therefore be encouraged. It is known that *'soccer culture'* can be quite conservative [14], indicating that the added value of information derived from tracking data should be very clear to the stakeholders within the club. The transition from more traditional metrics to new metrics resulting from tracking data may be accelerated by combining both types of metrics. The Playing Styles-metrics from **Study I** provide a clear example on how novel metrics provide different insights, by providing more context to ball possession in general. Addition of context is also useful when analysing physical game activity profiles [12]. Currently, physical game activity profiles are still provided in isolation, by e.g. reporting the total distance covered by a team or player without providing any context, which leads to one-dimensional insights into physical match performance [12], resembling the notational approach. It has been suggested that an integrated approach, which focusses on the activity profile in relation to key tactical and technical activities, is needed to allow a more complete understanding of movements from a physical perspective [12, 24]. Using an integrated approach, the activity profile is contextualized in relation to key tactical activities on a team and position-specific level [12], which could aid the understanding of the physical activity profile in relation to tactical roles and instructions provided to the players [12, 47]. Overall, a more holistic approach, in which (new) metrics are viewed in relation to more traditional metrics and contextual factors, can be utilized to support the use of new metrics and provide insights into how new metrics relate to those that are already used.

Opportunities

Tracking technology results in huge amounts of data, potentially entailing interesting information. To extract this information, those involved with the data are challenged to effectively translate data into information. To *ease* the use of tracking data, enhancing the *usability* of tracking data could be interesting from a technological, commercial and sport-scientific view. Specifically, problems with the usability can be identified (*which problems do the users encounter?*), for which (technological) solutions can be created. Companies and academics could in this light cooperate to improve usability and thereby potentially increase the use of tracking data.

In **Study I**, games were predicted as either a win or a loss, and subsequently compared to the true game outcome. When predicting game outcome in soccer, not only correct predictions can be interesting, incorrect predictions may provide useful insights as well. Incorrect predictions may indicate that the game outcome was undeserved for the team that was, according to the model, superior. Predictions may also indicate whether, based on the performance indicators, the game was close or that one team clearly outperformed the other. The possibility for local predictions [54], demonstrated in **Study I**, may aid the understanding of the assessment of game outcome. A factor that may have played a role with regard to game outcome, and should not be disregarded, is the factor *chance*. In practice, chance refers to as something uncontrollable and unforeseeable, not intended and not caused by skill or tactics [49]. In many sports, the line between skill and chance cannot be easily drawn, and sport-related actions may reflect varying elements of both skill and chance [74]. It is however stated that the influence of chance on game outcome is greater in low-scoring sports, such as soccer, compared to high(er)-scoring sports such as handball and basketball [30]. And although the factor chance tends to even out in the long run, during a single game or over a series of games, the superior soccer team may not always win [30]. The role of chance in soccer has been previously studied by Lames [49], who found that the closer the game, the higher the percentage of goals associated with chance are scored. This is in line with one of the findings of **Study I**, where predictive accuracy was smaller for victories or losses with a small margin. The influence of chance, also sometimes referred to as luck or randomness, in soccer has been repeatedly stated. Even more, the study that is often referred to as the beginning of modern performance analysis in soccer [49], is named *Skill and Chance in Association Football* [70]. This study did however rather contribute to the positivist paradigm that has shaped the evolution of performance analysis [56], where functional relations exist between causal and explanatory factors (independent variables) and outcomes (dependent variables) [67]. It assumes that human behaviour is measurable, causally derived and thus predictable and controllable [56]. Within the positivist paradigm, it is assumed that good performances are accompanied with good outcomes and vice versa, while this assumption may, because of the factor chance, not always hold. As in many sports, the line between skill and chance cannot be easily drawn in soccer [74], but when assessing performance and game outcome, it is important to consider the role of chance. Theoretically, within the positivist paradigm, the influence of the performance of both teams can be summarized as shown in Equation 1:

$$\begin{gathered}\text{Game outcome} = \text{Team Performance - Opponent Performance}\\ \text{Game outcome} = \Delta\text{Performance}\end{gathered} \tag{1}$$

Within the positivist paradigm, the factor chance is not accounted for, while it is clear that this factor plays its role in relation to game outcome, since it has been shown that almost half of all goals scored are associated with an indicator of chance [49]. Therefore, within a more holistic paradigm, the factor chance should be considered (Equation 2). Please note that the influence of chance can be either positive or negative.

$$\begin{gathered}\text{Game outcome} = \text{Team Performance - Opponent Performance} \pm \text{Chance}\\ \text{Game outcome} = \Delta\text{Performance} \pm \text{Chance}\end{gathered} \tag{2}$$

Based on Equation 2, a conceptual model can be drafted (Figure 1). The model illustrates the relation between the influence of the difference in performance in relation to the (possible) influence of chance.

This conceptual model is interesting for a better understanding of game outcome. It is widely known that contemporary soccer is a multi-billion euro industry [44] and that gaining a competitive advantage over rival teams can be significantly financially beneficial [68]. This indicates that the stakes for several stakeholders within the club are high. Either good or bad game outcomes evoke emotions [69], which may

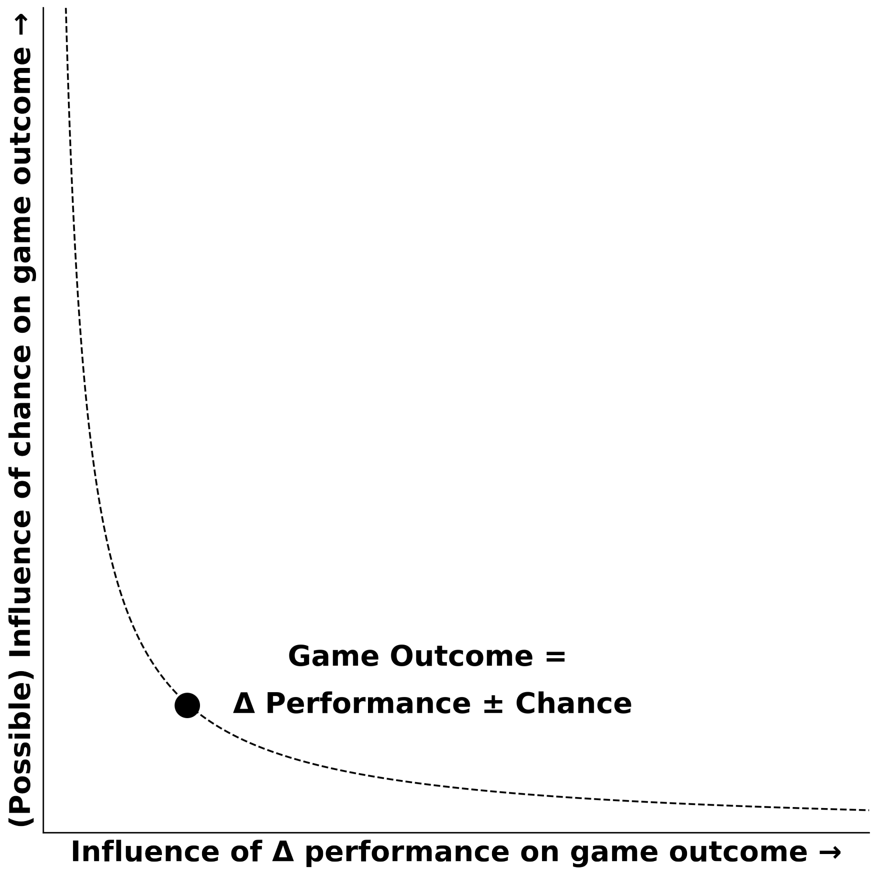


Figure 1: A conceptual model on the interplay between the influence of performance and the (possible) influence of chance on game outcome.

increase the bias that is connected to performance judgements [58]. Performance analysis is interesting for coaches and players, but for other stakeholders, such as people responsible for the management of the club, as well. From a management point of view, as an example, when performance and/or game outcomes are below expectations, those responsible have to correctly assess when a situation is sufficiently serious and attributable to the behaviour of the current leadership to justify immediate dismissal [28]. However, they often act out of panic and indulge in 'scapegoating' to appease stakeholders such as party members, shareholders or fans of the team [28]. A better understanding of performance indicators in soccer, both on a global and local level, could aid all stakeholders related to the team to correctly evaluate team and player performances [13]. Next to evaluating the performances of their team and the processes within the team that contribute to team performance, stakeholders should evaluate both the performances of the opponent(s) and whether the factor chance played a significant role during a single game or a series of games. Eventually, every club aims to make correct decisions in order to improve the performance of their own team, hoping to obtain favourable game outcomes. A better understanding of game outcome by the stakeholders within the club can bring them closer to this ambition.

An important aspect to improve performance is training [41], which can be applied to improve all aspects

Opportunities

Although the factor chance should be considered in soccer, the number of studies on this topic is currently limited. This may be because the concept of *chance* may be perceived as rather abstract, and may be deemed *unmeasurable*. To date, the majority of the studies that describe chance in soccer, discuss this factor in a more philosophical manner [30, 74]. The best attempt to empirically approach chance in soccer might be the study of Lames [49], which can be considered as a starting point for future studies focussing on the objectivation of the factor chance in soccer. A well-grounded theoretical substantiation on the factor chance may eventually help to increase the knowledge and awareness on this often decisive factor in soccer.

related to performance in soccer [76]. In **Study II** and **Study III**, the main subjects were training and monitoring, from a physical point of view. The results from **Study I**, and the findings from previous studies at the highest levels of international soccer demonstrating that the physical demands of elite soccer have increased over the years, confirm the importance of the physical aspect in professional soccer. At the same time, practitioners in soccer are faced with significant challenges with regard to the development or even maintenance of this physical aspect, such as congested fixtures schedules with games every 3-4 days, irregular game schedules with games starting at different times during the day, players that regularly go on international duty with their respective national team and (international) travel activities for away games. In practice, during the in-season period, teams work from game to game [79]. The previous and next game are arguably the most important reference points, first to recover from the previous game and second to prepare for the next game. The demanding sequences of games every 3-4 days, combined with taxing travelling activities, may for some players not be sufficient to restore homeostasis [64]. During the in-season period, the number of acquisition sessions from a physical viewpoint, being sessions with higher volume and/or intensity [79], may therefore for some teams be quite limited. Conclusively, during the in-season period, the goal from a physical perspective is to maintain general and specific fitness through general and game-specific training modalities, and increase game-specific condition through games [62]. After the in-season period, the off-season period allows players to recover from the accumulated load. The off-season period is followed by the preseason, which presents a unique window during the season. During this period, teams usually do not have to play competitive games, allowing opportunities for acquisition sessions. It is common that during the preseason a mix of different sessions are conducted [22, 45]. The preseason is often described as intense in terms of training load, in order to reach a fitness peak before the competitive season starts [62]. In **Study II**, the training characteristics of the preseason in a professional soccer team were extensively described. To the best of knowledge, however, there are no scientific studies to extensively report the structure of the preseason in soccer. It is therefore difficult to compare how the structure of the preseason in **Study II** relates to the preseason structure of other teams.

Opportunities

It is clear that the preseason is an important period during the season, given the congested fixture schedules during the in-season period [62], and consequently, limited opportunities for acquisition sessions. For practitioners, it may therefore be very informative to get detailed insights into the structure of the preseason in other professional teams, to optimally prepare their team for a new season. The information regarding preseason workload and structure in scientific literature is however limited. This may be related to the 'protectionist' approach applied by clubs, as they do not want to provide too much information to competitors on their *modus operandi* [16]. On the other hand, there may be researchers or practitioners who are willing to provide extensive descriptions on the preseason training structure and workload in the form of a case study, as the preseason period may not really be appropriate for an experimental design study. Unfortunately, case studies often do not adhere to the heavy requirements of peer-reviewed research [16], although these types of studies are perceived as important and valuable sources of information by practitioners [34]. In line with the statement of Buchheit [16], the use of case studies should therefore be promoted.

Study II also aimed to describe and compare different TID methods, using 4 combinations of intensity (heart rate (HR) and speed) and volume (time, distance) parameters. The TID is based on two intensity demarcation points, the first and second lactate/ventilatory threshold. These thresholds are physiologically substantiated, as they represent individual physiological mechanisms induced by exercise [1]. This is in contrast with the arbitrary and general intensity thresholds, usually expressed in speed or HR, that are currently often used. However, despite the physiological justification for the use of the first and second lactate/ventilatory threshold to subdivide exercise volume in different intensity zones, the number of studies describing the TID in soccer players is rather limited. This may be because of the intermittent nature of soccer, in which low- and high-speed locomotor activities are being randomly alternated, while the use of the TID is traditionally more associated with more steady state activities. Moreover, the intensity ranges in terms of speed thresholds may be quite narrow, as the difference in the first and second lactate/ventilatory threshold often ranges between 3-5 km/h. Moreover, the speed at the second lactate/ventilatory threshold generally ranges between 13-15 km/h in professional soccer players [5, 8, 83], while the maximum speed that players during games may attain approaches 35 km/h [18]. Thus, there exist a considerable speed range between the players' second lactate/ventilatory threshold and maximum speed, and in practice, there is great interest to quantify the distances in the higher speed zones [2]. Similarly, the speed range below the first lactate/ventilatory threshold could for example be subdivided into walking and jogging, separated by a speed that close to the general walk-to-run transition (around 7 km/h) [29]. From the point of individualization, the thresholds determined during a single moment of the season may not be representative of the thresholds during the season, as the off-season results in cardiorespiratory- and vascular, metabolic and muscular detraining effects [61]. In many clubs, players conduct a maximal incremental running test in a laboratory environment, prior to the start of the preseason. Based on the maximal incremental test, the first and second lactate/ventilatory threshold can be determined. It has however been shown that lactate thresholds increased directly after preseason [23, 57] and after a full season [55]. A test directly after the preseason may be a solution, as it has been shown that in professional youth soccer players, the lactate thresholds increased during the preseason, but remained stable afterwards [59]. This may however not be practically feasible because of the start of the competition. Also, although the lactate thresholds may generally remain stable throughout the season, individual alterations may occur for reasons such as injury or reduced workloads. Alterations of individual intensity thresholds can thus be regarded as a main disadvantage of individualized intensity thresholds. To accurately determine the TID, the lactate thresholds may therefore be re-tested during several occasions of the season, although this may not be practically feasible because of the expenses and the time-consuming nature of laboratory testing [77]. Field tests to assess the first and second lactate/ventilatory threshold could therefore be conducted, such as the Conconi test (only the second lactate/ventilatory threshold) [25]) or the RABIT®-test (Running Advisor Billat Training) [35]. Reduced costs, minimal use of equipment and the time-efficient manner with which field tests can be conducted make them theoretically more convenient to use throughout the season [77], but it is also important that field tests are performed in standardized conditions to ensure repeatability of the data. It should also be acknowledged that the validity of the Conconi-test has been criticized [11, 46], and that the detection of the first lactate/ventilatory threshold using the RABIT®-test is compromised [35]. Factors such as the playing schedule, and weather and field conditions further complicate the execution of standardized field tests during the season. The use of individualized first and second lactate/ventilatory thresholds may thus be practically difficult, however, although depending on the method of determination, it can be argued that within a group of players, the average speed at the first and second lactate threshold ranges between speeds at respectively 9-13 km/h [5] and 13-15 km/h [5, 8, 83] in professional soccer players. Overall speed demarcation points on a group level for the first and second lactate/ventilatory threshold, for example, respectively 11 and 14 km/h, may therefore be used to (at least) estimate the TID.

Opportunities

In **Study II**, the TID was determined during the preseason period. Future research could include both the preseason as well as (parts of) the in-season period into their research protocol, to quantify possible fluctuations in the TID throughout the season. In the current research protocol, this was not possible, because the GNSS-/HR-sensors used during the studies were, to our understanding, not approved by the FIFA for use during official games.

Nowadays, monitoring both external and internal workload is common practice in soccer. A widely used parameter to monitor internal training load is HR, but its use is criticized, because HR tends to underestimates or overestimates the intensity during intermittent activities [4, 52]. Moreover, HR-monitoring requires both technical and physiological expertise to make an appropriate analysis [3]. The best feasible alternative for internal load monitoring in soccer, the sRPE, is a simple and practical tool that represents the players' own perceptions of training stress, which include both physiological and psychological stress [40]. The sRPE has shown to be a valid indicator of intensity in soccer [10, 33, 38, 40] and a more valid marker of exercise intensity in soccer over a broad range of activities compared to HR-monitoring [52]. The sRPE originates from the RPE, which was first introduced by Gunnar Borg in the early 1960's [9], and the introduction of the RPE was accompanied by the following statement: 'Since man reacts to the world as he perceives it and not as it "really" is, it is important to know more about the relation between objective and subjective measures of physical stress'. This statement illustrates the early interests in investigating the relationship between objective and subjective measures of physical stress and summarizes the goal of **Study III**, namely to identify the most important determinants of the sRPE.

In contrast to much of the earlier research with regard to determinants of the sRPE, in which isolated variables were related to the sRPE, **Study III** included a wide range of external and internal load parameters, individual characteristics and supplementary variables to predict the sRPE using machine learning. The accumulated importance of the external training load parameters accounted for the largest proportion of the prediction, with over 60% of the total importance. The inclusion of internal load parameters, individual characteristics and supplemental variables is nevertheless also useful, as almost 40% of the total importance was determined by these factors. The results also showed that there are differences in the interpretation of the sRPE-scale, as some players consistently rate a training session higher or lower than the group mean, which was previously described by Bartlett et al. [6]. Recently, Foster et al. [31] stated that about 10% of people struggle to effectively use the sRPE method. These individuals have problems to indicate the sRPE of an entire session or have trouble to acknowledge that harder exercise is indeed becoming hard. The validity of a workload measure is typically examined through its agreement with a criterion which represents the true value [80]. In the case of the sRPE, the true value is however hard to determine. The prediction of the sRPE may however provide a more objective view on the sRPE and may be used to detect individuals who struggle to effectively use the sRPE method.

The sRPE is an interesting concept, as it summarizes the perceived workload of a complete session into a single index. In other fields related to soccer, 'single-index'-approaches are used as well, to summarize performances or qualities. Well known approaches are those by videogames such as FIFA [63], websites such as InStatSport [42] and academic approaches such as PlayeRank [66]. The indexes allow comparisons between players or teams. However, the indexes from the FIFA games, InStatSport or PlayeRank do not have a ground truth, meaning they are calculated based on inference rather than direct observation. The advantage of the sRPE is that it can be 'measured' and thus provides a ground truth to which other measures can be related. This further illustrates the potential of the sRPE, as a concept that is easy to interpret and use, lacks the requirement for an anchoring to a maximal exercise test and technical requirements to monitor training load [31].

Opportunities

One of the restrictions of the sRPE is a lack of precision. This problem may be partly countered by an interesting extension of the sRPE, in which the sRPE is subdivided into a 'local' or 'muscular' and a 'central' or 'respiratory' sRPE. This subdivision is based on the statement that the complex perception of exertion seems to be mainly composed of perceptions from the musculature and system of circulations [65]. A higher perceived exertion of the muscular sRPE, compared to respiratory sRPE, has been reported after matches [53]. Future studies may focus on obtaining insights into the determinants of the muscular and respiratory sRPE, similarly to the way that **Study III** and previous studies [6, 43, 71] studied the determinants of the sRPE.

1.3 Context, synthesis and conclusions

This thesis was conducted within the project KAA Gent - UGent Performance Center. This project aimed to improve training processes within KAA Gent, with special reference to the physical aspect. Within this project, it was possible to gain more insight into the domains of performance, training and monitoring in the applied setting of a professional soccer team. It is clear that these domains should not be viewed in isolation, as they are all interrelated. The application of technology and data is also a similarity between the three domains. In this final part of the thesis, the insights from the conducted studies, scientific literature and the applied setting are synthesised.

As previously mentioned, a commonality between **Study I**, **Study II** and **Study III** is the application of technology. The results of the study of Akenhead and Nassis [2] clearly demonstrate that at the highest levels of professional soccer, technology, and tracking technology specifically, is widely used to monitor players. Tracking technology is generally used to monitor the workload of players, to comprehensively understand the accumulation of workload over a certain period [2]. When implementing monitoring strategies using tracking technology, practitioners are faced with significant challenges, such as squad size and compliance of the players and coaching staff [19, 21]. Furthermore, there may be differences in the use of tracking technology between training sessions and games, so alignment of training and game data is important [2]. Monitoring applies to the tracking of physical workload, but practitioners in soccer nowadays also need to implement monitoring strategies to perform (amongst others) physical performance testing, and assessments of body composition, urine and perceived wellness. For these monitoring strategies, practitioners also need to be able to handle other technological devices and process the resulting data. Consequently, the use of technology applies to a bigger context than solely tracking technology. In this respect, it is important to remember that technology per se should not be the focus; helping the players and other staff members should always be the focus [78]. It important to remember that "*once the measure becomes a target, it ceases to be a good measure.*", citing Goodhart's Law [32].

The emergence of technology within soccer also results in greater amounts of data, and it can be expected that the following years the volume of data keeps increasing [37]. To minimize the risk of omitting valuable information that could contribute to explaining relationships, and consequently relevant insights into the training or game approach [80], the application of a different research paradigm may be necessary. In sports science, (experimental) research designs typically aim to confirm or reject a hypothesis, and the available data is often analyzed using a univariate approach. However, in the current practical reality of quantitative performance analysis or workload monitoring, it is common practice to collect multiple variables concurrently, often not within a carefully controlled experimental design [20]. The current practices on collecting data may therefore not be in line with the requirements of experimental research. In computer science, a different research paradigm is applied, in which patterns in the data are sought without formulating explicit hypotheses on these patterns [37]. This research paradigm may therefore better fit current practices and provide new research opportunities. For both **Study I** and **Study III**, machine learning was used to identify the best predictors, of respectively game outcome and the sRPE. Because of the number of input variables and the possible presence of complicated non-linear interactions [20], the use of machine learning may be more suitable for multidimensional entities such as game outcome and sRPE, rather than seeking relationships in a univariate approach.

Overall, a better understanding of game outcome could aid all stakeholders related to the team. Improved quantitative analyses, through an improved understanding of performance indicators in soccer, both on a global and local level, could aid all stakeholders related to the team to correctly evaluate team and opponent performances [13]. This does however not undermine the importance of qualitative performance analysis, as it should be acknowledged that the exceptional moments of the game, those moments that regularly determine game outcome, are often not reflected in the data [36]. In this respect, the combination of the subjective coaches' eye with scientific data may buffer the mutual weaknesses of these different approaches, as stated by Sieghartsleitner et al. [73]. Furthermore, the factor chance should also be acknowledged and accounted for when analysing game outcome and performance. Eventually, every club aims to make correct decisions, preferably by using a combination of thorough qualitative and quantitative performance analysis, to improve the processes focused on improving team performance, rather than needlessly disrupting productive ongoing processes through emotional decisions. Strictly, team performance is the only factor that can be influenced to increase the probability of satisfying game outcome, and depends on a myriad of technical, tactical, physical and mental aspects [76]. Based on qualitative and quantitative performance analysis, and the determinants and influencing factors, training can be appropriately scheduled. Technological advancements allow to monitor whether training is performed as scheduled and how the players respond to the workload, thus providing opportunities to adjust training accordingly. This helps the team to improve technical, tactical, physical and mental abilities, aiming to maximise team performance and hence the probability of desirable game outcomes. A graphical illustration on the interrelations between performance, training and monitoring is depicted in Figure 2.

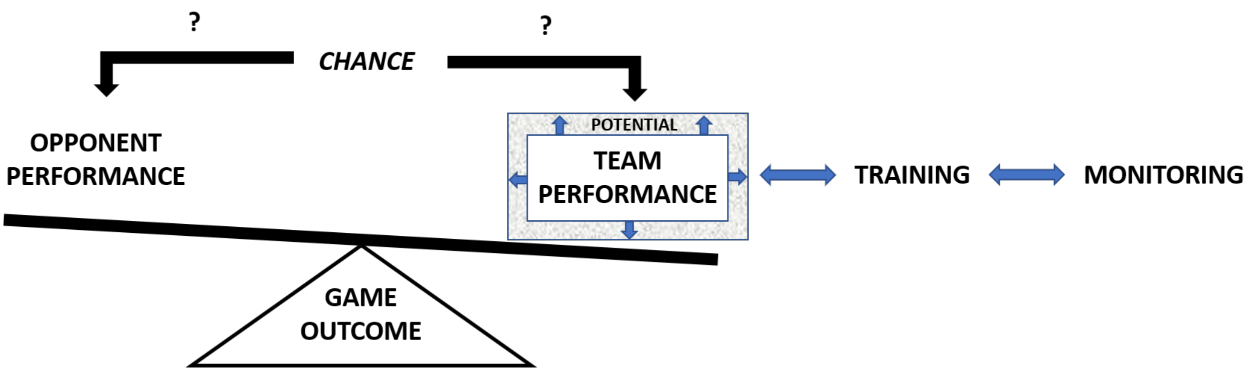


Figure 2: Theoretical framework on the interrelations between performance, training and monitoring. Game outcome is determined by the balance between team and opponent performance, and the factor chance, which theoretically varies in *weight* and *side* of the balance. To increase the probability of favourable game outcome, hypothetically, teams can/should only focus on improving team performance. Training is an important aspect in improving team performance, and can be optimized by monitoring.

Technology, and the resulting application of data should ultimately lead to useful *information*, that can be used to aid decision-making. Essentially, two processes can be identified. The first is the adoption of technology, which should be considered in light of available evidence, coaching philosophy, player compliance and resources [19]. A recent article by Windt et al. [81] provides an excellent framework for the adoption of (new) technology, in which it is important to evaluate whether the information resulting from the technology is 1) helpful and 2) trustworthy. Furthermore, it should be evaluated if 3) the data can integrated, managed and analysed appropriately and 4) whether technology can be effectively implemented in sports practice. Lastly, 5) the technology should be affordable. Practitioners should make these evaluations *prior* to purchasing and using new technology, to make sure that the available resources are usefully applied [17]. Secondly, the data has to be accessed, stored and processed. At this point, data are *just* discrete and objective facts (numbers, symbols, figures) without context and interpretation. There are many different models describing the *knowledge management*, such as the classic DIKW-pyramid [72] and its many variants, which all describe the translation from *D*ata to *I*nformation, *K*nowledge and/or *W*isdom (hence the term *DIKW*). All models nevertheless share their purpose, namely to support decision-making. In Figure 3, the framework for the adoption of technology by Windt et al. [81] and the knowledge management model by Snowden [75] were used to suggest a new structure in which technology and data can be systematically and productively used and assessed.

Within an applied setting, more and more clubs employ a sports scientist, which job is, amongst others, to support to coaching staff in the areas of performance, training and monitoring [60]. The positivist

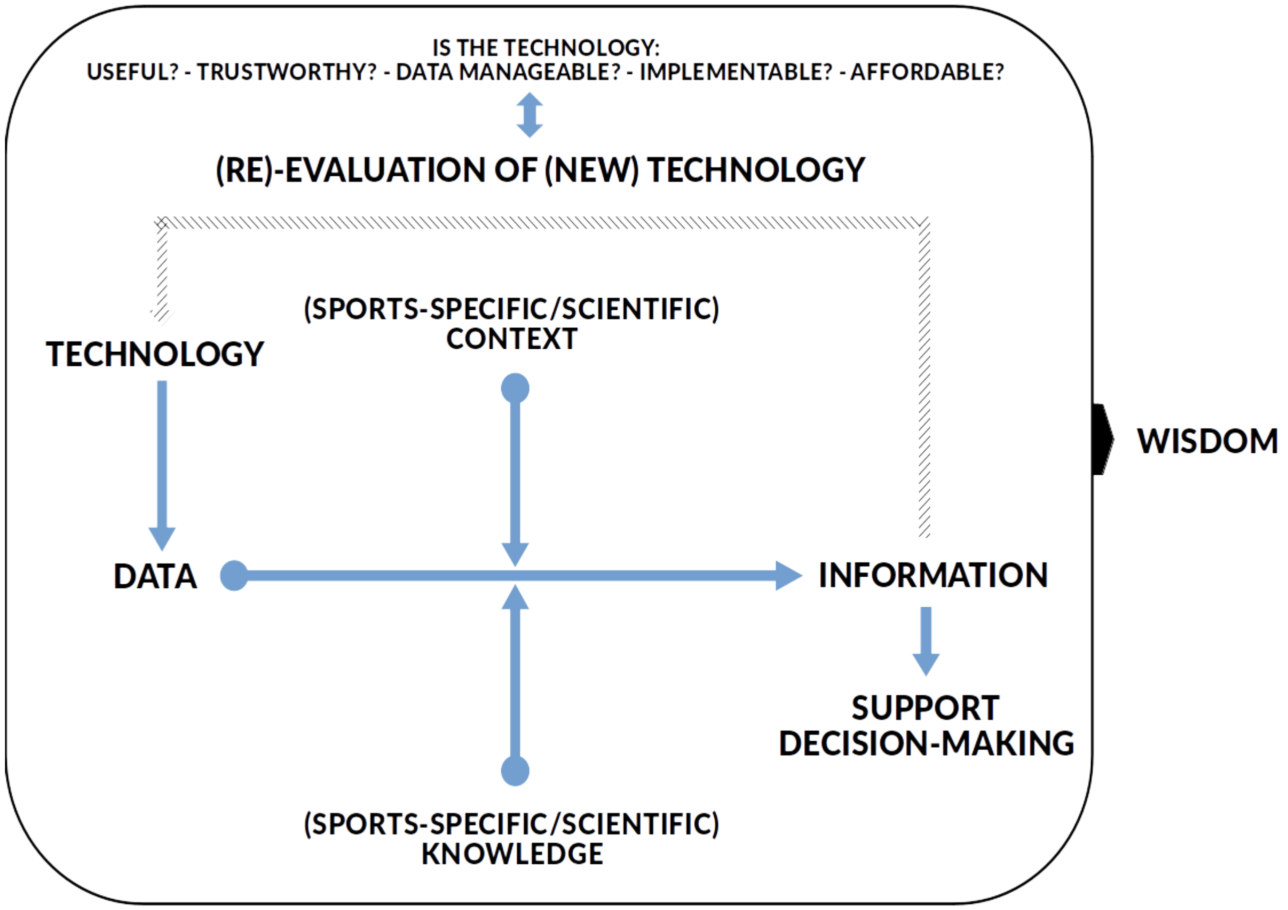


Figure 3: Proposed framework, integrating the adoption of technology [81] within adjusted knowledge management model by [75].

paradigm applied by many practitioners, leverages considerable challenges for the sports scientist. In practice, stakeholders generally want sports scientists, or rather academics in general, to provide them with studies and results that can provide the team a competitive advantage [26, 50]. Given that areas such as discussed in this thesis, being performance, training and monitoring, are holistic concepts, it may be more realistic to expect sports scientists to help better understand these concepts within the specific context. This also applies to the use of technology and data, in which its narrative use, concerning the contextualization of measurements or data within a bigger picture [32, 68], is more in line with its main properties, namely to expose patterns. Furthermore, technology should be used for a purpose, namely to make a job/task more efficient and effective, as well as to support others in making informed decisions [60]. Innovation and the application of new technology is encouraged, because it can provide new insight, but should always be critically reviewed before being applied. Foremost, it is important for sports scientists to understand the context in which they work, find their place in the organization and adjust to the high performance context they work in. Next to obtaining sufficient football-specific knowledge [14], it is also advised for modern sports scientists to develop their data management and analytical skills to keep up with the emerging possibilities [26, 37].

Statements such as 'sports scientists and athletes do not belong to the same species' [16] and 'sports scientists often answer questions that are not asked' [15] illustrate that there is still progression to be made with regard to the integration of academia/science/sports scientists into sports practice. A recently published study by Coutts [27] provided the sport-science matrix, depicted in Figure 4, which implied that successful integration of sport science relies on *scientific practitioners*. Collaborations between clubs and universities, such as the project KAA Gent - UGent Performance Center, and the employment of social, well-educated and critical sports scientists, may therefore be very useful to narrow the gap between soccer and science [16].

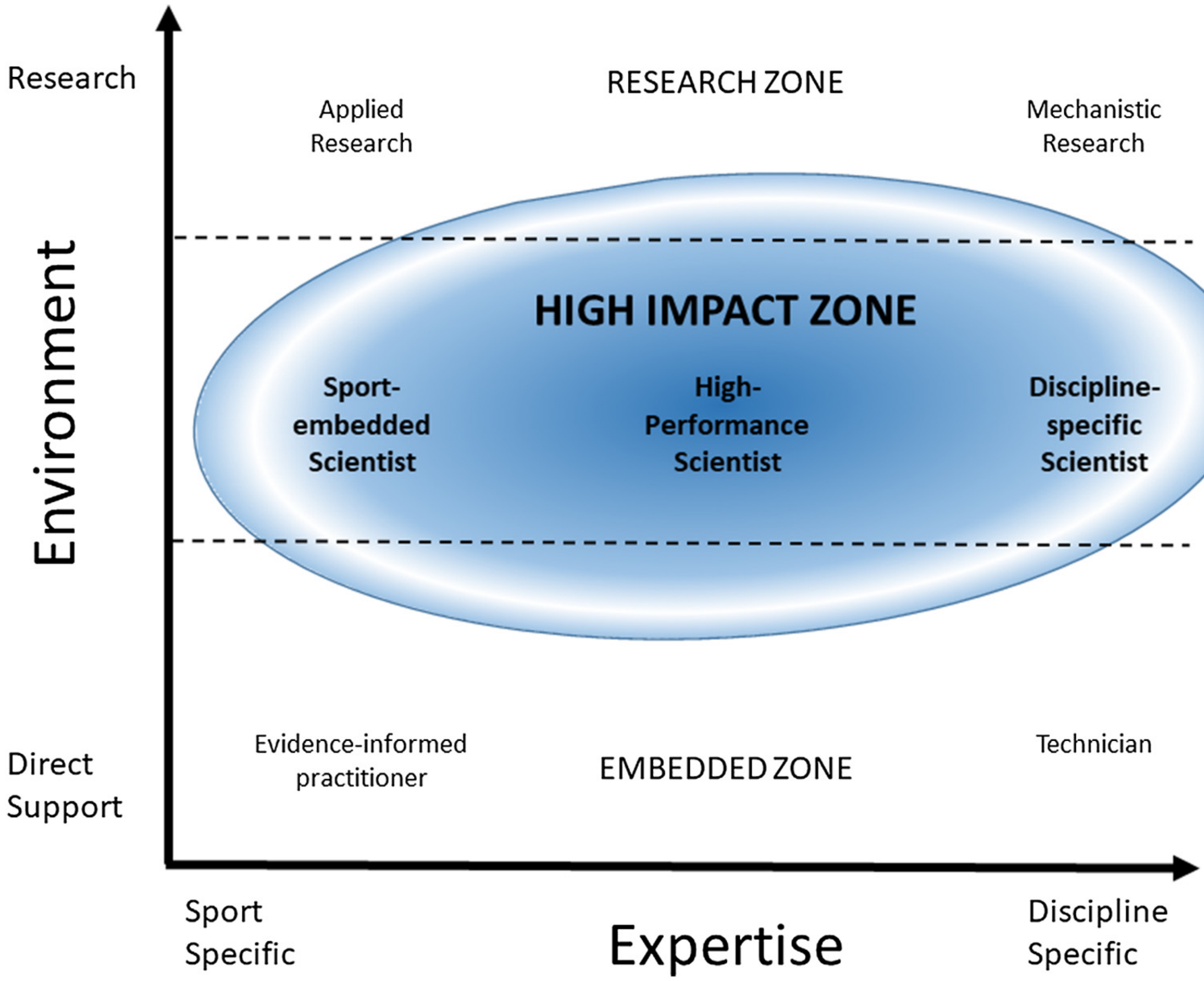


Figure 4: The *sport-science matrix*, designed by Coutts [27].

References

[1] Abt G and Lovell R. The use of individualized speed and intensity thresholds for determining the distance run at high-intensity in professional soccer. *Journal of Sports Sciences*, 27(9):893–898, 2009.

[2] Akenhead R and Nassis GP. Training load and player monitoring in high-level football: Current practice and perceptions. *International Journal of Sports Physiology and Performance*, 11(5):587–593, 2016.

[3] Alexiou H and Coutts AJ. A comparison of methods used for quantifying internal training load in women soccer players. *International Journal of Sports Physiology and Performance*, 3(3):320–330, 2008.

[4] Algrøy EA, Hetlelid KJ, Seiler S, and Pedersen JI. Quantifying training intensity distribution in a group of norwegian professional soccer players. *International Journal of Sports Physiology and Performance*, 6(1):70–81, 2011.

[5] Altmann S, Neumann R, Woll A, and Härtel S. Endurance Capacities in Professional Soccer Players: Are Performance Profiles Position Specific? *Frontiers in Sports and Active Living*, 2(September), 2020.

[6] Bartlett JD, O'Connor F, Pitchford N, Torres-Ronda L, and Robertson SJ. Relationships between internal and external training load in team-sport athletes: Evidence for an individualized approach. *International Journal of Sports Physiology and Performance*, 2017.

[7] Bialkowski A, Lucey P, Carr P, Yue Y, Sridharan S, and Matthews I. Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 725–730, Washington DC, United States, 2014. IEEE.

[8] Boone J, Vaeyens R, Steyaert A, Vanden Bossche L, and Bourgois JG. Physical Fitness of Elite Belgian Soccer Players by Player Position. *Journal of Strength and Conditioning Research*, 26(8):2051–2057, 2012.

[9] Borg GAV. Perceived Exertion as an Indicator of Somatic Stress. *Scandinavian Journal of Rehabilitation Medicine*, 2(3): 92–98, 1970.

[10] Borresen J and Lambert MI. Quantifying Training Load: A Comparison of Subjective and Objective Methods. *International Journal of Sports Physiology and Performance*, 3(1):16–30, 2008.

[11] Bourgois J, Coorevits P, Danneels L, Witvrouw E, Cambier D, and Vrijens J. Validity of the heart rate deflection point as a predictor of lactate threshold concepts during cycling. *Journal of Strength and Conditioning Research*, 18(3):498–503, 2004.

[12] Bradley PS and Ade JD. 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, 2018.

[13] Brechot M and Flepp R. Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals. *Journal of Sports Economics*, 21(4):335–362, 2020.

[14] Brink MS, Kuyvenhoven JP, Toering T, Jordet G, and Frencken WGP. What do football coaches want from sport science? *Kinesiology*, 50:150–154, 2018.

[15] Buchheit M. Buchheit (2016) - chasing the 0.2. *International Journal of Sports Physiology and Performance*, 11(1): 417–418, 2016.

[16] Buchheit M. Houston, we still have a problem. *International Journal of Sports Physiology and Performance*, 12(8): 1111–1114, 2017.

[17] Buchheit M and Simpson BM. Player-tracking technology: Half-full or half-empty glass? *International Journal of Sports Physiology and Performance*, 12(2):35–41, 2017.

[18] Buchheit M, Simpson BM, Hader K, and Lacome M. Occurences of near-to-maximal speed running bouts in elite soccer: insights for training prescription and injury migitation. *Science and Medicine in Football*, 1(1):1–6, 2020.

[19] Burgess DJ. 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(2):136–141, 2017.

[20] Bzdok D, Altman N, and Krzywinski M. Points of Significance: Statistics versus machine learning. *Nature Methods*, 15 (4):233–234, 2018.

[21] Carling C, Lacome M, McCall A, Dupont G, Le Gall F, Simpson BM, and Buchheit M. Monitoring of post-match fatigue in professional soccer: welcome to the real world. *Sports Medicine*, 48(12):2695–2702, 2018.

[22] Castagna C, Impellizzeri FM, Chaouachi A, Bordon C, and Manzi V. Effect of Training Intensity Distribution on Aerobic Fitness Variables in Elite Soccer Players: A Case Study. *Journal of Strength and Conditioning Research*, 25(1):66–71, 2011.

[23] Castagna C, Impellizzeri FM, Chaouachi A, and Manzi V. Preseason variations in aerobic fitness and performance in elite-standard soccer players: A team study. *Journal of Strength and Conditioning Research*, 27(11):2959–2965, 2013.

[24] Castellano J, Alvarez-Pastor D, and Bradley PS. Evaluation of research using computerised tracking systems (amisco®and prozone®) to analyse physical performance in elite soccer: A systematic review. *Sports Medicine*, 44(5):701–712, 2014.

[25] Conconi F, Ferrari M, Ziglio PG, Droghetti P, and Codeco L. Determination of the anaerobic threshold by a noninvasive field test in runners. *Journal of Applied Physiology*, 15(4):869–873, 1982.

[26] Coutts AJ. In the Age of Technology, Occam's Razor Still Applies. *International Journal of Sports Physiology and Performance*, 9(6):741, 2014.

[27] Coutts AJ. Working fast and working slow: The benefits of embedding research in high-performance sport. *International Journal of Sports Physiology and Performance*, 11(1):1–2, 2016.

[28] de Dios Tena J and Forrest D. Within-season dismissal of football coaches: Statistical analysis of causes and consequences. *European Journal of Operational Research*, 181(1):362–373, 2007.

[29] Diedrich FJ and Warren WH. Why Change Gaits? Dynamics of the Walk-Run Transition. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1):183–202, 1995.

[30] Dixon N. On winning and athletic superiority. *Journal of the Philosophy of Sport*, 26(1):10–26, 1999.

[31] Foster C, Boullosa DA, McGuigan MR, Fusco A, Cortis C, Arney BE, Orton B, Dodge C, Jaime S, Radtke K, van Erp T, de Koning JJ, Bok D, Rodríguez-Marroyo JA, and Porcari JP. 25 Years of Session Rating of Perceived Exertion: Historical Perspective and Development. *International Journal of Sports Physiology and Performance*, pages 1–10, 2021.

[32] Gamble P, Chia L, and Allen S. The illogic of being data-driven: reasserting control and restoring balance in our relationship with data and technology in football. *Science and Medicine in Football*, 4(4):338–341, 2020.

[33] Gaudino P, Alberti G, and Iaia M. Estimated metabolic and mechanical demands during different small-sided games in elite soccer players. *Human Movement Science*, 2014.

[34] Gerring J. What is a case study and what is it good for? *American Political Science Review*, 98(2):341–354, 2004.

[35] Giovanelli N, Scaini S, Billat V, and Lazzer S. A new field test to estimate the aerobic and anaerobic thresholds and maximum parameters. *European Journal of Sport Science*, 20(4):437–443, 2020.

[36] Goes FR, Kempe M, and Lemmink KAPM. Predicting match outcome in professional Dutch football using tactical performance metrics computed from position tracking data. In *MathSport International*, number June, pages 105–115, 2019.

[37] Goes FR, Meerhoff RLA, Bueno MJO, Rodrigues DM, Moura FA, Brink MS, Elferink-Gemser MT, Knobbe AJ, Cunha SA, and Lemmink KAPM. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science*, pages 1–16, 2020.

[38] Haddad M, Chaouachi A, Wong DP, Castagna C, Hambli M, Hue O, and Chamari K. Influence of fatigue, stress, muscle soreness and sleep on perceived exertion during submaximal effort. *Physiology & Behavior*, 119:185–189, 2013.

[39] Hirschberg J and Manning CD. Advances in natural language processing. *A Companion to Cognitive Science*, 349(6245): 226–234, 2008.

[40] Impellizzeri FM, Rampinini E, Coutts AJ, Sassi A, and Marcora SM. Use of RPE-based training load in soccer. *Medicine and Science in Sports and Exercise*, 36(6):1042–1047, 2004.

[41] Impellizzeri FM, Marcora SM, and Coutts AJ. Internal and external training load: 15 years on. *International Journal of Sports Physiology and Performance*, 14(2):270–273, 2019.

[42] Instatsport. Objective Performance Rating for Players & Teams, 2021. URL `https://instatsport.com/football/instat{_}index`.

[43] Jaspers A, Op De Beeck T, Brink MS, Frencken WGP, Staes F, Davis JJ, and Helsen WF. Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn From Machine Learning. *International Journal of Sports Physiology and Performance*, 13(5):625–630, 2017.

[44] Jaworska M. European football at the cross-roads. *Journal of Physical Education and Sport*, 20(2):1188–1194, 2020.

[45] Jeong TS, Reilly T, Morton JP, Bae SW, and Drust B. Quantification of the physiological loading of one week of "pre-season" and one week of "in-season" training in professional soccer players. *Journal of Sports Sciences*, 29(11): 1161–1166, 2011.

[46] Jones AM and Doust JH. The Conconi test is not valid for estimation of the lactate turnpoint in runners The Conconi test is not valid for estimation of the lactate turnpoint in runners. *Journal of Sports Sciences*, 15(4):385–394, 1997.

[47] Ju W, J Lewis C, Evans M, Laws A, and S Bradley P. The validity and reliability of an integrated approach for quantifying match physical-tactical performance. *Biology of Sport*, 39(2):253–261, 2022.

[48] Lacome M, Simpson BM, and Buchheit M. Part 2: Increasing coach 'buy-in' with good data visualisation. *Aspetar Sports Medicine Journal*, 1(August):64–66, 2018.

[49] Lames M. Chance involvement in goal scoring in football – an empirical approach. *German Journal of Exercise and Sport Research*, 48(2):278–286, 2018.

[50] Lepschy H, Wäsche H, and Woll A. How to be Successful in Football: A Systematic Review. *The Open Sports Sciences Journal*, 11(1):3–23, 2018.

[51] Lepschy H, Wäsche H, and Woll A. Success factors in football: an analysis of the German Bundesliga. *International Journal of Performance Analysis in Sport*, 20(2):150–164, 2020.

[52] Little T and Williams AG. Measures of Exercise Intensity During Soccer Training Drills With Professional Soccer Players. *Journal of Strength and Conditioning Research*, 21(2):367–371, may 2007.

[53] Los Arcos A, Yanci J, Mendiguchia J, and Gorostiaga EM. RATING OF MUSCULAR AND RESPIRATORY PERCEIVED EXERTION IN PROFESSIONAL SOCCER PLAYERS. *Journal of Strength and Conditioning Research*, 28(11):3280–3288, 2014.

[54] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, and Lee SI. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67, 2020.

[55] Macfadyen AM. Edwards, A.M.; Clark, N. & Macfadyen, A.M. *Journal of Sports Science and Medicine*, 2:23–29, 2003.

[56] MacKenzie R and Cushion C. Performance analysis in football: A critical review and implications for future research'. *Journal of Sports Sciences*, 32(1):2–7, 2013.

[57] Manzi V, Bovenzi A, Impellizzeri FM, Carminati I, and Castagna C. Individual training-load and aerobic-fitness variables in premiership soccer players during the precompetitive season. *Journal of Strength and Conditioning Research*, 27(3): 631–636, 2013.

[58] Mark MM, Mutrie N, Brooks DR, and Harris DV. Causal Attributions of Winners and Losers in Individual Competitive Sports: Toward a Reformulation of the Self-Serving Bias. *Journal of Sport Psychology*, 6:184–196, 1984.

[59] McMillan K, Helgerud J, Grant SJ, Newell J, Wilson J, Macdonald R, and Hoff J. Lactate threshold responses to a season of professional British youth soccer. *British Journal of Sports Medicine*, 39(7):432–436, 2005.

[60] Meur YL. 10 Challenges Facing Today ' s Applied Sport Scientist, 2019.

[61] Mujika I and Padilla S. Detraining: Loss of Training-Induced Physiological and Performance Adaptations. Part I. *Sports Medicine*, 30(3):79–87, 2000.

[62] Mujika I, Halson S, Burke LM, Balagué G, and Farrow D. An Integrated, Multifactorial Approach to Periodization for Optimal Performance in Individual and Team Sports. *International Journal of Sports Physiology and Performance*, 13(1): 538–561, 2018.

[63] Murphy R. FIFA player ratings explained: How are the card number & stats decided?, 2019. URL `https://www.goal.com/en-ae/news/fifa-player-ratings-explained-how-are-the-card-number-stats/1hszd2fgr7wgf1n2b2yjdpgynu`.

[64] Nédélec M, Halson S, Abaidia AE, Ahmaidi S, and Dupont G. Stress, Sleep and Recovery in Elite Soccer: A Critical Review of the Literature. *Sports Medicine*, 45(10):1387–1400, 2015.

[65] Pandolf KB. Differentiated ratings of perceived exertion during physical exercise. *Medicine & Science in Sports & Exercise*, 14(5):397–405, 1982.

[66] Pappalardo L, Cintia P, Ferragina P, Massucco E, Pedreschi D, and Giannotti F. PlayeRank: Multi-dimensional and role-aware rating of soccer player performance. *ACM Transactions on Intelligent Systems and Technology*, 10(5):1–27, 2018.

[67] Park YS, Konge L, and Artino AR. The Positivism Paradigm of Research. *Academic Medicine*, 95(5):690–694, 2020.

[68] Perin C, Vuillemot R, Stolper CD, Stasko JT, Wood J, and Carpendale S. State of the Art of Sports Data Visualization. *Computer Graphics Forum*, 37(3):663–686, 2018.

[69] Plessner H and Haar T. Sports performance judgments from a social cognitive perspective. *Psychology of Sport and Exercise*, 7(6):555–575, 2006.

[70] Reep C and Benjamin B. Skill and Chance in Association Football. *Journal of the Royal Statistical Society*, 131(4): 581–585, 1968.

[71] Rossi A, Perri E, Trecroci A, Savino M, Alberti G, and Iaia M. GPS data reflect players' internal load in soccer. In *IEEE International Conference on Data Mining Workshops, ICDMW*, pages 890–893, New Orleans, LA, USA, 2017. IEEE.

[72] Rowley J. The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2):163–180, 2007.

[73] Sieghartsleitner R, Zuber C, Zibung M, and Conzelmann A. 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 and Medicine*, 18(1):32–43, 2019.

[74] Simon R. Deserving to be lucky: Reflections on the role of luck and desert in sports. *Journal of the Philosophy of Sport*, 34(1):13–25, 2007.

[75] Snowden D. Complex Knowledge. In *Gurteen Knowledge Conference*, London, 2003.

[76] Stolen T, Chamari K, Castagna C, and Wisloff U. Physiology of Soccer. *Sports Medicine*, 35(6):501–536, 2005.

[77] Svensson M and Drust B. Testing soccer players. *Journal of Sports Sciences*, 23(6):601–618, 2005.

[78] Torres-Ronda L and Schelling X. Critical process for the implementation of technology in sport organizations. *Strength and Conditioning Journal*, 39(6):54–59, 2017.

[79] Walker GJ and Hawkins RD. Structuring a Program in Elite Professional Soccer. *Strength and Conditioning Journal*, 40 (3):72–82, 2018.

[80] Weaving D, Jones B, Till K, Abt G, and Beggs C. The case for adopting a multivariate approach to optimize training load quantification in team sports. *Frontiers in Physiology*, 8(DEC):1–3, 2017.

[81] Windt J, MacDonald K, Taylor D, Zumbo BD, Sporer BC, and Martin DT. "To tech or not to tech?" A critical decision-making framework for implementing technology in sport. *Journal of Athletic Training*, 55(9):902–910, 2020.

[82] Zhao B. *Encyclopedia of Big Data*. Springer International Publishing AG, 2020.

[83] Ziogas GG, Patras KN, Stergiou N, and Georgoulis AD. Velocity at lactate threshold and running economymust also be considered alongwith maximal oxygen uptake when testing elite soccer players during preseason. *Journal of Strength and Conditioning Research*, 25(2):414–419, 2011.

Part IV

Addenda

A Description of variables used in Study I

Variable	Description
Shot	A clear attempt to score [2].
Shots on Target	1) Any goal attempt that goes into the net regardless of intent, or 2) a clear attempt to score that would have gone into the net but is being saved by the goalkeeper or 3) is stopped by a player who is the last-man with the goalkeeper having no chance of preventing the goal (last line block) [2].
Shots not on Target	Any clear attempt to score that 1) goes over or wide of the goal without making contact with another player or 2) would have gone over or wide of the goal but is being stopped by a goalkeeper's save or by an outfield player or 3) directly hits the frame of the goal and a goal is not scored. Blocked shots are not counted as shots off target [2].
Shot from inside attacking penalty box	A clear attempt to score from inside the penalty box.
Shot on Target from inside attacking penalty box	1) Any goal attempt from inside the attacking penalty box that goes into the net regardless of intent, or 2) a clear attempt from inside the attacking penalty box to score that would have gone into the net but is being saved by the goalkeeper or 3) shot from inside the attacking penalty box stopped by a player who is the last-man with the goalkeeper having no chance of preventing the goal (last line block) [2].
Shot not on Target from inside attacking penalty box	Any clear attempt from inside attacking penalty box to score that 1) goes over or wide of the goal without making contact with another player or 2) shot from inside the attacking penalty box that would have gone over or wide of the goal but is being stopped by a goalkeeper's save or by an outfield player or 3) shot from inside the attacking penalty box that directly hits the frame of the goal and a goal is not scored [2].
Expected Goals	Measures the quality of a shot based on several variables such as assist type, shot angle and distance from goal, whether it was a headed shot and whether it was defined as a big chance. Adding up a player's or team's expected goals can give us an indication of how many goals a player or team should have scored on average, given the shots they have taken [1, 2].
Shots from Maintenance (Playing Styles)	Shots resulting from Maintenance phase. Maintenance captures possessions in which a team looks to maintain and secure possession of the ball within the defensive area of the pitch [3].
Shots from Build Up (Playing Styles)	Shots resulting from Build Up phase. Build Up captures long and controlled ball possessions – but is aimed at periods of play where a team is looking for opportunities to attack. [3].

Table 1 – continues on next page

Table 1 – continued from previous page

Variable	Description
Shots from Sustained Threat (Playing Styles)	Shots resulting from Sustained Threat phase. The Sustained Threat Playing Style focus lies on possessions in the attacking third of the pitch. [3].
Shots from Fast Tempo (Playing Styles)	Shots resulting from Fast Tempo phase. The objective of the Fast Tempo Playing Style is to capture when the team is moving the ball quickly on the opposing half to increase the tempo and speed of the game [3].
Shots from Direct Play (Playing Styles)	Shots resulting from Direct Play phase. Direct play captures instances of play in which teams attempt to move the ball quickly towards the opposition's goal through the use of long passes rather than sustained possession [3].
Shots from Counter Attack (Playing Styles)	Shots resulting from Counter Attack phase. A counter attack occurs once a team regains possession and moves the ball into an attacking area via passes, dribbles or a combination of both [3].
Shots from Crossing (Playing Styles)	Shots resulting from Crossing phase. As a playing style, crossing occurs if the ball is delivered from a wide area of the pitch with the intention of finding a teammate [3].
Shots from High Press (Playing Styles)	Shots resulting from High Press phase. High press captures regaining of possession higher than 5 metres before the halfway line, when the opponent is in possession for at least 10 seconds [3].
Duels won	Frequency in which a 50-50 contest between two players of opposing sides in the match is won [2].
Possession won	Frequency in which possession over the ball is gained [2].
Total passes	Total frequency of intentional played balls from one player to another. Passes include open play passes, goal kicks, corners and free kicks played as pass – but exclude crosses, keeper throws and throw-ins [2].
Forward passes	Total frequency of intentional played balls from one player to another, in an angle of -45° to 45° relative to the opponent goal [2].
Sideways passes	Total frequency of intentional played balls from one player to another, in an angle of -135° to -45°, and 45° to 135° relative to the opponent goal [2].
Backward passes	Total frequency of intentional played balls from one player to another, in an angle of -135° to 135° relative to the opponent goal [2].
Successful passes	Total frequency of passes in which the pass goes to a team mate directly without a touch from an opposition player [2].
Total passes to attacking half	Total frequency of intentional played balls from one player to another with the attacking half as destination [2].

Table 1 – continues on next page

Table 1 – continued from previous page

Variable	Description
Forward passes to attacking half	Total frequency of intentional played balls from one player to another, in an angle of -45° to 45° relative to the opponent goal with the attacking half as destination [2].
Sideways passes to attacking half	Total frequency of intentional played balls from one player to another, in an angle of -135° to -45°, and 45° to 135° relative to the opponent goal with the attacking half as destination [2].
Backward passes to attacking half	Total frequency of intentional played balls from one player to another, in an angle of -135° to 135° relative to the opponent goal with the attacking half as destination [2].
Total passes to attacking third	Total frequency of intentional played balls from one player to another with the attacking third as destination [2].
Forward passes to attacking third	Total frequency of intentional played balls from one player to another, in an angle of -45° to 45° relative to the opponent goal with the attacking third as destination [2].
Sideways passes to attacking third	Total frequency of intentional played balls from one player to another, in an angle of -135° to -45°, and 45° to 135° relative to the opponent goal with the attacking third as destination [2].
Backward passes to attacking third	Total frequency of intentional played balls from one player to another, in an angle of -135° to 135° relative to the opponent goal with the attacking third as destination [2].
Dribbles	This is an attempt by a player to beat/bypass an opponent when they have possession of the ball [2].
Successful dribbles	A successful dribble means the player beats/bypasses the defender while retaining possession [2].
Ball touches in attacking penalty box	Total frequency of a player touching the ball inside the attacking penalty box [2].
Passes <10 meter	Total frequency of intentional played balls from one player to another with a length <10 meter [2].
Passes <25 meter	Total frequency of intentional played balls from one player to another with a length <25 meter [2].
Passes >25 meter	Total frequency of intentional played balls from one player to another with a length >25 meter [2].
Possession in Maintenance (Playing Styles)	Total frequency of ball possession in Maintenance phase. Maintenance captures possessions in which a team looks to maintain and secure possession of the ball within the defensive area of the pitch [3].

Table 1 – continues on next page

Table 1 – continued from previous page

Variable	Description
Possession in Build Up (Playing Styles)	Total frequency of ball possession in Build Up phase. Build Up captures long and controlled ball possessions – but is aimed at periods of play where a team is looking for opportunities to attack. [3].
Possession in Sustained Threat (Playing Styles)	Total frequency of ball possession in Sustained Threat phase. The Sustained Threat Playing Style focus lies on possessions in the attacking third of the pitch. [3].
Possession in Fast Tempo (Playing Styles)	Total frequency of ball possession in Fast Tempo phase. The objective of the Fast Tempo Playing Style is to capture when the team is moving the ball quickly on the opposing half to increase the tempo and speed of the game [3].
Possession in Direct Play (Playing Styles)	Total frequency of ball possession in Direct Play phase. Direct play captures instances of play in which teams attempt to move the ball quickly towards the opposition's goal through the use of long passes rather than sustained possession [3].
Possession in Counter Attack (Playing Styles)	Total frequency of ball possession in Counter Attack phase. A counter attack occurs once a team regains possession and moves the ball into an attacking area via passes, dribbles or a combination of both [3].
Possession in Crossing (Playing Styles)	Total frequency of ball possession in Crossing phase. As a playing style, crossing occurs if the ball is delivered from a wide area of the pitch with the intention of finding a teammate [3].
Possession in High Press (Playing Styles)	Total frequency of ball possession in High Press phase. High press captures regaining of possession higher than 5 meters before the halfway line, when the opponent is in possession for at least 10 seconds [3].
Total ball possession	Total time in possession relative to the opponent [2].
Ball possession on defensive half	Possessions (one or more sequences in a row belonging to the same team, without opposition gaining control of the ball), starting in the defensive half.
Ball possession on attacking half	Possessions (one or more sequences in a row belonging to the same team, without opposition gaining control of the ball), starting in the attacking half.
Ball possession on attacking third	Possessions (one or more sequences in a row belonging to the same team, without opposition gaining control of the ball), starting in the attacking third.
Ball possession on attacking penalty box	Possessions (one or more sequences in a row belonging to the same team, without opposition gaining control of the ball), starting in the attacking penalty box
Total distance	Average total distance covered by all field players.
Distance between 0-6 km/h	Average distance covered at a speed between 0-6 km/h by all field players.

Table 1 – continues on next page

Table 1 – continued from previous page

Variable	Description
Distance between 6-15 km/h	Average distance covered at a speed between 6-15 km/h by all field players.
Distance between 15-20 km/h	Average distance covered at a speed between 15-20 km/h by all field players.
Distance between 20-25 km/h	Average distance covered at a speed between 20-25 km/h by all field players.
Distance >25 km/h	Average distance covered >25 km/h by all field players.
Distance at accelerations >2ms	Average distance covered when accelerating at >2ms by all field players.
Distance at decelerations >2ms	Average distance covered when decelerating at >2ms by all field players.
Distance at accelerations >3ms	Average distance covered when accelerating at >3ms by all field players.
Distance at decelerations >3ms	Average distance covered when decelerating at >3ms by all field players.
Actions at speed >15 km/h	Average total number of actions performed at a speed >15 km/h by all field players.
Actions at speed >25 km/h	Average total number of actions performed at a speed >25 km/h by all field players.
Accelerations >2ms	Average number of accelerations at >2ms by all field players.
Decelerations >2ms	Average number of decelerations at >2ms by all field players.
Accelerations >3ms	Average number of accelerations at >3ms by all field players.
Decelerations >3ms	Average number of decelerations at >3ms by all field players.
Fouls	Any infringement that is penalised as foul play by a referee [2].
Fouls at attacking half	Any infringement that is penalised as foul play by a referee on the attacking half [2].
Yellow cards	Number of yellow cards awarded by the referee [2].
Red cards	Number of red cards awarded by the referee [2].
Offside	Awarded to the player deemed to be in an offside position where a free kick is awarded [2].

Table 1 – continues on next page

Table 1 – continued from previous page

Variable	Description
Corners	Frequency of awarded corner kicks [2].
Penalties	Frequency of awarded penalties [2].
Free kick on attacking third	Number of free kicks on the attacking third [2].
Throw-in on attacking third	Number of throw-ins on the attacking third [2].
Line-up estimated total transfer value	Total estimated transfer value of the players in line-up, data derived from Transfermarkt.com.
Bench current estimated total transfer value	Total estimated transfer value of the benched players, data derived from Transfermarkt.com.
Line-up estimated paid total transfer	Total estimated paid transfer fees of the players in line-up, data derived from Transfermarkt.com.
Bench estimated paid total transfer	Total estimated paid transfer fees of the benched players, data derived from Transfermarkt.com.
Line-up average age	Average age of players in line-up, data derived from Transfermarkt.com.
Bench average age	Average age of benched players, data derived from Transfermarkt.com.
ClubELO	A club's ELO rating is an estimation of its strength based on past results allowing predictions for the future. Data derived from Clubelo.com.
Form	Form was defined as the difference in a clubs' current ELO-rating compared to the ELO-rating before their previous game. Data derived from Clubelo.com.
Days between games	Total number of games between the current and previous game.
Home/away	Binary statement on whether the team played at home or at the opponent venue.

References

[1] STATS Perform. Expected Goals. URL: https://www.statsperform.com/opta-analytics

[2] STATS Perform. Opta Event Definitions. URL: https://www.statsperform.com/opta-event-definitions

[3] STATS Perform. Playing Styles - An Introduction. URL: https://www.statsperform.com/resource/stats-playing-styles/introduction

List of publications

A1 **Y. Geurkink**, G. Vandewiele, M. Lievens, F. De Turck, F. Ongenae, S. P. Matthys, J. Boone, and J. G. Bourgois, "Modelling the Prediction of the Session Rating of Perceived Exertion in Soccer: Unraveling the Puzzle of Predictive Indicators," *International Journal of Sports Physiology and Performance,*. vol. 14, no. 6, pp. 1-6, 2019.

➔ Design of the work, acquisition of data, analysis and interpretation of data, drafting the work.

A1 **Y. Geurkink**, J. Boone*, S. Verstockt* and J. G. Bourgois*, "Machine Learning-Based Identification of the Strongest Predictive Variables of Winning and Losing in Belgian Professional Soccer," *Applied Sciences,*. vol. 11, no. 5, pp. 2378, 2021.
* These authors share last authorship

➔ Design of the work, acquisition of data, analysis and interpretation of data, drafting the work.

C1 G. Vandewiele, **Y. Geurkink**, M. Lievens, F. Ongenae, F. De Turck, and J. Boone, "Enabling Training Personalization by Predicting the Session Rate of Perceived Exertion (sRPE)," in *Machine Learning and Data Mining for Sports Analytics ECML/PKDD,* September 2017, pp. 1-12.

➔ Acquisition of data, critically revising the work, final approval of the version to be published.

Currently in submission

A1 **Y. Geurkink**, J. Boone, S. P. J. Matthys, J.G. Bourgois. "Training Characteristics and Training Intensity Distribution of the Preseason in a Professional Soccer Team," in *Journal of Exercise Science & Fitness*.

➔ Design of the work, acquisition of data, analysis and interpretation of data, drafting the work.

Curriculum Vitae

Education
2012-2017, Ghent University, Ghent, Belgium
Movement and Sports Sciences

Other certificates
Royal Belgian Football Association (KBVB):
◦ UEFA B, Physical Coach Level 1

IPVW-ICES:
◦ Statistics with Python, Machine Learning with Python

Experience

◦ September 2017 - November 2017, *Liaison UGent - VTS*, Ghent University
Temporary replacement, as a mentor for students with regard to sport-specific and professional internships.
◦ November 2017 - October 2021, *PhD-student*, Ghent University
PhD-student at project KAA Gent - UGent Performance Center. Within this project I am working as a sports scientist in the first team of KAA Gent on a daily basis. Specific responsibilities are/were:

- Preparing, processing and analyzing GPS-data (STATS Perform / Polar).
- Organization and reporting of anthropometric and physical testing of youth teams KAA Gent (U7-U21).
- Taking daily/weekly/periodical measures, such as weight, hydration, NordBord, Groinbar, sprint-& jumptests, ..
- Preparation and guidance of strenght sessions (in group & individually).
- Preparing and leading individual sessions on and off the pitch.
- Processing and reporting physical & technical/tactical game data.
- Management ProSoccerData/SoccerLab
- Management COVID-testing Champions League & Europa League.

◦ November 2021 - ..., *Physical and scientific coordinator*, Twente/Heracles Academy

Academic output

◦ A1 - Modelling the Prediction of the Session Rate of Perceived Exertion in Soccer Unravelling the Puzzle of Predictive Indicators. - *International Journal of Sports Performance and Physiology*.
◦ A1 - Machine Learning based Identification of the Strongest Predictive Variables of Winning and Losing in Belgian Professional Soccer. - *Applied Sciences*.
◦ A1 - Training Characteristics and Training Intensity Distribution of the Preseason in a Professional Soccer Team - *Submitted to Journal of Exercise Science and Fitness*.
◦ Reviews voor PlosONE en Biology of Sport.