

Energiemonitoring van voetbalspelers

Theo Vandeportaele

Studentennummer: 02008498

Promotoren: prof. dr. Steven Verstockt, dr. ir. Joachim Taelman

Begeleiders: dr. ir. Maarten Slembrouck, dr. Jan Van Haaren (Club Brugge)

Masterproef ingediend tot het behalen van de academische graad van
Master of Science in de industriële wetenschappen: informatica

Academiejahr 2023-2024

Dankwoord

Eerst wil ik Prof. Dr. Steven Verstockt bedanken voor de kans die hij me gaf om mijn eigen voorstel voor mijn masterproef te maken. Ik heb bij hem mijn idee voorgelegd en hij heeft ervoor gezorgd dat het werkelijkheid werd. Ook wil ik Dr. ir. Joachim Taelman en Dr. ir. Maarten Slembrouck bedanken voor hun waardevolle bijdrage aan mijn onderzoek. Hun kritische blik en constructieve feedback hebben mijn presentaties en geschreven teksten aanzienlijk verbeterd.

Verder wil ik graag mijn oprechte dank uitspreken aan Club Brugge en in het bijzonder aan Dr. Jan van Haaren. Mijn twee stageperiodes bij Club Brugge hebben een cruciale rol gespeeld in mijn interesse in data analyse in voetbal. Zonder deze ervaringen zou deze masterproef niet tot stand zijn gekomen. Bovendien heeft Jan altijd voorzien in de nodige data en expertise wanneer ik die nodig had. Zijn ondersteuning is van onschatbare waarde geweest voor mijn onderzoek en persoonlijke groei.

Als laatste wil ik mijn ouders bedanken voor hun onvoorwaardelijke steun gedurende mijn hele schoolcarrière en in het bijzonder tijdens het maken van mijn masterproef. Ze hebben altijd achter me gestaan en gezorgd dat ik me enkel met mijn studies moest bezig houden wanneer dat nodig was. Een speciale vermelding gaat uit naar mijn mama, die al mijn teksten heeft nagelezen op taal- en typfouten.

Theo Vandeportaele, mei 2024

Toelichting in verband met het masterproefwerk

Deze masterproef vormt een onderdeel van een examen. Eventuele opmerkingen die door de beoordelingscommissie tijdens de mondelinge uiteenzetting van de masterproef werden geformuleerd, werden niet verwerkt in deze tekst.

Abstract

Nederlands

Deze masterproef presenteert een aanpak voor het monitoren van de energie van voetballers tijdens wedstrijden. Het doel is om de vermoeidheid bij spelers op te sporen tijdens wedstrijden. Vermoeidheid kan leiden tot verminderde fysieke en technische prestaties en een verhoogd risico op blessures. Dit wilt men te allen tijde vermijden. Om de energieniveaus van de spelers te bepalen wordt gebruik gemaakt van fysieke parameters van spelers. Er wordt gebruik gemaakt van data in verband met de afgelegde afstand en data in verband met de snelheid. Deze twee soorten data worden berekend uit de trackingdata aangeboden door StatsPerform. De studie omvat de ontwikkeling van zowel een offline als online programma. Het offline programma werkt op trackingdata van StatsPerform die na de wedstrijd werd aangeboden en postprocessed is. Dit werd ontwikkeld om conclusies te kunnen vormen over vermoeidheid op basis van de fysieke parameters van spelers, alsook om te zien welke andere factoren invloed hebben op de fysieke parameters. Het online programma bestaat uit een dashboard en werd ontwikkeld voor live-gebruik tijdens wedstrijden. Zowel het online en offline programma berekent data voor het opstellen van twee grafieken die vermoeidheid kunnen weergeven. De eerste geeft de richtingscoëfficiënt van de cumulatieve afgelegde afstand weer en de tweede het aantal frames per minuut met een snelheid hoger dan het dubbele van de gemiddelde snelheid.

Aan de hand van het offline programma, worden een aantal resultaten bekomen over welke invloed bepaalde wedstrijdssituaties op vermoeidheid hebben. Dit wordt gedaan door te kijken hoe de grafieken met fysieke data van spelers zich gedragen ten opzichte van de gemiddelde fysieke data (berekend voor elke speler over de 30 wedstrijden uit de reguliere competitie heen) in bepaalde situaties. Bij korte invalbeurten van minder dan 15 minuten is de data niet uitgebreid genoeg om deftige conclusies te kunnen nemen. Bij invalbeurten tussen de 15 en 45 minuten is er wel voldoende data en zullen de grafieken van de spelers in 71% van de gevallen boven het gemiddelde eindigen. Wanneer spelers invielen bij een verschil in de stand van 2 doelpunten of meer, eindigden zelf 92% van de grafieken boven het gemiddelde. Wanneer een speler een rode kaart krijgt, zullen zijn teamgenoten gemiddeld 75% minder lopen na het krijgen van de rode kaart, bij een voorsprong. Bij een gelijkspel zal 64% van de spelers minder lopen en bij een achterstand is het gelijk verdeeld (50% zal meer lopen en 50% minder). De invloed van expected op de afgelegde afstand van spelers is positieafhankelijk. De invloed van balbezit is niet positieafhankelijk: spelers zullen gemiddeld meer lopen bij lager balbezit. Uit het online programma kunnen er ook conclusies genomen worden omtrent de bruikbaarheid van de data. Slechts voor 11 van de 30 wedstrijden in de reguliere competitie van Club Brugge is er complete trackingdata die live tijdens wedstrijden gebruikt kan worden.

English

This master's thesis presents an approach for the energy monitoring of football players during matches. The

objective is to detect player fatigue during matches, which can result in a decreased physical and technical performance and an increased risk of injuries. This is something to be avoided at all times. To determine the players' energy levels, physical parameters of players are utilized. Data related to distance covered and data related to speed are calculated from the tracking data provided by StatsPerform. The study includes the development of both an offline and online program. The offline program operates on tracking data from StatsPerform that was provided after the match and has been post-processed. It was developed to draw conclusions about fatigue based on the physical parameters of players and also to detect which other factors influence the physical parameters. The online program consists of a dashboard that can be used live during games and is based on the live tracking data of StatsPerform. The results of this research provide valuable insights into the dynamics of player fatigue during football matches. Both the online and offline programs calculate data to create two graphs that can indicate fatigue. The first graph shows the directional coefficient of the cumulative distance covered and the second one shows the number of frames per minute with a speed higher than twice the average speed.

Using the offline program, several results were obtained regarding the influence of certain match situations on fatigue. This was done by observing how the players' physical data graphs behave relative to the average physical data (calculated for each player over 30 regular-season matches) in specific situations. For short substitutions of less than 15 minutes, the data is not extensive enough to draw reliable conclusions. For substitutions between 15 and 45 minutes, there is sufficient data and the players' graphs will end above the average in 71% of cases. When players were substituted in with a score difference of 2 goals or more, 92% of the graphs ended above average. When a player receives a red card, their teammates will run an average of 75% less after the red card is issued if they are winning. In the case of a tie, 64% of the players will run less, and if their team is losing, it is evenly split (50% will run more and 50% will run less). The influence of expected goals on the distance covered by players is position-dependent. The influence of ball possession is not position-dependent: players will generally run more with lower ball possession. Conclusions can also be drawn from the online program regarding the usability of the data. Only 11 out of 30 regular-season matches of Club Brugge had complete tracking data that could be used live during matches.

Energy monitoring of football players

Theo Vandeportaele
Ghent University
Ghent, Belgium
theo.vandeportaele@ugent.be

Steven Verstockt
IDLab, Ghent University
Ghent, Belgium
steven.verstockt@ugent.be

Jan van Haaren
Club Brugge
Brugge, Belgium
jan.vanhaaren@clubbrugge.be

Maarten Slembrouck
IDLab, Ghent University
Ghent, Belgium
maarten.slembrouck@ugent.be

Joachim Taelman
IDLab, Ghent University
Ghent, Belgium
joachim.taelman@ugent.be

Abstract—This master’s thesis presents an approach for the monitoring of the energy of football players during matches. The objective is to detect player fatigue during matches, which can result in decreased physical and technical performance and an increased risk of injuries. This is something to be avoided at all times. To determine the players’ energy levels, physical parameters of players are utilized. Data related to distance covered and data related to speed are calculated from the tracking data provided by StatsPerform. The study includes the development of both an offline and online program. The offline program operates on tracking data from StatsPerform that was provided after the match and has been post-processed. This was developed to draw conclusions about fatigue based on the physical parameters of players and also to detect which other factors influence the physical parameters. The online program consists of a dashboard that can be used live during games and is based on the live tracking data of StatsPerform. The results of this research provide valuable insights into the dynamics of player fatigue during football matches.

Index Terms—trackingdata, fatigue, StatsPerform, energy monitoring, football

I. INTRODUCTION

In recent years, the usage of data has been on a rise in the football world. The use of data has become essential in the daily operations of a football team. Data is used, for example, to determine players’ physical parameters during matches and training sessions. Based on these physical parameters, player fatigue can be assessed. As the number of matches per season increases and they become more frequent, it is crucial to monitor players’ physical parameters during a match to intervene on time when fatigue occurs.

The objective of this thesis is to design an application that analyses players’ physical parameters during matches. It also examines the conclusions that can be drawn from these physical parameters, regarding fatigue. To calculate the physical parameters, tracking data from StatsPerform is utilized.

Initially, an offline program is developed, which operates on the post-processed tracking data from StatsPerform to draw conclusions from trends in players’ physical data. After that, an online program is created. This includes a dashboard and can be used live during a match to track player fatigue.

II. BACKGROUND

A. Introduction

The rapid professionalization and commercialization of football entails that seasons are becoming increasingly longer, consisting of more matches and these matches consequently follow each other more closely. Professional teams play an average of between 50 and 80 matches per season [1]. One risk factor associated with playing so many matches is fatigue. Fatigue is a risk factor that leads to a higher likelihood of injuries. It also results in diminishing physical and technical performances of players. A tired player has a higher chance of poor control or passing, and will also run less and at a slower pace on average [2].

B. Fatigue

Fatigue can be described as a decrease in physical, mental, and technical performance associated with an increase in the actual intensity or difficulty of the task or exercise that needs to be performed. In the world of sports, the stimulus for fatigue is primarily physical activity [3]. The fatigue experienced by players during training and matches can be categorized into two different types of loads. Physiological stressors occurring during training and matches are referred to as internal load. An example of physiological stressors is a player’s elevated heart rate during the match. Kinematic activities that create physiological stressors are termed external load. Some examples of kinematic activities include the distance covered during a match or the number of sprints [4]

C. Measure fatigue

Figure 1 shows the average distance covered by a team in defense per minute during a match. It is clear that teams cover much fewer meters towards the end of the match compared to the beginning. This naturally applies to individual players as well. This decrease may be due to fatigue or the match result. In either case, it indicates a decline in performance level.

However, only considering the distance covered may not provide a total picture of a player's fatigue. A player might cover a large distance during a match but at a very low intensity, or cover a shorter distance but at a much higher intensity. Therefore, other external load parameters are also necessary. One of the other parameters that can be used is the distance covered at high intensity per minute. Figure 2 demonstrates that, similar to the total distance covered (blue line), the total distance covered at high intensity (orange line) can also indicate fatigue. The high intensity distance decreases even more strongly and directly than the total distance covered, thus showing signs of fatigue even sooner [5].

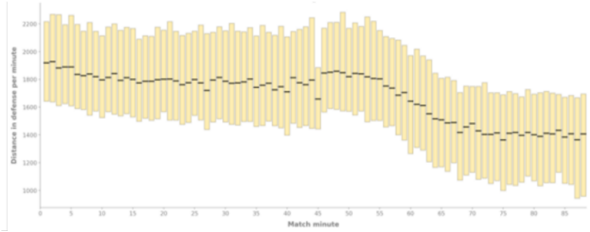


Fig. 1. The average distance covered in defense during a football match

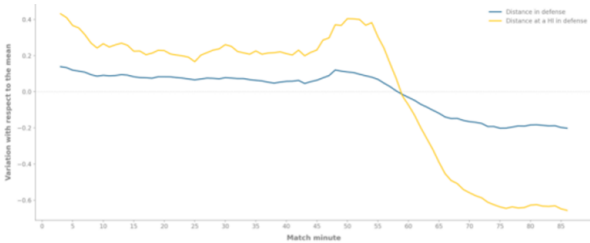


Fig. 2. The average distance covered at high intensity during a football match

III. DATA EXPLORATION

For the thesis, two types of data are used: tracking data and API data from StatsPerform. Tracking data includes the positional coordinates of players and the ball during matches. There are two types of tracking data: the post-processed tracking data, which is made available after the match via the StatsPerform API and the non-post-processed tracking data, which is made available live during the match via the FTP server of StatsPerform. The API data consists of various endpoints, with each type of data identified by an ID and available in JSON or XML format. The API endpoints include data on match statistics, ball possession, expected goals, player

ratings, and win probability. This data is used to gain insights into the performances of teams and players during matches, for both analysis purposes and the online program.

IV. DATA PROCESSING

The tracking data comprises a large number of rows, at least 135,000. To process this data more easily, the external library Floodlight [6] is used, which has built-in functions to read the data and automatically convert it into a floodlight XY object. This object contains positional coordinates of the players throughout the match. Functions can be applied to convert these coordinates into physical data. This allows for the calculation of cumulative distance (distance model) or velocity (velocity model) with ease.

The distance model calculates the distance a player travels between two frames, which can be easily converted into the total cumulative distance covered by a player during a match. To obtain a clear overview of how the cumulative distance changes per minute, the slope of the cumulative distance is calculated. This is shown in figure 3 and is one of the two types of graphs used to determine player fatigue.

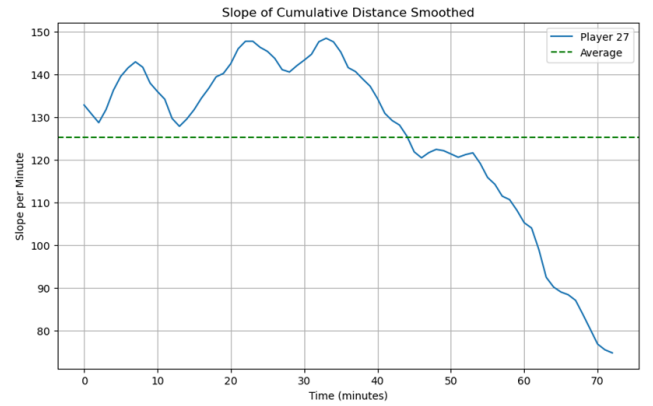


Fig. 3. Graph of the slope of cumulative distance

The velocity model calculates the velocity at which a player moves between two frames. It is not possible to use the same approach as the distance model. Therefore, the decision was made to calculate the number of frames per minute where a player achieved a velocity above a threshold value. The chosen threshold value is twice the player's average velocity. This way, the threshold value is different for each player and better aligns with the player's physical parameters. The trend of the number of frames per minute above that threshold value (as seen in Figure 4) is the second type of graph used to determine fatigue.

V. THE FUNCTIONING OF THE OFFLINE PROGRAM

An offline program has been developed that utilizes tracking data to generate graphs displaying players' physical data. This data is obtained via the StatsPerform API and undergoes post-processing, ensuring minimal to no errors and has a significantly higher data quality than the live tracking data.



Fig. 4. Graph of the number of frames above a certain threshold per minute

The offline program generates two types of graphs for each player: one depicting the slope coefficient of cumulative distance and another illustrating the number of frames with a velocity exceeding twice the average velocity. These graphs are generated based on tracking files from individual matches using the Floodlight library.

A. The quality of the offline program

In this section, the quality of data generated by the offline program is analyzed, focusing solely on the quality of data regarding the slope coefficient of cumulative distance, as it is the same as for the number of frames exceeding twice the average velocity. The offline data (which is based on post-processed tracking data) is compared with that of StatsPerform and the online program, both of which are based on non-post-processed tracking data. The average percentage difference between the physical data from StatsPerform and the offline data is 8.12%, while the average difference with the online data is only -0.72%. This is because both the online data and the physical data from StatsPerform utilize the online tracking data, whereas the offline data is based on post-processed offline tracking data.

Since the online tracking data is provided live during a match and it is therefore important that it is created at a high velocity, the quality of the data will suffer. This data will be significantly less accurate than the post-processed offline tracking data. Graphs generated by the offline and online programs exhibit similar trends, but with slight differences in the average trend. Figure 5 and 6 provides an example of this. The graph from the online program generally shows a higher cumulative distance. Also, the peaks and troughs will have respectively higher maxima and lower minima for the online program.

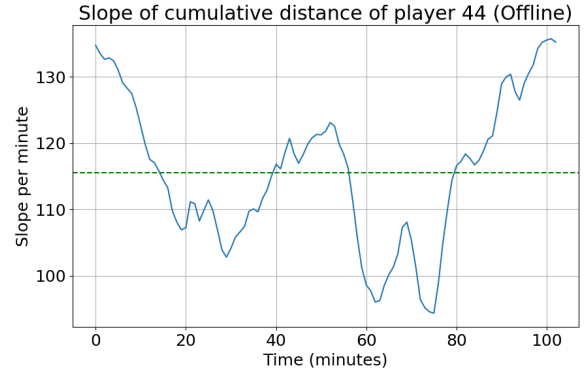


Fig. 5. Graph of the number of frames above a certain threshold per minute

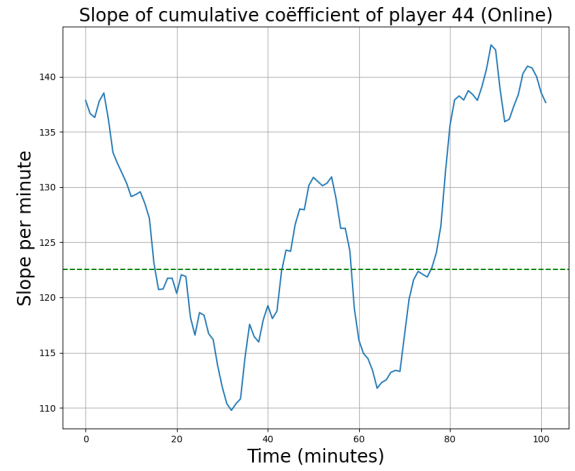


Fig. 6. Graph of the number of frames above a certain threshold per minute

B. Special cases and exceptions

This section discusses special cases and exceptions that can influence the trajectory of graphs with physical data of players, generated by an offline program. The two situations covered are short substitutes and the impact of red cards.

For short substitutes, the usefulness of monitoring physical data to track fatigue for players who play less than 45 minutes in a match is investigated. This is done by examining whether the graphs of the slope of the cumulative distance of players ends above or below the average, that is calculated based on previous matches. The data shows that for substitutes of less than 15 minutes, the results are evenly split between graphs ending above and below the average (50% each). There is insufficient data for these graphs to draw clear conclusions about fatigue. For substitutes of 15 to 35 minutes, 75% of players have a graph ending above the average. For these players, there is sufficient data available. For substitutes of 35 to 45 minutes, this percentage is slightly lower (69%). These results can be seen in table I.

The impact of the score at the time of substitution was also

TABLE I

NUMBER OF SUBSTITUTE PLAYERS WITH A GRAPH SHOWING A TREND THAT ENDS ABOVE OR BELOW THE AVERAGE VALUE PER TIME INTERVAL

Time interval	Total number of players	Above average	Below average	Percentage above average
0-15	58	29	29	50.00
15-35	47	35	12	74.47
35-45	29	20	9	68.97

investigated. When the score difference is 2 goals or more, 91.67% of players have a positive graph for substitutes of 15 to 35 minutes. This is because the match is less intense with a larger score difference. With a smaller score difference, the percentages are less extreme but follow the same trend.

The influence of red cards on players' physical data was examined using the 51 red cards from the regular season of the Jupiler Pro League. The analysis focused on the difference in distance covered by players in the 15 minutes before and after a red card, as shown in figure 7. About 64% of players covered more distance after their team received a red card. However, this percentage varied depending on the timing of the card, as seen in table II. Red cards in the first quarter of the second half resulted in more than 80% of players covering more distance, while cards in the last quarter of the first half resulted in only 44% of players covering more distance.

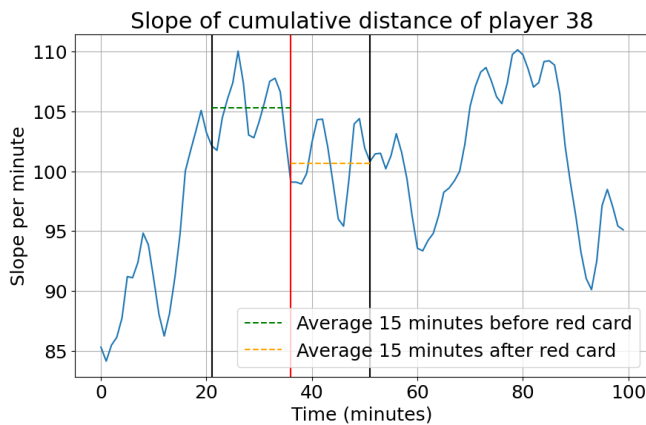


Fig. 7. Explanation of the procedure for the impact of red cards

TABLE II

NUMBER OF PLAYERS WHO COVERED MORE AND LESS DISTANCE AFTER A RED CARD FOR DIFFERENT TIME INTERVALS

Time interval	Number above	Percentage above	Number below	Percentage below
0-15	16	72,73%	6	27,27%
16-30	34	66,67%	17	33,33%
31-45	34	44,16%	43	55,84%
46-60	49	81,67%	11	18,33%
61-80	47	68,11%	22	31,88%
81+	72	67,92%	34	32,08%

Additionally, the impact of the score at the time of the red

card was examined. The results are shown in table III. If a team was ahead, 75% of players covered less distance after the red card, presumably because the opposing team increased pressure. In a tie, 64% of players covered less distance. If the team that received the red card was losing, there was an equal chance that players would cover more or less distance (50% each).

TABLE III

NUMBER OF PLAYERS WHO RUN MORE OR LESS AFTER A RED CARD FOR DIFFERENT SCORELINES

Score	Number more	Percentage more	Number less	Percentage less
Winning	29	24,37	90	75,63
Drawing	52	35,62	94	64,38
Losing	51	49,51	52	50,49

C. Other influences on physical data

In this chapter, the influence of two aspects on the physical performances of football players has been investigated: expected goals and ball possession. Firstly, the correlation between the number of expected goals (xG) and players' physical data was examined. The data was analysed per position, as the impact of xG varies across different positions. The results of this analysis can be found in table IV.

TABLE IV

AGGREGATE OF PHYSICAL DATA FOR EXPECTED GOALS INTERVALS OF THE FIRST AND SECOND HALF

ID	xG < 0.50	0.50 < xG < 1	1 < xG
Keeper	110,31	108,52	104,17
Central defender	206,11	213,04	213,79
Fullback	215,32	224,39	210,36
Midfielder	229,07	236,61	243,31
Wide attacker	219,58	217,04	214,57
Wide attacker	203,37	215,26	236,41
Forward	222,52	221,29	235,26
Forward	215,81	203,95	206,27

Overall, it was observed that at lower xG values, goal-keepers covered more distance on average. This is because they are more involved in the game when there are fewer expected goals. Central defenders covered more distance when the team's xG was higher, indicating they were more involved in attacks in matches with more goal-scoring opportunities. For fullbacks, there was little difference in physical data between different xG intervals. Midfielders generally exhibited higher physical performances at higher xG values, while the results for wide attackers and forwards varied depending on the player type.

Additionally, the influence of ball possession on players' physical performances was examined. No significant difference was found between different positions. Generally, players exhibited higher physical performances when the team's ball possession was lower. This is logical, as players need to cover more ground when possession is lower. The results of the influence of ball possession can be found in Table V.

TABLE V
AGGREGATE OF PHYSICAL DATA FOR BALL POSSESSION INTERVALS OF
THE FIRST AND SECOND HALF

ID	possession < 45	45 < possession < 55	55 < possession
22	110.16	107.74	105.08
44	219.60	207.40	207.24
14	223.80	219.41	211.19
15	240.32	230.58	232.62
32	216.97	216.62	156.24
99	217.01	206.72	206.02

VI. THE FUNCTIONING OF THE ONLINE PROGRAM

A. Processing of the online data

This chapter describes the online program developed as part of a master's thesis to track player fatigue during football matches. Every minute, a tracking file containing data is uploaded to the FTP server of StatsPerform, containing tracking data from the previous minute. The online program retrieves this tracking data and processes it to generate graphs displaying the players' physical parameters. Processing online tracking data posed several challenges, such as issues with player identification. Most problems were resolved by tracking the players' positions. However, the main issue is that the data on the FTP server is not always complete. Only 11 out of the 30 matches from Club Brugge's regular season have complete data. The other 19 matches are either missing or have incomplete data.

B. The functioning of the dashboard

The final product of the master's thesis is a dashboard that can be used during matches. This dashboard is built using the Python library Dash [7]. It consists of a left and right window. The left window contains 2 dropdown lists where players from either the own team or the opponent can be selected, as shown on figure 8. For the 2 selected players, graphs showing the slope of the cumulative distance and the number of frames with a velocity higher than 2 times the average velocity are displayed. In the right window, there is an option to display a list where players are ranked according to fatigue (figure 9) or a list of statistics that can influence the course of the graphs with players' physical data (figure 10). Some of these statistics include the score [8], expected goals, and ball possession (whose influence was discussed in the previous chapter). All data displayed on the dashboard is automatically updated every minute.

CONCLUSION

Previous research has shown that players' physical parameters decrease when fatigue occurs. Monitoring these parameters during a match can serve as a method to track player fatigue. For this master's thesis, both the distance covered and the velocity of the players were monitored using tracking data from StatsPerform. This data is used to calculate the physical parameters of players based on their positional coordinates during the match.

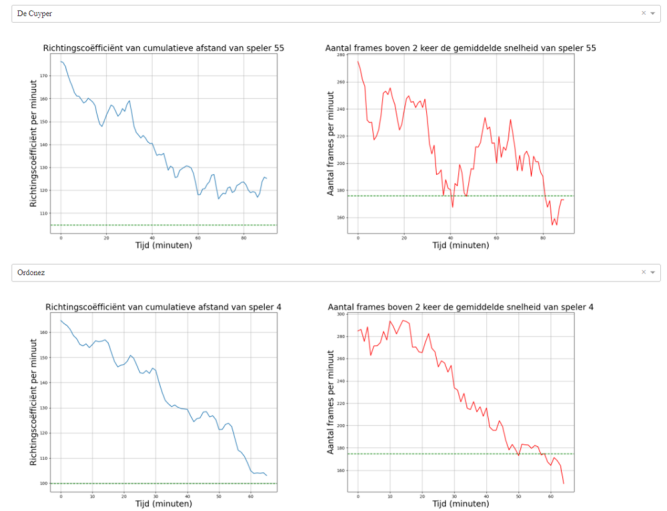


Fig. 8. Picture of graphs with physical data in dashboard



Fig. 9. Picture of fatigue list in dashboard

Score: 1-1
Scoring Attempts: 7-13
On-Target Scoring Attempts: 4-3
Possession: 24.9 - 75.1
Expected goals: 0.58 - 1.08

Ratings:

Player	Rating
Jutglà	83.92537
De Cuyper	80.25544
Meijer	77.16919
Mechele	64.36298
Spileers	60.19833
Skóras	59.23875
Mignolet	57.38798
Vanaken	34.84602
Thiago	24.12721

Fig. 10. Picture of match stats in dashboard

However, it is important to note that a lower distance covered does not automatically mean that a player is fatigued, as other factors can also influence this. The course of the match, such as the score, ball possession, and expected goals, can all affect players' physical parameters. Factors such as red cards can also influence players' running performance, but this does not necessarily mean they are more fatigued as a result.

For this master's thesis, tracking data from StatsPerform was used. There were some issues with this data. The main problem was that complete tracking data was available for only 11 of the 30 regular season matches of Club Brugge. This means that for 19 of the 30 matches, fatigue cannot be determined or cannot be optimally determined due to missing data. Since the availability and quality of the data cannot be guaranteed to improve by StatsPerform, it cannot be ensured that this program will consistently perform well for every match. An important task for StatsPerform is to provide qualitative data for each match.

REFERENCES

- [1] R. Julian, R. M. Page, and L. D. Harper, "The Effect of Fixture Congestion on Performance During Professional Male Soccer Match-Play: A Systematic Critical Review with Meta-Analysis," 2020.
- [2] F. Dambroz, F. M. Clemente, and I. Teoldo, "The effect of physical fatigue on the performance of soccer players: A systematic review," 2022.
- [3] J. F. Tornero-Aguilera, J. Jimenez-Morcillo, A. Rubio-Zarapuz, and V. J. Clemente-Suárez, "Central and Peripheral Fatigue in Physical Exercise Explained: A Narrative Review," 2022.
- [4] Z. Akyildiz, Y. Ocak, F. M. Clemente, Y. Birgonul, M. Günay, and H. Nobari, "Monitoring the post-match neuro-muscular fatigue of young Turkish football players," 2022.
- [5] S. Llana, B. Burriel, P. Madrero, and J. Fernández, "Is it worth the effort? Understanding and contextualizing physical metrics in soccer," 2022.
- [6] D. Raabe, H. Biermann, M. Bassek, M. Wohlan, R. Komitova, R. Rein, T. K. Groot, and D. Memmert, "floodlight - a high-level, data-driven sports analytics framework," *Journal of Open Source Software*, vol. 7, no. 76, p. 4588, 2022. [Online]. Available: <https://doi.org/10.21105/joss.04588>
- [7] Dash user guide & documentation," accessed: May 24, 2024
- [8] T. Modric, S. versic, I. Jukic, and D. Sekulic, "Physical performance discriminating winning and losing in uefa champions league: a full-season study," 2024.

Inhoudsopgave

Abstract	iv
Lijst van figuren	xiv
Lijst van tabellen	xvi
1 Inleiding	1
2 Achtergrond	3
2.1 Inleiding	3
2.2 Vermoeidheid	4
2.2.1 Definitie	4
2.2.2 Vermoeidheid in voetbal	6
2.3 Meten van vermoeidheid	7
2.3.1 Fysieke parameters	7
2.3.2 Invloed van fysieke parameters	7
2.3.3 Afhankelijkheid per positie en speler	8
2.4 Optical Tracking data vs. gps-data	11
2.4.1 Tracking data	11
2.4.2 gps-data	12
2.4.3 Conclusie	12
2.5 Verwerken van tracking data	13
2.5.1 Floodlight	13
2.5.2 Kloppey	13
3 Dataverkenning	15
3.1 Trackingdata	15
3.2 StatsPerform API data	16
3.2.1 Algemene informatie wedstrijd (MA1)	16
3.2.2 Wedstrijdstatistieken MA2	17

3.2.3	Overige API data	17
4	Dataverwerking	18
4.1	Distance model	18
4.2	Velocity model	21
5	Werking van het offline programma	26
5.1	Werking offline programma	26
5.2	Kwaliteit niet-live programma	28
5.2.1	Werkwijze	28
5.2.2	Conclusie	29
5.3	Speciale gevallen en uitzonderingen	33
5.3.1	Korte invalbeurten	33
5.3.2	Invloed van rode kaarten	37
5.4	Andere invloeden op fysieke data	40
5.4.1	Invloed van expected goals	42
5.4.2	Invloed van balbezit	46
6	Werking van het online programma	48
6.1	Verwerking van de online trackingdata	48
6.1.1	Problemen met online trackingdata	49
6.2	Werking van het dashboard	53
	Conclusie	57
	Ethische en maatschappelijke reflectie	57
	Referenties	59

Lijst van figuren

2.1	Aantal blessures tijdens een bepaalde periode in een voetbalwedstrijd.	5
2.2	Gemiddelde afstand in verdediging afgelegd tijdens een voetbalwedstrijd.	8
2.3	Gemiddelde afstand afgelegd aan hoge intensiteit tijdens een voetbalwedstrijd	8
2.4	Op welke manier spelers op bepaalde positie gemiddeld hun afstand tijdens wedstrijden afleggen.	9
2.5	Percentage van flankspelers' hoge intensiteit inspanningen in aanval met of zonder bal.	10
2.6	Kaart van de gemiddelde hoge intensiteit inspanningen van Serge Gnabry tijdens 2 verschillende voetbalwedstrijden in 2 verschillende rollen.	10
4.1	Cumulatieve afstand afgelegd tijdens eerste helft	19
4.2	Cumulatieve afstand afgelegd tijdens volledige wedstrijd	19
4.3	Richtingscoëfficiënt van cumulatieve afstand tijdens volledige wedstrijd	20
4.4	Afgeflakte grafiek van richtingscoëfficiënt van cumulatieve afstand	21
4.5	Snelheid van speler tijdens wedstrijd	22
4.6	Aantal keer 50 opeenvolgende frames boven bepaalde 4 m/s per minuut	23
4.7	Aantal keer 20 opeenvolgende frames boven bepaalde 5 m/s per minuut	23
4.8	Aantal frames boven bepaalde drempelwaarde per minuut	24
4.9	Afgeflakte grafiek van aantal frames boven bepaalde drempelwaarde per minuut	25
4.10	Afgeflakte grafiek van richtingscoëfficiënt van cumulatieve afstand van Skov Olsen	25
5.1	Grafiek met richtingscoëfficiënt van cumulatieve afstand van Casper Nielsen	27
5.2	Grafiek met aantal frames met snelheid boven 2 keer gemiddelde snelheid van Casper Nielsen	27
5.3	Vergelijking van richtingscoëfficiënt van cumulatieve afstand gegenereerd door offline en online programma's	32
5.4	Verloop van richtingscoëfficiënt van cumulatieve afstand van speler met invalbeurt tussen 0 en 15 minuten	35
5.5	Verloop van richtingscoëfficiënt van cumulatieve afstand van speler met invalbeurt tussen 0 en 15 minuten	35
5.6	Verloop van richtingscoëfficiënt van cumulatieve afstand van speler met invalbeurt tussen 15 en 35 minuten	36

5.7	Verloop van richtingscoëfficiënt van cumulatieve afstand van speler met invalbeurt tussen 35 en 45 minuten	37
5.8	Uitleg werkwijze invloed van rode kaarten	39
5.9	Grafiek met richtingscoëfficiënt van cumulatieve afstand na rode kaart bij voorsprong	40
5.10	Grafiek met richtingscoëfficiënt van cumulatieve afstand na rode kaart bij gelijkstand	41
5.11	Grafiek met richtingscoëfficiënt van cumulatieve afstand na rode kaart bij achterstand	41
6.1	Vergelijking van richtingscoëfficiënt van cumulatieve afstand gegenereerd door offline en online programma's	52
6.2	Dashboard: linkervenster met grafieken van fysieke data	54
6.3	Dashboard: rechtervenster met vermoeidheidslijst	55
6.4	Dashboard: rechtervenster met wedstrijdstatistieken	55

Lijst van tabellen

3.1	Overzicht van soorten data, inhoud en formaten	16
5.1	Cumulatieve afstand in meter voor verschillende soorten data	29
5.2	Totale cumulatieve afstanden in meter van de 9 complete wedstrijden voor verschillende soorten data	29
5.3	Totale cumulatieve afstanden in meter en percentage van verschil met StatsPerform data	30
5.4	Percentueel gemiddelde en standaardafwijking van verschil tussen fysieke StatsPerform data en respectievelijk offline data en online data	31
5.5	Aantal ingevallen spelers die een grafiek met verloop van richtingscoëfficiënt hebben die eindigt boven of onder gemiddelde waarde per tijdsinterval	33
5.6	Aantal ingevallen spelers die een grafiek met verloop van richtingscoëfficiënt hebben die eindigt boven of onder gemiddelde waarde per tijdsinterval wanneer verschil in score groter of gelijk aan 2 was	34
5.7	Aantal ingevallen spelers die een grafiek met verloop van richtingscoëfficiënt hebben die eindigt boven of onder gemiddelde waarde per tijdsinterval wanneer verschil in score kleiner dan 2 was	34
5.8	Tijdverdeling van rode kaarten	38
5.9	Aantal spelers die meer en minder afstand aflegden na rode kaart	38
5.10	Aantal spelers dat meer en minder afstand aflegden na rode kaart voor verschillende tijdsintervallen	39
5.11	Aantal spelers die meer of minder lopen na rode kaart voor verschillende tussenstanden	40
5.12	Fysieke data voor expected goals intervallen in de eerste helft	43
5.13	Fysieke data voor expected goals intervallen in de tweede helft	44
5.14	Som van fysieke data voor expected goals intervallen van eerste en tweede helft	45
5.15	Fysieke data voor balbezit intervallen in de eerste helft	46
5.16	Fysieke data voor balbezit intervallen in de tweede helft	47
5.17	Som van fysieke data voor balbezit intervallen van eerste en tweede helft	47
6.1	Kwaliteitscontrole van de online trackingdata per wedstrijd	50

1

Inleiding

De laatste jaren is het gebruik van data in de voetbalwereld aan een enorme opmars bezig. Een opvallend voorbeeld van deze trend is de toenemende inzet van data-analyse in het scoutingproces binnen de voetbalwereld. Toen er vroeger een speler gescout werd, moest een scout een aantal wedstrijden van die speler bekijken om zo een conclusie te trekken over het niveau van de speler en in welke positie en rol hij het best zou gebruikt worden. Tegenwoordig wordt er enorm veel data verzameld uit verschillende wedstrijden in verschillende landen, zodat de scouts binnen enkele seconden, op basis van deze data, een beeld kunnen vormen van deze speler.

Data analyse valt niet meer weg te denken binnen het dagelijkse opereren van de meeste professionele voetbalploegen. Het heeft een groot aantal toepassingen. Zoals eerder verteld wordt data gebruikt voor scouting. Niet enkel op basis van data worden potentieel interessante spelers om het team te versterken gescout, maar ook andere teams worden gescout. Zo kan er voor de wedstrijd een beter idee gevormd worden over de zwakke en sterke punten van tegenstanders.

Naast scouting wordt data ook gebruikt om fysieke parameters en voetbal specifieke statistieken van spelers tijdens wedstrijden en trainingen te bepalen. Het is bijna onmogelijk om met het blote oog te bepalen hoeveel een speler tijdens een training gelopen heeft of hoeveel passes een speler tijdens een wedstrijd heeft gegeven. Het is dus belangrijk om dergelijke zaken te automatiseren. Aan de hand van deze fysieke parameters kan vermoeidheid van spelers bepaald worden. Aangezien er steeds meer wedstrijden per seizoen gespeeld worden en de belasting dus steeds groter wordt, is het heel belangrijk om deze fysieke parameters te monitoren tijdens wedstrijden, zodat er tijdens een wedstrijd tijdig ingegrepen kan worden wanneer er vermoeidheid bij spelers optreedt.

De doelstelling van deze thesis is het ontwerpen van een applicatie die tijdens wedstrijden de fysieke parameters van spelers analyseert. Ook wordt er bekeken welke conclusies i.v.m. vermoeidheid, er getrokken kunnen worden uit deze fysieke parameters. Om de fysieke parameters te berekenen wordt er gebruik gemaakt van de tracking data van StatsPerform. De beelden van de wedstrijd worden omgezet in coördinaten die de positie van de spelers weergeven op iedere moment tijdens de wedstrijd. Op basis daarvan kan berekend worden hoe ver de spelers zich verplaatsen, in welke richting en aan welke snelheid.

1 Inleiding

Deze thesis begint met een beschrijving van de data die beschikbaar is. Er zijn verschillende soorten data van StatsPerform die gebruikt kunnen worden, maar deze zijn niet allemaal even nuttig voor deze thesis. Er wordt kort uitgelegd wat de data inhoudt en waarvoor de data bruikbaar is voor deze thesis. Daarna zal uitgelegd worden hoe deze data verwerkt wordt. De tracking data bestaat uit grote dataframes met heel veel positionele coördinaten en moet er dus op een snelle, efficiënte manier verwerkt worden. Verder worden de werking en resultaten van de niet-live applicatie beschreven. Deze werd ontwikkeld om zo betere inzichten te krijgen in het bepalen van vermoeidheid van de spelers. Deze inzichten worden hier dan ook besproken. Als laatste zal de live-applicatie besproken worden. Voordat de live-applicatie ontwikkeld werd gebruik makend van de data op de FTP server, werd er eerst een simulatie van de FTP server gemaakt om zo makkelijker te kunnen testen. Ook wordt hier het dashboard, dat gebruikt kan worden tijdens wedstrijden, besproken.

2

Achtergrond

In dit hoofdstuk wordt de literatuur besproken die de basis van de masterproef vormt. In een inleidend deel wordt kort de situatie van de masterproef geschetst, waarna er wat dieper wordt ingegaan op het begrip vermoeidheid. Eerst zal vermoeidheid gedefinieerd worden. Daarna wordt vermoeidheid met betrekking tot voetbal besproken. Het derde deel gaat over het meten van vermoeidheid tijdens voetbalwedstrijden. Hierop volgt een deel over de persoonsafhankelijkheid en positieafhankelijkheid van vermoeidheid. De vraag of vermoeidheid verschillend moet gedefinieerd en gemeten worden van persoon tot persoon en positie tot positie wordt hier beantwoord. Er wordt afgesloten met de verschillen tussen gps-data en visuele data.

2.1 Inleiding

De snelle professionalisering en commercialisatie van het voetbal brengt met zich mee dat seizoenen steeds langer duren, uit meer matches bestaan en deze matches, bijgevolg, steeds dichter en dichter op elkaar volgen. Per seizoen spelen professionele ploegen gemiddeld tussen de 50 en 80 wedstrijden. Dit komt overeen met 40 weken per jaar waar er vaak cyclussen zijn van 2 wedstrijden per week [1]. Dit zorgt er voor dat er tussen wedstrijden niet veel recuperatietijd is. Bovendien wordt voetbal gekarakteriseerd door lange speeltijd, een hoog tempo en intense fysieke concentraties [2]. Tijdens officiële wedstrijden, maar ook tijdens trainingen, worden voetballers blootgesteld aan grote fysieke verwachtingen, zoals lopen, springen, sprinten, versnellen, vertragen en snel veranderen van richting. Tijdens een wedstrijd van 90 minuten, lopen professionele spelers gemiddeld 10 km met een gemiddelde intensiteit die dicht bij de anaerobe drempel ligt (80-90% van de maximale hartslag) [3]. Het is dus belangrijk om vermoeidheid te tracken tijdens wedstrijden.

De combinatie van deze factoren leidt tot een hoog aantal wedstrijd blessures bij professionele voetballers. Voorgaande studies tonen aan dat een voetbalploeg per speler ongeveer 2 blessures per seizoen kan verwachten. Bij een ploeg met 25 spelers zijn er gemiddeld 3 spelers per wedstrijd niet beschikbaar door een blessure. Dit betekent dat gemiddeld 12% van de spelers niet beschikbaar zijn voor trainingen en wedstrijden omwille van blessures. Hierbij zijn spierblessures het meest voorkomende type. Blessures hebben niet enkel invloed op de persoonlijke prestaties van de spelers, maar ook een invloed op de totale positie van een ploeg in de uiteindelijke

2 Achtergrond

rangschikking. Om het potentieel van een ploeg zo goed mogelijk te benutten, wilt men dus zo weinig mogelijk blessures [4].

Ook zijn er enkele risicofactoren die een hogere kans op blessures voor professionele voetballers met zich meebrengen. Een eerste studie heeft aangetoond dat oudere, krachtigere atleten met een lagere range of motion een groter risico lopen op blessures. Beperkte flexibiliteit is dus een eerste risicofactor [5]. Een tweede risicofactor is veel sprinten tijdens een wedstrijd. Een toename in sprintactiviteit tijdens wedstrijden brengt een hogere kans op blessures met zich mee [6]. Een toename van wedstrijdbelasting leidt ook tot een verhoogd risico op blessures. Hoe meer wedstrijden spelers spelen op korte termijn, hoe hoger de kans op blessures. Dit staat recht tegenover het aantal trainingen. Spelers met een hoge blootstelling aan fysieke prikkels tijdens training zullen beter beschermd zijn tegen wedstrijdbleesures dan spelers die weinig tot niet getraind hebben. Dit komt omdat de intensiteit tijdens een training en een wedstrijd verschillend is. In een training zal de intensiteit lager liggen [7].

De laatste risicofactor is vermoeidheid. Figuur 2.1 geeft het aantal blessures in een bepaald tijdslot in een wedstrijd. De duidelijk trend in deze grafiek is dat het grootste deel van de blessures het meest voorkomen tegen het einde van een speelhelft, omdat er op dat moment vermoeidheid optreedt bij de spelers. Vermoeidheid is dus de laatste risicofactor die leidt tot een hogere kans op blessures. Deze laatste factor zal ook de basis vormen voor het onderzoek van de masterproef [8].

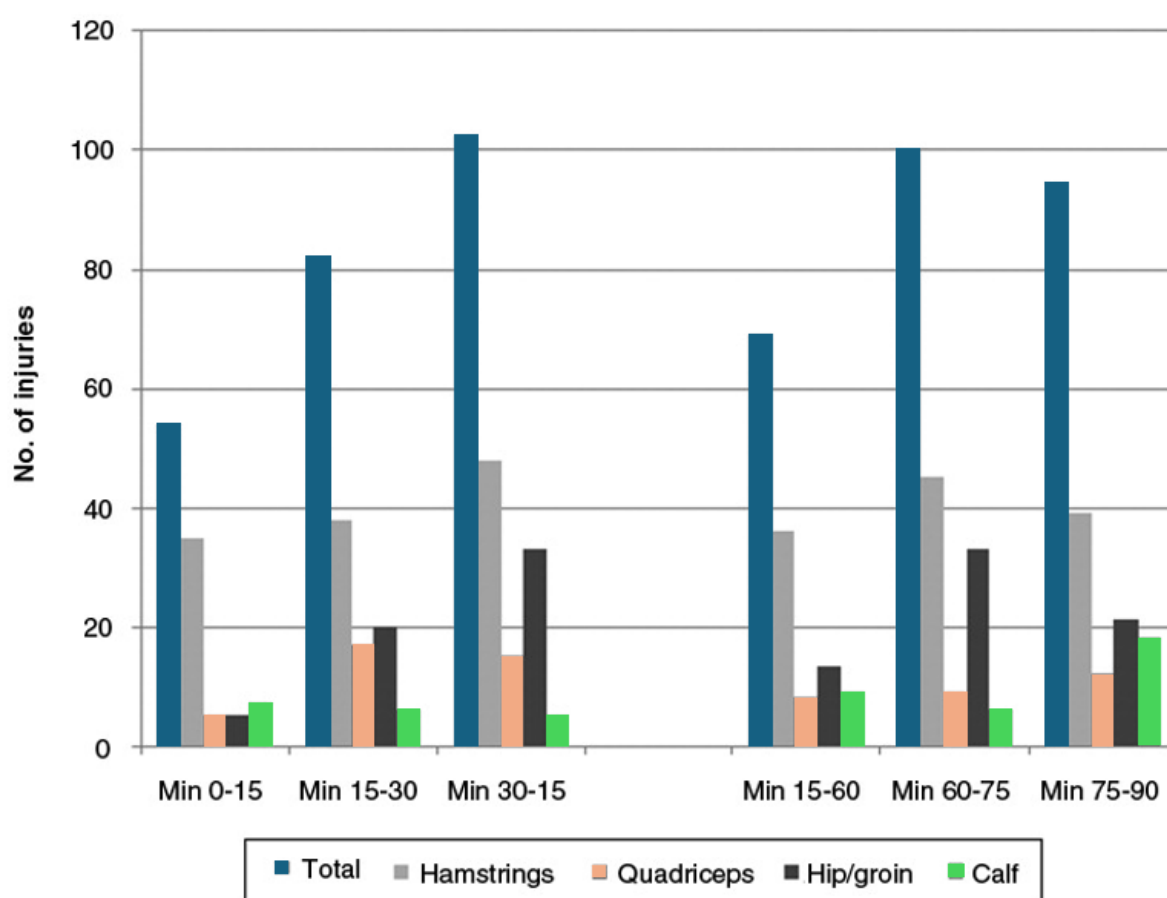
Vermoeidheid heeft niet enkel een invloed op de kans op blessures van voetballers, maar brengt ook lagere fysieke en technische prestaties met zich mee. Een vermoeide speler heeft een hogere kans op fouten tijdens het passen, dribbelen en schieten van de bal. Wanneer een speler vermoeid is, zal dit zich ook weerspiegelen op de fysieke parameters van een speler. Vermoeidheid zorgt ervoor dat spelers gemiddeld veel minder sprints afleggen dan wanneer ze niet vermoeid zijn. Ook zullen vermoeide spelers deze sprints aan een lagere snelheid en intensiteit uitvoeren [9].

2.2 Vermoeidheid

2.2.1 Definitie

Vermoeidheid wordt gedefinieerd als een interne verstoring van de homeostase, veroorzaakt door een toename van de energieproductie die wordt geëist door een externe prikkel. Algemener kan vermoeidheid omschreven worden als een afname in fysieke, mentale en technische prestaties in verband met een toename van werkelijke intensiteit of moeilijkheid van de taak of oefening die uitgevoerd moet te worden. In de sportwereld is de prikkel voor vermoeidheid voornamelijk een fysieke activiteit [10].

Vermoeidheid kan ook gezien worden als de afname in maximale vrijwillige spierkracht als gevolg van lichaamsbeweging. Er zijn twee mogelijke verklaringen die tot deze afname leiden. De eerste is spierversmoeidheid. Spier-



Figuur 2.1: Aantal blessures tijdens een bepaalde periode in een voetbalwedstrijd.

2 Achtergrond

vermoeidheid treedt op de wanneer de spieren moe worden door de fysieke inspanning zelf. Dit komt door het samentrekken van spiercellen. De tweede mogelijke verklaring is de centrale vermoeidheid. Dit is afname van spierkracht door een verminderde neurale aansturing van de motorische cortex (hersenen) naar de motorische eenheden. Op basis van deze twee verklaringen, worden de fysieke en tactische gedragingen van spelers beïnvloed tijdens periodes van spier vermoeidheid [11].

2.2.2 Vermoeidheid in voetbal

Trainingen zorgen voor gemiddeld veel minder blessures dan wedstrijden. Ook is de vermoeidheid en de verwachte hersteltijd na een training meestal veel lager dan na een wedstrijd. Dit komt doordat de intensiteit tijdens wedstrijden een stuk hoger ligt dan tijdens trainingen. Het feit dat er op training vaak oefeningen zijn waar geen fysiek contact is en duels zijn tussen spelers en er tijdens een wedstrijd altijd duels zijn, is ook een van de mogelijke oorzaken hiervoor. Daarom wordt er tijdens deze masterproef gefocust op wedstrijden.

De prestaties van voetballers verslechteren 1 tot 4 dagen na de wedstrijd. Een voetbalwedstrijd veroorzaakt kortdurende spierschade en een tijdelijke ontstekingsreactie. Dit komt door vermoeidheid na wedstrijden. De acute ontstekingsreactie in het voetbal lijkt hetzelfde patroon te volgen als bij andere vormen van lichaamsbeweging. Dit toont de noodzaak van voldoende hersteltijd voor professionele voetballers na een wedstrijd, aan [12]. Zware inspanningen tijdens een voetbalmatch leiden tot vermoeidheid die gekoppeld is aan een combinatie van uitdroging, uitputting van glycogeen en mentale vermoeidheid. De omvang van vermoeidheid die optreedt na een voetbalwedstrijd kan worden beïnvloed door extrinsieke factoren en/of intrinsieke factoren zoals slaapttekort, stress, ..., die de hersteltijd beïnvloeden [13].

Er zijn ook een aantal factoren die invloed hebben op hoe een persoon vermoeidheid ervaart. Geslacht is een van de aspecten. De interpretatie van inspanning en stress verschilt tussen mannen en vrouwen, waarbij de perceptie van de inspanning bij een gelijke inspanning, groter is bij vrouwen. Wanneer een man en een vrouw dezelfde fysieke taak krijgt, zal volgens de perceptie van de vrouw, de opdracht zwaarder zijn. Op dezelfde manier is ook de perceptie van vermoeidheid groter bij vrouwen. Ook persoonlijke problemen, angst en een afname in activatie kunnen ervoor zorgen dat de vermoeidheid hoger ligt dan normaal. De laatste factor die de perceptie van vermoeidheid beïnvloedt, zijn externe factoren zoals stress of stressvolle prikkels en slaapgebrek. Deze leiden tot een verstoringen in de waarnemingen wat ervoor zal zorgen dat de subjectieve perceptie van vermoeidheid hoog is. Deze factoren zorgen er voor dat vermoeidheid heel erg afhankelijk is van de persoon en van de situatie waarin die persoon zich bevindt. Ook is dit een redenen waarom het moeilijk is om vermoeidheid te bepalen [10].

De vermoeidheid die voorkomt bij spelers tijdens trainingen en wedstrijden kan gecategoriseerd worden in twee verschillende soorten belasting. Fysiologische stressoren die voorkomen tijdens trainingen en wedstrijden worden de interne belasting genoemd. Een voorbeeld van fysiologische stressoren is een hoge hartslag van een speler tijdens de wedstrijd. Kinematische activiteiten die fysiologische stressoren creëren worden de externe belasting

2 Achtergrond

genoemd. Enkele voorbeelden van kinematische activiteiten zijn de afgelegde afstand tijdens een wedstrijd of het aantal sprints [11].

2.3 Meten van vermoeidheid

2.3.1 Fysieke parameters

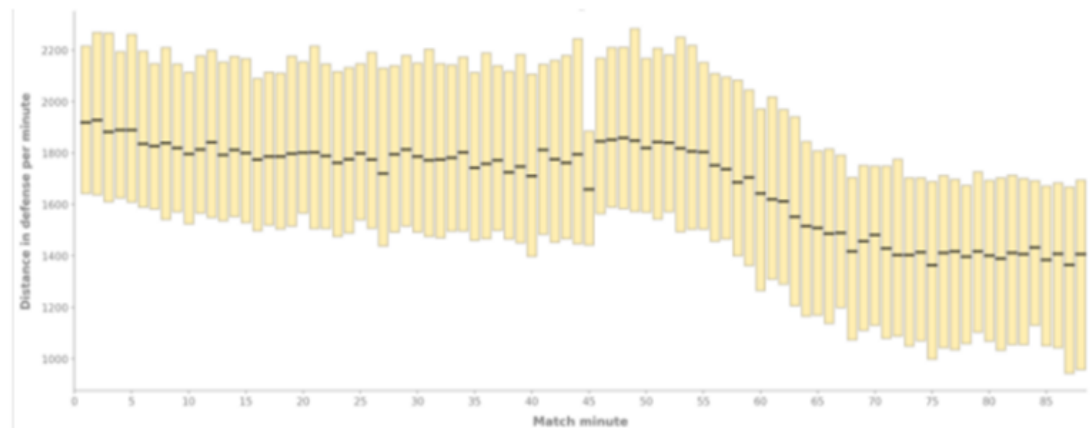
Er worden 6 verschillende fysieke parameters gedefinieerd die betrekking hebben tot de afgelegde afstand van de spelers. Deze verschillende parameters vertellen elk iets over de snelheid waarmee de afstand wordt afgelegd. De eerste is staande intensiteit (≤ 11 km/u) wat overeenkomt met een snelheid tussen stilstaan en joggen. Verder heb je verschillende soorten intensiteiten van lopen: lage intensiteit lopen (>11 en ≤ 14 km/u), gemiddelde intensiteit lopen (>14.1 en ≤ 17 km/u), hoge intensiteit lopen (>17.1 en ≤ 21 km/u) en zeer hoge intensiteit lopen (> 21 km/u). Alles wat zich boven een snelheid van 24 km/u bevindt, noemen we sprints. Er kan ook een zevende fysieke parameter gedefinieerd worden, namelijk de versnelling. Dergelijke parameters kunnen gemonitord worden om vermoeidheid te detecteren [14].

2.3.2 Invloed van fysieke parameters

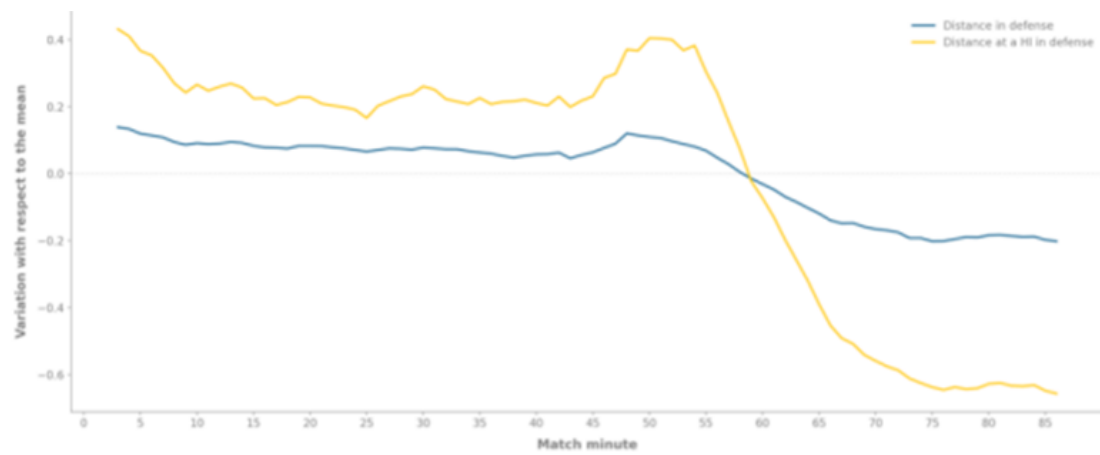
Nadat er in het vorig onderdeel werd uitgelegd wat de fysieke parameters zijn, zal er in dit stukje besproken worden wat de invloed van deze fysieke parameters is op de vermoeidheid van spelers. Figuur 2.2 toont de gemiddelde afstand afgelegd door een team in verdediging per minuut in een wedstrijd. Hieruit valt duidelijk te zien dat teams naar het einde van de wedstrijd veel minder meters afleggen dan in het begin van de wedstrijd. Dit geldt logischerwijs ook voor individuele spelers. Dit kan komen door het feit dat er vermoeidheid begint op te treden of door het scoreverloop van de wedstrijd. In beide gevallen betekent dit dat het prestatieniveau aan het dalen is. De grafiek heeft een grote dip op minuut 45 (de laatste minuut van de eerste speelhelft). Dit komt omdat de waarde per minuut wordt berekend op basis van de 2 voorgaande minuten, de minuut zelf en de twee volgende minuten. Aangezien we hier soms met minder dan 2 minuten blessuretijd zitten, zal dit een grote invloed hebben op deze dip [12].

Enkel de afgelegde afstand zal geen eenduidig beeld geven over de vermoeidheid van een speler. Het kan zijn dat een speler een grote afstand afgelegd tijdens een wedstrijd, maar aan een heel lage intensiteit of een grotere afstand, maar aan een veel hogere intensiteit. Er zijn dus ook nog andere externe loadparameters nodig. Een van de andere parameters die gebruikt kan worden, is de afgelegde afstand aan hoge intensiteit per minuut. Figuur 2.3 toont aan dat net als de totale afstand afgelegd (blauwe lijn), ook de totale afstand afgelegd aan hoge intensiteit (oranje lijn) aangeeft wanneer spelers vermoeider beginnen te worden. De hoge intensiteit neemt zelf nog sterker en directer af dan de totale afgelegde afstand, dus toont al iets sneller tekenen van vermoeidheid aan. Beide curves werden genormaliseerd door het gemiddelde van 5 minuten te nemen (2 minuten voor en 2

2 Achtergrond



Figuur 2.2: Gemiddelde afstand in verdediging afgelegd tijdens een voetbalwedstrijd.



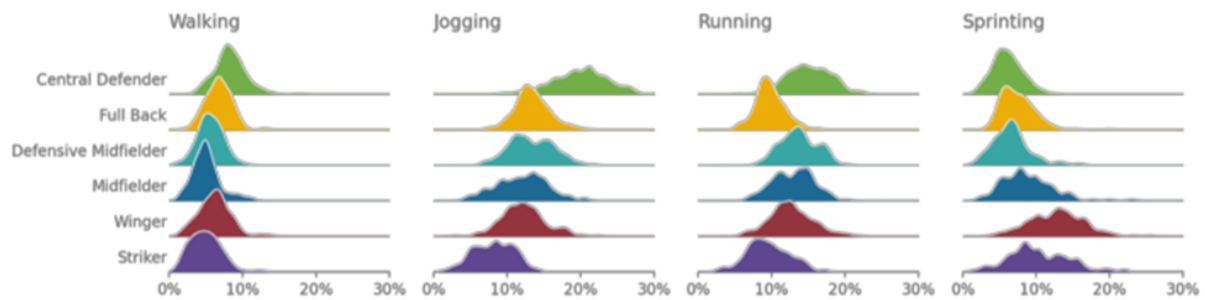
Figuur 2.3: Gemiddelde afstand afgelegd aan hoge intensiteit tijdens een voetbalwedstrijd

minuten na de afgebeelde minuut) om de totale (hoge intensiteit) afstand per minuut nauwkeuriger weer te kunnen geven [12].

2.3.3 Afhankelijkheid per positie en speler

Figuur 2.4 toont duidelijk aan dat verschillende posities op een totaal andere manier en intensiteit hun afstand afleggen tijdens wedstrijden [12]. Het is bewezen dat fysieke parameters zoals: afgelegde afstand, afgelegde afstand aan hoge intensiteit en afgelegde sprintafstand, positieafhankelijk zijn. Zo zijn centrale middenvelders, vleugelmiddenvelders en flankverdedigers de posities die gemiddeld de hoogste afstand aan hoge intensiteit afleggen per wedstrijd. Terwijl spitsen en centrale verdedigers een veel lagere afstand aan hoger intensiteit zullen afleggen. Ook leggen vleugelmiddenvelders en flankverdedigers de grootste gemiddelde afstand al sprintend af per wedstrijd. Het is dus duidelijk dat de positie van de speler bepalend is voor hoeveel een speler loopt of sprint

2 Achtergrond



Figuur 2.4: Op welke manier spelers op bepaalde positie gemiddeld hun afstand tijdens wedstrijden afleggen.

tijdens en wedstrijd [13].

Onderzoeken tonen ook aan dat de fysieke parameters van spelers tijdens verschillende matches niet heel veel variëren. Dit betekent dat spelers die op dezelfde positie blijven spelen, meestal ongeveer dezelfde afstand, hoge intensiteit afstand en sprintafstand zullen hebben tijdens verschillende wedstrijden, tenzij er sprake is van grote tactische ingrepen of positionele verandering van de speler. Om vermoeidheid te bepalen kan er dus teruggekoppeld moeten worden naar de resultaten van de fysieke parameters uit vorige wedstrijden [15].

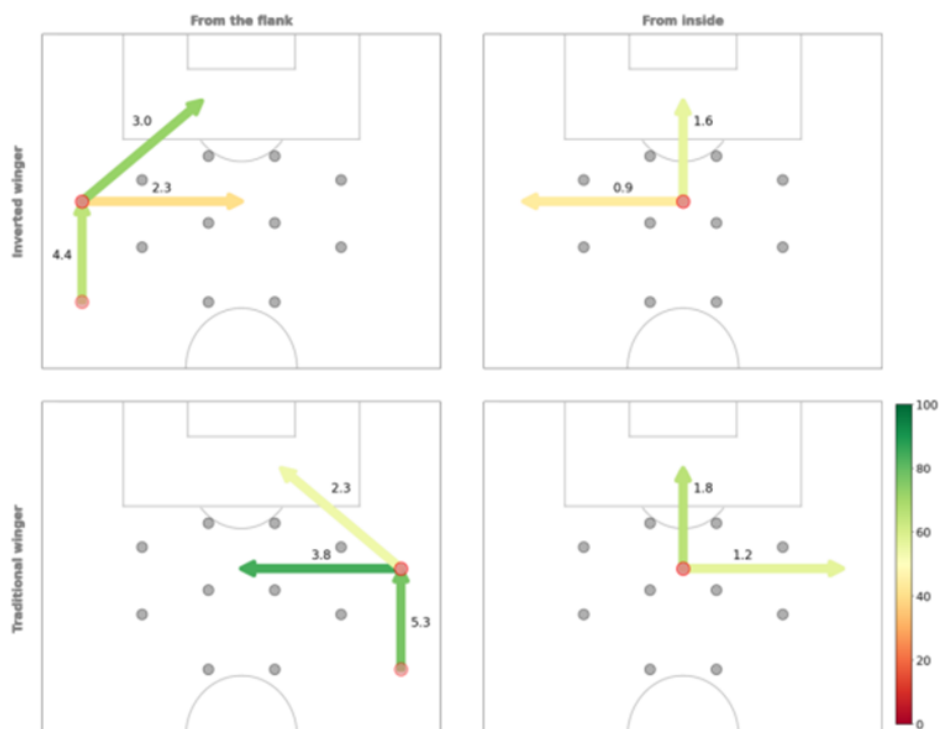
Ook is de gemiddelde afstand afgelegd aan hoge intensiteit per wedstrijd afhankelijk van de speler zelf. Figuur 2.5 toont het aantal loopacties aan hoge intensiteit en hoeveel van deze loopacties met de bal aan de voet worden afgelegd. Zoals te zien op figuur sparen spelers zoals Messi, Grealish en Mbappe hun energie tot ze een grote kans hebben om de bal te krijgen. Terwijl spelers zoals Adama Traoré en Allan Saint-Maximin veel vaker met de bal lopen aan hoge intensiteit per wedstrijd. Ook kunnen we uit deze figuur concluderen dat flankaanvallers gemiddeld 60 tot 80% van hun totale afstand aan hoge intensiteit, afleggen zonder de bal zelf te hebben [12].

Nadat al eerder aangetoond werd dat fysieke parameters positieafhankelijk en spelersafhankelijk zijn, toont figuur 2.6 aan dat deze parameters ook afhankelijk kunnen zijn van de rol die de speler invult binnen het team. De figuur toont 5 soorten loopacties voor flankspelers en hoe vaak dit soort loopactie aan hoge intensiteit gemiddeld plaats vond per wedstrijd (hoe donkerder de pijl, hoe meer deze plaatsvond). Van de 5 soorten loopacties zijn er 3 die ingezet worden vanop de flank en 2 vanuit het centrum. In figuur 2.6 worden de inspanningen aan hoge intensiteit van Serge Gnabry (spelend voor Bayern München en Duitsland) in kaart gebracht. Hieruit valt duidelijk te zien dat de speler tijdens 2 wedstrijden in 2 verschillende rollen speelde: als traditionele flankspeler en geïnverteerde flankspeler. Als traditionele flankspeler, legt Gnabry een hogere afstand af aan hoge intensiteit dan als geïnverteerde flankspeler. Ook de types van loopacties zijn erg verschillend. Als traditionele flankspeler liep hij veel vaker aan hoge intensiteit naar binnen, terwijl hij als geïnverteerd flankspeler veel vaker de diepte opzocht [12]. Voor deze thesis is er gekozen om de fysieke parameters als spelersafhankelijk te bepalen en niet als positieafhankelijk. Dit omdat verschillende spelers die op dezelfde positie staan een andere rol aannemen binnen het team.

2 Achtergrond



Figuur 2.5: Percentage van flankspelers' hoge intensiteit inspanningen in aanval met of zonder bal.



Figuur 2.6: Kaart van de gemiddelde hoge intensiteit inspanningen van Serge Gnabry tijdens 2 verschillende voetbalwedstrijden in 2 verschillende rollen.

2.4 Optical Tracking data vs. gps-data

De meest voorkomende manieren van data tracking in het voetbal zijn gps-data en visuele data. Als er in deze thesis over tracking data wordt gesproken, kan er van uitgegaan worden dat dit om optische tracking data gaat. In het volgende hoofdstuk worden deze twee beschreven, vergeleken met elkaar en worden van beide de voor- en nadelen opgelijst.

2.4.1 Tracking data

Optical tracking is het verzamelen van tracking data door gebruik te maken van beelden die opgenomen worden met camera's. Hier worden meerdere camera's gebruikt om de nauwkeurigheid van de data te verhogen. De beelden verkregen door deze camera's worden gebruikt om aan een hoge snelheid, positionele data van elke speler op het veld en de positie van de bal te genereren. Door de opeenvolgende posities van de spelers te vergelijken, kunnen parameters zoals snelheid, versnelling, afgelegde afstand, ... berekend worden. Aangezien het niet eenvoudig is om deze beelden om te zetten in positionele data, wordt de verwerking van de data vaak uitgevoerd door externe, gespecialiseerde bedrijven, zoals StatsPerform [16].

Voordelen

Een groot voordeel van optische tracking is dat de spelers er zelf niks van merken. Dit wordt volledig gedaan a.d.h.v. camera's en spelers zelf hoeven dus geen gps of dergelijke bij te houden tijdens de wedstrijd [17]. Ook heeft tracking data een hoge positionele nauwkeurigheid. De posities van de spelers zullen dus op enkele centimeters na nauwkeurig zijn [16].

Nadelen

Het is heel moeilijk om de grote hoeveelheid tracking data om te zetten naar een duidelijk overzicht van data die bruikbaar zijn. Bij tracking data heb je een heel groot aantal frames. Uit elke frame wil je uiteindelijk de positie van elke speler op het veld uithalen om zo snelheden en afstanden te berekenen. Dit is niet simpel om te doen zonder grote expertise in dit specifieke vakgebied. Gps-data is veel makkelijker te verwerken dan tracking data [18, 19]. Ook zal tracking data soms een overschatting geven van de afgelegde afstand van de spelers. Waar de positie van de spelers meestal heel nauwkeurig is, zal de data van de afgelegde afstand triviale afwijkingen tonen. Deze afwijkingen zijn echter vrij klein, waardoor ze geen tot weinig invloed hebben op de conclusie of het gebruik van de gegevens [20].

2.4.2 gps-data

Gps of Global Positioning Tracking maakt gebruik van satellieten die de signalen verzenden naar een ontvanger die gedragen wordt door de speler. Een schatting van de positie van de speler kan worden berekend door het verschil tussen meerdere satelliet signalen. Tegenwoordig wordt gps-data bij heel veel professionele voetbalploegen gebruikt [21].

Voordelen

Een van de grote voordelen van gps-data is het mogelijk gebruik ervan tijdens training. Voor gps-data zijn er geen camera's nodig, maar moeten de spelers enkel een GPS bij zich hebben. Voor tracking data zijn er verschillende camera's nodig. Ook is het simpeler om gps-data te verwerken dan tracking data [19, 21]. Gps-data hebben ook als voordeel dat ze heel nauwkeurige resultaten terug geven. Deze resultaten zullen ook zo goed als onmiddellijk beschikbaar zijn [22].

Nadelen

Veel teams gebruiken gps-tracking enkel tijdens trainingen. Dit komt onder andere door wat we bij optische tracking ook als voordeel bespraken. Voor optische tracking heb je enkel camera's nodig, maar zorg je niet voor overlast bij de spelers. Als elke speler een gps moet bijhouden, kan het zijn dat deze overlast wel aanwezig is [17]. Ook zullen gps-signalen vaak kunnen nauwkeurigheid verliezen bij hoge snelheden. Dus wanneer spelers sprinten zal de nauwkeurigheid van de gps-data verminderen. Vroeger werd gps-data niet toegelaten door UEFA tijdens wedstrijden. Dit is in de recentere jaren pas goedgekeurd [19, 21]. Gps-data heeft net als tracking data last van een systematische overschatting. De totale afgelegde afstand zal meestal overschat worden, maar de afwijking op de conclusie of resultaten van een onderzoek is minimaal [22].

2.4.3 Conclusie

Zowel gps-data als tracking-data hebben hun sterke punten en beperkingen. De keuze tussen beide hangt af van specifieke behoeften, voorkeuren en de context van gebruik. Ook kunnen ze naast elkaar gebruikt worden. Daar bestaan tegenwoordig verschillende methodes voor, wat de data analisten een grote flexibiliteit biedt. In beide systemen zal er een overschatting gemaakt worden voor bepaalde data. Deze overschattingen zijn echter verwaarloosbaar en zullen dus slechts een kleine invloed hebben op de conclusie en het totale resultaat. Aangezien er sprake is van een systematische overschatting, kan deze foutmarge ook makkelijk in rekening gebracht

worden [22].

2.5 Verwerken van tracking data

Voor deze masterproef wordt er gebruik gemaakt van de tracking data van StatsPerform. Er bestaan verschillende manieren om deze positionele data om te zetten naar fysieke parameters zoals afgelegde afstand, snelheid en versnelling. Voor deze masterproef wordt er in Python gewerkt. In dit hoofdstuk worden er twee Python packages besproken die vaak gebruikt worden voor het verwerken van tracking data in het voetbal: Floodlight and Kloppey.

2.5.1 Floodlight

Het Python-pakket Floodlight is een op data gericht sportanalyseframework dat is ontworpen om teamsportgegevensanalyse te ondersteunen en te automatiseren. Floodlight is gebouwd om trackingsgegevens, evenementgegevens en andere metagegegevens van het spel te verwerken. Ook biedt het een flexibele aanpak voor prestatieanalyse. Een van de grote voordelen aan het gebruik van een framework zoals Floodlight, is dat het parsers voorziet voor bepaalde providers. Zo is er een manier om een tracking data file van StatsPerform makkelijk om te zetten naar een NumPy-array die de xy-coördinaten voor alle spelers aan een bepaalde frequentie voorziet. Ook zijn er een groot aantal functies voorzien om uit de data van de xy-coördinaten fysieke parameters te berekenen. Enkele voorbeelden van fysieke parameters zijn: afgelegde afstand tussen twee frames, cumulatieve afgelegde afstand tussen twee frames, snelheid en versnelling van een speler tussen twee frames, De gegevensstructuren van Floodlight zijn onderliggend gebaseerd op NumPy en Pandas en ze bieden een basis voor wetenschappelijke berekeningen en geoptimaliseerde gegevensmanipulatie [23].

2.5.2 Kloppey

Kloppey is een Python-pakket dat de uitdagingen aanpakt die worden veroorzaakt door de verscheidenheid aan gegevensformaten tussen verschillende data providers. Het streeft ernaar fundamentele bouwsteen te zijn voor het verwerken van voetbal tracking- en eventgegevens. Net als Floodlight zijn er verschillende parsers voorzien om de data van een bepaalde provider om te zetten naar een gestandaardiseerde Kloppey dataframe. Ook biedt Kloppey gestandaardiseerde gegevensmodellen, filters en transformers aan, waardoor het werken met tracking-data van verschillende leveranciers eenvoudig wordt. Kloppey bestaat uit 3 grote delen. Het laden van data, het processen van data en pattern matching. Het laden van data kan zoals eerder vermeld door middel van het parsen van data van een bepaalde provider, maar indien men niet beschikt over deze data, kan er ook gebruik gemaakt worden van een publieke dataset. Het processen van data zorgt dat je makkelijk je dataset kan transformeren naar om het even welk formaat. Zo kan Kloppey de dimensies van het veld flexibel van het ene formaat naar het andere aanpassen. Het laatste deel van Kloppey is het herkennen van patronen. Kloppey kan op basis van de tracking data, een pas, schot richting doel, voorzet, ... herkennen. Dit zorgt ervoor dat de tracking data makkelijk kan

2 Achtergrond

omgezet worden naar event data [24].

3

Dataverkenning

In dit hoofdstuk worden de verschillende soorten data die gebruikt worden voor de thesis, besproken. Alle data die besproken wordt, is afkomstig van wedstrijden van Club Brugge en wordt aangeboden door StatsPerform. Er zijn twee soorten data die zowel tijdens een wedstrijd als na een wedstrijd beschikbaar zijn. De eerste soort data is de trackingdata. Dit is een text bestand waar om de 0,04 seconden (dus aan een frequentie van 25 Hz) de positionele coördinaten van alle spelers tijdens een wedstrijd weergegeven worden. Deze data is tijdens een wedstrijd beschikbaar via de StatsPerform FTP server en na de wedstrijd via de StatsPerform API. De tweede soort data, is de StatsPerform API data. Ook deze kan tijdens of na een wedstrijd opgevraagd worden. Zowel tijdens als na de wedstrijd wordt deze data opgehaald via de StatsPerform API. De StatsPerform API bestaat uit een aantal endpoints, waar elke soort data verkrijgbaar is via een andere endpoint. De verschillende soorten API data, hebben verschillende namen en worden gekenmerkt door een ID, genaamd MA plus een getal. In totaal zijn er een vijftigtal verschillende soorten API data.

3.1 Trackingdata

De trackingdata vormt de basis voor deze thesis. Dit .txt bestand bestaat uit een groot aantal rijen met data. Elke rij bevat informatie over de positie op het veld van de spelers, de bal en de scheidsrechters. Deze rijen worden tijdens een wedstrijd gegenereerd aan een frequentie van 25 Hz. Dit komt overeen met 1 rij per 0,04 seconden of 1500 rijen per minuut. Er zijn twee verschillende versies van de trackingdata. De eerste versie is de data die achteraf wordt aangeboden via de StatsPerform API. Deze data is postprocessed. Hij is dus volledig en bevat weinig tot geen fouten. Een rij bestaat hier uit 6 verschillende kolommen: de timestamp, de frameCount, de matchPeriod (eerste of tweede helft), de match status (gepauzeerd of niet), de spelers en scheidsrechters (2 dimensionele array met alle spelers en scheidsrechters op het veld met hun positie en de positie van de bal op een gegeven moment), extra geneste data over de spelers en scheidsrechter en een kolom met de positie van de bal.

De tweede soort trackingdata is de data die live tijdens een wedstrijd aanwezig is. Tijdens een wedstrijd is er elke minuut een bestand beschikbaar op een FTP server van StatsPerform. Aangezien dit live is, wordt de data niet postprocessed en bevat deze data dus meer fouten. Ook gebeurt het vaak dat er voor een aantal minuten geen

3 Dataverkenning

nieuwe bestanden op de FTP server beschikbaar zijn en er dus veel ontbrekende data is. Het formaat van deze data komt grotendeels overeen met die van trackingdata die achteraf aangeboden wordt, maar aangezien deze data niet postprocessed is, bevat deze o.a. geen match status.

3.2 StatsPerform API data

Zoals vermeld in de inleiding, kunnen er verschillende soorten data verkregen worden via de StatsPerform API. Deze soorten data zijn gelabeld met een naam en een MA identifier. Wanneer er een call gedaan worden naar de StatsPerform API, krijg je een antwoord in JSON of XML formaat. Dit is afhankelijk van de soort data (sommige endpoints geven enkel XML data terug, andere enkel JSON data) en van de voorkeur die opgegeven wordt tijdens het maken van de call. In tabel 3.1 zie je een overzicht van de soorten API data die er voor deze thesis werd gebruikt, waarvoor de data gebruikt kan worden en in welk formaat ze opgehaald kunnen worden.

Soort data	Inhoud data	Formaat
Wedstrijd (MA1)	- Data over wedstrijd (ploegen, tijdstip, ...) - Startopstelling beide ploegen (enkel XML)	JSON (beperkter) XML (uitgebreider)
Wedstrijdstatistieken (MA2)	- Algemene statistieken van de wedstrijd (doelpunten, kaarten, kansen, ...)	JSON & XML
Balbezit (MA5)	- Balbezit van beide teams per 5, 15, 30 of 45 minuten	JSON & XML
Expected goals (MA12)	- Alle pogingen met het aantal verwachte doelpunt in een wedstrijd	XML
Speler ratings (MA20)	- Ratings van spelers gebaseerd op data van wedstrijd	XML
Winst waarschijnlijkheid (MA21)	- Winst waarschijnlijkheid op basis van huidige en voorgaande wedstrijden (wordt dynamisch geüpdatet tijdens wedstrijd)	JSON & XML
Wedstrijd fitness (MA24)	- Fysieke parameters van spelers tijdens wedstrijd (afgelegde afstand, aantal sprints, ...)	JSON & XML

Tabel 3.1: Overzicht van soorten data, inhoud en formaten

3.2.1 Algemene informatie wedstrijd (MA1)

Deze endpoint geeft data terug die een overzicht van de algemene informatie van de wedstrijd geeft. Deze data bevat informatie over de competitie waarin de wedstrijd gespeeld wordt, de ploegen die deelnemen aan de wedstrijd, de locatie en het tijdstip waar de wedstrijd doorgaat en informatie over de aangestelde scheidsrechter

3 Dataverkenning

en zijn/haar assistenten. Indien er tijdens het maken van de call gekozen wordt voor XML als voorkeursformaat, zullen ook de opstelling van beide ploegen en de gemaakte wissels teruggegeven worden. De MA1 data wordt in deze thesis gebruikt voor het automatisch bepalen en weergeven van de ploegen die de wedstrijd spelen en voor het bijhouden van de spelers die tijdens de wedstrijd gewisseld worden.

3.2.2 Wedstrijdstatistieken MA2

Deze endpoint geeft data terug met betrekking tot de basisstatistieken voor zowel de wedstrijd, de teams als de individuele spelers. Ook kan er hier informatie over de gescoorde doelpunten en de gele en rode kaarten teruggevonden worden. Per spelers en per team is er een lijst met statistieken voorzien. Hoeveel keer trapte speler X op doel, hoeveel reddingen deed speler Y en hoeveel geslaagde passes gaf team 2 tijdens de wedstrijd? Deze data is opgebouwd in uit name-value pairs. Een aantal voorbeelden van dergelijke statistieken zijn: totalPass (aantal passes door een bepaalde speler in een wedstrijd), ballRecovery (aantal balveroveringen), touches (aantal balcontacten), Per speler worden er een 50-tal statistieken weergegeven. De MA2 data wordt tijdens de thesis vooral gebruikt om in het live-programma enkele statistieken weer te geven die kunnen aantonen welk team het overwicht heeft in de wedstrijd.

3.2.3 Overige API data

Verder zijn er nog vier verschillende API endpoints die gebruikt worden voor deze thesis, maar minder informatie bevatten dan de voorgaande, dus ook minder uitgebreid zullen besproken worden. De eerste is de balbezit (MA5) endpoint. Zoals de naam al doet vermoeden geef deze endpoint het balbezit van beide teams terug. Niet enkel het totale balbezit wordt teruggegeven, maar ook het balbezit per 5, 15, 30 of 45 minuten. De tweede endpoint is de expected goals (MA12) endpoint. Deze endpoint geeft een lijst terug met alle kansen tijdens een wedstrijd en het aantal expected goals dat deze kansen met zich meebrachten. Ook de niet-penalty expected goals kunnen hier teruggevonden worden. De derde endpoint is de spelers ratings (MA20) endpoint. Een call naar deze endpoint moet als voorkeursformaat XML hebben. Indien als voorkeursformaat toch JSON wordt meegegeven, zullen er geen ratings beschikbaar zijn. Deze endpoint geeft voor elke speler een getal op 100 terug, dat weergeeft hoe goed een speler aan het spelen is op basis van data. De vierde endpoint is de winst waarschijnlijkheid (MA21) endpoint, die live tijdens de wedstrijd de winst waarschijnlijkheid van beide teams teruggeeft op basis van de voorgaande wedstrijden en events die plaatsvonden in de huidige wedstrijd, zoals kaarten, doelpunten, kansen, Deze drie endpoints worden gebruikt in het live programma om een beter idee te geven over welk team het best in de wedstrijd zit. De laatste endpoint is de wedstrijd fitness (MA24) endpoint. Deze geeft voor elke speler enkele fysieke statistieken terug, zoals de afgelegde afstand, het aantal sprints, Deze data wordt gebruikt om te vergelijken of de data die gegenereerd werd in deze thesis, wel enigszins klopt en overeen komt met de data van StatsPerform (hierover meer in hoofdstuk 5).

4

Dataverwerking

In dit hoofdstuk wordt de dataverwerking besproken. De trackingdata bestaat uit een heel groot aantal rijen (minimaal 135.000). Om deze data makkelijker te verwerken, wordt er gebruikt gemaakt van een externe library: floodlight. Floodlight heeft een aantal ingebouwde functies die ervoor zorgen dat de data makkelijk kan worden ingelezen en deze automatisch kan worden omgezet naar een floodlight XY object. Op dit XY object, kunnen er een aantal functies uitgevoerd worden om de fysieke statistieken van spelers te berekenen uit de trackingdata. De drie functies die het meest interessant leken zijn: distance model (om de afgelopen afstand te berekenen), velocity model (om de snelheid waaraan de afstand werd gelopen, te berekenen) en het metabolic power model (dit geeft aan hoeveel energie een speler nodig heeft om een bepaalde actie uit te voeren).

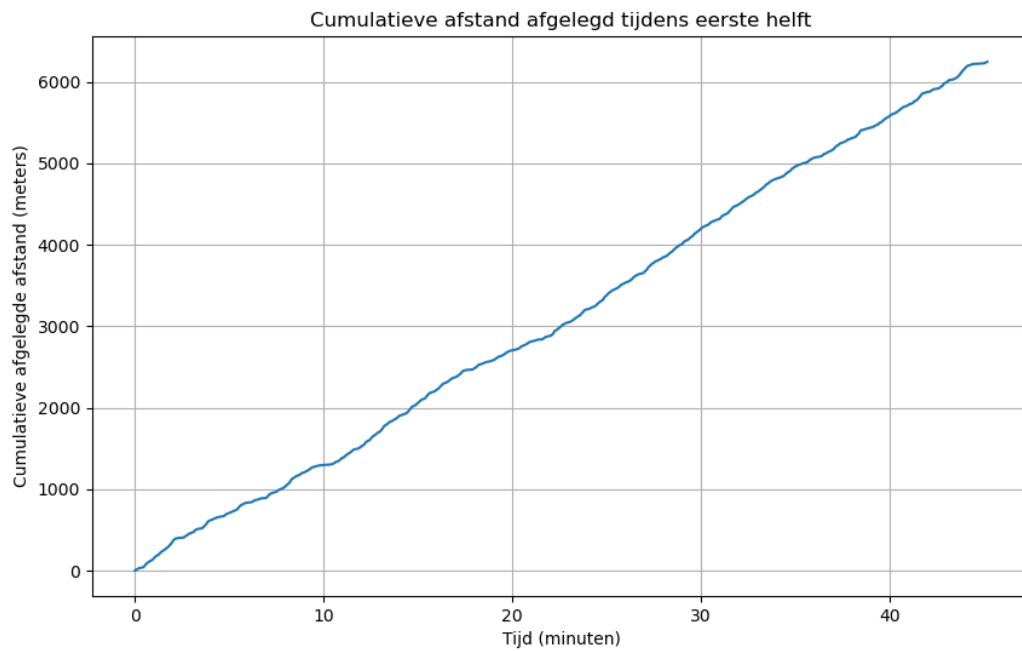
4.1 Distance model

Het distance model berekent de afstand die een speler aflegt tussen 2 frames. Om een duidelijker beeld te krijgen van het verloop van de afgelopen afstand van een speler tijdens een wedstrijd, wordt de data van de afgelopen afstand uitgezet op een grafiek met als y-as de afgelopen afstand in meters en als x-as de tijd in minuten. Figuur 4.1 geeft de afgelegde afstand van Club Brugge speler Casper Nielsen tijdens de eerste helft van de wedstrijd Club Brugge – STVV weer. Figuur 4.2 geeft de afgelegde afstand van de speler weer tijdens de volledige wedstrijd. Zoals duidelijk te zien op deze figuur wordt Nielsen rond minuut 70 gewisseld, aangezien vanaf dat moment de gelopen afstand niet meer omhoog gaat.

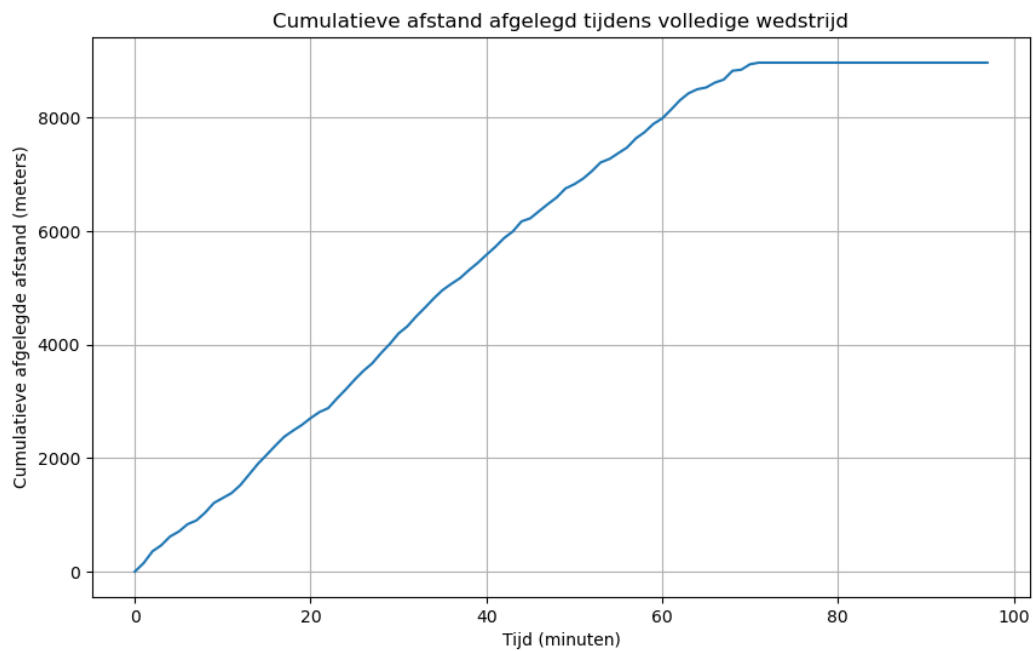
Het is moeilijk om op basis van deze grafiek conclusies te trekken. Rond minuut 60 stijgt de grafiek iets minder snel dan in bijvoorbeeld de eerste helft, maar het is moeilijk om af te leiden hoe groot het verschil is tussen beide waarden. Om het verschil in de cumulatieve afstand wat duidelijker weer te geven werd per minuut de richtingscoëfficiënt van de data berekend. Dit komt overeen met het aantal gelopen meters per minuut. In figuur 4.3 wordt deze afgebeeld en is het al een stuk duidelijker hoe de afgelopen afstand van een speler tijdens de wedstrijd evolueert.

Het volgende probleem is dat de grafiek heel erg gevoelig is voor een kleine wijziging. Indien de wedstrijd een

4 Dataverwerking

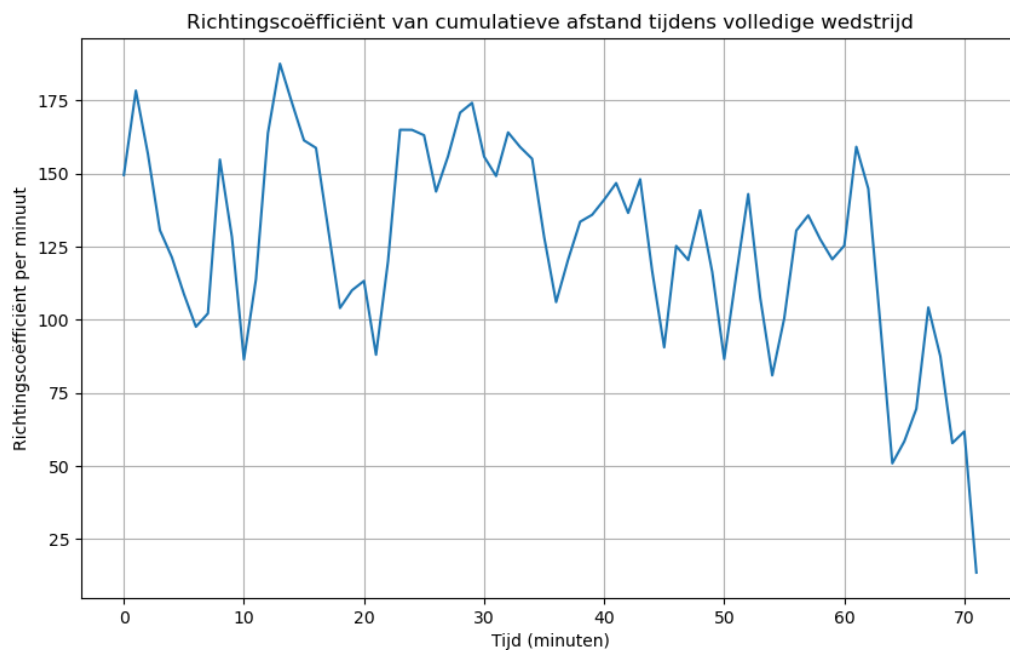


Figuur 4.1: Cumulatieve afstand afgelegd tijdens eerste helft



Figuur 4.2: Cumulatieve afstand afgelegd tijdens volledige wedstrijd

4 Dataverwerking

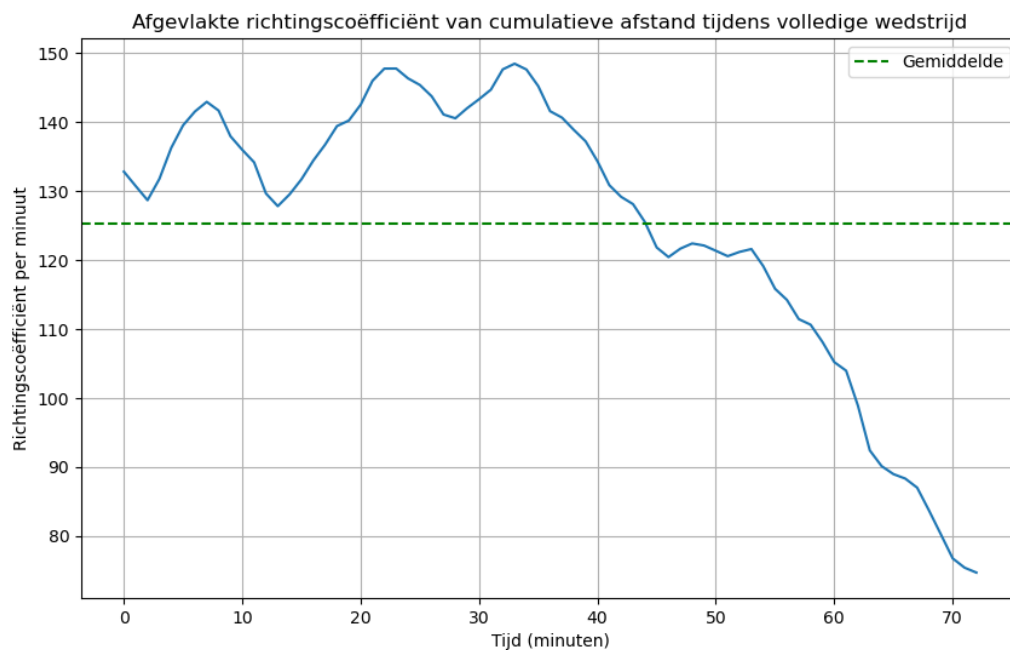


Figuur 4.3: Richtingscoëfficiënt van cumulatieve afstand tijdens volledige wedstrijd

minuut stil ligt, zal dit resulteren in een dal in de waarden van de grafiek van figuur 4.3. Om dit te voorkomen wordt er gebruik gemaakt van een afvlakkingsfunctie. Elke waarde in de grafiek van figuur 4.4 bestaat uit het gewogen gemiddelde van de 5 voorgaande, de 5 volgende waarden en de waarde zelf. Met gewogen gemiddelde wordt er bedoelt dat de waarden die verder van de waarde zelf liggen, minder hard zullen doorwegen dan de waarden die er dichterbij liggen en de waarde zelf het meest zal doorwegen. Dit zorgt ervoor dat de grafiek minder last heeft van ruis in de data. Ook werd er nog een lijn toegevoegd aan de grafiek die de gemiddelde waarde weergeeft. Dit maakt het makkelijker om met het blote oog te zien hoe de grafiek evolueert.

Zoals hierboven vermeld wordt voor het maken van de grafiek het gewogen gemiddelde van de waarde zelf, de 5 voorgaande en de 5 volgende waardes genomen. Dit zorgt ervoor dat een daling in de grafiek eigenlijk al kan voorspelt worden, juist voor die effectief begint plaats te vinden (en er dus proactief gewerkt kan worden). Tijdens een live-wedstrijd is dit echter niet mogelijk, aangezien er geen toekomstige waarden beschikbaar zijn. Daarom wordt het gewogen gemiddelde genomen van de waarde zelf en zijn 5 voorgaande waardes. Dit zorgt ervoor dat de grafiek realtime zal zakken, als de speler minder afstand aflegt.

Casper Nielsen werd rond minuut 70 gewisseld. Op basis van de data berekend uit de afgelegde afstand lijkt dit een juiste keuze. Nielsen liep in de tweede helft een stuk minder dan in de eerste helft. Dit kan aantonen dat Nielsen vermoeid begon te worden in de loop van de tweede helft. Er zijn echter nog andere factoren die de



Figuur 4.4: Afgeflakte grafiek van richtingscoëfficiënt van cumulatieve afstand

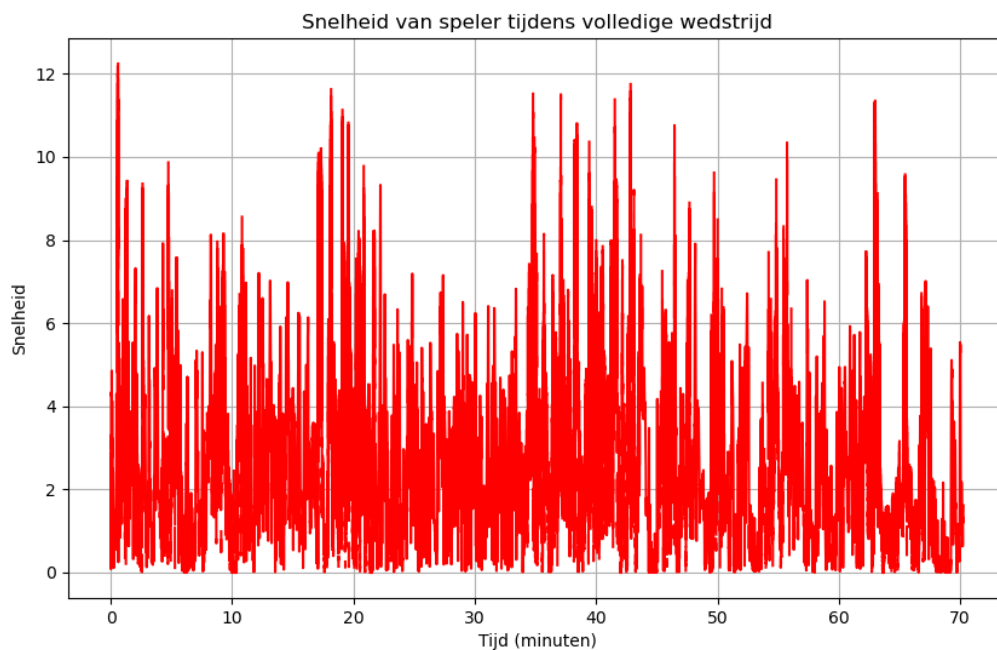
afgelegde afstand kunnen beïnvloeden. Hierover later meer.

4.2 Velocity model

Het velocity model berekent hoe snel een speler beweegt tussen 2 verschillende frames. De eerste manier om een beter inzicht te krijgen in de snelheid waarmee spelers zich voortbewegen tijdens de wedstrijd, is door simpelweg de snelheden weer te geven op een grafiek. Zoals te zien in figuur ?? is dit totaal niet overzichtelijk en is het onmogelijk om hier een duidelijke conclusie uit te trekken. Dit komt omdat er om de 0.04 seconden een waarde op de grafiek afgebeeld wordt en de waarden dus heel dicht bij elkaar liggen. Ook voor de grafieken van het velocity model wordt de data van Casper Nielsen uit de wedstrijd Club Brugge – STVV gebruikt.

De eerste werkwijze die dit probleem mogelijks oplost, is door te kijken hoe veel keer per minuut een speler een bepaald aantal seconden boven een threshold snelheid zit. Voor deze werkwijze werden er verschillende waarden uitgetest voor snelheden en voor aantal seconden. Zoals te zien op figuur 4.6 en figuur 4.7 is er over het algemeen wel een dalende trend zichtbaar, maar deze grafiek is absoluut niet duidelijk genoeg om hier consistente conclusies over de vermoeidheid van de spelers uit te trekken. Figuur 4.6 en figuur 4.7 zijn twee grafieken die bovenstaande werkwijze volgen, maar beiden met verschillende parameters geconfigureerd zijn. Figuur 4.6 kijkt hoeveel keer per minuut er een langere inspanning wordt gedaan, maar aan een minder hoge intensiteit dan in

4 Dataverwerking



Figuur 4.5: Snelheid van speler tijdens wedstrijd

figuur 4.7. Figuur 4.7 geeft weer hoeveel keer per minuut er een kortere inspanning wordt gedaan, maar aan een nog hogere intensiteit dan in figuur 4.6.

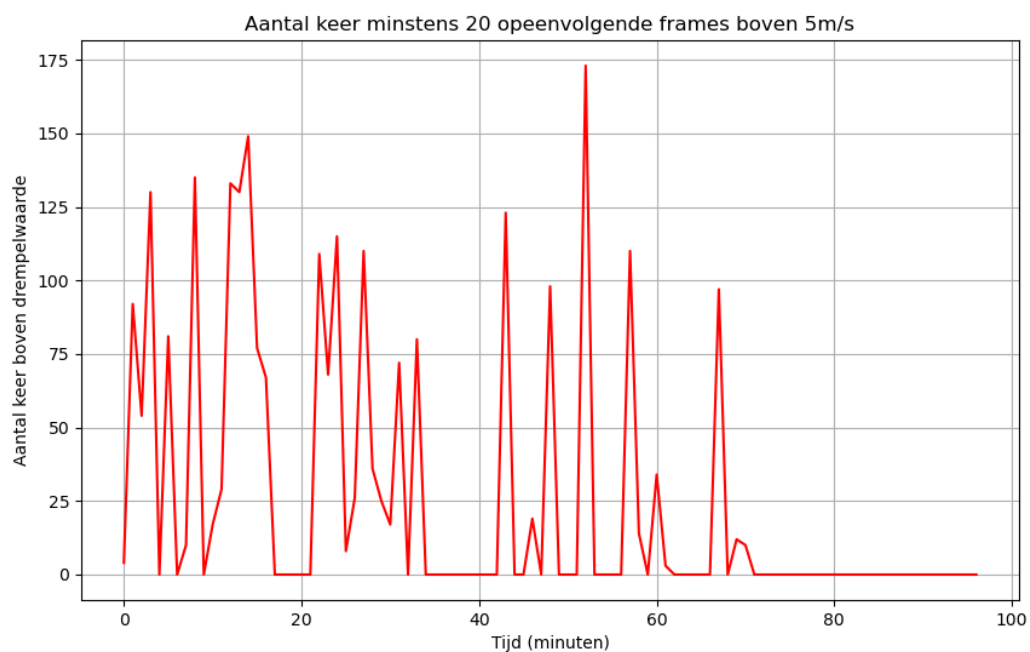
Aangezien er uit figuur 4.6 en figuur 4.7 geen duidelijke conclusies getrokken kunnen worden, werd er nog naar een andere manier gezocht om de snelheid te kunnen gebruiken als een parameter om vermoeidheid te bepalen. De volgende manier berekent hoeveel frames er per minuut boven een bepaalde snelheid zitten. Alle waarden worden eerst in buckets van 1500 frames verdeeld. Dit komt overeen met buckets van een minuut. Daarna wordt er per bucket geteld hoeveel van deze 1500 frames er boven een bepaalde drempelwaarde zitten. Deze waarden worden vervolgens afgebeeld op een grafiek zoals te zie in figuur 4.8.

In figuur 4.8 kunnen we een duidelijke trend zien in de data. Er zijn echter nog 3 problemen met deze werking. Het eerste probleem is dat het verloop van de grafiek heel erg gevoelig is voor kleine veranderingen in de data. Dat was ook het geval bij het distance model. Zoals te zien op figuur 4.8 zijn er heel veel pieken en dalen. Om dit op te lossen werd er gebruikgemaakt van dezelfde oplossing die gebruikt werd bij het distance model. De data wordt afgevlakt door het gewogen gemiddelde van de omliggende waarden en de waarde zelf te nemen. Verder is het ook een probleem om met een universele drempelwaarde te werken. Een centrale verdediger zal tijdens een wedstrijd minder vaak aan hoge snelheden lopen dan een snelle flankaanvaller. Om dit probleem op te lossen, werd gebruik gemaakt van een aparte drempelwaarde per speler. Hier werd gekozen om het dubbele te nemen van de gemiddelde snelheid van de speler tijdens de volledige wedstrijd. Ten slotte werd er bij het eindresultaat

4 Dataverwerking

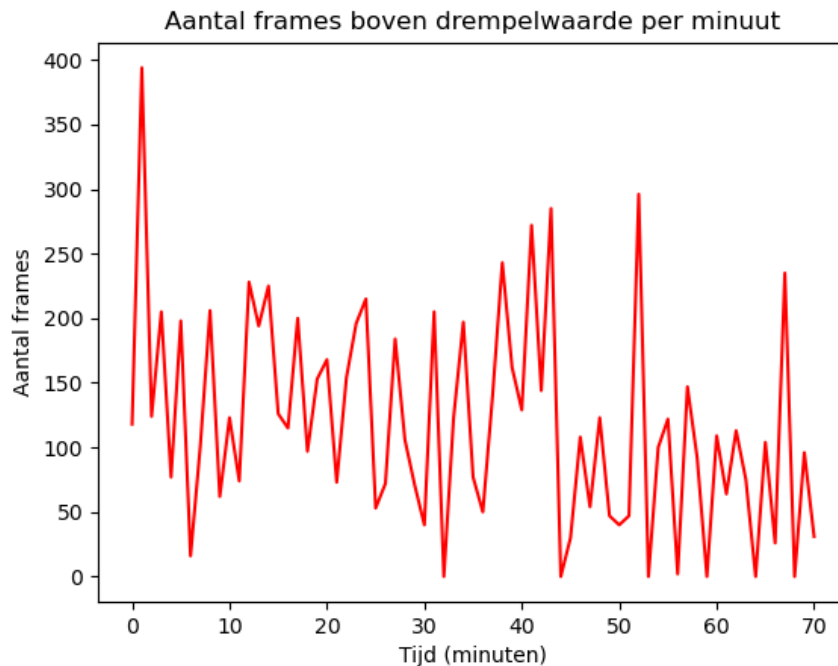


Figuur 4.6: Aantal keer 50 opeenvolgende frames boven bepaalde 4 m/s per minuut



Figuur 4.7: Aantal keer 20 opeenvolgende frames boven bepaalde 5 m/s per minuut

4 Dataverwerking



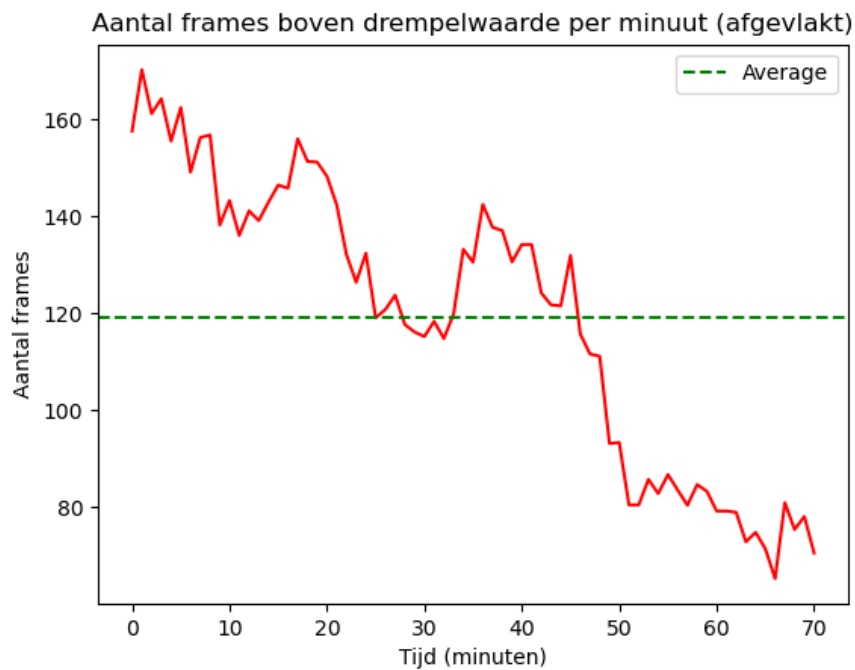
Figuur 4.8: Aantal frames boven bepaalde drempelwaarde per minuut

ook een gemiddelde lijn toegevoegd om zo een duidelijker beeld te kunnen schetsen van het verloop van de waarden.

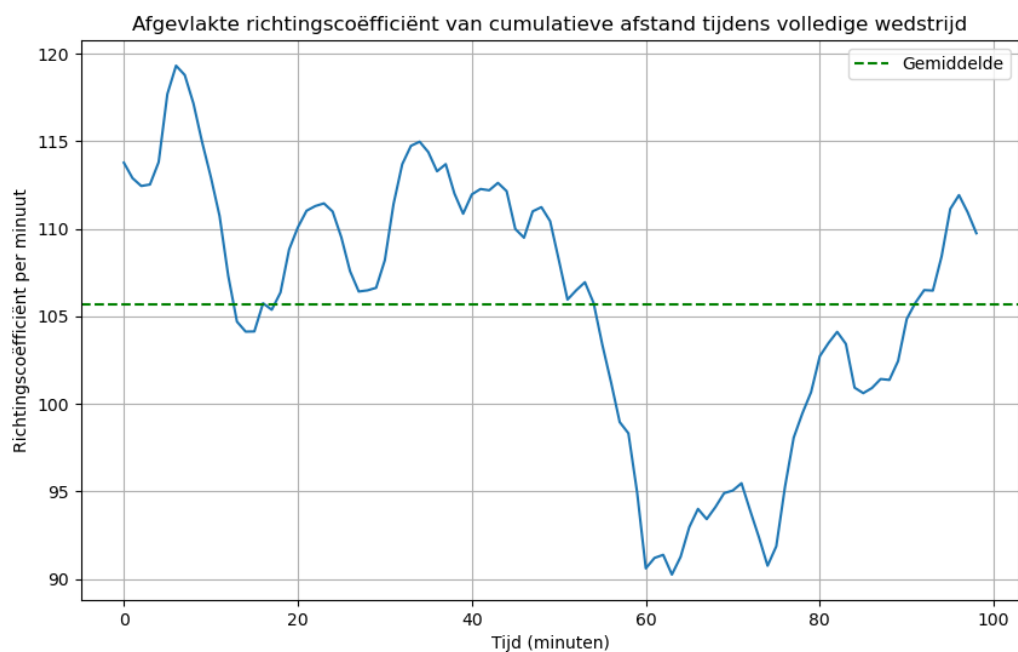
Figuur 4.9 geeft de nieuwe grafiek weer met de drie bovenstaande aanpassingen. Zoals te zien op de grafiek is er opnieuw een dalende trend met een heel grote daling tijdens de eerste minuten van de tweede helft. Dit komt overeen met de grafiek die opgesteld werd voor de afgelegde afstand. Hieruit kan geconcludeerd worden dat ook het aantal acties aan hoge intensiteit een indicator kan zijn voor vermoeidheid en kan deze dus ook gebruikt worden in het vervolg van deze thesis. Net zoals bij de grafieken met betrekking tot de afgelegde afstand, wordt tijdens het live programma het gewogen gemiddelde genomen van de waarde zelf en de 5 voorgaande waarden. Ook hier worden de toekomstige waarden logischerwijs niet meegenomen.

Zoals te zien in figuur 4.4 gaan de fysieke parameters van Nielsen al naar beneden vanaf minuut 40-45. Echter wordt deze speler maar rond minuut 70 gewisseld. Dit komt doordat er nog andere factoren zijn die een invloed kunnen hebben op het verloop van de fysieke parameters van spelers (vb. score en balbezit). Zoals te zien in figuur 4.10 gaat ook bij Skov Olsen de grafiek van de afgelegde afstand omlaag. Rond minuut 60 à 65 begint deze grafiek echter weer te stijgen, terwijl deze bij Nielsen enkel maar afneemt. Het is dus belangrijk om bij het bepalen van vermoeidheid ook die externe factoren die de fysieke parameters beïnvloeden, mee te nemen en de vergelijking te maken met andere spelers.

4 Dataverwerking



Figuur 4.9: Afgevlakte grafiek van aantal frames boven bepaalde drempelwaarde per minuut



Figuur 4.10: Afgeflakte grafiek van richtingscoëfficiënt van cumulatieve afstand van Skov Olsen

5

Werking van het offline programma

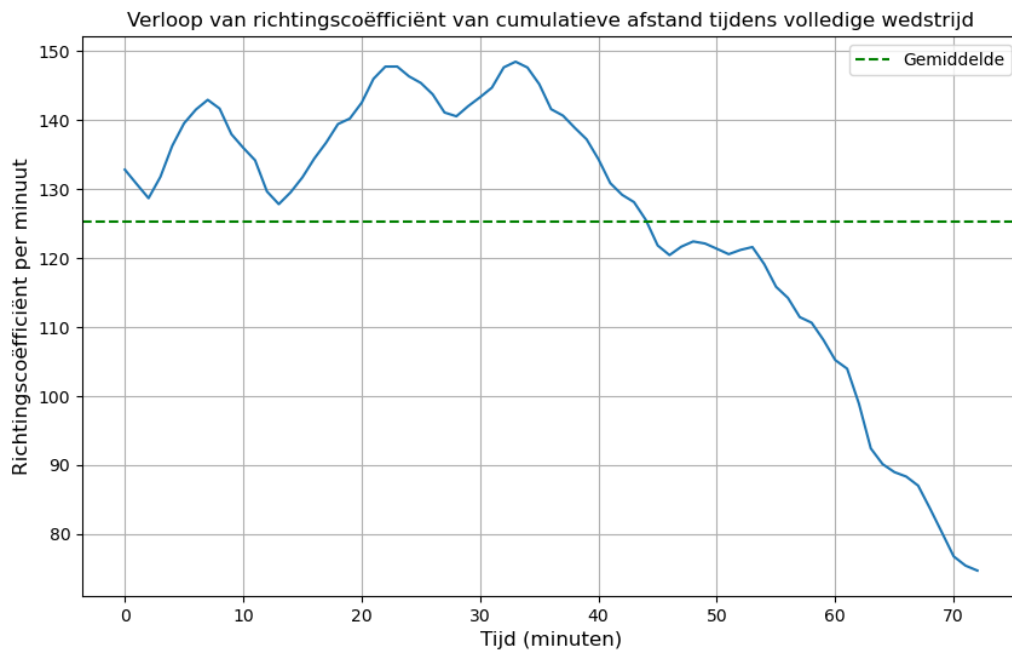
Nadat er in het vorig hoofdstuk een manier is gevonden om de trackingdata om te zetten in overzichtelijke fysieke data, wordt er in dit hoofdstuk dieper ingegaan op de besluiten die uit deze data genomen kunnen worden en de manier waarop deze besluiten tot stand komen. Om deze besluiten te kunnen nemen, werd er een programma ontwikkeld dat op basis van de trackingdata, de verschillende grafieken met fysieke data van de spelers opstelt. Hiervoor wordt de trackingdata gebruikt die na de wedstrijd beschikbaar wordt gesteld via de StatsPerform API. Het verschil tussen de trackingdata die StatsPerform na de wedstrijd en tijdens de wedstrijd beschikbaar stelt, is dat data na de wedstrijd postprocessed is. Deze data heeft dus als voordeel dat het amper tot geen fouten bevat, er geen ontbrekende samples of spelers zijn en dus de kwaliteit over het algemeen een stuk beter zal zijn dan die van de data die live tijdens de wedstrijden beschikbaar is.

5.1 Werking offline programma

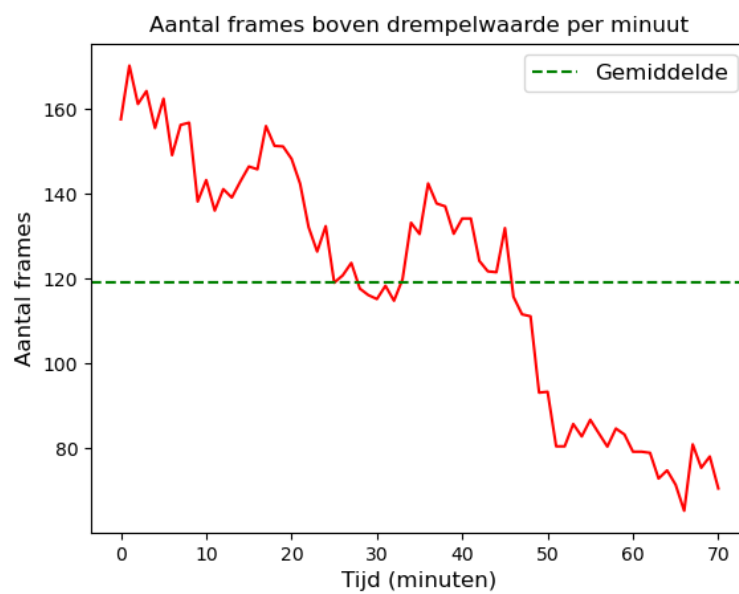
De applicatie voor offline werking is ontworpen om op basis van de ruwe trackingdata, grafieken op te stellen die de fysieke data van de spelers weergeven. Zoals reeds besproken in het vorige hoofdstuk omtrent dataverwerking, worden er voor elke speler twee grafieken opgesteld. De eerste grafiek stelt het verloop van de richtingscoëfficiënt van de cumulatieve data voor. De tweede stelt het aantal frames met een snelheid hoger dan 2 keer de gemiddelde snelheid voor (waarbij de gemiddelde snelheid voor elke speler berekend werd op basis van de voorgaande wedstrijden, waar hij minstens 45 minuten in gespeeld heeft).

De input van de offline applicatie is een tracking file die een bepaalde wedstrijd voorstelt. Deze tracking file bevat de trackingdata van alle spelers die tijdens die wedstrijd gespeeld hebben. Door gebruik te maken van de library Floodlight, worden voor al deze spelers de twee grafieken opgesteld en wordt de bijhorende data berekend. Figuur 5.1 toont een voorbeeld van een grafiek met het verloop van de richtingscoëfficiënt van de cumulatieve afstand en figuur 5.2 toont een voorbeeld van het aantal frames met een snelheid 2 keer hoger dan de gemiddelde snelheid. Beide grafieken werden opgesteld met behulp van de offline applicatie.

5 Werking van het offline programma



Figuur 5.1: Grafiek met richtingscoëfficiënt van cumulatieve afstand van Casper Nielsen



Figuur 5.2: Grafiek met aantal frames met snelheid boven 2 keer gemiddelde snelheid van Casper Nielsen

5 Werking van het offline programma

5.2 Kwaliteit niet-live programma

In dit onderdeel zal er worden gekeken naar de kwaliteit van de door het offline programma gegenereerde data. Enkel de data met betrekking tot de afstand zal hier besproken worden, aangezien het ophalen en verwerken van deze data grotendeels hetzelfde is als voor data met betrekking tot de snelheid (ophalen en verwerken van data wordt op dezelfde manier gedaan; enkel laatste stap verschilt om verschillende types grafieken te kunnen weer-geven). Ook zal de data gegenereerd door het offline programma vergeleken worden met de data aangeboden door StatsPerform en de data van het online programma (die beiden niet postprocessed zijn).

5.2.1 Werkwijze

De data die door het offline programma wordt gegenereerd, is gebaseerd op de ruwe trackingdata die na de wedstrijd door StatsPerform wordt aangeboden. Deze ruwe trackingdata is postprocessed. Om de kwaliteit van de data, gegenereerd door het offline programma, te testen, wordt deze vergeleken met de fysieke data die StatsPerform aanbiedt. Deze fysieke data is tijdens en na de wedstrijd verkrijgbaar via hun API en is gebaseerd op de ruwe trackingdata die tijdens de wedstrijd beschikbaar is. Die trackingdata is dus niet postprocessed. Ook het online programma (zie hoofdstuk 6) maakt gebruik van de niet postprocessed trackingdata. De data die het online programma genereert zal hier ook vergeleken worden met de fysieke data van StatsPerform en de data geproduceerd door het offline programma.

Het eerste probleem dat optreedt bij het vergelijken van deze data, is dat er voor slechts 11 wedstrijden van Club Brugge, in de reguliere competitie, complete online trackingdata beschikbaar is. De online trackingdata wordt tijdens de wedstrijd op een FTP-server van StatsPerform geplaatst en is ook nadien via deze FTP-server beschikbaar. Voor de 19 andere wedstrijden is er incomplete of geen trackingdata beschikbaar op de FTP-server (de gedetailleerde info over de data op de FTP-server wordt in het volgende hoofdstuk besproken). De fysieke data die door StatsPerform aangeboden wordt is wel compleet voor alle 30 wedstrijden. Dit betekent dat de online trackingdata dus wel effectief voor alle wedstrijden bestaat, maar gewoon niet iedere keer correct op de FTP-server wordt geplaatst. De FTP-server is de enige manier om deze online trackingdata op te vragen. Van de 11 wedstrijden die wel complete online trackingdata op de FTP-server hebben, zijn er 2 wedstrijden die geen complete postprocessed trackingdata hebben (die gebruikt wordt voor het offline programma). Om de postprocessed trackingdata op de halen moet er een API call gemaakt worden naar de juiste endpoint van de StatsPerform API. Voor 2 wedstrijden geeft deze call incomplete trackingdata terug.

Voor het vergelijken van de verschillende data, zal, voor de 9 wedstrijden die complete data hebben, de totale cumulatieve afstand berekend worden voor alle Club Brugge spelers die in deze wedstrijd gespeeld hebben. De data van de cumulatieve afstand vormt de basis voor één van de grafieken die door het offline en online programma gecreëerd worden. De cumulatieve afstanden berekend door de verschillende programma's, worden vergeleken in het volgende onderdeel.

5 Werking van het offline programma

5.2.2 Conclusie

Nadat de verschillende data met betrekking tot de cumulatieve afstand, voor de 9 complete wedstrijden, werden opgehaald, zijn ze samengebracht in een tabel. Tabel 5.1 toont een segment van deze tabel. Om de data te kunnen vergelijken, werd voor elke soort data, voor elke speler, de totale cumulatieve afstand van de 9 wedstrijden berekend. In tabel 5.2 wordt een segment weergegeven van deze totale cumulatieve afstanden.

Tabel 5.1: Cumulatieve afstand in meter voor verschillende soorten data

Wedstrijd	Mignolet			De Cuyper		
	StatsPerform	Offline	Online	StatsPerform	Offline	Online
Anderlecht vs Club Brugge	5758.70	5227.99	5523.05	12151.40	11031.59	11796.55
AS Eupen vs Club Brugge	5711.20	5207.63	5657.88	12007.30	10916.36	11974.80
Club Brugge vs Genk	5312.60	4861.13	5300.63	12212.10	11224.67	15774.04

Tabel 5.2: Totale cumulatieve afstanden in meter van de 9 complete wedstrijden voor verschillende soorten data

	Stats	Offline	Online
Mignolet	50637.991	46652.044	50790.875
De Cuyper	91877.531	83845.689	95487.654

Nadat voor elke speler de totale cumulatieve afstand van de 9 complete wedstrijden werd berekend, voor zowel de fysieke data van StatsPerform, de offline data en de online data, werd het verschil berekend tussen deze verschillende waarden. Eerst werd, per speler, het percentueel verschil tussen de fysieke data van StatsPerform en de online data berekend. Daarna werd, per speler, het percentueel verschil tussen de fysieke data van StatsPerform en de offline data berekend. Deze waarden worden getoond in tabel 5.3. De kolom met als naam 'online verschil' geeft het percentueel verschil weer tussen de fysieke data van StatsPerform en de online data. De kolom 'offline verschil' doet hetzelfde voor de fysieke data van StatsPerform en de offline data.

Om nu een duidelijke conclusie te kunnen vormen over de verschillen tussen de data, werd het gemiddelde percentuele verschil en de standaardafwijking berekend. Tabel 5.4 geeft deze weer voor het verschil tussen de fysieke data van StatsPerform en de offline data en voor de fysieke data van StatsPerform en de online data. Wat direct opvalt is dat het gemiddelde percentueel verschil tussen de fysieke data van StatsPerform en de online data heel klein is. Dit is logisch aangezien deze gebruik maken van hetzelfde type ruwe trackingdata: de niet postprocessed trackingdata die tijdens de wedstrijd wordt aangeboden. De standaardafwijking ligt wel hoger dan die van de offline data. Het gemiddeld percentueel verschil tussen de fysieke data van StatsPerform en de offline data ligt een stuk hoger (8.12). Dit komt omdat de offline data gebruikmaakt van de postprocessed trackingdata. Deze

5 Werking van het offline programma

Tabel 5.3: Totale cumulatieve afstanden in meter en percentage van verschil met StatsPerform data

Speler	StatsPerform	Offline	Offline verschil	Online	Online verschil
Balanta	15030.81	13825.72	8.02	14724.24	2.04
Buchanan	44851.52	41239.64	8.05	44083.62	1.71
De Cuyper	91877.53	83845.69	8.74	95487.65	-3.93
Jutgla	32535.35	29729.83	8.62	32034.59	1.54
Mechele	103495.69	94773.85	8.43	108371.88	-4.71
Meijer	18295.72	16879.55	7.74	17171.96	6.14
Mignolet	50637.99	46652.04	7.87	50790.88	-0.30
Nielsen	41701.41	38609.68	7.41	42503.19	-1.92
Nusa	34841.57	31802.60	8.72	35728.03	-2.54
Odoi	32439.00	29798.50	8.14	33283.82	-2.60
Onyedika	84939.08	77616.39	8.62	84875.15	0.08
Rits	12049.44	11311.45	6.12	11898.78	1.25
Sabbe	28290.54	26263.47	7.17	27461.85	2.93
Skoras	18970.54	17719.20	6.60	18393.02	3.04
Skov Olsen	92111.44	83891.61	8.92	100429.37	-9.03
Spileers	94901.73	86988.67	8.34	95058.04	-0.16
Thiago	75204.10	68748.50	8.58	77000.94	-2.39
Vanaken	108979.87	99799.15	8.42	111834.31	-2.62
Vetlesen	68592.65	62683.23	8.62	70637.97	-2.98
Zinckernagel	59099.90	53601.82	9.30	59148.44	-0.08

5 Werking van het offline programma

postprocessed trackingdata heeft als voordeel dat ze nauwkeuriger is en amper tot geen fouten bevat. Hieruit kan dus geconcludeerd worden dat fysieke data die berekend wordt op basis van de niet-postprocessed trackingdata, een overschatting van de effectief afgelegde afstand zal maken. Dit komt doordat deze data minder nauwkeurig is en meer fouten bevat.

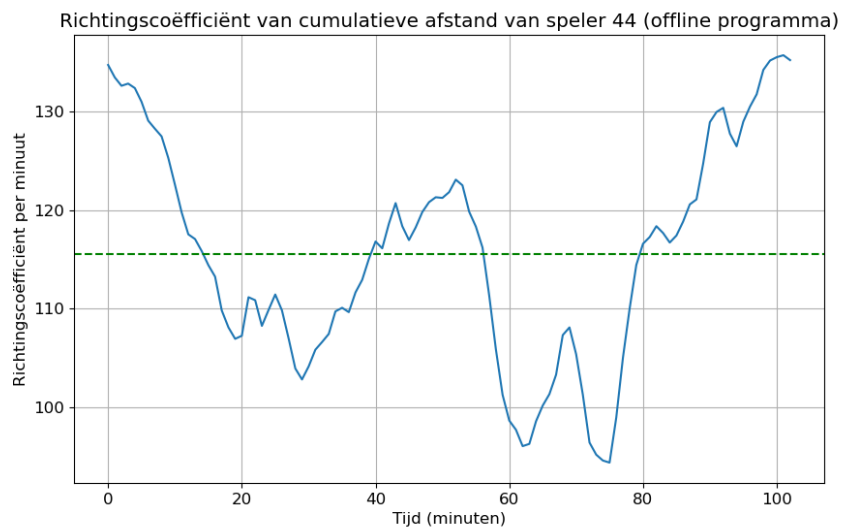
	Gemiddelde	Standaardafwijking
Vershil fysieke data StatsPerform en online data	-0.72	3.27
Vershil fysieke data StatsPerform en offline data	8.12	0.77

Tabel 5.4: Percentueel gemiddelde en standaardafwijking van verschil tussen fysieke StatsPerform data en respectievelijk offline data en online data

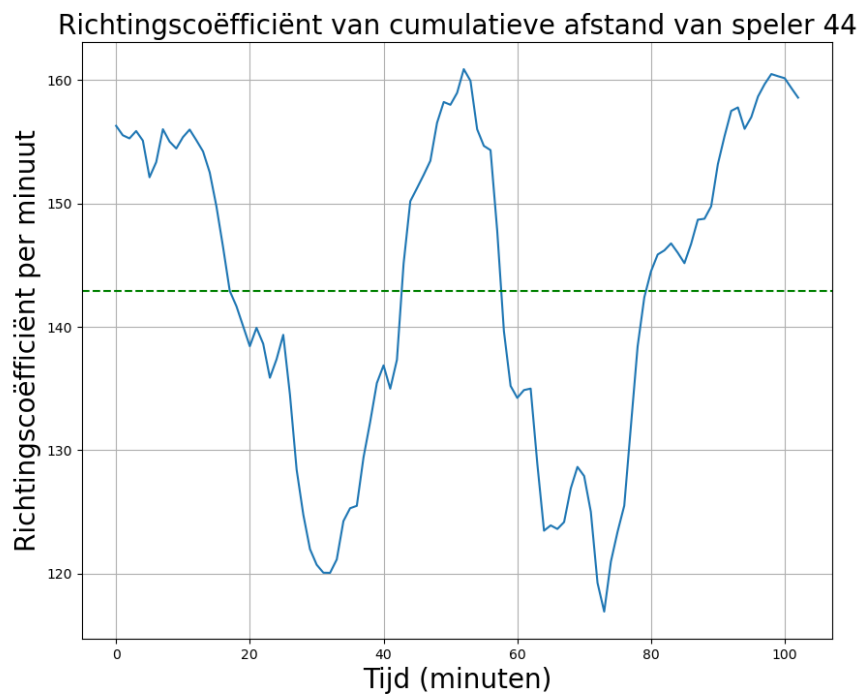
De reden waarom, tijdens het verwerken van de niet postprocessed trackingdata, de cumulatieve afstand overschat wordt, is omwille van onnauwkeurigheden in de data. De exacte onnauwkeurigheden en problemen van deze trackingdata zullen uitgebreid besproken worden in het volgende hoofdstuk, maar hier zal het probleem van de nauwkeurigheid van de x,y-coördinaten van de trackingdata besproken worden. StatsPerform transformeert tijdens een wedstrijd de beelden aan 25 Hz (1500 frames per seconde) in positionele coördinaten waarbij de positie van elke speler op het veld uitgedrukt wordt in x,y-coördinaten. Deze x,y-coördinaten, met extra metadata toegevoegd (zoals onder andere de frequentie waarmee deze gegenereerd zijn), vormen samen de trackingdata. Deze trackingdata wordt dan elke minuut op de StatsPerform FTP-server geplaatst. Aangezien deze data elke minuut stipt op de FTP-server geplaatst moet worden is de snelheid heel belangrijk. Dit zorgt ervoor dat de kwaliteit zal verminderen.

Uiteindelijk wordt deze cumulatieve afstand door het offline en online programma gebruikt om de grafiek met het verloop van de richtingscoëfficiënt van de cumulatieve afstand op te stellen. Figuur 5.3 geeft een voorbeeld van deze grafiek voor het offline programma en voor het online programma. Wat direct opvalt is dat de grafieken ongeveer hetzelfde verloop hebben, maar er toch nog een aantal verschillen inzitten. De pieken en dalen van de online grafiek zullen respectievelijk hogere lokale maxima en lagere lokale minima hebben. Ook valt direct op dat het gemiddelde van de grafiek, gecreëerd door het online programma, een stuk hoger ligt dan bij het offline programma. Dit betekent dat de cumulatieve afstand doorheen de wedstrijd gemiddeld sneller stijgt dan bij de online grafiek. Dit is logisch aangezien, zoals daarnet al aangetoond, de totale cumulatieve afstand berekend door het online programma hoger ligt dan deze berekend door het offline programma. In het hoofdstuk 6 zullen er een aantal technieken besproken worden om het verloop van de grafiek gegenereerd door het online programma, nauwer te doen aansluiten bij die van het offline programma.

5 Werking van het offline programma



(a) Offline programma



(b) Online programma

Figuur 5.3: Vergelijking van richtingscoëfficiënt van cumulatieve afstand gegenereerd door offline en online programma's

5 Werking van het offline programma

5.3 Speciale gevallen en uitzonderingen

In dit onderdeel zullen er een aantal speciale gevallen en uitzonderingen besproken worden. Dit zijn situaties die invloed kunnen hebben op het verloop van de grafieken met fysieke data die door het offline programma gecreëerd worden. Eerst zullen korte invalbeurten besproken worden. Wat gebeurt er als een speler 30 minuten voor het einde mag invallen? Hoe zal de fysieke data van die spelers eruit zien. Daarna zal de invloed van rode kaarten op de fysieke data besproken worden. Gaan spelers nadat een speler uit hun team een rode kaart krijgt meer lopen om deze rode kaart te compenseren?

5.3.1 Korte invalbeurten

Is het monitoren van de fysieke data van spelers om vermoeidheid te tracken wel nuttig bij spelers die minder dan 45 minuten invallen tijdens een wedstrijd? Deze vraag zal beantwoord worden in dit onderdeel. Om deze vraag te kunnen beantwoorden, werden de grafieken met het verloop van de richtingscoëfficiënt van de cumulatieve afstand van alle spelers die minder dan 45 minuten hebben gespeeld in een wedstrijd, bestudeerd. Dit werd gedaan voor alle wedstrijden van Club Brugge in de reguliere competitie. Tabel 5.5 geeft de absolute waarden en het percentage weer van de spelers waarvan de grafiek met verloop van de richtingscoëfficiënt van de cumulatieve afstand, boven of op het gemiddelde eindigt. Indien dit het geval is wordt er gesproken van een positieve grafiek. Wanneer de grafiek onder het gemiddelde eindigt wordt er gesproken van een negatieve grafiek. De gemiddelde waarde waarmee vergeleken wordt, is een waarde die uniek is voor elke speler. Voor elke wedstrijd uit de reguliere competitie waar een speler 45 minuten of meer gespeeld heeft, werd het gemiddelde genomen van de grafiek met het verloop van de richtingscoëfficiënt van de cumulatieve afstand. Voor elke speler werd vooraf het gemiddelde berekend, dat gebruikt werd voor het vergelijken van de data bij korte invalbeurten.

Tijdsinterval	Totaal aantal spelers	Aantal positief	Aantal negatief	Positief percentage (%)
0-15	58	29	29	50.00
15-35	47	35	12	74.47
35-45	29	20	9	68.97

Tabel 5.5: Aantal ingevallen spelers die een grafiek met verloop van richtingscoëfficiënt hebben die eindigt boven of onder gemiddelde waarde per tijdsinterval

Omdat enkel het aantal minuten niet alles zegt over een invalbeurt, werd er ook gekeken naar de score op het moment dat de speler inviel. Er wordt een onderscheid gemaakt tussen 2 verschillende soorten invalbeurten. De eerste is een invalbeurt wanneer één van beide teams 2 doelpunten of meer voorstaat (zie tabel 5.6). De tweede is een invalbeurt wanneer het verschil in score tussen beide teams maximaal 1 doelpunt is (zie tabel 5.7).

Wat direct opvalt uit bovenstaande tabellen, is dat, voor spelers die minder dan 15 minuten invallen, de resultaten

5 Werking van het offline programma

Tijdsinterval	Totaal	Aantal positief	Aantal negatief	Positief percentage (%)
0-15	58	13	13	50.00
15-35	47	11	1	91.67
35-45	29	0	0	-

Tabel 5.6: Aantal ingevallen spelers die een grafiek met verloop van richtingscoëfficiënt hebben die eindigt boven of onder gemiddelde waarde per tijdsinterval wanneer verschil in score groter of gelijk aan 2 was

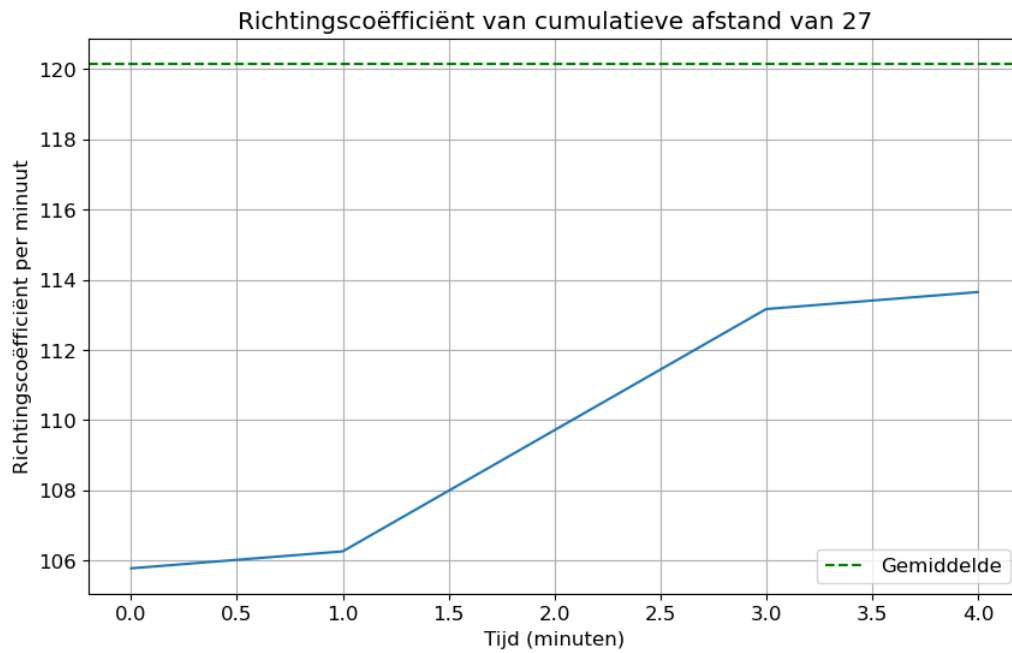
Tijdsinterval	Totaal	Aantal positief	Aantal negatief	Positief percentage (%)
0-15	58	16	16	50.00
15-35	47	24	9	72.73
35-45	29	20	9	68.97

Tabel 5.7: Aantal ingevallen spelers die een grafiek met verloop van richtingscoëfficiënt hebben die eindigt boven of onder gemiddelde waarde per tijdsinterval wanneer verschil in score kleiner dan 2 was

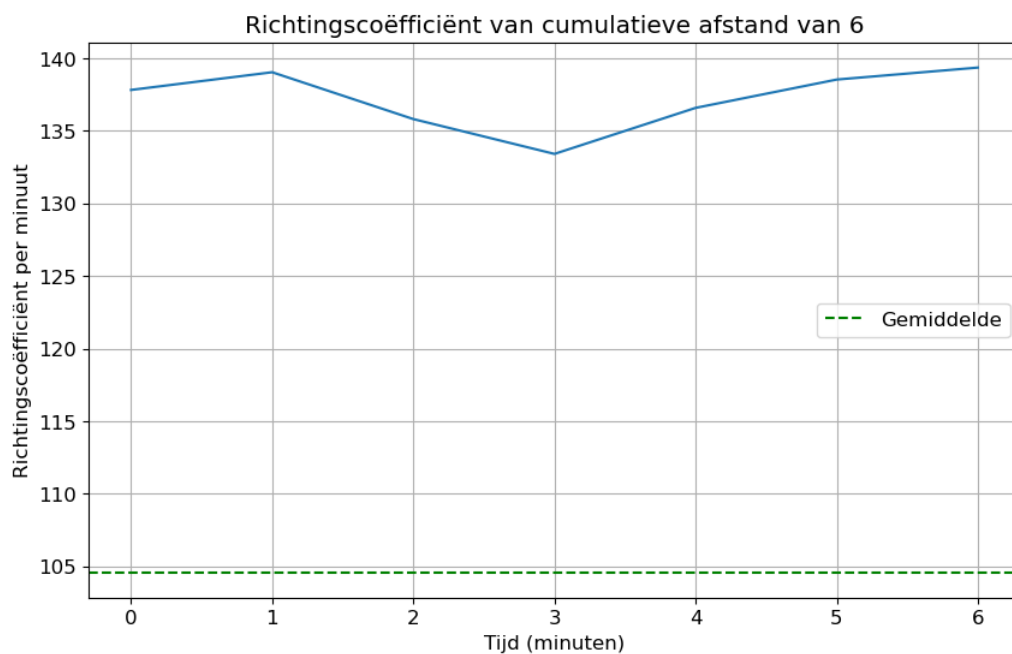
uiteenlopend zijn. De helft van de ingevallen spelers eindigt met een grafiek onder het gemiddelde, de andere helft eindigt met een grafiek boven of gelijk aan het gemiddelde. Ook de score heeft geen invloed bij dit tijdsinterval. Voor zowel wedstrijden waar het verschil in score tussen beide ploegen minder dan 2 is of het verschil groter is of gelijk aan 2, eindigt 50% van de grafieken die het verloop van de richtingscoëfficiënt van de cumulatieve afstand weergeven, onder het gemiddelde en 50% boven of gelijk aan het gemiddelde. Dit komt omdat er te weinig data is om een deftige grafiek op te stellen. Figuur 5.4 en 5.5 tonen beiden de grafiek met het verloop van de richtingscoëfficiënt van de cumulatieve afstand van spelers die respectievelijk 4 en 6 minuten zijn ingevallen. Bij figuur 5.4 was de score tijdens invallen 0-1 voor het team van deze speler en bij figuur 5.5 was de score 3-2 voor het team van deze speler. Bij invalbeurten die minder dan 15 minuten duurden, kan er dus geen duidelijke conclusie met betrekking tot vermoeidheid gemaakt worden, omdat er te weinig data is. Dit is echter geen probleem, omdat spelers die slechts in de laatste 15 minuten gewisseld worden, niet in die mate vermoeid zullen zijn dat een nieuwe wissel aan de orde is.

Wanneer er gekeken wordt naar de invalbeurten tussen de 15 en 35 minuten, ligt de verhouding tussen grafieken boven of gelijk aan het gemiddelde en onder het gemiddelde, volledig anders. Bijna 75% van deze spelers zullen eindigen met een grafiek die boven of gelijk aan het gemiddelde is. Dit is logisch, aangezien er nu wel genoeg data is om een correcte grafiek te kunnen maken. Ook is het logisch dat spelers die slechts 15 tot 35 minuten spelen in een wedstrijd, vaker niet vermoeid zullen zijn op het einde van de wedstrijd, dan dat ze wel vermoeid zullen zijn. Wat opvalt bij tabel 5.6 is dat, wanneer het verschil tussen de score van beide teams meer dan twee doelpunten bedraagt, de speler die inviel meer dan 90% van de tijd met een grafiek boven of gelijk aan het gemiddelde eindigt. Dit is logisch aangezien de intensiteit dan meestal minder laag zal liggen, omdat de wedstrijd al beslist is. Figuur 5.6 geeft een voorbeeld van een speler die inviel wanneer de wedstrijd al beslist was.

5 Werking van het offline programma

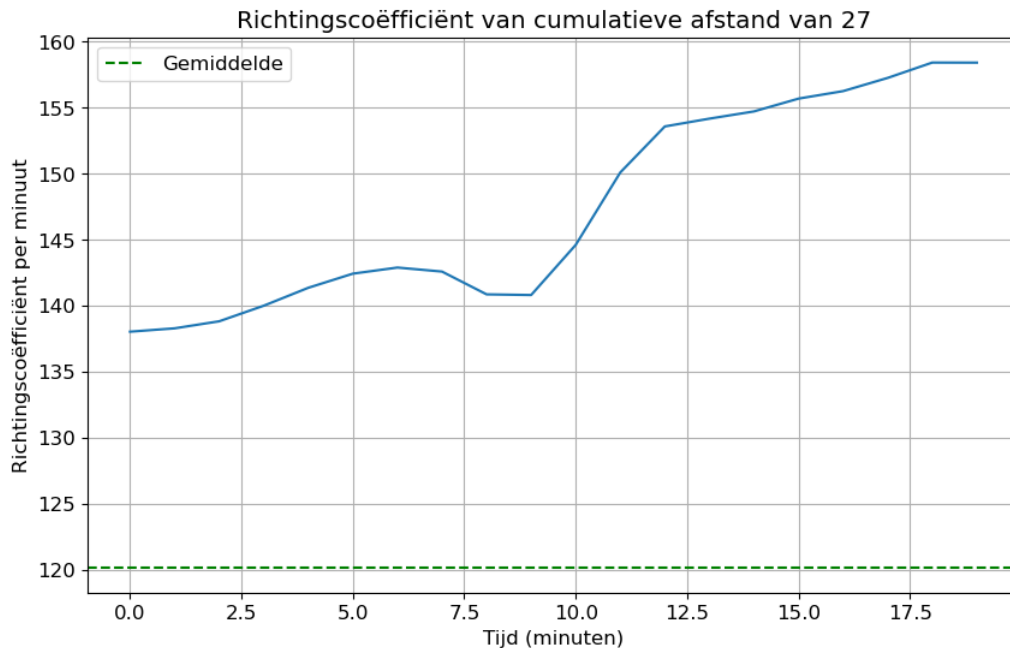


Figuur 5.4: Verloop van richtingscoëfficiënt van cumulatieve afstand van speler met invalbeurt tussen 0 en 15 minuten



Figuur 5.5: Verloop van richtingscoëfficiënt van cumulatieve afstand van speler met invalbeurt tussen 0 en 15 minuten

5 Werking van het offline programma

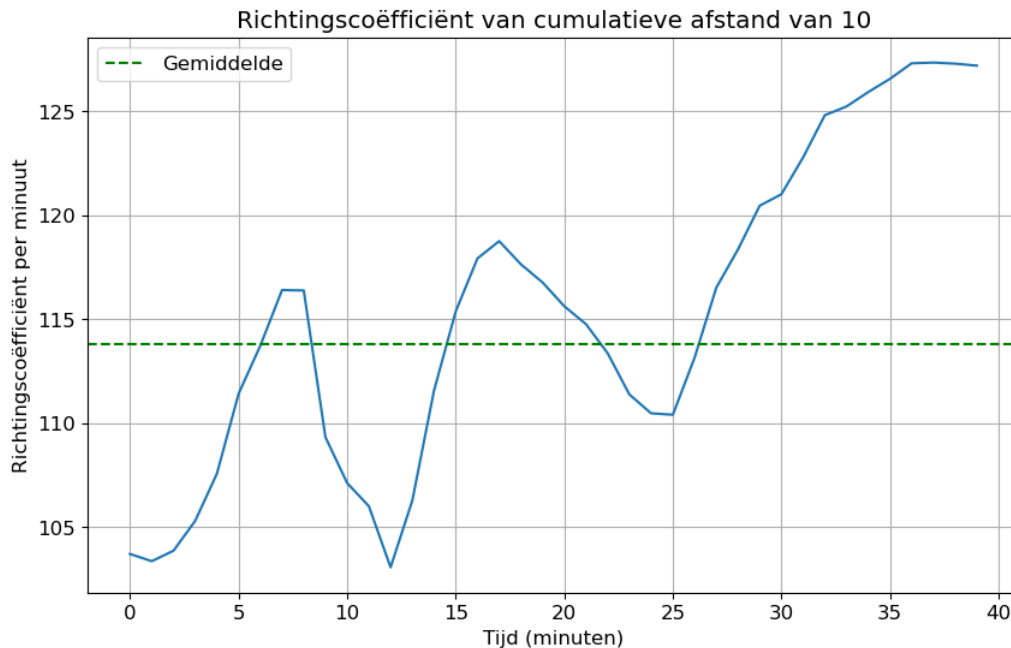


Figuur 5.6: Verloop van richtingscoëfficiënt van cumulatieve afstand van speler met invalbeurt tussen 15 en 35 minuten

Voor invalbeurten tussen de 35 en 45 minuten, kan ongeveer dezelfde conclusie getrokken worden als voor invalbeurten tussen 15 en 35 minuten. Ook hier is er genoeg data en zullen er dus correcte grafieken kunnen opgesteld worden. Waar het percentage voor grafieken die boven of gelijk aan het gemiddelde zijn, bij invalbeurten tussen de 15 en 35 minuten iets minder dan 75% bedraagt, ligt dat voor invalbeurten tussen de 35 en 45 minuten lager en is het gelijk aan 69%. Dit is logisch aangezien een speler die langer speelt, een hogere kans op vermoeidheid heeft en de grafieken dus lager zullen liggen. Opvallend is dat geen enkele spelers die tussen de 36 en 45 minuten inviel, dit deed op een moment dat de score 2 doelpunten of meer verschilde. Figuur 5.7 geeft een voorbeeld van een speler die inviel wanneer de wedstrijd al beslist was.

Concluderend kan er gesteld worden dat invalbeurten minder dan 15 minuten niet genoeg data bevatten, om conclusies te kunnen trekken. Vanaf hoeveel minuten er wel genoeg data is, kan niet exact bepaald worden. Dit is een grijze zone. Vanaf minuut 15 is er meestal wel voldoende data om een deftige grafiek te kunnen opstellen en dus conclusies hieruit te kunnen trekken. Vanaf het moment dat er langer dan 15 minuten ingevallen wordt, zal er in de meeste gevallen geëindigd worden met een grafiek die boven of gelijk aan het gemiddelde is. Hoe langer de speler speelt, hoe hoger de kans dat deze grafiek onder het gemiddelde zal eindigen. Spelers die invallen in wedstrijden, waar de score tussen beide ploegen veel verschilt (2 doelpunten of meer), zullen minder vaak onder het gemiddelde eindigen, dan wanneer de score dicht bij elkaar ligt en de wedstrijd dus nog spannend is.

5 Werking van het offline programma



Figuur 5.7: Verloop van richtingscoëfficiënt van cumulatieve afstand van speler met invalbeurt tussen 35 en 45 minuten

5.3.2 Invloed van rode kaarten

In dit deel zal de invloed van rode kaarten op de fysieke data van spelers besproken worden. Hoe bruikbaar is het offline programma, wanneer er een team met een man minder komt te staan? Gaan de spelers meer of minder afstand afleggen? Om deze vragen te kunnen beantwoorden, werden alle wedstrijden in de reguliere competitie van de Jupiler Pro League, waar minstens 1 rode kaart viel, onderzocht. In totaal werden er 51 rode kaarten bestudeerd. Als eerste stap, werd de verdeling van het tijdstip waarop deze kaarten werden gegeven, onderzocht. Zoals te zien in tabel 5.8 wordt er meer dan 40% van de rode kaarten gegeven na minuut 80. Iets minder dan drie kwart van de rode kaarten, worden getrokken in de tweede helft.

Om de invloed van rode kaarten op het verloop van de fysieke data (berekend door het niet-live programma) te bestuderen, werd gekeken naar het verschil tussen de data van de spelers van het team dat de rode kaart kreeg, voor de rode kaart en na de rode kaart. Voor elk van deze spelers werd de data berekend, die het verloop van de richtingscoëfficiënt van de afgelegde afstand voorstelt. Vervolgens werd voor elke speler, aan de hand van deze data, de som van de data in de 15 minuten voorafgaand aan de rode kaart en de 15 minuten die volgden op de rode kaart, berekend. Een voorbeeld hiervan is te zien in figuur 5.8. De rode lijn in deze figuur stelt de minuut voor wanneer de rode kaart gevallen is. De zwarte lijnen de 15 minuten voor en na de rode kaart. De stippellijnen stellen het gemiddelde van de 15 minuten voor en de 15 minuten na de rode kaart voor. Deze gemiddeldes worden vergeleken. In sommige gevallen valt een rode kaart te vroeg of te laat in de wedstrijd, waardoor er geen volledige

5 Werking van het offline programma

Tabel 5.8: Tijdverdeling van rode kaarten

Tijdsinterval	Aantal rode kaarten	Percentage (%)
0-15	2	3.92
16-30	5	9.80
31-45	7	13.73
46-60	7	13.73
61-80	9	17.65
81+	21	41.18
Totaal	51	100.00

15 minuten voorafgaand aan of na de kaart beschikbaar zijn. In dergelijke gevallen werd de daadwerkelijke tijd die voorafging of volgde op de kaart gebruikt voor de berekeningen. Indien de daadwerkelijke tijd minder dan 5 minuten bedroeg, werd deze data niet meegenomen in het onderzoek.

Door de sommen van de data die voor de rode kaart en na de rode kaart plaatsvond, te vergelijken, kan er geconcludeerd worden of een speler na de rode kaart meer of minder gelopen heeft. Zoals te zien in tabel 5.9 liep ongeveer 64% van de spelers meer nadat zijn team een rode kaart had gekregen. Ongeveer 36% liep minder na het krijgen van een rode kaart.

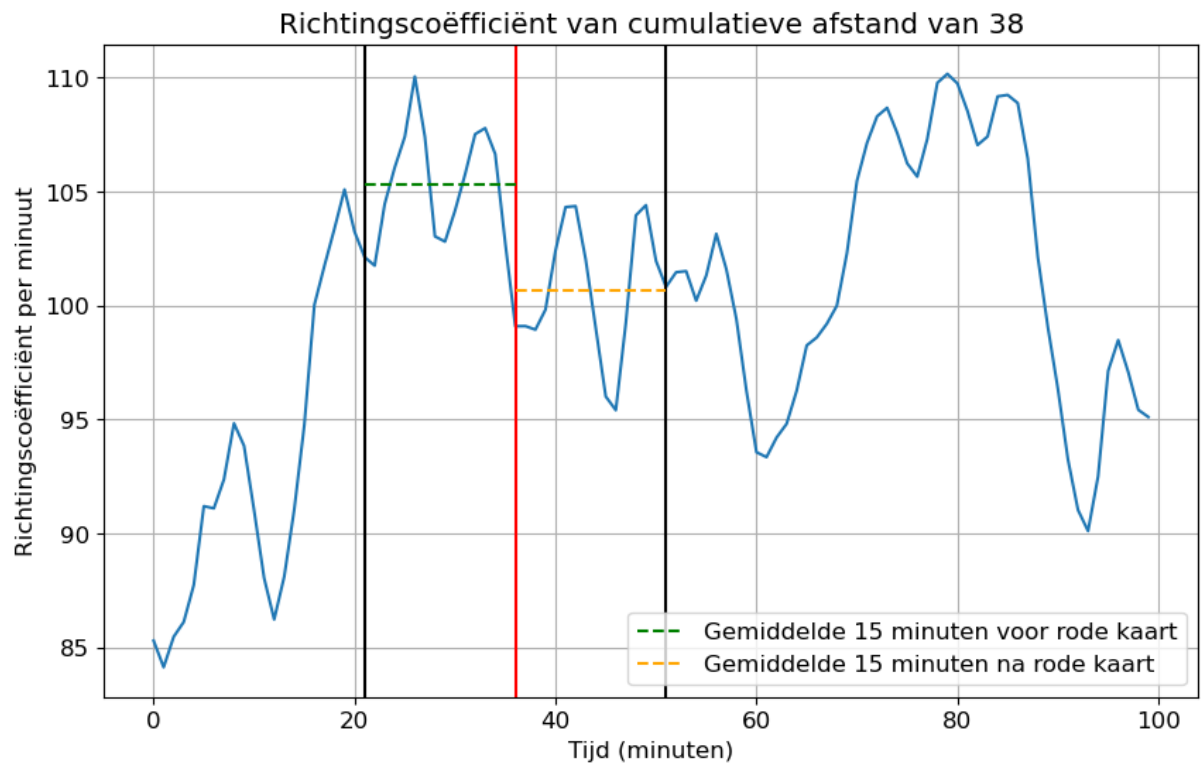
Tabel 5.9: Aantal spelers die meer en minder afstand aflegden na rode kaart

	Aantal	Percentage
Spelers die meer afstand aflegden	236	64.13%
Spelers die minder afstand aflegden	132	35.87%
Totaal	368	100.00%

Tabel 5.10 is een combinatie van tabel 5.8 en 5.9. Hier wordt er per tijdsinterval weergegeven hoeveel spelers (en hoeveel procent) er meer of minder afstand afleggen nadat hun team een rode kaart gekregen heeft. Wat opvalt is dat, voor de meeste tijdsintervallen, ongeveer 65 tot 75% van de spelers meer loopt nadat hun team een rode kaart gekregen heeft. Wanneer de rode kaart in het eerste kwartier van de tweede helft valt, zal meer dan 80% van de spelers meer lopen na de rode kaart. Als de rode kaart in het laatste kwartier van de eerste helft valt, zal slecht 44% van de spelers meer lopen nadat de kaart gegeven is.

Om nog een beter beeld te krijgen van de invloed van rode kaarten, op de fysieke data, werd er ook rekening gehouden met de tussenstand, wanneer de rode kaart plaatsvond. In tabel 5.11 wordt weergegeven hoeveel spelers meer of minder lopen, nadat hun team een rode kaart gekregen heeft. Hier wordt ook rekening gehouden met het feit of hun team aan het winnen, gelijk spelen of verliezen is. Wanneer een team op voorsprong staat, lopen 75%

5 Werking van het offline programma

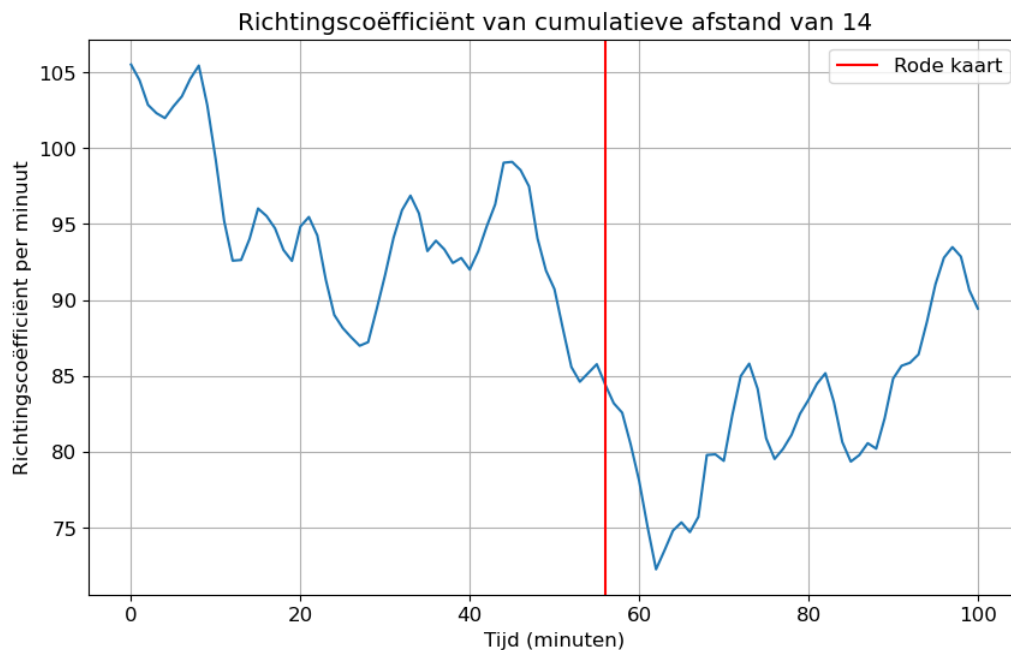


Figuur 5.8: Uitleg werkwijze invloed van rode kaarten

Tabel 5.10: Aantal spelers dat meer en minder afstand aflegden na rode kaart voor verschillende tijdsintervallen

Tijdsinterval	Aantal hoger	Percentage hoger	Aantal lager	Percentage lager
0-15	16	72,73%	6	27,27%
16-30	34	66,67%	17	33,33%
31-45	34	44,16%	43	55,84%
46-60	49	81,67%	11	18,33%
61-80	47	68,11%	22	31,88%
81+	72	67,92%	34	32,08%

5 Werking van het offline programma



Figuur 5.9: Grafiek met richtingscoëfficiënt van cumulatieve afstand na rode kaart bij voorsprong

van de speler minder na de rode kaart. Figuur 5.9 toont hier een voorbeeld van. Dit komt waarschijnlijk omdat de tegenstander na deze rode kaart de druk opvoert en op zoek gaat naar de gelijkmaker. Bij een gelijkspel zien we een heel gelijkaardig verloop. 64% van de spelers lopen minder nadat hun team een rode kaart krijgt. De reden hiervoor is gelijkaardig aan de vorige. De tegenstander zal na de rode kaart proberen door te drukken en het winnend doelpunt te maken. Figuur 5.10 toont hier een voorbeeld van. Als het team dat de rode kaart krijgt aan het verliezen is, zijn er 50% spelers die minder gaan lopen en 50% spelers die meer gaan lopen. Een voorbeeld hiervan kan teruggevonden worden in figuur 5.11.

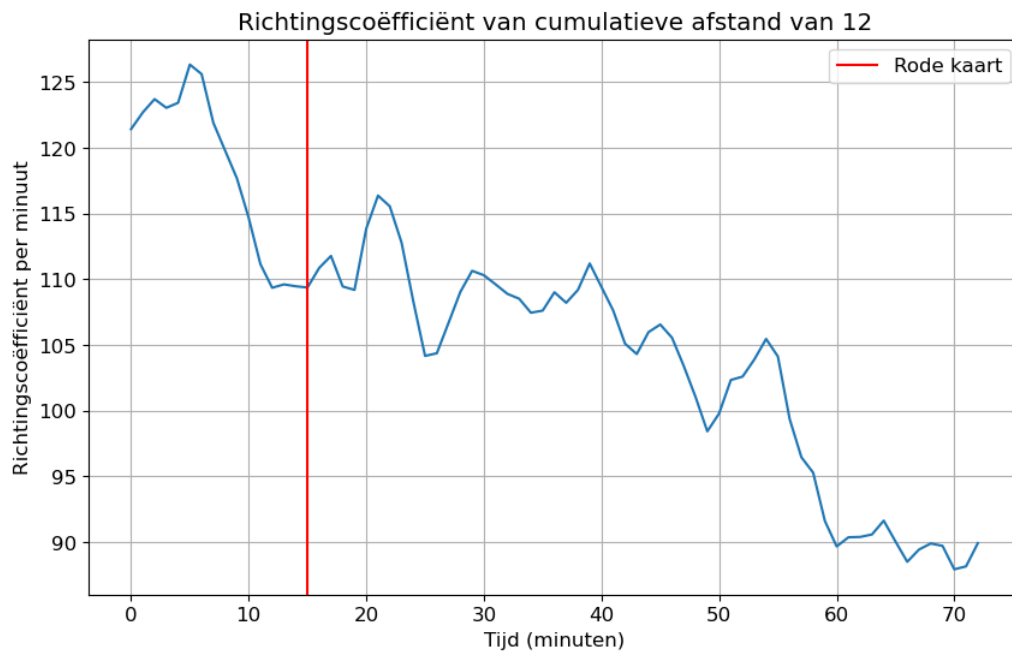
Tabel 5.11: Aantal spelers die meer of minder lopen na rode kaart voor verschillende tussenstanden

Stand	Aantal meer	Percentage meer	Aantal minder	Percentage minder
Voorsprong	29	24,37	90	75,63
Gelijkspel	52	35,62	94	64,38
Achterstand	51	49,51	52	50,49

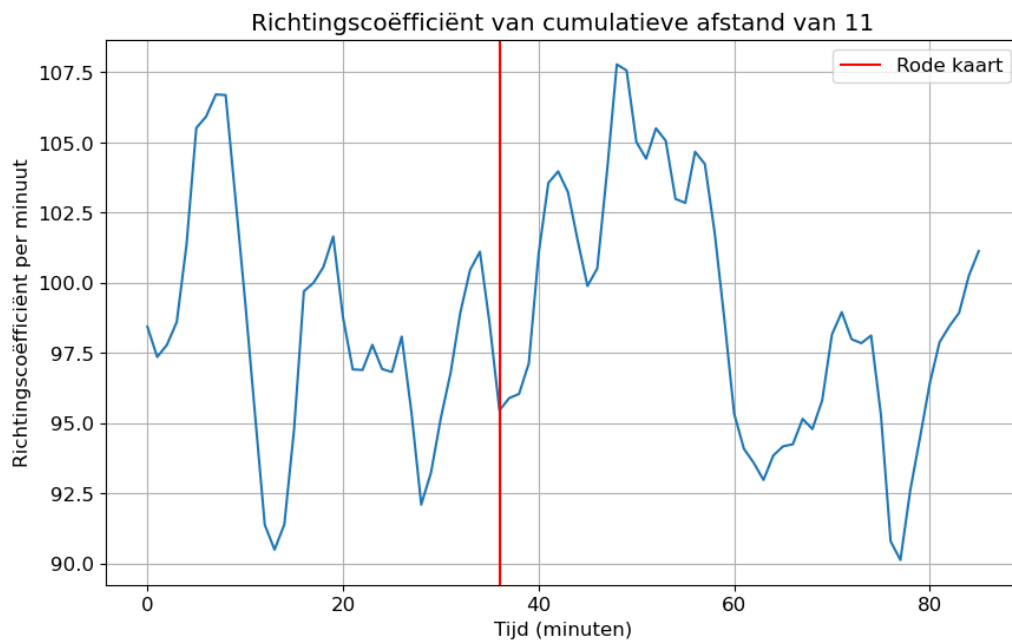
5.4 Andere invloeden op fysieke data

In dit hoofdstuk zal er gekeken worden naar enkele andere zaken die invloed kunnen hebben op het verloop van de grafieken met fysieke data van voetballers. De twee zaken die zullen besproken worden, zijn de expected

5 Werking van het offline programma



Figuur 5.10: Grafiek met richtingscoëfficiënt van cumulatieve afstand na rode kaart bij gelijkstand



Figuur 5.11: Grafiek met richtingscoëfficiënt van cumulatieve afstand na rode kaart bij achterstand

5 Werking van het offline programma

goals en het balbezit. Eerst zal er gekeken worden of het hebben van een hoog aantal expected goals, een invloed zal hebben op de fysieke data van de spelers. Daarna zal het verschil tussen fysieke data bij een hoog balbezit percentage en een laag balbezit percentage bestudeerd worden.

5.4.1 Invloed van expected goals

Om de invloed van de expected goals (xG) te bepalen, zijn gegevens gebruikt van alle wedstrijden in de reguliere competitie van Club Brugge. Eerst is via de StatsPerform API de xG-data voor deze wedstrijden opgehaald. Vervolgens is voor elke helft (inclusief extra speeltijd, dus de helften kunnen ongelijk zijn) berekend hoeveel expected goals Club Brugge had. Er wordt hier per helft gewerkt, omdat een speler in het begin van de wedstrijd doorgaans meer zal lopen dan op het einde van de wedstrijd.

Voor elke speler in elke wedstrijd werd per helft de fysieke data van de richtingscoëfficiënt van de afgelegde afstand berekend en het gemiddelde hiervan genomen. Het gemiddelde werd genomen omdat de totale som van de gegevens niet representatief is, aangezien de duur van beide helften verschillend kan zijn. Dit gemiddelde werd vervolgens gekoppeld aan de eerder berekende xG-score per helft. Voor drie intervallen van verwachte doelpunten (tussen 0 en 0,5, tussen 0,5 en 1, en groter dan 1) werd de gemiddelde waarde van de fysieke data per speler berekend. Dus elke speler heeft verschillende gemiddelde waarden voor deze intervallen voor de eerste en tweede helft. Alleen spelers die voldoende hebben gespeeld zodat ze voor elk interval van elke helft gegevens hebben, worden meegenomen in het onderzoek.

Tabel 5.12 toont de data voor de spelers in de eerste helft en tabel 5.13 toont de data voor de spelers in de tweede helft. Tabel 5.14 toont de som van de data van beide helften. Zoals duidelijk te zien uit de waarden van deze tabellen, wordt er over het algemeen minder gelopen in de tweede helft, dan in de eerste helft. Om een conclusie te trekken omtrent de invloed van expected goals op het verloop van de fysieke data van spelers, wordt er per positie gekeken, want de invloed van expected goals is anders per positie.

Als de expected goals lager liggen, zal de keeper gemiddeld meer afstand afleggen. Dit komt doordat de keeper dan meer betrokken wordt in het spel, dan wanneer de expected goals hoger liggen. De centrale verdedigers legden meer afstand af wanneer de expected goals van het team hoger lagen. Voor flankverdedigers hadden de expected goals niet echt een invloed op de fysieke data. Zowel bij hoge als bij lage expected goals lag de data ongeveer even hoog. Bij de middenvelders lag de fysieke data meestal hoger wanneer het aantal expected goals hoger lag. Bij zowel flankaanvallers als spitsen waren de resultaten afwisselend. Dit hangt af van het type speler. De ene zullen hogere fysieke data hebben bij hoge expected goals, de andere wanneer het aantal expected goals lager ligt.

5 Werking van het offline programma

ID	$xG < 0,50$	$0,50 < xG < 1$	$1 < xG$
Keeper			
22	57,92	57,71	48,48
Centrale verdedigers			
44	104,01	111,27	108,86
58	106,36	104,10	103,72
Flankverdedigers			
55	109,94	110,79	104,18
14	109,55	115,81	105,45
Middenvelder			
10	118,77	115,42	114,51
15	118,98	124,52	127,93
20	117,55	119,54	114,66
Flankaanvaller			
7	112,15	115,26	112,63
32	105,93	114,52	132,18
Spits			
9	115,94	112,75	122,46
99	111,70	109,64	102,74

Tabel 5.12: Fysieke data voor expected goals intervallen in de eerste helft

5 Werking van het offline programma

ID	$xG < 0,50$	$0,50 < xG < 1$	$1 < xG$
Keeper			
22	52,40	50,82	55,69
Centrale verdedigers			
44	102,10	101,77	104,93
58	97,28	100,30	106,46
Flankverdedigers			
14	105,77	108,58	104,91
55	101,98	101,26	107,91
Middenvelder			
10	106,25	106,71	109,01
15	110,09	112,09	115,39
20	106,28	107,79	115,19
Flankaanvaller			
7	107,43	101,78	101,93
32	97,44	100,75	104,23
Spits			
9	106,58	108,54	112,80
99	104,11	94,31	103,53

Tabel 5.13: Fysieke data voor expected goals intervallen in de tweede helft

5 Werking van het offline programma

ID	$xG < 0,50$	$0,50 < xG < 1$	$1 < xG$
Keeper			
22	110,31	108,52	104,17
Centrale verdedigers			
44	206,11	213,04	213,79
58	203,64	204,41	210,18
Flankverdedigers			
14	215,32	224,39	210,36
55	211,92	212,05	212,08
Middenvelder			
10	225,02	222,13	223,52
15	229,07	236,61	243,31
20	223,83	227,34	229,85
Flankaanvaller			
7	219,58	217,04	214,57
32	203,37	215,26	236,41
Spits			
9	222,52	221,29	235,26
99	215,81	203,95	206,27

Tabel 5.14: Som van fysieke data voor expected goals intervallen van eerste en tweede helft

5 Werking van het offline programma

5.4.2 Invloed van balbezit

De invloed van het balbezit op de fysieke data van spelers wordt op dezelfde manier berekend als de invloed van de expected goals. Opnieuw wordt de wedstrijd verdeeld in 2 helften en wordt het gemiddelde van het verloop van de richtingscoëfficiënt van de cumulatieve afstand berekend voor een aantal intervallen. In tabel 5.15 worden de resultaten van de eerste helft weergegeven en in tabel 5.16 de resultaten van de tweede helft. Tabel 5.17 toont de som van de waarden van de twee helften. Opnieuw valt uit de tabellen af te lezen dat over het algemeen de fysieke data in de tweede helft lager ligt dan in de eerste helft. In tabel 5.17 zijn er 10 spelers (van de 13 spelers) die hogere fysieke data hebben als het balbezit lager ligt. Dit is logisch, aangezien spelers bij lager balbezit meer achter de bal zullen moeten lopen. De invloed van balbezit op de fysieke data is niet positieafhankelijk.

Tabel 5.15: Fysieke data voor balbezit intervallen in de eerste helft

Index	balbezit < 45	45 < balbezit < 55	55 < balbezit
22	57.63	56.94	54.29
44	116.08	108.50	103.64
58	115.41	103.81	104.85
55	124.76	109.65	103.94
14	115.63	114.42	108.71
6	114.21	108.07	106.80
20	122.88	117.86	114.39
15	126.42	123.74	116.06
10	117.43	116.44	117.70
7	111.93	114.48	114.06
32	113.54	115.22	105.45
99	111.67	111.69	106.41
9	119.43	117.27	114.23

5 Werking van het offline programma

Tabel 5.16: Fysieke data voor balbezit intervallen in de tweede helft

Index	balbezit < 45	45 < balbezit < 55	55 < balbezit
22	52.53	50.80	50.79
44	103.52	98.90	103.60
58	99.41	100.53	97.57
14	108.17	104.99	102.48
55	99.23	104.75	108.69
6	94.82	103.07	96.32
10	104.75	109.89	104.67
15	113.90	106.84	116.56
20	106.65	106.73	108.82
7	101.83	105.70	112.54
32	103.43	101.40	90.79
9	112.34	108.36	109.87
99	105.34	95.03	99.61

Tabel 5.17: Som van fysieke data voor balbezit intervallen van eerste en tweede helft

Index	balbezit < 45	45 < balbezit < 55	55 < balbezit
22	110.16	107.74	105.08
44	219.60	207.40	207.24
58	214.82	204.34	202.42
14	223.80	219.41	211.19
55	223.99	214.40	212.63
6	209.03	211.14	203.12
20	229.53	224.59	223.21
15	240.32	230.58	232.62
10	222.18	226.33	222.37
7	213.76	220.18	226.60
32	216.97	216.62	156.24
9	231.77	225.63	224.10
99	217.01	206.72	206.02

6

Werking van het online programma

Het einddoel van deze masterproef is om een programma te creëren dat live tijdens wedstrijden de vermoeidheid van de spelers trackt. In dit hoofdstuk zal het online programma besproken worden. Daarnaast zullen de problemen besproken worden, die tijdens het creëren van het online programma naar voor gekomen zijn. In het eerste deel van dit hoofdstuk zal de verwerking van de online trackingdata besproken worden. Er wordt dieper ingegaan op het verschil met de offline trackingdata, de problemen met de online data en hoe die problemen opgelost worden. In het tweede deel zal het effectieve eindproduct, het dashboard dat gebruikt kan worden tijdens wedstrijden, besproken worden.

6.1 Verwerking van de online trackingdata

Via de StatsPerform API kan een bestand opgehaald worden met de offline trackingdata van de volledige wedstrijd. Voor de online data is dit anders. De online trackingdata is tijdens wedstrijden beschikbaar via de FTP-server van StatsPerform. Tijdens een wedstrijd wordt er elke minuut een bestand op de FTP-server geplaatst met de trackingdata van de afgelopen minuut. Telkens wanneer een nieuw bestand op de FTP-server wordt gevonden, moet de data uit dat bestand worden samengevoegd met de data van de voorgaande minuten om de fysieke data van een speler tot dat punt te kunnen berekenen.

Om het verwerken van de verschillende trackingfiles per minuut, te testen, werd er gekozen om een simulatie van de FTP-server te maken om het testen te vergemakkelijken. Deze simulatie plaatst elke minuut een bestand in een lokale folder. Deze folder stelt de folder op de FTP-server voor, waar de data normaal gezien tijdens de wedstrijd in verschijnt. Door te kijken wanneer er nieuwe bestanden aan deze folder worden toegevoegd, kan het online programma getest worden, wanneer er geen live wedstrijd is. Eenmaal het verwerken van de verschillende bestanden werkte, kon er getest worden op de FTP-server van StatsPerform zelf.

Het effectieve live programma werkt natuurlijk wel met data van de FTP-server. Voor de wedstrijd begint, moet aangegeven worden om welke wedstrijd het gaat, om te weten in welke folder op de FTP-server de data zal verschijnen. Vervolgens zal tijdens de wedstrijd, om de minuut, de nieuwste tracking file van de FTP-server opgehaald

6 Werking van het online programma

en verwerkt worden. Voor elke speler die start in de wedstrijd of invalt, zal de data van de richtingscoëfficiënt van de cumulatieve afstand en het aantal frames met een snelheid hoger dan 2 keer boven de gemiddelde snelheid per minuut, alsook hun bijhorende grafieken, berekend worden.

Voor het verwerken van de data werd gebruikt gemaakt van de library Floodlight [23]. Floodlight werd ook gebruikt voor het offline programma. De trackingdata van het offline programma en van het online programma hebben een licht verschillend formaat. Het grootste verschil tussen de twee is dat de spelers in de online trackingdata niet geordend zijn. Bij de offline trackingdata zal voor iedere rij de volgorde van de spelers in die rij, dezelfde zijn. Floodlight verwacht ook dat de input data, geordende trackingdata is. Na het ophalen van de online trackingdata van de FTP-server, moet de trackingdata dus eerst geordend worden.

6.1.1 Problemen met online trackingdata

Zoals al eerder vermeld, wordt de online trackingdata niet aangepast door manuele interventies bij StatsPerform. Dit heeft als gevolg dat de data veel gebreken bevat. In dit hoofdstuk zullen alle fouten die in de online trackingdata voorkomen, besproken worden. Tevens wordt uitgelegd hoe deze fouten werden aangepakt tijdens het verwerken van de online trackingdata. Eerst zullen de algemene problemen met betrekking tot de bestanden en de FTP-server besproken worden. Daarna zal er wat dieper worden ingegaan op de problemen in de trackingdata zelf.

Problemen met FTP-server

Het grootste probleem is er een in verband met de kwantiteit van de data op de FTP-server. Zoals reeds vermeld in het vorige hoofdstuk, zijn er slechts 11 van de 30 wedstrijden in de reguliere competitie van Club Brugge, die complete online data hebben op de FTP-server. In tabel 6.1 wordt een overzicht gegeven van alle wedstrijden, het percentage aanwezige data en het aantal ontbrekende bestanden in de eerste en tweede helft. Zoals te zien in de tabel zijn er soms slechts een klein aantal ontbrekende bestanden (bijvoorbeeld voor de wedstrijd Union - Club Brugge zijn er slechts 3 ontbrekende bestanden). In dit geval zal het online programma wel bruikbaar zijn. Indien er bijvoorbeeld een volledige helft ontbreekt (zoals bij Genk - Club Brugge), is het logisch dat deze data niet bruikbaar is voor het online programma. De ontbrekende bestanden zijn niet per se bestanden van opeenvolgende minuten. Iets wat direct opvalt uit tabel 6.1 is dat data uit de eerste 10 wedstrijden van de reguliere competitie veel completer is dan in de laatste 20 van de reguliere competitie.

Problemen met trackingdata

Het eerste probleem bij de online tracking data is de identificatie van spelers. Bij de offline data werd er met een StatsPerform ID gewerkt om spelers te identificeren. Bij de online trackingdata wordt er echter gewerkt met een random gegenereerd ID, dat voor elke nieuwe tracking file (dus elke minuut) verandert. Er is geen manier om

6 Werking van het online programma

Wedstrijd	Percentage aanwezige bestanden	Aantal missende bestanden eerste helft	Aantal missende bestanden tweede helft
Club Brugge - KV Mechelen	100%	-	-
Westerlo - Club Brugge	100%	-	-
Eupen - Club Brugge	100%	-	-
Club Brugge - RWDM	100%	-	-
AA Gent - Club Brugge	97%	0	3
Club Brugge - Charleroi	100%	-	-
Anderlecht - Club Brugge	100%	-	-
Club Brugge - Genk	100%	-	-
Club Brugge - STVV	100%	-	-
Standard - Club Brugge	100%	-	-
Kortrijk - Club Brugge	100%	-	-
Club Brugge - Antwerp	0%	46	49
Union - Club Brugge	97%	0	3
Club Brugge - Cercle	0%	47	48
OH Leuven - Club Brugge	75%	13	11
Club Brugge - Standard	0%	47	49
KV Mechelen - Club Brugge	75%	11	13
Club Brugge - Gent	72%	13	14
RWDM - Club Brugge	0%	49	51
Club Brugge - Union	74%	12	14
Club Brugge - Westerlo	74%	12	13
Charleroi - Club Brugge	100%	-	-
Club Brugge - Kortrijk	0%	47	52
Antwerp - Club Brugge	47%	12	0
Club Brugge - Eupen	75%	37	35
Cercle - Club Brugge	87%	13	0
Club Brugge - Anderlecht	0%	50	48
Genk - Club Brugge	39%	9	51
Club Brugge - OH Leuven	86%	47	36
STVV - Club Brugge	0%	50	51
Complete data: 11/30			
Geen data: 7/30			
Incomplete data: 12/30			

Tabel 6.1: Kwaliteitscontrole van de online trackingdata per wedstrijd

6 Werking van het online programma

op basis van het ID spelers te linken over verschillende online tracking files heen, aangezien er ook geen link is tussen deze random gegenereerde IDs. Wel zijn er andere oplossingen.

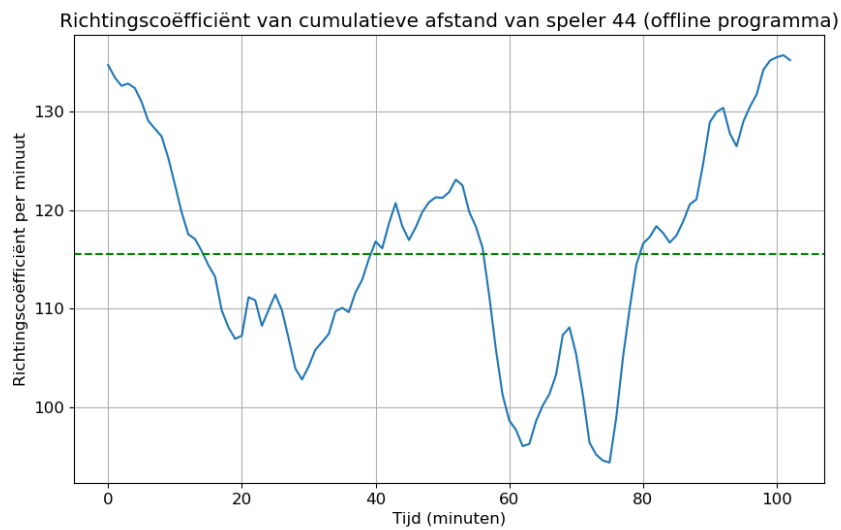
De manier waarop de spelers wel geïdentificeerd kunnen worden, is aan de hand van hun rugnummer. Hier treden er echter ook een paar problemen op. Ongeveer 5 tot 10 minuten voor de wedstrijd begint, plaatst StatsPerform al data op de server. Deze data is echter nog onbruikbaar, aangezien de spelers nog niet op het veld staan. Voor alle spelers zal het rugnummer in deze 5 tot 10 minuten gelijk zijn aan -1. Dit betekent dat er nog geen rugnummer gevonden is of dat er een ander probleem opgetreden is met het rugnummer. Wanneer de wedstrijd begint, worden de rugnummers dan automatisch veranderd van -1 naar het effectieve rugnummer van de spelers. Het probleem hierbij is dat sommige rugnummers nog niet direct veranderen en dit soms enkele minuten kan duren. Om dit op te lossen, wordt de data van de eerste minuten voor alle spelers op het veld bijgehouden en worden de rugnummers later (wanneer de -1 verandert in het rugnummer van de speler) toegewezen aan de juiste data. Hetzelfde probleem treedt op na de rust en ook daar wordt dezelfde oplossing gebruikt.

Soms gebeurt het ook tijdens de wedstrijd dat het rugnummer van 1 of meerdere spelers verdwijnt en er -1 in de plaats komt. Dit gebeurt meestal voor enkele frames en zal meestal niet lang duren. Er zijn echter uitzonderingen waarbij dit meerdere minuten kan duren. Voor een speler waarbij het rugnummer -1 is, kan aan de hand van het vorige frame het correcte rugnummer bepaald worden. Aangezien er slechts 0,04 seconden tussen de twee frames zitten, zal de afstand tussen de positie van de speler van 2 opeenvolgende frames, niet veel verschillen. Door de afstand te berekenen tussen de posities van de spelers tussen 2 frames, kan het correcte rugnummer bepaald worden.

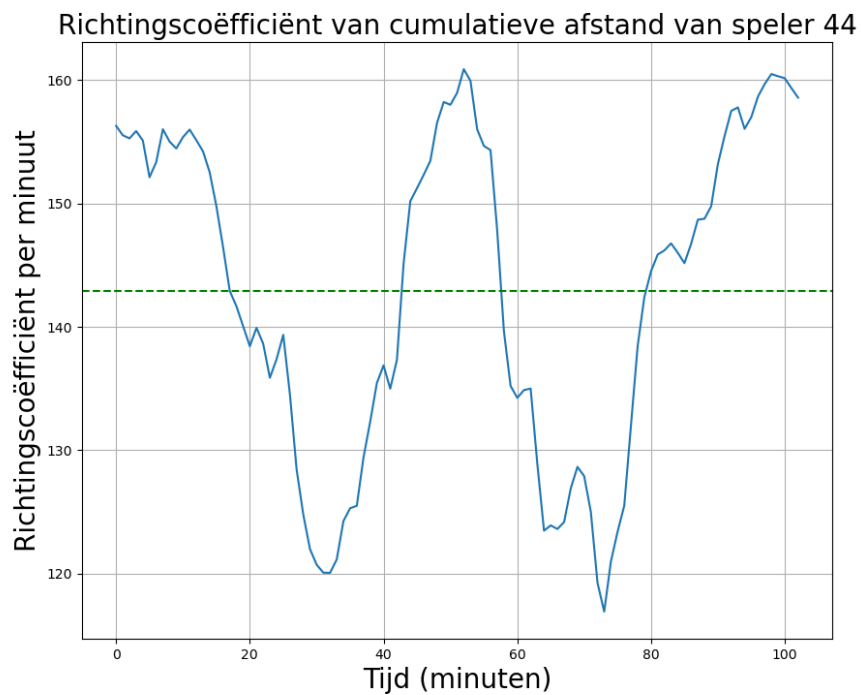
Het volgende probleem dat voorkomt bij de online trackingdata is dat een rij van de data soms 10 of 12 spelers van een bepaalde ploeg weergeven, terwijl er wel degelijk 11 spelers per ploeg op het veld staan. Dit gebeurt echter niet vaak en als het gebeurt, is het slechts voor 1 rij in de data (niet voor meerdere rijen na elkaar). Aangezien dit niet vaak gebeurt en ook (in de gebruikte dataset) niet voor een langere periode na elkaar, wordt de frame met een ontbrekende speler of een speler te veel, genegeerd.

Het laatste probleem is de precisie waarmee de positie van de spelers op het veld bepaald wordt. Aangezien de online trackingdata near-realtime op de FTP-server moet worden gezet en er hier dus amper delay op mag zitten (op de eerste lijn van elk bestand zal er telkens 1 minuut delay zijn), zal de kwaliteit van de trackingdata hieronder lijden. De afwijking is niet heel groot, maar zorgt er wel voor dat de grafieken gecreëerd door het online programma, gemiddeld hogere waarden aannemen dan hun equivalenten gecreëerd door het offline programma. Beide grafieken zullen wel hetzelfde verloop hebben, maar de pieken en dalen van de grafieken van het online programma zullen respectievelijk hogere lokale maxima en lagere lokale minima aannemen. Dit wordt getoond op figuur 6.1.

6 Werking van het online programma



(a) Offline programma



(b) Online programma

Figuur 6.1: Vergelijking van richtingscoëfficiënt van cumulatieve afstand gegenereerd door offline en online programma's

6 Werking van het online programma

6.2 Werking van het dashboard

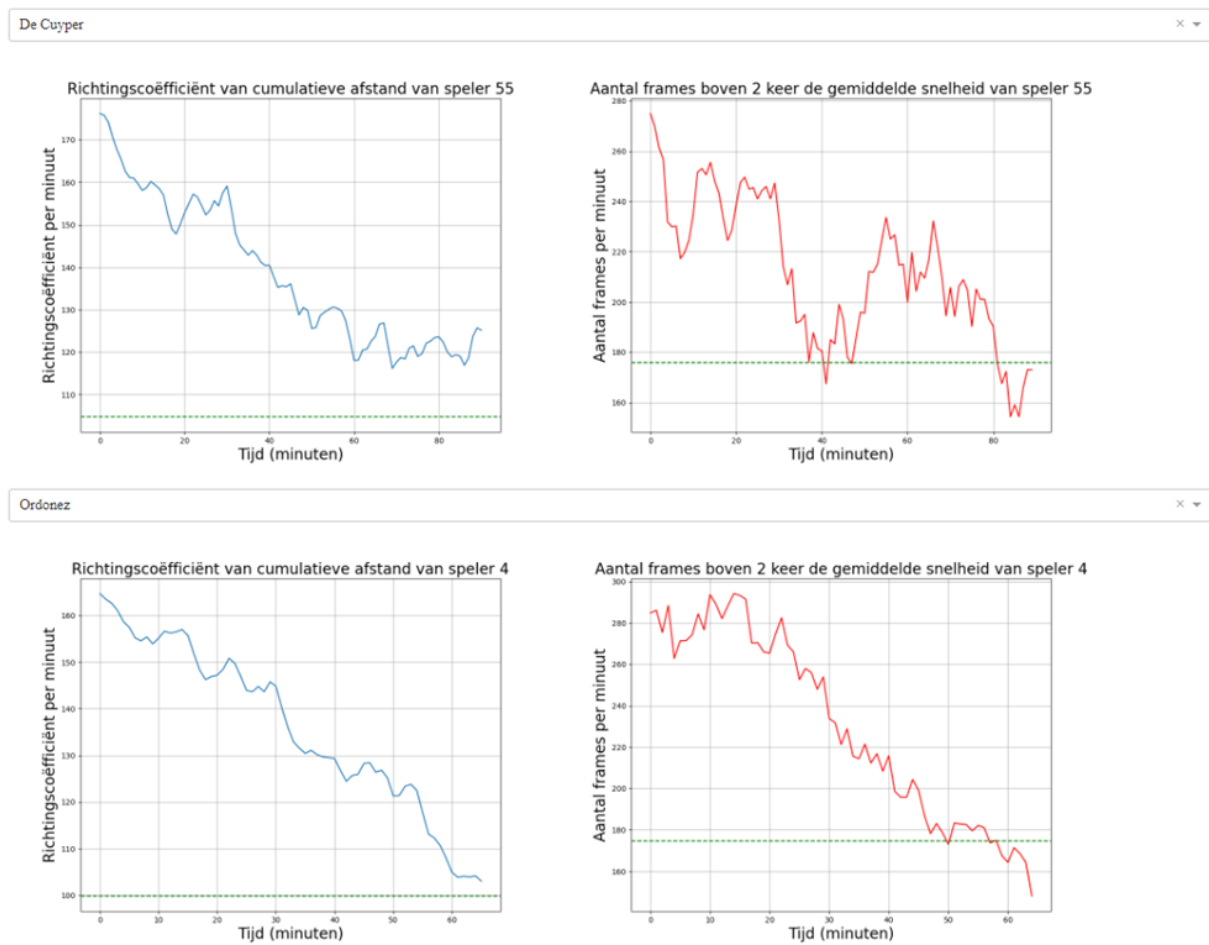
Het eindproduct van deze masterproef is een dashboard dat tijdens de wedstrijd gebruikt kan worden om de fysieke data van de spelers op een overzichtelijke manier te kunnen weergeven. Het dashboard is opgebouwd, gebruik makend van de library Dash [25] en bestaat uit een linker- en rechtervenster. Er zijn twee verschillende versies gemaakt van het online programma waarbij het linkervenster verschilt voor beiden. Een eerste versie die enkel de data van de spelers van Club Brugge weergeeft en een tweede versie die ook in staat is om de fysieke data van de tegenstander weer te geven. Het linkervenster bevat alle data en grafieken in verband met de fysieke data van de spelers. Zoals te zien op figuur 6.2 bestaat de linkerhelft uit twee dropdownlijsten met daaronder de grafieken die de richtingscoëfficiënt van de afgelegde afstand en het aantal frames met een snelheid hoger dan 2 keer de gemiddelde snelheid, weergeven. In deze dropdownlijsten kunnen alle spelers (van Club Brugge) die op het veld staan, geselecteerd worden en wordt hun data weergegeven. De grafiek wordt elke minuut automatisch geüpdatet, aangezien er elke minuut nieuwe data op de FTP-server komt. Wanneer een speler gewisseld wordt, wordt de invaller automatisch toegevoegd als keuze in de dropdownlijst.

Het rechtervenster bestaat uit een dropdownlijst, waaruit er 2 opties geselecteerd kunnen worden. De eerste optie is de vermoeidheid. Hier wordt een lijst teruggegeven van alle spelers die op het veld staan, geordend volgens vermoeidheid. Figuur 6.3 toont hier een voorbeeld van. Bij een speler die in het rood staat, zullen beide grafieken onder het gemiddelde liggen. Bij een speler in het oranje zal er slechts 1 grafiek onder het gemiddelde liggen. Bij spelers in het geel zullen beide grafieken boven het gemiddelde liggen, maar er niet ver boven. Als beide grafieken ruim boven het gemiddelde liggen, zal de speler in het groen staan.

Indien in de dropdownlijst van het rechtervenster 'match stats' wordt geselecteerd, worden er statistieken die invloed kunnen hebben op het verloop van fysieke data van de spelers of invloed op een tactische wissel, weergegeven. Een voorbeeld wordt weergegeven in figuur 6.4. Zoals eerdere onderzoeken reeds hebben uitgewezen [26] heeft het feit of je wint of verliest een invloed op de fysieke parameters van een speler. In hoofdstuk 6 werd de invloed van expected goals en balbezit op de fysieke parameters van spelers reeds besproken. Ook worden de ratings van de spelers weergegeven. Dit zijn ratings op 100 die aangeven hoe goed een speler speelt, berekend op basis van data (aangekomen passen, gemiste kansen, gescoorde doelpunten, ...). Al deze statistieken worden opgevraagd via de StatsPerform API en worden elke minuut geüpdatet. Wanneer er tussen de 2 opties in de dropdownlijst wordt gewisseld, duurt het een 20-tal seconden om de data op te vragen. Om dit op te lossen wordt er met caching gewerkt.

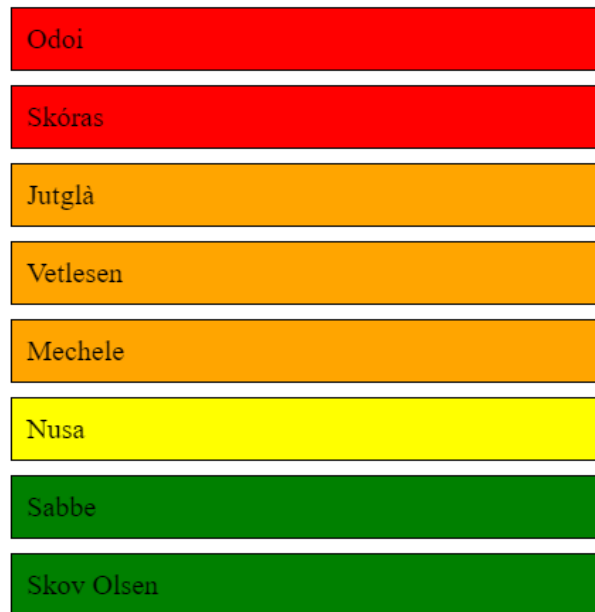
Concluderend kan er besloten worden dat er een groot aantal fouten in de trackingdata zit, maar de meeste fouten opgelost kunnen worden tijdens het verwerken van de data door het online programma. Er is één probleem dat met deze werkwijze niet kan opgelost worden en dat is de compleetheid van de data op de FTP-server. Voor de meeste wedstrijden van Club Brugge in de reguliere competitie van de Jupiler Pro League is niet alle data van de volledige wedstrijd aanwezig. De grafieken die het verloop van de richtingscoëfficiënt van de cumulatieve data

6 Werking van het online programma



Figuur 6.2: Dashboard: linkervenster met grafieken van fysieke data

6 Werking van het online programma



Odoi
Skóras
Jutglà
Vetlesen
Mechele
Nusa
Sabbe
Skov Olsen

Figuur 6.3: Dashboard: rechternvenster met vermoeidheidslijst

Score:	1-1
Scoring Attempts:	7-13
On-Target Scoring Attempts:	4-3
Possession:	24.9 - 75.1
Expected goals:	0.58 - 1.08

Ratings:

Player	Rating
Jutglà	83.92537
De Cuyper	80.25544
Meijer	77.16919
Mechele	64.36298
Spileers	60.19833
Skóras	59.23875
Mignolet	57.38798
Vanaken	34.84602
Thiago	24.12721

Figuur 6.4: Dashboard: rechternvenster met wedstrijdstatistieken

6 Werking van het online programma

en het aantal frames met een snelheid hoger dan 2 keer het gemiddelde voorstellen, hebben een gelijkaardig verloop voor zowel het offline en online programma, maar de grafieken van het online programma zullen hogere pieken en lagere dalen hebben dan bij het offline programma. De grafieken gecreëerd door het online programma worden, net als de data horend bij deze grafieken en andere data opgehaald via de StatsPerform API, gebruikt door het dashboard.

Conclusie

Zoals reeds bewezen door voorgaande onderzoeken, gaan de fysieke parameters van spelers achteruit wanneer er vermoeidheid optreedt. Omgekeerd kunnen de fysieke parameters dus ook gebruikt worden om vermoeidheid te bepalen. Het monitoren van de fysieke parameters van spelers tijdens een wedstrijd kan dus gebruikt worden om vermoeidheid van spelers te monitoren. Voor deze masterproef is er gekozen om zowel de afgelegde afstand als de snelheid van de speler te monitoren. Er wordt gebruik gemaakt van de trackingdata van StatsPerform. Op basis van de positionele coördinaten van de spelers tijdens de wedstrijd, worden de fysieke parameters van deze spelers berekend.

Het is natuurlijk niet omdat een speler minder loopt, hij sowieso last heeft van vermoeidheid. Er zijn nog een aantal andere factoren die invloed hebben hierop. Een van deze factoren is het verloop van de wedstrijd. Zaken zoals de tussenstand, het balbezit en de expected goals hebben allemaal invloed op de fysieke parameters van de spelers. Wanneer de wedstrijd al beslist is, zal de intensiteit lager liggen en zal er dus minder gelopen worden. Dit betekent niet dat de spelers sowieso vermoeid zijn. Een andere factor is rode kaarten. Wanneer een team een rode kaart krijgt en met een man minder valt, zullen de andere 10 spelers, afhankelijk van de score, meer of minder lopen. Dit betekent niet per se dat de spelers daarom meer of minder vermoeid zijn na de rode kaart.

Zoals eerder vermeld werd er voor deze masterproef gebruik gemaakt van de (tracking)data van StatsPerform. Er waren echter enkele problemen met deze data. Het grootste probleem is dat er maar voor 11 van de 30 wedstrijden in de reguliere competitie van Club Brugge, complete online trackingdata was. Dit betekent dat voor 19 van de 30 wedstrijden de vermoeidheid niet of niet goed bepaald kan worden, omdat er ontbrekende data is. Aangezien de data slechts enkele minuten op voorhand op de ftp server begint te verschijnen, is er ook geen enkele manier om te weten of er online trackingdata gaat zijn voor een bepaalde wedstrijd. Ook kan er niet op voorhand gezegd worden of de data compleet zal zijn of niet.

Ethische en maatschappelijke reflectie

In dit onderdeel zal de ethische en maatschappelijke reflectie van de masterproef besproken worden. Het onderzoek draagt bij aan twee Sustainable Development Goals (SDGs). Deze zijn SDG 3 (Goede gezondheid en welzijn) en SDG 9 (Industrie, innovatie en infrastructuur). Het onderzoek draagt bij aan SDG 3 (Goede gezondheid en welzijn) door de fysieke parameters van spelers tijdens wedstrijden van Club Brugge te monitoren. Door de afgelegde afstand en snelheid van spelers tijdens wedstrijden te analyseren, kan er gezorgd worden dat het tijdstip van vermoeidheid van spelers gedetecteerd wordt. Dit draagt bij aan het bevorderen van hun welzijn en creëert een gezondere sportomgeving.

Bovendien ondersteunt het onderzoek SDG 9 (Industrie, innovatie en infrastructuur) door het gebruik van tracking-technologieën en data-analyse binnen de sportwereld. Door deze innovaties kunnen nieuwe methoden worden

6 Conclusie

ontwikkeld om de prestaties van spelers te verbeteren. Het benutten van deze technologieën draagt bij aan de verbetering van wedstrijdstrategieën, wat resulteert in een duurzamere sportindustrie.

Referenties

- [1] R. Julian, R. M. Page, and L. D. Harper, "The Effect of Fixture Congestion on Performance During Professional Male Soccer Match-Play: A Systematic Critical Review with Meta-Analysis," 2020.
- [2] Z. Jiang, Y. Hao, N. Jin, and Y. Li, "A Systematic Review of the Relationship between Workload and Injury Risk of Professional Male Soccer Players," 2022.
- [3] T. Stølen, K. Chamari, C. Castagna, and U. Wisløff, "Physiology of Soccer," 2012.
- [4] J. Ekstrand, M. Hägglund, K. Kristenson, H. Magnusson, and M. Waldén, "Fewer ligament injuries but no preventive effect on muscle injuries and severe injuries: an 11-year follow-up of the UEFA Champions League injury study," vol. 47, 2013.
- [5] G. Henderson, C. A. Barnes, and M. D. Portas, "Factors associated with increased propensity for hamstring injury in English Premier League soccer players," 2010.
- [6] W. Gregson, V. D. Salvo, M. C. Varley, M. Modonutti, A. Belli, K. Chamari, M. Weston, L. Lolli, and C. Eirale, "Harmful association of sprinting with muscle injury occurrence in professional soccer match-play: A two-season, league wide exploratory investigation from the Qatar Stars League," 2020.
- [7] A. Arnason, S. B. Sigurdsson, A. Gudmundsson, I. Holme, L. Engebretsen, and R. Bahr, "Risk Factors for Injuries in Football," vol. 32, 2004.
- [8] J. Ekstrand, M. Hägglund, and M. Waldén, "Epidemiology of Muscle Injuries in Professional Football (Soccer)," vol. 39, 2011.
- [9] F. Dambroz, F. M. Clemente, and I. Teoldo, "The effect of physical fatigue on the performance of soccer players: A systematic review," 2022.
- [10] J. F. Tornero-Aguilera, J. Jimenez-Morcillo, A. Rubio-Zarapuz, and V. J. Clemente-Suárez, "Central and Peripheral Fatigue in Physical Exercise Explained: A Narrative Review," 2022.
- [11] Z. Akyildiz, Y. Ocak, F. M. Clemente, Y. Birgonul, M. Günay, and H. Nobari, "Monitoring the post-match neuromuscular fatigue of young Turkish football players," 2022.
- [12] S. Llana, B. Burriel, P. Madrero, and J. Fernández, "Is it worth the effort? Understanding and contextualizing physical metrics in soccer," 2022.
- [13] S. Altmann, L. Forcher, L. Ruf, A. Beavan, T. Groß, P. Lussi, A. Woll, and S. Härtel, "Match-related physical performance in professional soccer: Position or player specific?" 2021.
- [14] J. Castellano, A. Blanco-Villaseñor, and D. Alvarez, "Contextual variables and time-motion analysis in soccer," 2011.

6 Referenties

- [15] M. D. Bush, D. T. Archer, R. Hogg, and P. S. Bradley, "Factors Influencing Physical and Technical Variability in the English Premier League," vol. 10, p. 865–872, 2014.
- [16] D. Linke, D. Link, and M. Lames, "Football-specific validity of TRACAB's optical video tracking systems," 2020.
- [17] L. Torres-Ronda, E. Beanland, S. Whitehead, A. Sweeting, and J. Clubb, "Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis," 2022.
- [18] J. Fang, A. Swain, R. Unni, and Y. Zheng, "Decoding Optical Data with Machine Learning," 2020.
- [19] D. Carrilho, M. S. Couceiro, J. Brito, P. Figueiredo, R. J. Lopes, and D. Araújo, "Using Optical Tracking System Data to Measure Team Synergic Behavior: Synchronization of Player-Ball-Goal Angles in a Football Match," 2020.
- [20] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Romasi, "Performance Measures and a Data Set for Multi-target, Multi-camera Tracking," 2016.
- [21] J. Mara, S. Morgan, K. Pumpa, and K. G. Thompson, "The Accuracy and Reliability of a New Optical Player Tracking System for Measuring Displacement of Soccer Players," 2017.
- [22] E. Pons, T. Garcia-Calvo, R. Resta, H. Blanco, R. L. del Campo, J. D. Garcia, and J. J. Pulido, "A comparison of a GPS device and a multi-camera video technology during official soccer matches: Agreement between systems," 2019.
- [23] D. Raabe, H. Biermann, M. Bassek, M. Wohlan, R. Komitova, R. Rein, T. K. Groot, and D. Memmert, "floodlight - a high-level, data-driven sports analytics framework," *Journal of Open Source Software*, vol. 7, no. 76, p. 4588, 2022. [Online]. Available: <https://doi.org/10.21105/joss.04588>
- [24] "Kloppy: standardizing soccer tracking and event data," 2020.
- [25] "Dash user guide & documentation," accessed: May 24, 2024.
- [26] T. Modric, S. versic, I. Jukic, and D. Sekulic, "Physical performance discriminating winning and losing in uefa champions league: a full-season study," 2024.