



Referee! It is getting hot in here, can we get a break?

Mathieu Leng

Dissertation written under the supervision of professor Miguel
Nogueira

Dissertation submitted in partial fulfilment of requirements for the MSc in
Business Analytics, at the Universidade Católica Portuguesa, 28/12/2023.

Abstract

The sports industry is one of the most important industries in the world. In 2023 only, it generated globally more than \$512.14 billion in revenue (Kiko 2023). For Spain more specifically, during the 2021-22 season, professional football generated 8.39 €billion in taxes, contributing to 1.44% of the country's Gross Domestic Product (GDP).

Extreme heat is one of the consequences of climate change and has a big impact on the sports and football industries.

In this study, we address the situation in Spanish male football first division (LaLiga) regarding heat stress exposure for the past two decades and the coming two decades. For both periods, we calculate the Universal Thermal Climate Index (UTCI) and evaluate the heat stress categories, as well as the policies that should be taken. We also evaluate the situation across hours, months, and different locations grouped by similar climatic data by using the K-median algorithm. The two periods are compared to one another to estimate the evolution. To conclude the analysis, the optimal schedule with the least overlapping and the smallest heat stress is calculated.

By using ERA5 and CORDEX data, we observe that LaLiga was exposed 15.5% to strong heat stress conditions and 2.6% to very strong heat stress conditions over the past two decades. In the future, it could increase between 16% and 17.7%, and 2.8% up to 4%, depending on human emission scenarios. It is shown that schedule optimization can significantly reduce heat stress exposure by up to 5%.

Title: Referee! It is getting hot in here, can we get a break?

Author: Mathieu Leng

Keywords: Climate Change, Extreme Heat, Professional Football, Spain, Universal Thermal Climate Index (UTCI), Sports Industry

Sumário

A indústria do desporto é uma das indústrias mais importantes do mundo. Só em 2023, gerou globalmente mais de 512,14 mil milhões de dólares em receitas (Kiko 2023). Mais especificamente em Espanha, durante a época de 2021-22, o futebol profissional gerou 8,39 mil milhões de euros em impostos, contribuindo para 1,44% do Produto Interno Bruto (PIB) do país.

O calor extremo é uma das consequências das alterações climáticas e tem um grande impacto na indústria do desporto e na indústria do futebol.

Neste estudo, abordamos a situação na primeira divisão Espanhola de futebol masculino (LaLiga) no que respeita à exposição ao stress térmico nas duas últimas décadas e nas duas décadas seguintes. Calculamos o Índice Universal de Clima Térmico (UTCI) e avaliamos as categorias de stress térmico, bem como as políticas que devem ser adotadas. Também avaliamos a situação entre horas, meses e diferentes locais agrupados por dados climáticos semelhantes utilizando o algoritmo K-median. Os dois períodos são comparados entre si para estimar a evolução.

Utilizando os dados ERA5 e CORDEX, verificamos que a LaLiga esteve exposta a 15,5% a condições de stress térmico forte e 2,6% a condições de stress térmico muito forte nas últimas duas décadas. No futuro esta exposição poderá aumentar entre 16% e 17,7%, e 2,8% até 4%, dependendo do cenário de evolução das emissões antropogénicas. É demonstrado aqui que uma optimização do calendário dos jogos pode reduzir significativamente a exposição ao stress térmico até 5%.

Título: Árbitro! Está a ficar calor aqui, podemos fazer uma pausa?

Autor: Mathieu Leng

Palavras-Chave: Alterações Climáticas, Calor Extremo, Futebol Profissional, Espanha, Índice Universal de Clima Térmico (UTCI), Indústria desportiva

Acknowledgments

First of all, I would like to thank Professor Miguel Nogueira for his help and supervision throughout the semester. Professor Nogueira has been a constant help and has greatly contributed to the quality of this work, by introducing me to a field I did not know very well and allowing me to feel comfortable in it. I'm convinced that without his help, it would have been impossible for me to turn in work like this.

I would also like to thank my family. First and foremost my parents, who have been of immense support, not only financially but above all psychologically, emotionally and pedagogically, not only during this thesis semester but throughout my entire schooling. I would also like to thank my brother who has always been an example to me and who has enabled me to become the man I am today.

Finally, I'd like to thank Catolica Lisbon University for allowing me to take up this challenge under extraordinary conditions, and all the teachers who have helped me to progress in this field that I appreciate so much.

Contents

- 1 Introduction** **8**

- 2 Literature Review** **10**
 - 2.1 Climate Change and its Impact 10
 - 2.2 Climate Change and Sports Industry 10
 - 2.3 Business Analytics and Data Science for Climate Risk Management 13
 - 2.4 Adaptive Strategies in Sports Industry for Climate Change 15

- 3 Methodology** **16**
 - 3.1 Research Design 16
 - 3.2 Data Collection 17
 - 3.3 Data Analysis 20
 - 3.3.1 Descriptive Analysis 20
 - 3.3.2 Predictive Analysis 21
 - 3.3.3 Prescriptive Analysis 21

- 4 Results and Discussion** **24**
 - 4.1 Results 24
 - 4.1.1 Historical Analysis 24
 - 4.1.2 Future Analysis 28
 - 4.1.3 Comparison between the 2 periods 32
 - 4.1.4 Optimisation Problem 33
 - 4.2 Discussion 34
 - 4.2.1 Interpretation of the Research Findings 35
 - 4.2.2 Correspondence with Literature Review 37
 - 4.2.3 Appropriateness of the Methodology and Data Collection 37
 - 4.2.4 Effect on Professional and Academic Practice 38
 - 4.2.5 Generalization 39

- 5 Limitations** **39**

- 6 Conclusions** **39**
 - 6.1 Main Findings and Conclusions 39
 - 6.2 Future Research 42

- Appendix A Appendix** **50**

- Appendix B Code** **55**

List of Figures

- 1 UTCI assessment scale of thermal stress, corresponding physiological responses and possible protection measure (Di Napoli, Pappenberger, and Cloke 2019) 15
- 2 Histogram of UTCI values of every location at every timeslot for the 2002-2022 period 18
- 3 UTCI comparisons with other thermal indexes 20
- 4 Box plots of UTCI values for the 2002-2022 period, colored by optimal clustering 25
- 5 Stress Categories and Policies percentages for the 2002-2022 period 26
- 6 Stress Categories by clusters 27
- 7 Stress Categories by hours and months 28
- 8 Bias correction on UTCI values for the 2024-2043 years 29
- 9 Percentages of the Heat Stress Categories for the Different Scenarios 30
- 10 Percentages of Heat Stress Categories across Hours under 3 different scenarios 31
- 11 UTCI distribution comparison between the past and the 3 future scenarios 33
- 12 Number of Games by Location 50
- 13 Silhouette Analysis 51
- 14 Percentages of Heat Stress Categories across Optimal Clusters under 3 different scenarios 53
- 15 Percentages of Heat Stress Categories across Months under 3 different scenarios 54

List of Tables

1	Descriptive Statistics 2002-2022	25
2	Descriptive Statistics for 2024-2043	29
3	Thermal Stress Conditions Past vs Future	33
4	Dataset overview	50
5	Number of Games by Hour	51
6	Standardized Residuals	52

1 Introduction

The sports industry is one of the largest industries in the world. In 2023 only, the global sports industry generated more than \$512.14 billion in revenue (Kiko 2023), and this is expected to continue to grow. As a comparison, the airline industry generated \$896 billion in 2023 (IATA 2023). By generating \$50 billion a year globally, football (soccer) represents the most profitable sport (Kiko 2023). Regarding Spain, more specifically, professional football generated €8.39 billion in taxes during the 2021-22 season, representing 1.44% of the country's Gross Domestic Product (GDP) (Corkhill 2023).

Some of the most memorable scenes from the 2022 World Cup held in Qatar were the massive cooling systems to refresh supporters, staff, and players in open-roofed stadiums. Those scenes, yet surprising at first sight, might become more usual as time passes. Climate change is considered one of the biggest challenges that Humanity will face during the coming century. It has severe consequences on different industries, such as agriculture and tourism, but also on the sports industry. Sports organizations face multiple challenges due to climate change, such as lowered revenue (Kay and Vamplew 2006), damage and destruction of facilities (Elsasser and Bürki 2002), event delays and cancellations (Fairley, Ruhanen, and Lovegrove 2015, Filo, Cuskelly, and Wicker 2015), and an overall decline in interest in a sport (Dawson, Scott, and Havitz 2013). Those challenges can lead to dramatic economic consequences for sports organizations. As a result, it is utterly important for organizers to be able to proactively assess and manage the risks of climate change for their organizations (Orr and Inoue 2019). In addition, for many sports organizations, adaptation to climate change is complex and hard to implement. It is thus very important to understand, as soon as possible, the vulnerability and take the needed actions before it is too late.

Until very recently, the relationship between the sports industry and climate change was not studied as a bidirectional relationship but rather as the impact of sports on climate change. Nonetheless, as recent huge sports events, such as the latest Summer Olympics and Paralympics in Tokyo and the 2022 FIFA World Cup in Qatar, took place in locations where extreme heat conditions are common, the bidirectional relationship between organized competitive sports and climate change has been gaining popularity recently (Gerrett et al. 2019, Sofotasiou, Hughes, and Calautit 2015).

Climate change poses a wide range of risks to the sports industry. These risks can be divided into two main categories: direct and indirect health-related effects. Direct health-related effects, which represent the focus of this research, are primarily caused by extreme temperatures and other weather conditions, such as heatwaves, extreme weather,

and ultraviolet radiation (Schneider 2021). Indirect consequences result from climate-induced changes to our ecosystem, such as air pollutants, allergens, viruses, bacteria, and the associated vectors and natural reservoirs (Schneider 2021). The consequences of these effects on the sports industry can vary depending on the specific conditions and, also, from one human being to another, but they include heat illness, heatstroke, heat exhaustion, and hyperthermia (Parsons 2014, Grantham et al. 2010, Brotherhood 2008, Gerrett et al. 2019, A. J. Grundstein et al. 2021, Honjo et al. 2018, Kerr et al. 2019).

The present work will investigate the situation in Spain, which is located in the Mediterranean Region, that is considered a climate change hotspot due to its vulnerability to global warming impacts, namely the occurrence of heat-related extreme events, such as drought, heat waves, and fires (Diffenbaugh and Giorgi 2012). In addition, the importance of the professional football industry for the Spanish economy increased our attention. As a result, the scope of this research is to understand the impact of extreme heat on the male professional football industry in Spain, focusing on the first division (henceforth denoted LaLiga). This problem statement substantiates itself in the following 3 research questions:

- RQ1: What is the current impact of extreme heat on LaLiga?
- RQ2: What will be the impact of extreme heat on LaLiga for the next two decades and different climate change evolution scenarios?
- RQ3: Is it possible to adapt LaLiga to increasing extreme heat stress conditions by optimizing the match schedule?

To answer these 3 research questions, we gathered data from state-of-the-art high-resolution climate models and observations, providing both historical climate data for the past two decades and projections for the upcoming two decades under different greenhouse gas emissions scenarios. This climate data included a widely used human thermal heat stress index, the Universal Thermal Climate Index (UTCI) for the past, but also for the projected future under different anthropogenic greenhouse gas emission scenarios. Our analysis is restricted to the boreal months of June, August, and September, corresponding to the period where extreme heat conditions occur in Spain, and its evolution for the coming two decades. Notice that the month of July is absent from the analysis as it corresponds to the inter-season break and no games are played. After some data preprocessing, we assessed the situation in the past by using UTCI stress categories based on UTCI values and computed UTCI values for the future based on the state-of-the-art climate model output data. We corrected the bias that exists in the climate model projections by employing Quantile Delta Mapping (QDM) between the two past data, a widely used bias correction algorithm in climate applications which preserves relative changes in

data. Following, we assessed the situation in the future by using UTCI stress categories. Finally, we used Linear Programming to suggest the optimal schedule in the future.

The present manuscript is divided into 5 sections and has the following structure: Section 2 reviews the key literature; Section 3 presents the methods and datasets used in the study; Section 4 introduces the results and their interpretation; Section 5 highlights the limitations that may have arisen from this research, and Section 6 serves as a conclusion and proposes future research.

To the researchers' knowledge, as the subject of the relationship between climate change and the sports industry in both ways is still in its pioneering phase, there are no previous quantitative studies on the impact of extreme heat on the professional male football industry, neither in Spain nor somewhere else.

2 Literature Review

2.1 Climate Change and its Impact

Climate change, its urgency, increasing severity, and associated impacts are well established now (P. C. Change 2018, I. C. Change 2014), and by having consequences on the environment, society, economy, and people's health, it represents the greatest threat for Humanity for the upcoming century (Eichinger 2019, Townsend et al. 2003). Some effects were already measured, such as on the global average surface temperature that increased by $0.6 \pm 0.2^\circ\text{C}$ over the 20th century, and this trend is expected to strengthen as a further increase between 1.4°C and 5.8°C , depending on the scenario, is projected by 2100 (Houghton et al. 2001, I. P. O. C. Change 2007). Spain and more specifically, the southern of Spain has already been particularly affected by climate change, witnessing increased temperature and reduced rainfall (García-Mozo, Mestre, Galán, et al. 2010), and according to the most probable scenarios, these consequences of climate change will continue (Meehl and Tebaldi 2004, Tebaldi et al. 2006, I. P. O. C. Change 2007), with even more intense extreme heat events expected in the Iberian Peninsula (Adopted 2014, Fischer and Schär 2009, King and Karoly 2017).

2.2 Climate Change and Sports Industry

Sports can be defined as the combination of 3 components: competition, physical activity, and structures underpinned by rules and laws (Guttman 2004). The sports industry is massive and generated in 2023 only, more than \$512.14 billion in revenue (Kiko 2023). Sports involve billions of people worldwide, through both participants and audiences (T. Miller et al. 2001), and impact whole economies (Flyvbjerg and A. Stewart 2012), thanks

to the contribution to the GDP, to the media rights revenue, and sales of sporting goods ranging from dozens to hundreds of billions of dollars per year (Andreff and Szymanski 2006).

Organized sports (e.g., sports competitions and organizations that operate them) show degrees of vulnerability at the organizational, team, athletes, and facility levels, and it is a recent understanding that the conditions of the natural environment surrounding sports facilities are not granted (Orr and Inoue 2019), and that possible changes in those conditions will impact the natural environment (McCullough, Orr, and Kellison 2020). The potential consequences of climate change for sports organizations are multiple and include lowered revenues (Kay and Vamplew 2006), damage and destruction of facilities (Elsasser and Bürki 2002), event delays and cancellations (Fairley, Ruhanen, and Lovegrove 2015, Filo, Cuskelly, and Wicker 2015), and an overall decline in interest in a sport (Dawson, Scott, and Havitz 2013). Some examples exist in Spain, where LaLiga organizers already had to take the harsh decision to reschedule games due to intense heat in recent years, such as 2016, 2018, 2020, and 2023.

The relationship between sports and the natural environment is bidirectional. On the one hand, sports have an impact on the natural environment via waste production for example (Chard and Mallen 2012, Trendafilova, Babiak, and Heinze 2013). On the other hand, the natural environment provides natural resources such as field of play and weather conditions (Kay and Vamplew 2006). As a result, it impacts sports (Orr and Inoue 2019, McCullough, Orr, and Kellison 2020). Research on organized competitive sport and climate change was until 2021 in a 'pioneering phase', with hints that it is now beginning a 'growth phase' (Orr, Inoue, et al. 2022), but research of the impact of the sports industry on climate changes has been far more researched than the reverse (Orr and Inoue 2019, Dingle and B. Stewart 2020). Yet, the topic of the impact of climate change on organized sports competitions has gained popularity recently, as the latest summer Olympics and Paralympics and FIFA World Cup took place in locations, Tokyo and Qatar respectively, regularly experiencing extreme heat conditions (Gerrett et al. 2019, Sofotasiou, Hughes, and Calautit 2015), which led researchers to be concerned by the potential risks of thermal stress among athletes and spectators (Honjo et al. 2018, Kakamu et al. 2017, Olya 2019, Tipton et al. 2019, Vanos et al. 2019).

Sport-specific risks jeopardized by climate change can be divided into two categories: direct health-related and indirect health-related effects (Eis et al. 2011, Mücke and Matzarakis 2020). Our focus is on direct consequences, which are primarily caused by extreme temperature and other weather conditions, such as heatwaves, extreme weather, and ultraviolet radiation (Mücke and Matzarakis 2020, Schneider 2021). In general, increased outdoor

temperature causes increased stress on the cardiovascular, respiratory, and metabolic systems (Mücke and Matzarakis 2020). The core body of humans is mostly close to 37°C (Lemke and Kjellstrom 2012) and an increase in temperature is acceptable up to 39°C, after which it can lead to health risks, such as heat illness and heatstroke (Parsons 2014, Grantham et al. 2010, Brotherhood 2008), but also heat exhaustion, and hyperthermia (Gerrett et al. 2019, A. J. Grundstein et al. 2021, Honjo et al. 2018, Kerr et al. 2019). There are 4 ways to transfer heat to/from the body: by convection, conduction, radiation, and evaporation of water (sweat) (Lemke and Kjellstrom 2012). Those can be affected by the characteristics of the surrounding environment, such as air temperature, wind speed, humidity, and heat radiation sources, which ultimately determine the heat stress (Parsons 2014). When doing sports in an environment having both high temperature and high humidity, the usual mechanisms allowing human beings to limit the rise in core temperature by more than 2 or 3°C during exercise are less efficient (R. Maughan et al. 2010, Brotherhood 2008, Hanna et al. 2011), exposing even more people to heat stress and endangering their health (Brotherhood 2008, Kappas 2009, Townsend et al. 2003). Athletes may be at higher risk of experiencing heat-related illnesses compared to spectators or the general population due to the physical exertion involved in sport and prolonged exposure (Brocherie, Girard, and Millet 2015, Olya 2019), and they are even considered as a risk group regarding climate change-related health risks (Hanna et al. 2011, Mücke and Matzarakis 2020) even if well-trained athletes can tolerate increases in core body temperature that is about 0.9°C higher than untrained individuals (Hanna et al. 2011), Europa 2019). Nonetheless, also amateurs, spectators, and officials will be directly confronted with the implications of climate change on our health shortly (Schneider 2021).

Regarding football more specifically, while it is mostly considered a low-intensity sport, because for >70% of the game athletes perform low-intensity activities, it has a high total demand of energy, as suggested by measurements of heart rate and body temperature (Bangsbo, Mohr, and Krstrup 2006). This high energy demand mainly comes from the repeated high-intensity efforts performed by the players. Moreover, football is also considered an endurance game as top-class players perform on average 150-250 brief intense actions during a game, and male outfield players usually cover about 10-13km (Bangsbo, Mohr, and Krstrup 2006). When looking at other endurance sports, footballers have all the reasons to be concerned as many examples of serious, although rarely fatal, heat illness during endurance events held in hot weather are known (R. Maughan et al. 2010), as well as examples from American football where heat-related fatalities are less uncommon (Bergeron, McKeag, et al. 2005).

2.3 Business Analytics and Data Science for Climate Risk Management

Due to the escalation of climate change, which is posing a threat to sports competitions (Orr 2020, Scott, Steiger, Rutty, and P. Johnson 2015, Scott, Steiger, Rutty, and Fang 2019), it is essential to assess as soon as possible the potential risks facing host cities, in order to minimize these risks through technological adaptations or other new facilities or policy options (Orr, Inoue, et al. 2022). To evaluate those risks, business analytics and data science offer researchers tools. The first thing required is a way to evaluate the situation. Amongst the different consequences of climate change, our focus here is on extreme heat. To have a quantitative idea of the impact of heat, it is necessary to have a heat stress index and create links between this index and the work-rest cycles needed to prevent any heat stress effects (Lemke and Kjellstrom 2012). The most accurate option would be to use specific methods based on individual measurements to determine heat strain in individuals, but we will use another index based on heat stress. There are more than 60 heat stress indices that have been suggested to evaluate high-temperature environments and predict the potential for heat damage, all of them having their benefits and drawbacks (Burton, Ebi, and McGregor 2009).

The Wet Bulb Globe Temperature (WBGT) could represent a good thermal index as it is the most commonly used index of heat stress (Zare et al. 2018). It was proposed more than 50 years ago, during the 1950s, and was used by the US Army and Machine Corps to reduce heat-related illnesses in the training camps (Budd 2008, Yaglou, Minaed, et al. 1957). This measure of environmental heat can be calculated based on meteorological data, with only 4 parameters being needed: air temperature 2 meters above the ground, relative humidity, wind speed, and radiation heat (Zare et al. 2018). Moreover, it has been recommended by the International Organization for Standardization (ISO) certification (Parsons 2006), American College of Sports Medicine (Sawka et al. 2007, Armstrong et al. 2007), International Olympic Committee (Bergeron, Bahr, et al. 2012), by the Federation Internationale de Football Association (FIFA)(Mountjoy et al. 2012) and has all the correct components to indicate physiological heat stress (Parsons 2014). Nonetheless, this index has been challenged. Firstly, the WBGT is often not used in the way it has been intended. As recommended by ISO and the American Conference of Governmental Industrial Hygienists, the WBGT index should only be used as a preliminary tool, while another index, the Predicted Heat Strain (PHS) (ISO 2004), should be used in more extreme conditions (Brocherie and Millet 2015). Moreover, the WBGT is accused of not reflecting correctly the severity of the weather/climate (Brocherie and Millet 2015) and of being too stringent for determining conditions when full workload should be reduced (P. K. Nag, A. Nag, and Ashtekar 2007, Brake and Bates 2002), even if it has been corrected by ISO (Lemke and Kjellstrom 2012). Because of those points, the validity,

accuracy, and applicability of the WBGT are questioned, leading to a necessary move to other metrics for a more adequate screening of heat stress during exercise (Alfano et al. 2014, Brocherie, Girard, Pezzoli, et al. 2014, Brocherie, Girard, and Millet 2015).

The Universal Thermal Climate Index (UTCI) was developed to have a standard criterion for assessing heat stress in the light of human meteorology (Baejczyk 2010). It is an index based on an advanced multi-node model of human heat transfer and thermoregulation (Fiala, Lomas, and Stohrer 2001, Fiala, Havenith, et al. 2012) and represents the equivalent temperature for the environment derived from a reference environment. This index is calculated based on 4 parameters, which are air temperature 2 meters above the ground, mean radiation temperature, relative humidity, and wind speed. The value of UTCI can be converted into 10 different heat stress categories based on its value (Young 2017). The different stress categories, the expected physiological impact, and the recommended protection measures can be seen in Figure 1. Even if UTCI is regarded as one of the most comprehensive indices for calculating heat stress in outdoor spaces (Blazejczyk 1994), it received some criticism as it does not take into account differences in metabolic rates during work, the impact of specialized protective clothing, or the constant change in position and movement during real work situations (Lemke and Kjellstrom 2012).

In order to calculate UTCI values, we need data from two different sources which are ERA5 and CORDEX data. CORDEX is a World Climate Research Programme (WCRP) project (Giorgi, Jones, Asrar, et al. 2009, Giorgi and Gutowski Jr 2015) which provides climate information covering the main continental areas of the world at a regional scale by means of coordinated Regional Climate Downscaling (RCD) techniques (Hewitson and Crane 1996). On the other hand, ERA5 is a combination of observational data and weather model, which fills the gaps existing in observation data. It is the 5th and the latest generation of ECMWF. CORDEX data are available for the 3 RCP scenarios, namely RCP2.6, RCP4.5, and RCP8.5. The 3 RCP are 3 different scenarios depicting 3 different concentrations of the greenhouse gases. RCP2.6, with a radiative forcing of $2.6 \text{ W}\cdot\text{m}^{-2}$, is the least pessimistic and leads to an average global warming of about 1°C by 2100, while RCP 8.5 equals 4°C by the same time horizon (Chou et al. 2014). Quantile Delta Mapping (QDM), which is a bias correction algorithm used specifically to preserve relative changes in data (Cannon, Sobie, and Murdock 2015), will be used to correct bias on CORDEX data. Finally, K-median optimal clustering will be used to group locations sharing similar climate characteristics. It is a derived version of K-means clustering where we use the median as a central point instead of the mean. K-means clustering is a data exploration technique used to group objects sharing similar characteristics together (Pham, Dimov, and Nguyen 2005) based on unsupervised learning. We use this method as K-means cluster analysis has already been used previously to group locations according to their climatic

characteristics (Carvalho et al. 2016), but we prefer the median as it is less sensitive to outliers. To find the best K, we will use silhouette analysis, an unsupervised method for scoring the performance of a clustering method (Shutaywi and Kachouie 2021).

UTCI range [°C]	Stress category	Physiological responses	Protection measures
Above +46	Extreme heat stress	Increase in rectal temperature time gradient. Steep decrease in total net heat loss. Averaged sweat rate $>650 \text{ g h}^{-1}$, steep increase.	Temporary body cooling and drinking $>0.5 \text{ L h}^{-1}$ necessary. No physical activity.
+38 to +46	Very strong heat stress	Low core–skin temperature gradient. Increase in rectal temperature at 30 min.	Temporary use of air condition. Shaded places necessary. Drinking $>0.5 \text{ L h}^{-1}$. Reduce physical activity.
+32 to +38	Strong heat stress	Averaged sweat rate $>200 \text{ g h}^{-1}$. Increase in rectal temperature at 120 min. Instantaneous change in skin temperature.	Use shaded places. Drinking $>0.25 \text{ L h}^{-1}$. Temporary reduce physical activity.
+26 to +32	Moderate heat stress	Change of slopes in sweat rate and rectal and skin (mean, face, hand) temperature. Occurrence of sweating at 30 min. Steep increase in skin wettedness.	Drinking $>0.25 \text{ L h}^{-1}$.
+9 to +26	No thermal stress	Averaged sweat rate $>100 \text{ g h}^{-1}$. Plateau in rectal temperature time gradient.	Physiological thermoregulation sufficient to keep thermal comfort
0 to +9	Slight cold stress	Local minimum of hand skin temperature.	Use gloves and hat.
–13 to 0	Moderate cold stress	Vasoconstriction. Face skin temperature at 30 min $<15^\circ\text{C}$ (pain).	Intensify activity and protect face and extremities against cooling.
–13 to –27	Strong cold stress	Numbness. Increase in core–skin temperature gradient.	Intensify activity and protect face and extremities against cooling. Use warmer clothing.
–27 to –40	Very strong cold stress	Frostbite, numbness, shivering. Steeper decrease in rectal temperature.	Intensify activity and protect face and extremities against cooling. Use warmer clothing. Reduce outdoor exposure time.
Below –40	Extreme cold stress	Frostbite. Decrease in rectal temperature time gradient.	Stay at home. If outdoor exposure is necessary, use heavy and wind protected clothing.

Figure 1: UTCI assessment scale of thermal stress, corresponding physiological responses and possible protection measure (Di Napoli, Pappenberger, and Cloke 2019)

2.4 Adaptive Strategies in Sports Industry for Climate Change

Since the sports industry possesses unique characteristics (Gyri and Balogh 2017), e.g central roles of managing competition (Trail, Anderson, and Fink 2000) and spectatorship (Funk and Jeff James 2001, JD James, Walker, and Kuminka 2009), sport management needs specific framework and strategies for assessing and managing climate change impact (Orr and Inoue 2019). Heat policies are a way to manage the degree of heat-related illness amongst athletes and spectators (A. Grundstein et al. 2013, A. J. Grundstein et al. 2021). Heat policies enable sports organizations and sports people to respond rapidly when temperatures soar to ensure the health and well-being of athletes and spectators (A. Grundstein et al. 2013). This adaptive measure often uses a temperature threshold that triggers a predetermined action, which can include taking a water break, changing equipment (e.g., decreasing the number of layers worn), or allowing athletes to use cooling agents (such as cooling towels and mist showers) (Kakamu et al. 2017, R. Maughan et al.

2010). Other measures are also recommended in the event of prolonged thermal stress, such as doing sports activities during morning or evening hours and regular intake of fluid before, during, and after activity (Wonisch and Ledl-Kurkowski 2017, Europa 2019), or acclimatization (R. Maughan et al. 2010). However, regarding acclimatization, the process usually takes longer than the thermal heat exposure (Schneider 2021), so it is advised to use appropriate indices and analyses to foresee issues and to have other solutions, such as the one mentioned before (Nybo et al. 2021). Concerning football, FIFA’s position on extreme heat is quite clear: the WBGT index is employed to assess thermal stress, and if the WBGT value exceeds 32°C, cooling breaks are mandatory in both halves of the match, while the decision on whether to suspend or cancel the match is up to competition organizers (FIFPRO n.d.). Nonetheless, FIFpro (the International Federation of Professional Footballers) does not agree with it and released an 11-point plan on what should be done to assess extreme heat, mentioning that cooling breaks should be mandatory when the WBGT is above 26°C, and that when WBGT is above 28°C game should be rescheduled, and mid-day and afternoon matches should be avoided (FIFPRO 2023).

Regarding Spain, more specifically, LaLiga is known to be aware of the risks of holding football fixtures during the hot months of summer and recognizes that it puts both spectators and players at risk (Vicente S. 2023). In 2020, because of COVID-19, the season was rescheduled, and while the season was expected to stop in May, it was extended and resumed on the 8th of June and finished on the 19th of July. As games were played during a hotter period than usual, LaLiga decided to add weather conditions alongside all of the programmed games, such as temperature forecasts and historical highs in all 20 locations (Vicente S. 2023), in order to predict and avoid playing in extreme heat conditions. Moreover, in recent years, Spanish football fans have been used to see fixtures being rescheduled later the same day due to unexpected extremely hot conditions.

3 Methodology

In this chapter, the goal is to present and explain the methodology used to study the subject at hand and to answer the research questions formulated in section 1. This section will be divided into 3 subsections namely: Research Design, Data Collection, and Data Analysis.

3.1 Research Design

The methodology used to achieve this goal and answer the research questions formulated in the first Section is a quantitative methodology. Indeed, numerical data from several sources were merged and analyzed in order to perform a thorough analysis.

3.2 Data Collection

First of all, data about the different stadiums were collected. The stadiums selected were the ones of the teams playing in the first division of male Football in Spain (LaLiga) during the 2022-23 season. The data collected about them included the name of the team, name of the stadium, latitude, longitude, and capacity of the stadium. Since no available datasets combined that information, they were created by the researchers using available material from the Internet (Wikipedia and Google Maps). Some cleaning and data manipulation were needed, especially on the latitude and longitude features, where they have been rounded up to the closest fourth to facilitate future merging.

Secondly, data about LaLiga fixtures during the 2022-23 season were collected in order to have information about the time and the location of the games played. As mentioned in Section 1, only the games played during summer were selected, so going from June until September. This narrows the 380 games played during the whole season to 71. In order to further analyze the impact of time, for every day there was a game at a certain location, extra data points were added at 12h, 15h, 18h, and 21h. Those hours were chosen to have knowledge about the situation during the whole afternoon, and because they cover the different timeslots for games during the summer. To avoid confusing true games with false ones, one additional binary column called 'True_Game' has been added, having 1 for the true ones. Besides, since the goal of this research is to have an idea of the change between the past two decades and the future two decades, we decided to extend the LaLiga schedule of the 2022-23 season for this whole period, namely from 2002 until 2043. So, we considered every season between 2002 and 2043 to have the exact same schedule as the 2022-23 season. This trick is required for simplicity, necessity, and comparability. For simplicity, because for previous seasons we could have used previous LaLiga schedules, but since teams change every year because of relegation and promotion, this would have added complexity. For necessity, because both the schedule and the teams for the future LaLiga schedule are unknown, as a result, it makes sense to copy the LaLiga schedule of the 2022-23 season and use it for the seasons from 2002 to 2043. Finally, for comparability as it allows us to study the impact of climate change, since the days and locations are kept constant. Those two datasets, one for each period, are then merged with the previous one in order to have the longitude and latitude for each game.

Furthermore, historical data regarding UTCI was gathered. The data came from the ERA5-HEAT (Human thErmAl comforT) dataset, which is produced by the European Centre for Medium-range Weather Forecasts (ECMWF) and represents the current state-of-the-art for bioclimatology data record production by using ERA5 reanalysis. This dataset is available on Copernicus.eu website. The data are gridded, with a horizon-

tal coverage of $0.25^\circ \times 0.25^\circ$ (corresponding at the equator to approximately 27.75 km x 27.75km), a vertical coverage at surface level, and an hourly temporal resolution. In this case, the hourly temporal resolution is very important since we want to have an idea of the UTCI during a football match, which only lasts for nearly 2 hours. We merged those data with the dataset of the historical period (2002-2022) in order to have UTCI values only for the locations and timeslots that we are interested in. The plot of the distribution of UTCI values for every location at every timeslot can be seen in Figure 2.

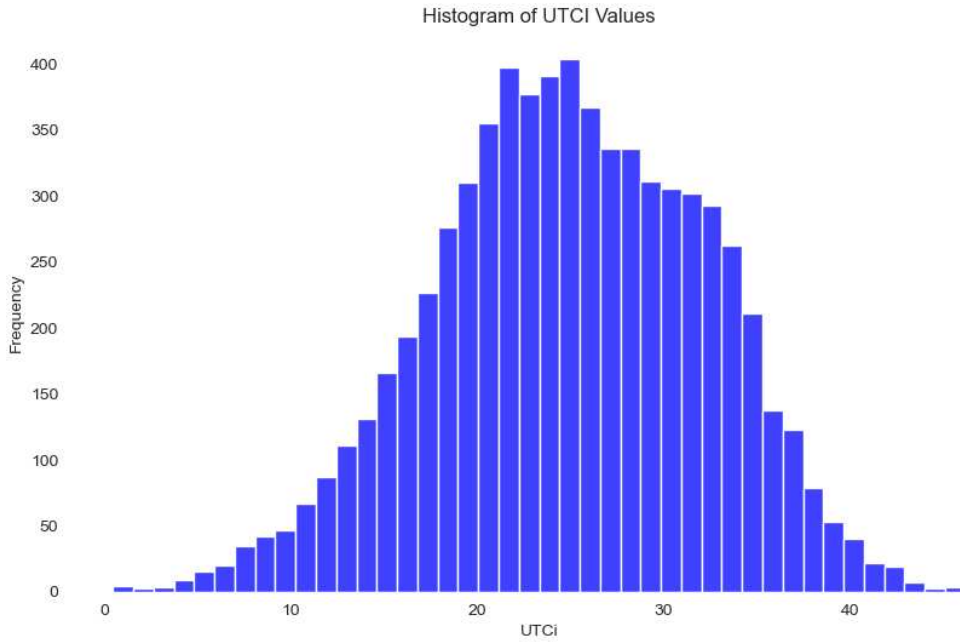


Figure 2: Histogram of UTCI values of every location at every timeslot for the 2002-2022 period

Since ERA5 is an observational product, it is not available in the future. Hence, future climate data was obtained from CORDEX regional climate models, a set of high-resolution climate simulations generated with state-of-the-art numerical climate models, providing a plausible future realization of the future climate under different future human emission scenarios (Kotlarski et al. 2014). This data was downloaded for the period 2002-2022 for a certain model in order to evaluate any bias in the model and correct it to be able to use this model for the projections for 2024-2043. The model used is for Europe, CNRM-CERFACS-CM5(France) for the Global Climate Model (GCM), KNMI-RACMO22E(Netherlands) for the Regional Climate Model (RCM), and r1i1p1 for the ensemble member. The data are available with a horizontal resolution of $0.11^\circ \times 0.11^\circ$ (approximately 12km x 12km over Europe) and a temporal resolution of 3 hours. Once again, the 3-hour temporal resolution is important as we want to assess the situation

during football games, which last approximately 2 hours. The variables downloaded are 2m air temperature, 2m relative humidity, 10m wind speed, surface solar radiation downwards, surface thermal radiation downward, and surface upwelling shortwave radiation. For data from 2002 until 2005, historical data were downloaded. Those data come from experiments in the past using GCM boundary conditions. For the data after this period, RCP 4.5 was used, as it is the stabilization scenario. Because the downloaded data concerned the whole of Europe for this whole period, we narrowed it down to the closest timestamp and location of the fake and real fixtures. Using those data we calculated UTCI values for the historical period. Moreover, using the same model as mentioned previously, CORDEX data were used for the period 2024-2043. We used projected data according to the 3 RCP scenarios, namely RCP 2.6, RCP 4.5, and RCP 8.5. The same variables were downloaded, and the UTCI was calculated the same way. The bias of the model was calculated by comparing the CORDEX and ERA5-HEAT data for the historical period, and the same correction was then applied using QDM on the UTCI values for the future period, as a bias might exist due to the fact that the data were generated by a model.

Last but not least, we created several extra features for our datasets. For both datasets, we use K-median clustering to group locations sharing similar characteristics, namely their median and their standard deviation of UTCI values. This is done in order to find locations that are similarly affected and to address the future evolution in those locations. This column is called "Optimal_Cluster". In order to use K-median clustering, you need to provide K (number of clusters) beforehand. To find the best K, we use the silhouette analysis. Moreover, another feature is created for each column containing UTCI values (1 in the dataset for the past, 3 for the dataset for the future), which is the 'Stress Category' and consists of the translation between UTCI values and the corresponding stress category. The translation can be seen in Figure 1. Finally, another translation layer is added by translating UTCI values into FIFA's and FIFpro's policies. In order to do so, we first find the equivalent values between UTCI and WBGT by using the results from Figure 3 and then translate UTCI values by using those threshold values.

Thermal perception	Indices					
	UTCI	WBGT	SET	PMV	PET	
Very cold ¹ (Extreme cold stress ^{1,2})	< -40			-3	<4	
(very strong cold stress ²)	-40 to -27					
Cold ¹ (Strong cold stress ^{1,2})	-27 to -13			-2.5	4-8	
Cool ^{1,3} (Moderate cold stress ^{1,2} / Moderate Hazard ³)	-13 to 0		<17	-1.5	8-13	
Slightly cool ¹ (Slight cold stress ^{1,2})	0 to +9			-0.5	13-18	
Comfortable ^{1,3} (No thermal stress ^{1,2} / No Danger ^{3,4})	+9 to +26	<18	17-30	0	18-23	
Slightly warm ¹ (Slight heat stress ¹)				0.5	23-29	
Warm ^{1,3,4} (Moderate heat stress ^{1,2} / Caution ^{3,4})	+26 to +32	18-23	30-34	1.5	29-35	
Hot ^{1,3,4} (Strong heat stress ^{1,2} / Extreme caution ^{3,4})	+32 to +38	23-28	34-37	2.5	35-41	
(very strong heat stress ²)	+38 to +46					
Very hot ^{1,3,4} (Extreme heat stress ^{1,2} / Danger ^{3,4})	> +46	28-30	>37	3	>41	
Sweltering ⁴ (extreme danger ⁴)		≥30				

¹ PET and PMV ²UTCI ³SET ⁴WBGT

Figure 3: UTCI comparisons with other thermal indexes

3.3 Data Analysis

The analysis is divided into three parts: what happened for the past 2 decades, what we can expect in the future, depending on different future anthropogenic emission scenarios, and the optimization problem. The goal is to have a clear idea of the current situation of the football games played in Spain and, furthermore, to see how it might evolve in the future, and finally, to evaluate the optimal schedule. It is expected that keeping the exact same schedule would result in a higher percentage of games played in heat-stress conditions, which might ultimately lead to using different policies, such as rescheduling.

In order to have a clear view, UTCI Stress categories will be used to assess the type of conditions according to the UTCI scale in Figure 1. The data will be used to have an idea of the locations more at risk, as well as the hours and months. The different policies according to FIFA and FIFPro will also be used.

The analysis will be divided into 3 parts: Descriptive Analysis, Predictive Analysis, and Prescriptive Analysis.

3.3.1 Descriptive Analysis

The goal of this section is to answer the question "What happened?". In this section, we will use descriptive statistics and visualization in order to have an idea of the range of UTCI values, the percentage of conditions in each stress category both at game time and not, the percentage of the different policies both at game time and not, the differences in UTCI values between hours and months, and the difference between the different clusters.

As we want to evaluate the climate, it needs to be done over a rather long period, two decades in our case, and that is why we use percentages over the whole period. It is also important to have an idea of the current schedule of LaLiga games regarding the effort made by Liga Nacional de Fútbol Profesional (LNFP), the LaLiga organizers, to avoid playing in heated conditions.

3.3.2 Predictive Analysis

The goal of this section is to answer the question "What will happen?". This part will be focused on using projections for different scenarios (RCP 2.6, RCP 4.5, and RCP 8.5) to analyze the situation in the different locations and compare the months and hours of games. These three future greenhouse gas emission scenarios represent respectively a low-emission Paris-aligned future, a middle-of-the-road scenario, and a business-as-usual high-emission scenario without relevant efforts for emission reduction. The predictive analysis includes visualization, using bar chart plots, tables, and statistical hypothesis tests. We will first analyze the future projections and compute the anomalies comparing the future with the past 2 decades in order to evaluate the evolution.

Since we will compare the different scenarios and periods based on percentages of each stress category, we will use a chi-squared test to assess differences in distribution. The chi-squared test is a statistical hypothesis test. It can be used to check whether two categorical variables have the same distribution. Indeed, we do not care if the mean or the median increased or decreased, but if the distribution changed, having more high UTCI values resulting in more heat stress conditions. This will ensure that the differences observed between heat stress categories are statistically significant. The differences in distribution will first be evaluated between the three scenarios and then between each scenario compared with the past period. There are four assumptions for the chi-squared test: the two variables must be categorical, all observations have to be independent, cells in the contingency table are mutually exclusive, and the expected value of cells should be 5 or greater in at least 80% of cells. They have all been checked, and the first 3 have been confirmed. Regarding the last one, it was not met when comparing the 3 scenarios together and RCP8.5 with the past. Nonetheless, we will still use this test as it allows us to verify that the distributions are different between at least two future scenarios and the past, and because we are not interested in the exact difference between the past and one scenario but the trend.

3.3.3 Prescriptive Analysis

The goal of this section is to answer the question "What should I do?". This section is decision-focused, based on an optimization problem. The goal of the optimization prob-

lem is to minimize the number of games played under moderate, strong, very strong, and extreme heat stress conditions and to minimize the audience concurrency due to overlapping games. Extra constraints are added on the locations of the games and the number of games per hour. As a result, this model would give us an optimal schedule for future seasons. In order to minimize two different values, we use a weighted sum between the two parts of Z . The weights have been chosen, so the two terms are on a similar scale. The hours of the games will be rounded to 15h, 18h, and 21h in order to simplify. As the same hours were not available in every location, keeping the original schedule would have led to more complex overlapping calculations and less flexibility.

Sets and Indices:

- $l \in Location$
- $m \in Month$
- $d \in Day$
- $h \in Hour$
- The tuples $T(l, d, m, h)$ where $l \in Location$, $m \in Month$, $d \in Day$, and $h \in Hour$ are not uniformly distributed across all locations. Each location l has a specific subset of tuples that are valid, which are determined by the true original schedule, in which each location welcomed games only on certain days.

Parameters:

- $risk_{l,m,h}$, the risk of heat stress exposure
- $capacity_l$, the capacity of the Stadium for the location l , in number of people
- $number_h$, the number of games per hour in the original schedule

Decision variables:

- $x_{l,d,m,h} \in [0, 1]$, is the fixture in location l , on day d of month m at hour h on?
- $overlapping_{l,d,m,h} \in [0, 1]$, equals 1 if there is an overlapping between two games on day d of month m on hour h

$risk_{l,m,h}$ is the hourly risk of thermal stress for each location, and each month; it is a continuous number between 0 and 4, based on the monthly conditions for the whole future period. $capacity_l$ is the capacity of location l . It is used as a proxy to address the audience of each game. $x_{l,d,m,h}$ is a binary variable at 1 if a game occurs in this location l , on the day d , of the month m , 0 otherwise. $overlapping_{l,d,m,h}$ is a binary variable put at 1 if the game happening in location l on day d of month m at hour h is overlapping with at least another game, on the same day and same hour, but on another location.

Minimize

$$Z = 0.99995 * \sum_{h \in Hour, d \in Day, m \in Month, l \in Location} x_{l,d,m,h} * risk_{l,m,h} + 0.00005 * \sum_{d \in Day, l \in Location, m \in Month, h \in Hour} overlapping_{l,m,d,h} * capacity_l \quad (1)$$

Subject to the constraints:

$$\sum_{h \in Hour} x_{l,m,d,h} = 1 \quad \forall l \in Location, \forall m \in Month, \forall d \in Day \text{ and } (l, m, d, h) \in T \quad (2)$$

$$\sum_{d \in Day, m \in Month, l \in Location} x_{l,m,d,h} = number_h \quad \forall h \in Hour, \text{ and } (l, m, d, h) \in T \quad (3)$$

$$\sum_{l2 \in Location} x_{l2,m,d,h} \leq 1 + M * overlapping_{l,m,d,h} \quad \forall m \in Month, \forall d \in Day, \forall h \in Hour \quad (4)$$

$$\sum_{l2 \in Location} x_{l2,m,d,h} \geq 2 - M * (1 - overlapping_{l,m,d,h}) \quad \forall m \in Month, \forall d \in Day, \forall h \in Hour \quad (5)$$

The objective function contains two members. The first one concerns the heat stress risks and is a sum of the risks for each game that happens. The second term concerns overlapping audiences and is the sum of overlapping capacities.

The first constraint, equation (2), ensures that for every fixture (gameday and location), one and only one game is being played to be sure at least one is played, which is what we want, but also that there is only one game being played. The second constraint, equation (3), ensures that the sum of games being played at every hour is the same as during the 2022-23 season. We added this constraint because it is very likely that the LNFP has a good reason to schedule games at certain hours and not all of them at 21h. The third and fourth constraints link the values of x and $overlapping$ and ensure that $overlapping$ is equal to 1 whenever there is an overlapping. M has to be large enough, in

this case at least the maximum of games that can be scheduled at the same timeslots. By adding all those constraints, we try to model as much as possible the constraints existing on the LaLiga schedule in order to have results that can be compared to each other.

4 Results and Discussion

4.1 Results

This section presents the results of the analysis. It is divided into 4 parts. The first part will be focused on analyzing the historical period in the range of values of UTCI, the percentage of the different stress categories, and the difference by cluster, by location and by month. For the future period, it will be focused on understanding what could be the results, and evaluating the same things as for the historical period. The third part is the differences between the future and the past two decades. The last part evaluate the results of the optimization problem.

4.1.1 Historical Analysis

The dataset contains 6888 rows. The range of values goes from the 4th of June 2002 up to the 30th of September 2022. A description of the numerical features can be seen in Table 1, the distribution of UTCI values can be seen in Figure 2, and on overview of the dataset can be seen in Table 4. The distribution of UTCI values is normal, and the mean and median are close to each other (24.89 and 24.82 respectively). The range of UTCI values is large as the minimum is 0.45 and the maximum is 46.19. Regarding the capacity, it ranges from 13450 to 99354 with a mean of 39168. For 'True_Game', the true fixtures represent 21.65% of all the data points, as the mean is equal to 0.2165. The number of different teams represented is 20, so every team in LaLiga 22/23 is present. Because our analysis only considers 3 months, a limited sample of the full LaLiga schedule, some teams welcomed more games (ranging between 2 and 5 games) (cf Supplementary Figure 12). Before any cleaning on the hours, the amount of timeslots is quite large. It exists 11 different timeslots, the first game beginning at 14h while the last one begins at 22h, more details can be found in the Appendix in Table 5.

	True_Game	Capacity	UTCI
count	6888	6888	6888
mean	0.216463	39168.356707	24.886114
std	0.411864	22047.645271	7.313227
min	0.000000	13450.000000	0.451074
25%	0.000000	23142.000000	19.981348
50%	0.000000	36017.000000	24.819666
75%	0.000000	53289.000000	30.373308
max	1.000000	99354.000000	46.188898

Table 1: Descriptive Statistics 2002-2022

As described in Section 1, we employed K-median clustering to group locations sharing similar climate characteristics. We decided to group locations based on the median and the standard deviation. In order to find the best value for K, we used the silhouette score. Having 5 clusters gave us the highest silhouette score which is equal to 0.44. The details of the silhouette analysis can be found in the Appendix on Figure 13. The results of K-median clustering can be seen in Figure 4. 5 different clusters have been created, and it appears that locations having similar median and distribution are grouped together, while locations having different median and distribution are separated groups.

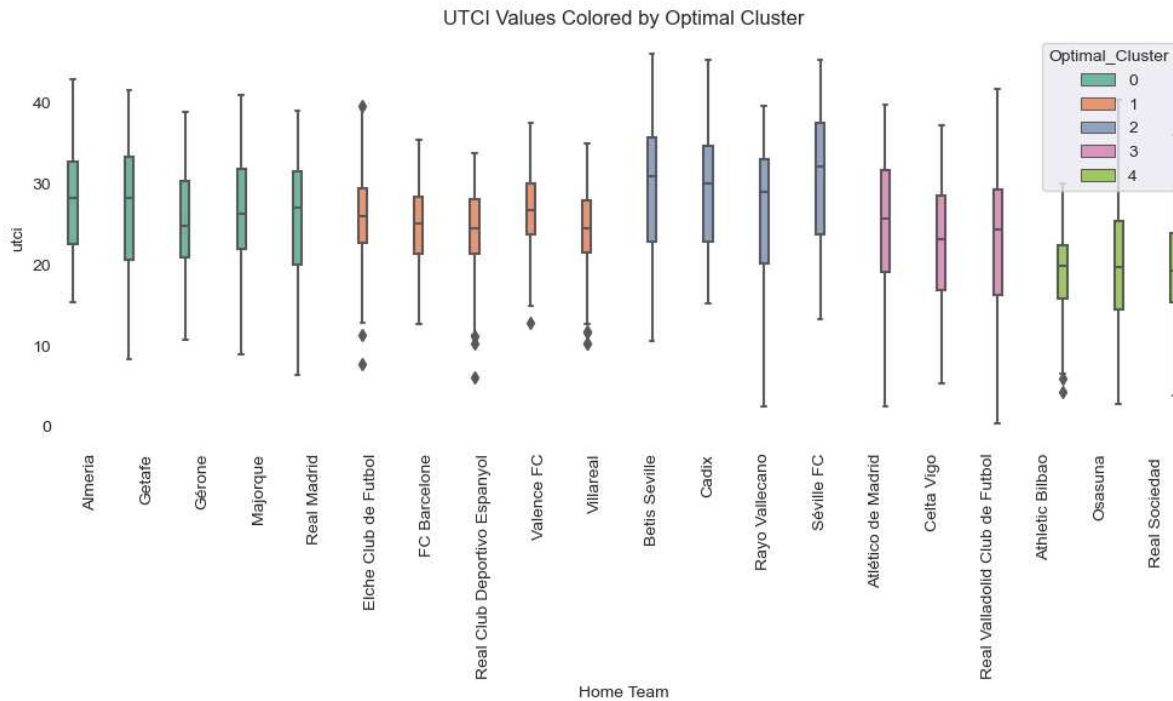


Figure 4: Box plots of UTCI values for the 2002-2022 period, colored by optimal clustering

Secondly, regarding heat stress categories during the afternoons in Spain during summer, we see that 54.1% of the time there is no thermal stress and that the combination of strong and very strong heat stress categories represents 18.4% of the conditions. When translated into FIFA and FIFPro policies, we observed that according to FIFpro policies, 2.6% of the conditions would mean that the game should be rescheduled and 3.4% should have mandatory cooling breaks, while very close to none would result in mandatory cooling breaks according to FIFA. On the other hand, when narrowing our analysis on games that truly happened (keeping the simplification on the schedule as mentioned before), we observed that 74.4% of the games were played with no thermal stress, and only 5.4% of the games were played with strong heat stress and 0.1% with very strong heat stress. When those results are translated into FIFA and FIFpro policies, according to FIFpro, 0.8% of the games should have had mandatory cooling breaks, and only 0.1% of the games should have been rescheduled. More details can be seen in Figure 5.

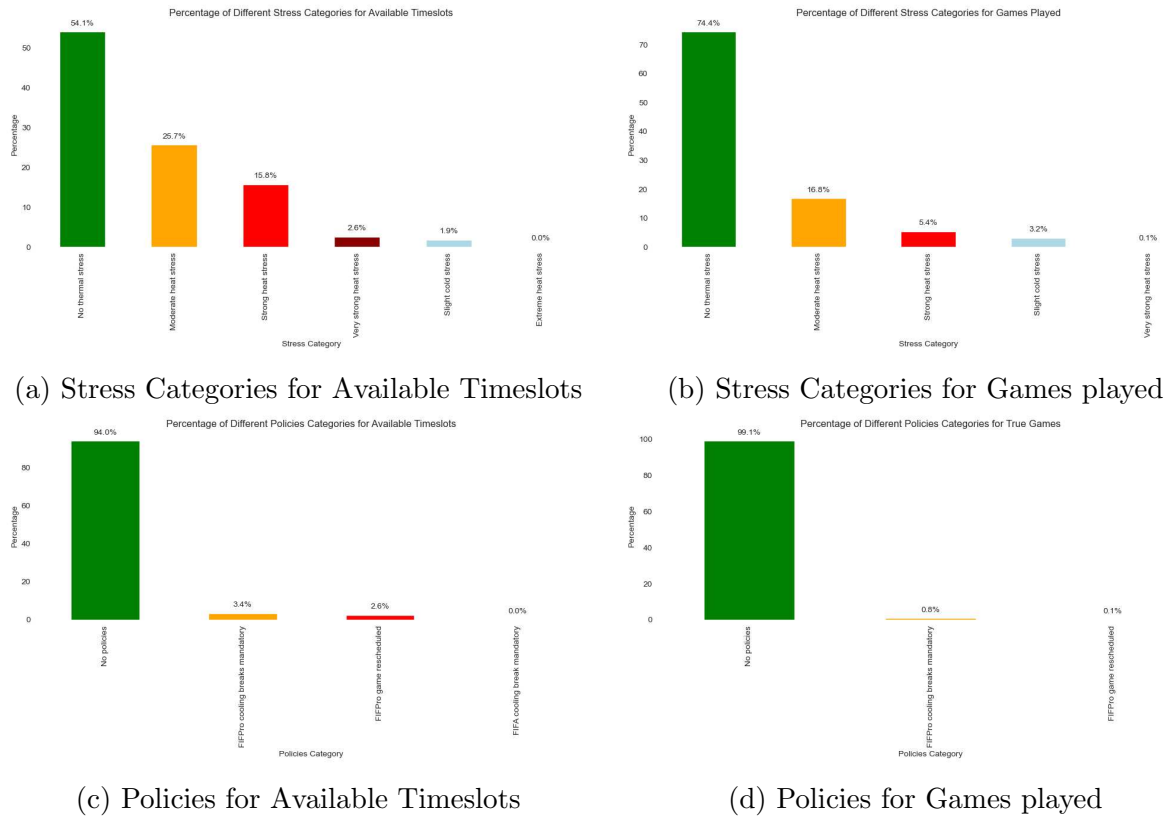


Figure 5: Stress Categories and Policies percentages for the 2002-2022 period

In addition, as observable in Figure 6, by dividing the analysis by optimal clusters, the situation is quite different amongst the different clusters. The locations in cluster number 4 (Athletic Bilbao, Real Sociedad, and Osasuna) were exposed to no thermal stress 81.5% of the time and to strong or very strong heat stress only 2.7% of the time, while the most exposed cluster, cluster number 2 (Betis Seville, Cadix, Rayo Vallecano, and Seville FC)

was exposed only 37.1% of the time to no thermal stress, 31.4% of the time to strong and 12% of the time to very strong heat stress. The results are different when looking at the conditions during games. Cluster number 4 is still the one exposed to the least heat stress, with 84.7% of the games being played with no thermal stress. But for the other 4 clusters, the percentage of conditions with no thermal stress increased a lot, from 16.7% for cluster number 3 to 32.1% for cluster number 2. Besides, cluster 0 is the one in which games were the most played in strong or very strong heat stress conditions.

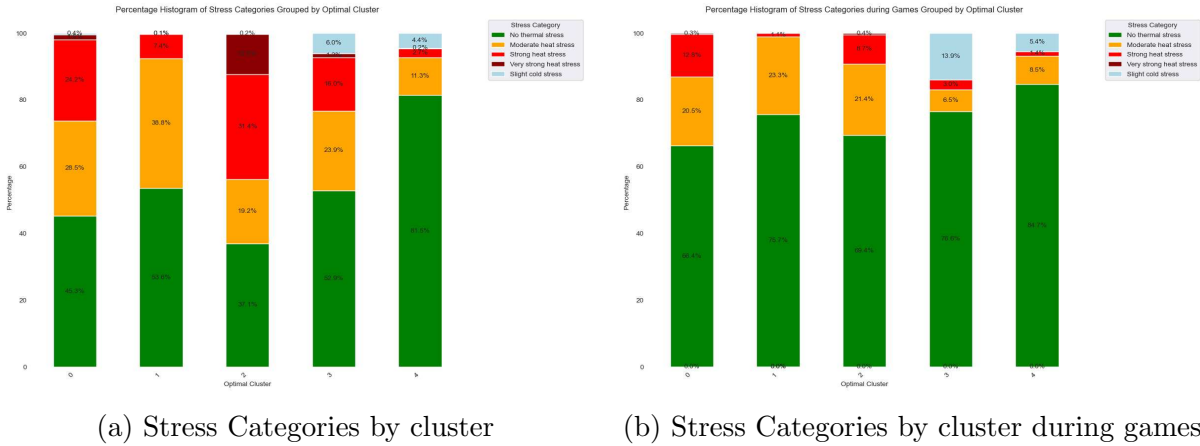


Figure 6: Stress Categories by clusters

Furthermore, the analysis of the differences in heat stress conditions by hour, on Figure 7a, highlighted that there is no heat stress after 20h, the hour with the smallest percentage of no thermal stress is 12h, and it is also the hour with the highest percentage of strong and very strong heat stress (35%). The three periods, 12h, 14h, and 15h, are very similar and are the most exposed to heat stress, with a combined percentage of strong and very strong heat conditions of 35%, 34.9%, and 34.5%, respectively.

In addition, the analysis of the monthly thermal stress situation gave us the results on Figure 7b,. The month of June is the one with the highest percentage of no thermal heat stress conditions, with a percentage of 70.1 %. Moreover, June is exposed 5.5 % of the time to strong heat stress. The month having the highest percentage of strong and very strong heat stress is August, with 24.7 %, followed by September at 16.4%. The month of August also has the smallest percentage of no thermal stress conditions, with 48.9 % while September has 53.5 %

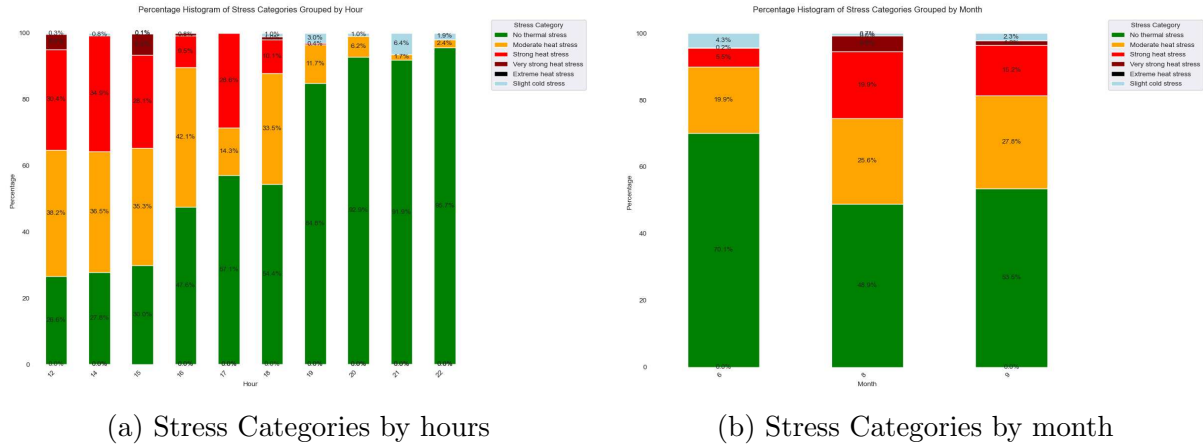


Figure 7: Stress Categories by hours and months

4.1.2 Future Analysis

For the future period, the data goes from the 4th of June 2024 until the 30th of September 2043. It contains 6560 entries; amongst them, 328 contain NA's for the value in UTCI under RCP2.6. As a result, those rows have been removed for every scenario. The details of the values can be seen in Table 2. The 'Capacity' and 'True_Game' columns have the same statistics as for the historical period. For the UTCI values of the 3 scenarios, the means are very close to each other. They are equal to 25.87 for RCP2.6, 25.71 for RCP 4.5, and 25.92 for RCP8.5. As for the historical period, the median and the means are very close; the median is equal to 25.75, 25.61, and 25.73, respectively. The range of UTCI values is different across the scenarios; as for RCP 2.6, the minimum value is 1.27, and the maximum value is 46.44; for RCP 4.5, it is 0.13 and 46.37, and for RCP8.5, it is -3.72 and 46.39. Outliers have been kept, and we created 2 different types of columns, each time for the 3 different scenarios: Stress Category and Policies. The values in the Stress Category columns represent the different stress categories. The values in Policies are based on the 2 policies according to FIFpro and the one according to FIFA, they are based on the UTCI value after using the equivalence between UTCI and WBGT values. The UTCI values have a normal distribution, as shown in Figure 11.

As mentioned before, QDM has been used to correct the existing bias in UTCI values projected. The results can be seen in Figure 8. We can observe that after the bias correction, the distribution of the UTCI values is narrower and that the median is higher for every scenario.

Moreover, in Figure 9, we see that the second scenario (RCP4.5) has the highest percentage of no thermal stress conditions, with 51.1% followed by RCP2.6, with 50.9% and topped by RCP8.5, with 50.4%. The highest percentage of both strong and very strong heat stress appears in the least optimistic scenario (21.7%), and they are the lowest in the

Table 2: Descriptive Statistics for 2024-2043

	True_Game	Capacity	UTCI_26	UTCI_45	UTCI_85
count	6232.000000	6232.000000	6232.000000	6232.000000	6232.000000
mean	0.216463	39168.356707	25.867316	25.710620	25.922416
std	0.411867	22047.813765	6.679556	6.890237	7.212438
min	0.000000	13450.000000	1.273134	0.134271	-3.716052
25%	0.000000	23142.000000	21.350342	21.121230	21.191384
50%	0.000000	36017.000000	25.747507	25.613697	25.727391
75%	0.000000	53289.000000	30.812346	30.649761	31.254653
max	1.000000	99354.000000	46.436747	46.368222	46.394194

second scenario (19.2%), even if, in this case, the percentage of very strong heat stress is higher than in RCP2.6. In RCP2.6, the cumulated percentage of both strong and very strong heat stress is 19.4%. The assumptions of the chi-squared analysis were not met so we can not ensure that the differences in distribution are statistically significant. By taking a look at the residuals, we can see that no absolute value exceeds 2. The details can be found in Table 6 in the Appendix.

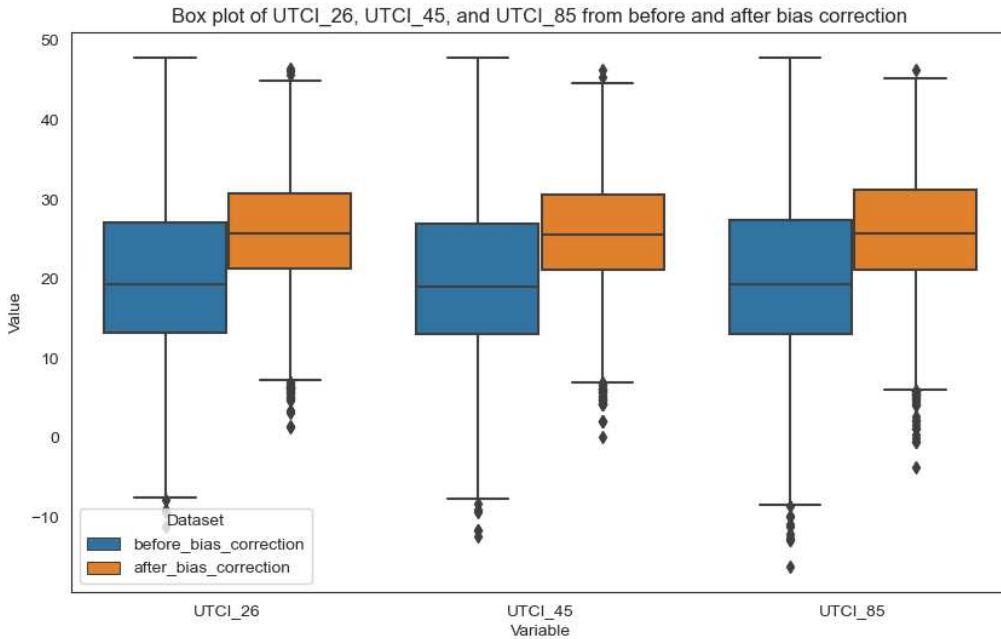


Figure 8: Bias correction on UTCI values for the 2024-2043 years

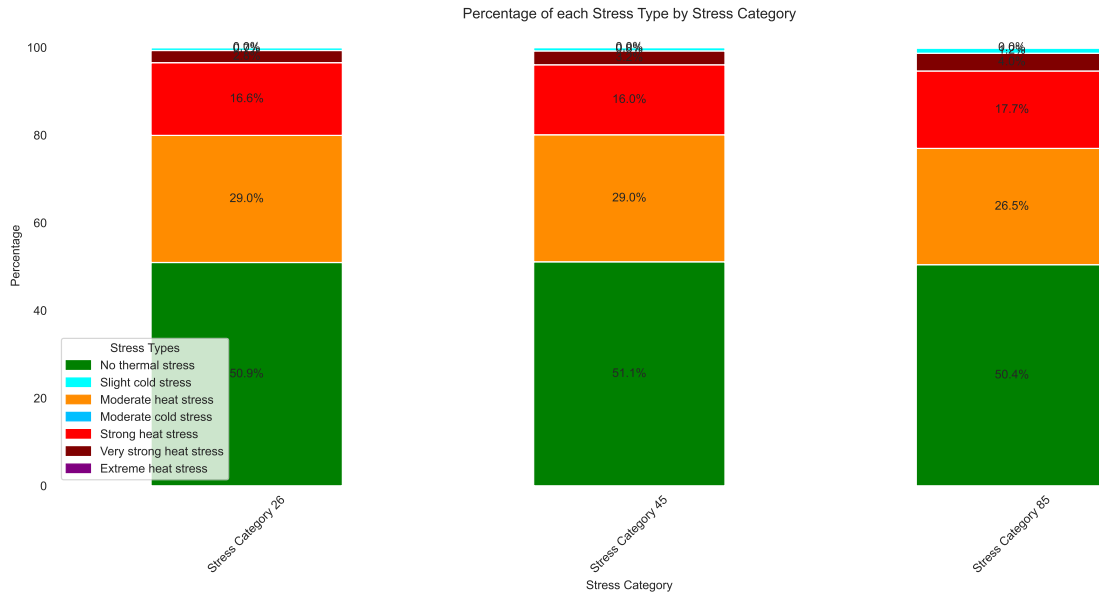
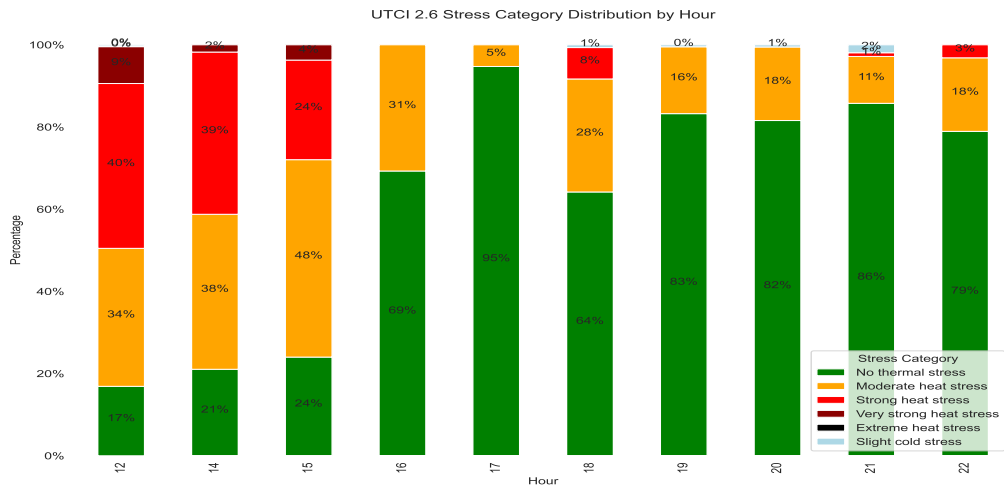
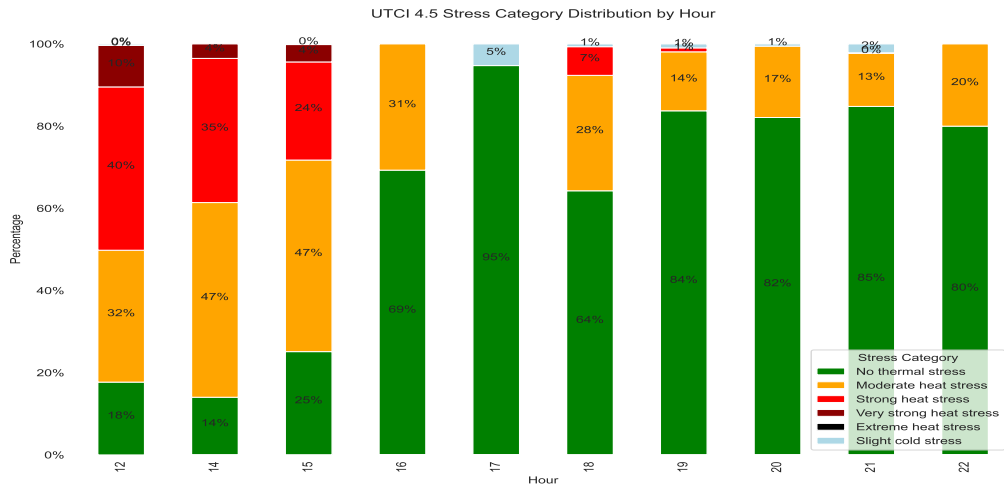


Figure 9: Percentages of the Heat Stress Categories for the Different Scenarios

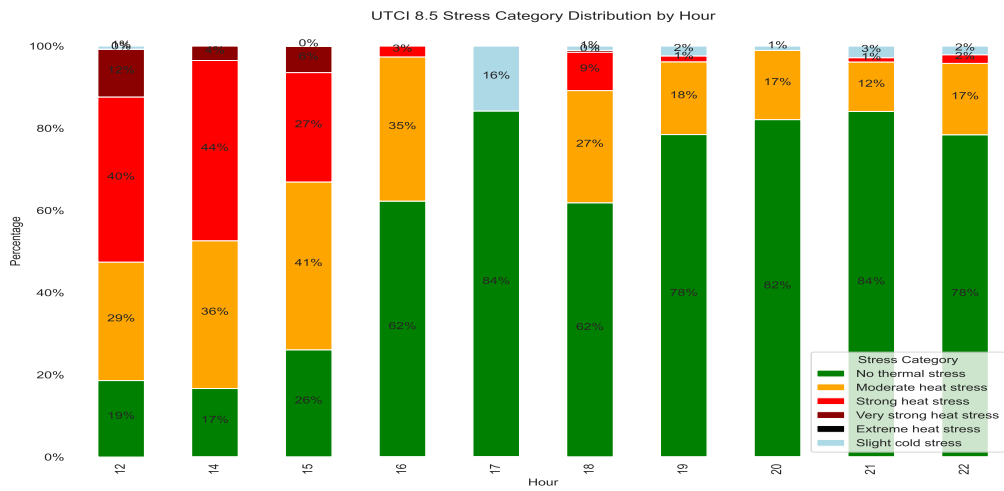
Figure 10 shows that the percentages of the heat stress categories across hours show similar patterns under the different scenarios. Our results showed that for the 3 scenarios, 12h, 14h, and 15h represent the hours with the highest percentages of both strong heat stress and very strong heat stress. In RCP2.6 and RCP4.5, 12h is linked with the highest percentage of strong heat stress, with 40%, while for RCP8.5, it is 14h, with 44%. The highest percentage of very strong heat stress is at 12h for each scenario, with 9%, 10%, and 12% respectively. The smallest percentage of no thermal stress is for each scenario at 14h.



(a) Percentages of Heat Stress Categories across Hours under RCP2.6



(b) Percentages of Heat Stress Categories across Hours under RCP4.5



(c) Percentages of Heat Stress Categories across Hours under RCP8.5

Figure 10: Percentages of Heat Stress Categories across Hours under 3 different scenarios

We found that cluster number 2 has the smallest percentage of no thermal stress in all scenarios, with around one-third of occurrences, and would be exposed to the most heat stress, with strong heat stress conditions being between 29.3% and 32.2% and very strong heat stress between 12.3% and 18.5%. Cluster number 4 has the highest percentage of no thermal stress, with 76.7%, 79.8%, and 79.7%, and the smallest percentage of strong heat stress, with 5.3%, 3.2%, and 3.7%. More details can be found in the Appendix on Figure 14.

Finally, regarding the differences across the months under the different scenarios, we observe that the 3 scenarios are similar (cf. Appendix on Figure 15). In the 3 scenarios, the 3 months have a similar percentage of no thermal stress conditions. Moreover, August is always the month with the highest percentages of both strong heat stress and very strong heat stress conditions. The highest values are obtained for August under RCP2.6 and RCP8.5 with 19% for strong heat stress, 4% for very strong heat stress, and 18% for strong heat stress, and 5% for very strong heat stress, respectively. Under each scenario, June has the highest percentage of no thermal stress with 53%, 53%, and 52% across the scenarios.

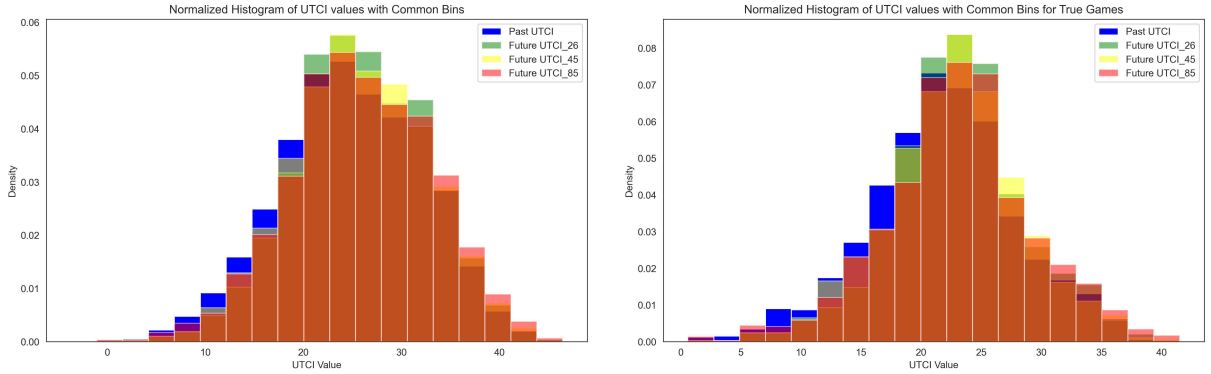
4.1.3 Comparison between the 2 periods

The evaluation of both periods leads to an assessment of the evolution of the situation between the two periods. The results can be found in Table 3. In the past, 54.05% of the conditions were in no thermal stress; this number decreased by 3.6% in the least optimistic scenario and 3.1% and 2.94% for RCP 2.6 and RCP 4.5, respectively. For the 4 other thermal stress conditions, namely moderate heat stress, strong heat stress, very strong heat stress, and extreme heat stress, all witnessed an increase in percentage. The increase in percentage for the extreme heat stress thermal category is the slightest, the maximum value being 0.01% for RCP 2.6 and the smallest being 0.001528% for the two others. Chi-square analysis was done to evaluate if those differences are statistically significant. The test was done between the past value of UTCI and each scenario separately. A p-value of 1.63×10^{-11} , 5.44×10^{-11} , and 6.38×10^{-9} were found for RCP2.6, RCP4.5, and RCP8.5, respectively, and all the assumptions were met except for RCP8.5. So the trends are found to be significant except for RCP8.5.

Table 3: Thermal Stress Conditions Past vs Future

Condition	Past	RCP_26 (in %)	RCP_45 (in %)	RCP_85 (in %)
No thermal stress	54.050523	-3.103796	-2.943334	-3.601229
Moderate heat stress	25.667828	3.327679	3.311633	0.856562
Strong heat stress	15.752033	0.839752	0.213950	1.914848
Very strong heat stress	2.627758	0.148236	0.533346	1.415887
Extreme heat stress	0.014518	0.017574	0.001528	0.001528

These future changes in UTCI for different scenarios is also illustrated in Figure 11 representing the histogram of UTCI values across historical and the different future scenarios. A clear shift to the right of the distribution is observable, with higher heat stress values for the higher emission scenarios. This result is even more noticeable when taking UTCI values during games, where, first of all, values are least extreme, and furthermore, the distribution change is bigger.



(a) UTCI distribution for the past and the 3 scenarios

(b) UTCI distribution for the past and the 3 scenarios during games

Figure 11: UTCI distribution comparison between the past and the 3 future scenarios

Notice that even if the increase in percentages does not seem large for some heat stress categories, by focusing on months, hours, and clusters, we observe that the situations are very different and that some are exposed to a massive increase in heat stress exposure.

4.1.4 Optimisation Problem

The solution of the Linear Problem gave us an optimal schedule for the 3 scenarios.

For RCP 8.5, by opting for the optimal schedule, 79.2% of the games would be played under no thermal stress, 17.5% under moderate heat stress, and only 1.6% of the games would be played in strong heat stress conditions. On the other hand, keeping the same schedule as before would mean having 70.3% of the games played in no thermal stress

conditions, 19.3% in moderate heat stress, 7.6% in strong heat stress conditions, and 0.7% in very strong heat stress conditions. Translating this into FIFA and FIFpro policies, by using the optimal schedule, we ensure that 100% of the games are played without cooling breaks nor rescheduling, while 4% of the conditions would require that the game is rescheduled according to FIFpro and 4.3% would require mandatory cooling breaks. By using the current schedule, those numbers decrease to 0.7% and 1.2% respectively.

For RCP 4.5, by opting for the optimal schedule, it is 79.5% of the games would be played under no thermal stress, while only 0.7% would be played in strong heat stress conditions and 18.8% under moderate heat stress. By using the current schedule, the percentage of no thermal stress conditions would decrease to 73.2%, moderate heat stress conditions would represent 20.3%, and 5% of the games would be played under strong heat stress and 0.3% under very strong heat stress. Under RCP4.5, 3.2% of the games would need to be rescheduled, and 4% would need mandatory cooling breaks, according to FIFpro. By using the current schedule, those numbers decrease to 0.3% and 1% respectively, and by using the optimal schedule, those decrease to 0.1% for the cooling breaks and 0% for the rescheduling.

Last but not least, for RCP 2.6, the current schedule would expose spectators, staff, and players to no thermal stress in 73% of the games, to moderate heat stress in 19.3% of the games, to strong heat stress in 6.5% of the games and to very strong heat stress 0.1% of the games. By opting for the optimal schedule, spectators, staff, and players would be exposed to no thermal stress in 80.6% of the games, to moderate heat stress in 16.3% of the games, to strong heat stress in 2.0% of the games and never to very strong heat stress. Moreover, 3.9% of the timeslots would mean mandatory cooling breaks and 2.8% would mean that the game should be rescheduled, both according to FIFpro. Keeping the current schedule would mean that 0.8% need mandatory cooling breaks, and 0.1% of the games would need to be rescheduled. Finally, by using the optimal schedule, both percentages would be 0.

4.2 Discussion

The goal of this section is to interpret the research findings, compare them with established theories and ideas, aim for generalization, and the effect on professional and academic practice.

4.2.1 Interpretation of the Research Findings

The objectives of this research were firstly to have an idea of the situation in Spain regarding heat stress exposure during gamedays in the summer, then more specifically, to address the situation during games. Secondly, the goal was to know what we could expect in the future in Spain, and more specifically if we keep the same schedule. Finally, we wanted to know if the future schedule could be optimized to reduce heat stress exposure, and if so, how much would it improve the situation for players, staff, and spectators.

Firstly, the results give us a good insight into the situation in Spain in the different football stadiums for the past 20 years. We understand that slightly more than half of the time during summer, there was no thermal stress in Spain, but that the combination of both strong and very strong heat stress conditions represented 18.4% of the conditions. It is important to understand that the time slots for the football games can not just be randomly picked to maximize the audience or for other financial reasons and LaLiga has to take into account the climatic characteristics of the environment to choose wisely the game schedule. Moreover, regarding the policies, we observed that 94% of the conditions would result in no cooling breaks nor rescheduling, 3.4% in cooling breaks, and 2.6% in rescheduling, according to FIFpro. Fortunately, by analyzing the conditions in which games were played in the past 20 years, it appears that the percentage of games played with no thermal stress is higher than the percentage of conditions without thermal stress during this period, 74.4% and 54.1%, respectively. In addition, the percentages of mandatory cooling breaks and rescheduling were lower. This serves as a key indicator that the LNFP is already choosing the time slots wisely in order to reduce the heat stress implied to players, staff, and spectators.

Moreover, grouping locations according to the median value of UTCI and the standard deviation and then using those optimal clusters to further our analysis helped us to get insight into the locations that were more at risk. The disparity between the different groups is very large. Some were almost never exposed to strong or very strong heat stress conditions, such as cluster number 4 grouping Athletic Bilbao, Osasuna, and Real Sociedad (all in northern Spain), while for others, such as cluster number 2 (including Betis Seville, Cadix, Rayo Vallecano, and Seville FC), it represented close to half of the conditions. August was found to be the summer month with the highest percentage of both strong and very strong heat stress, while June was the summer month with the highest percentage of no thermal stress conditions. This might be an indication that shifting the season to begin later in August or in September, and to end later in June might be interesting in the future. Finally, analyzing the situation hour by hour shows that games played before 15h are strongly exposed to heat stress, while after 20h there is no heat

stress. Once again, this might represent an alternative to the current schedule, even if, of course, shifting all the games later might mean more overlapping and, thus, reduced audiences. We think that this whole analysis gives us a good idea of the past two decades in Spain during summer in the football locations of LaLiga.

In addition, regarding the situation for the coming twenty years, we observe that the three scenarios, namely RCP2.6, RCP4.5, and RCP 8.5, depict a similar but slightly different situation. Firstly, the shift to higher heat stress conditions of the distribution of UTCI values between the past two decades and the future two decades supports the idea that LaLiga players, supporters, and staff will be subjected to more heat stress in the coming years. The fact that the three scenarios depict a situation in which the percentages of moderate heat stress, strong heat stress, very strong heat, and even extreme heat stress conditions increase supports the hypothesis that players, staff, and spectators will be exposed to more thermal heat stress. Furthermore, according to the 3 scenarios, cluster number 2 will remain the one being the most exposed to extreme temperatures, and cluster number 4 the one being the least exposed to extreme temperatures. For every cluster, the percentage of no thermal stress conditions will decrease, and depending on the scenario, some percentages of strong and very strong heat stress conditions decrease, but in general, they all increase. The increase in UTCI values is general and exposes every location to more heat stress. As a result, they all have to take into consideration that the situation is getting worse. In the future, the most exposed hour will be narrower than in the past but more extreme. According to the 3 scenarios, 12h, 14h, and 15h will be the most exposed hours to heat stress, with 12h being the most exposed. In addition, for the different months, the situation will change as June, August, and September will be more similar. June will be exposed to more heat stress, and August a bit less. This supports that in the future, it might be difficult to maintain playing games before 15h and in June, which is contradictory to the historical analysis. The schedule should shift to later hours and the season might be shortened to finish earlier.

Finally, it was possible to find an optimal schedule in the three different scenarios. It was surprising to find out that even when trying to reduce overlapping as much as possible and while keeping the same amount of games played at each hour as for the 2022-23 season schedule, it was possible to find a schedule with a low percentage of games played in strong or very strong heat stress, although the overlapping is higher than in the past with 35 games overlapping, compared to 21 in the past. Besides, we observed that using the optimal schedule over the current one for the future reduces significantly the exposure to strong heat stress conditions by approximately 5% for each scenario and that it leads to no cooling breaks or rescheduling of games. It supports that rescheduling is needed and effective to reduce heat stress exposure but it has a cost: more overlapping

will be induced. Moreover, the percentages of heat stress exposure were lower than the current schedule during the past two decades. It might be the case because there are other constraints that should be introduced and that could explain why the percentages of games played under strong or very strong heat stress conditions are so "high" for the past two decades.

4.2.2 Correspondence with Literature Review

It has been shown that climate change, via extreme heat, has an impact on the sports industry, forcing games to be rescheduled or to have mandatory cooling breaks, and here, this impact was observable for a specific case: laLiga in Spain. Moreover, the results for the future period match the expectations presented in the Literature Review, stating that more extreme heat events are expected as time passes. Knowing that this is expected and to what extent allows LaLiga organizers to anticipate and to already try to find solutions to apply this optimized schedule. Because it is good to know that this optimal schedule exists, that it is feasible, and that by using it, the percentage of games played under strong or very strong heat stress can be significantly reduced and that rescheduling might be avoided, it is something else to be able to introduce this new schedule, as there might constraints from LaLiga's point of view that we do not know. Nonetheless, thanks to this analysis, the LNFP can anticipate and find solutions to reduce heat stress in the coming years, even with all those constraints.

4.2.3 Appropriateness of the Methodology and Data Collection

The quality of the data is high since they come from trustworthy sources, although some things might lead to issues. First of all, the length of the period observed for both the past and the future. Those two periods were supposed to be 20 years long, but due to missing values, the future period is shorter, and it is considered by some climate scientists that in order to observe significant differences regarding the climate, a period of 30 years is necessary.

Moreover, the precision of the data could lead to inaccuracies. The temporal accuracy for the future is lower as the resolution is 3-hours and not hourly. In addition, the spatial accuracy can also lead to inaccuracies as the data we got do not concern exactly the location of the stadiums, and some stadiums were summarised to the same location. Furthermore, we base our analysis of the future on only one model, and the analysis would be more accurate by using different models, each one proposing one possible future. Additionally, considering that the schedule has been constant for the past 20 seasons has had some impact on the evaluation of the situation in the past two decades. Finally, the choice of the values for the equivalence between the WBGT and UTCI threshold is not

perfectly precise and might lead to different results, even if it might be negligible.

Concerning the methodology, we think that it was appropriate. We decided to use a model proposed from reliable sources, and by using QDM to correct it, we managed to handle the bias. As a result, the resulting model is free from bias, and the trend is kept, which is fundamental in climate science. Moreover, choosing K-median clustering as a tool for the grouping of locations makes sense, as using the median over the mean ensures that it is less sensitive to outliers. We could have opted for geographical clustering as close locations could have similar climates, but it might be the case that the situation between two close locations is very different. As a result, we decided to use K-median, topped with the analysis of standard deviation, because two locations might have a close median, but in our analysis, the standard deviation has a big impact, as the most extreme heat stress conditions arise from the most extreme UTCI.

Furthermore, using percentages makes sense in our case as for the future period, the model we use does not depict the climatic characteristics by day or by hour but ensures that the distribution of probability could look like that. As a result, it would not make sense to analyze the future by day or year, but we analyze the probability distribution of the climatic situation for the two coming decades in Spain. Regarding the past decades, we do have the data for each day, and those are (almost) real data, but first of all, since the schedule has been made up, it would not make sense to observe day by day. Besides, the goal of our analysis is to analyze the climate, and this is made over a long period of time and not day by day. For all those reasons, using percentages made the most sense to analyse both situations and the differences between the two.

Finally, we opted for a Linear Program to find the optimal schedule with a multi-objective function. This makes sense since it is one way to resolve optimization problems. Nonetheless, we might lack some constraints that we do not know regarding the way that the schedule is done. Moreover, the weights in the objective function have been chosen so the two terms have the same scale, but it is possible that the LNFP would focus more on one objective.

4.2.4 Effect on Professional and Academic Practice

The goal of this research is to have an impact on professional practice as we want to make sports competition organizers understand that because of climate change, organizations have to use different policies to adapt as it might put players, supporters, and staff at risk. We think that this practical case of LaLiga's exposition to extreme heat makes sense and that it will attract sports organizers' attention. On the other end, even if no new

technique was used in this research, we believe this might inspire researchers to analyze the situation for other football leagues or other sports. In that regard, we expect to have an academic effect by adding our research to a new field.

4.2.5 Generalization

Last but not least, those results can be generalized to, first of all, other sports competitions, not only in Spain but in the whole Mediterranean Region. Spain is known to welcome lots of different big sports events such as football, tennis, golf, F1, ... In that regard, we think the results obtained in this analysis can be generalized for other sports in Spain. Moreover, since the Mediterranean region is exposed to those important climate changes and in similar proportion, we expect those results to be relevant for the rest of this region.

5 Limitations

First of all, we based our analysis on the past 20 years and predict the future climate for the coming two decades while it is supported that 20 years is the minimum to observe significant changes in climate and that a period of 30 years might be more suitable. Since the same schedule has been applied for the past 20 years, but this is not the actual schedule that happened, conclusions taken on percentages of games in those conditions, or whether LaLiga already adapts or not, might be impacted.

Finally, We have some missing values in the future (about 2 years), and this might impact the results too. We only used one model to predict the future situation, which proposes one possible future. By using 2 or 3 models, the conclusions should be more supported. The data gathered have a lack of temporal and spatial accuracy, which could ultimately affect the results. UTCI is not a perfect index to address heat stress exposure, and its translation from WBGT could be improved

6 Conclusions

6.1 Main Findings and Conclusions

The goal of this research was to assess the current and future impact of extreme heat on the professional football industry in Spain, focusing on LaLiga - the man's football first division - and to evaluate if a rescheduling optimization strategy might be enough to limit

this impact.

First, we showed that the football locations in Spain have already been exposed to high heat stress events over the last two decades. We observed that the combination of strong and very strong heat stress represents 18.1 % of the conditions from 12h until 22h on LaLiga game days. Moreover, we analyzed that 5.4% of the games were played under strong or very strong conditions during the past two decades. In addition, we discovered that the situations are very different across the 20 locations and that those locations can be grouped into 5 clusters based on the median and the standard deviation of their UTCI values. The least heated group was only exposed 2.9% of the time to strong or very strong heat stress, while the most heated group was exposed almost half of the time to strong or very strong heat stress. This disparity is also observable for hours where the 12h-15h period represents the most at-risk period with around one-third exposition to strong or very strong heat stress, while there is no heat stress after 20h. Furthermore, August represents the most heat-exposed month, followed by July, and then by June. Game exposition is way lower than the general exposition, leading to the intuition that the LNFP already tries to reduce players', spectators', and staff's heat stress exposure. To sum up the situation in Spain for the past two decades, Spain has been exposed to high heat stress, but the situation has already been handled by La Liga's organizers by scheduling the games to better time slots, introducing mandatory cooling breaks, and rescheduling the game when needed. In other words, LaLiga's schedule is relatively well adapted to the current Spanish summer climate conditions.

Secondly, we evaluated the situation that might arise in the coming 2 decades. We based ourselves on three different scenarios, namely RCP2.6, RCP4.5, and RCP8.5 - respectively representing a Paris-aligned future emission scenario, and middle-of-the-road scenario, and a business-as-usual emission scenario without relevant efforts for emission reduction. We found that the situation is similar but different between the 3 scenarios. For each scenario, the percentage of moderate heat stress conditions, strong heat stress, and very strong heat stress was higher than in the past two decades, and the percentage of no thermal stress conditions decreased. Moreover, the percentages of mandatory cooling breaks and rescheduling would increase by 1% and 1.4%, respectively, in the worst-case scenario. The most exposed hours remain at 12h, 14h, and 15h, but the heat stress exposure increases for 12h and 14h and decreases for the others. For every month, the heat stress exposure increases, with the highest increase in June. Every cluster will also be exposed to more heat stress. As expected, the situation will worsen by exposing athletes, staff, and spectators to more heat stress in the coming years, and this in every location.

We solved, for each scenario, an optimization problem in order to find the optimal

schedule, reducing the exposure to heat stress as much as possible, as well as avoiding dilution of audiences. For each scenario, we managed to find an optimal schedule, never exposing players, staff, and spectators to very strong heat stress, and to strong heat stress 1.9%, 0.7%, and 1.6% of the times, for RCP 2.6, RCP4.5, and RCP 8.6 respectively.

By analyzing the situation in the past twenty years according to FIFA's and FIFpro's policies, we observed that 0.8% of the games should have had cooling breaks while 0.1% should have been rescheduled or canceled. This translates into approximately 2 games canceled or rescheduled and around 12 games with mandatory cooling breaks in 21 years. In the next 2 decades, those numbers, if keeping the same schedule, are between 0.8%, and 1.2%, and 0.1% and 0.7%, respectively, according to the scenarios, which shows that more and more games will need to apply those policies in order to protect players, staff, and spectators. Fortunately, by using the optimal schedule, the number of games being rescheduled or canceled, or with mandatory cooling breaks can decrease to 0.

This might indicate that rescheduling is enough to mitigate the impacts of climate change on the professional football industry in Spain, but we think that it has to be taken with a grain of salt. Indeed, it is possible that we are missing some constraints, making impossible to apply this optimal schedule. Moreover, we think that both FIFA's and FIFpro's recommendations might be too optimistic and not protecting athletes, staff, and supporters enough, as we can see that it is recommended to reduce physical activity when exposed to strong heat stress (Figure 1).

Extreme heat exposure can have several consequences, such as injuries or casualties, but also legal and financial. Financially, game rescheduling or overlapping leads to a revenue decrease, but also, for football clubs in the most exposed regions, it might get harder to have stadiums financed. Besides, legally, the LNFP is facing more risks if spectators, staff, and players are put into more endangering conditions. For all those reasons, our research gives an important insight into what can be expected in the near future for the Spanish professional football league. It will help organizers to make decisions with full knowledge of the facts and anticipate future problems that might arise.

Even if, according to the results, the situation appears under control and not too extreme, we want to strongly underline that it only takes once to become a massive event, and that could change the rules.

6.2 Future Research

Future research might be oriented towards doing the same analysis in other locations, especially in the Mediterranean Region, or a similar analysis can be done for other sports, such as tennis, which has been highly impacted by extreme heat events in recent history. In addition, it would be very interesting to deepen the financial impact of extreme heat on the professional football industry in Spain or to address the impact of extreme heat on performance, which would also have several financial consequences for the LNFP.

References

- Adopted, IPCC (2014). “Climate change 2014 synthesis report”. In: *IPCC: Geneva, Switzerland*, pp. 1059–1072.
- Alfano, Francesca R dAmbrosio et al. (2014). “WBGT index revisited after 60 years of use”. In: *Annals of Occupational Hygiene* 58.8, pp. 955–970.
- Andreff, Wladimir and Stefan Szymanski (2006). *Handbook on the Economics of Sport*. Edward Elgar Publishing.
- Armstrong, Lawrence E et al. (2007). “Exertional heat illness during training and competition”. In: *Medicine & Science in Sports & Exercise* 39.3, pp. 556–572.
- Bangsbo, Jens, Magni Mohr, and Peter Krstrup (2006). “Physical and metabolic demands of training and match-play in the elite football player”. In: *Journal of sports sciences* 24.07, pp. 665–674.
- Bergeron, Michael F, Roald Bahr, et al. (2012). “International Olympic Committee consensus statement on thermoregulatory and altitude challenges for high-level athletes”. In: *British journal of sports medicine* 46.11, pp. 770–779.
- Bergeron, Michael F, Douglas B McKeag, et al. (2005). “Youth football: heat stress and injury risk”. In: *Medicine and science in sports and exercise* 37.8, pp. 1421–1430.
- Blazejczyk, K (1994). “New Climatological-and-physiological Model of the Human Heat Balance Outdoor and its Applications in Bioclimatological Studies in Different Sites”. In: *Bioclimatic Research of the Human Heat Balance (ed.)*, *Zeszyty IGIPZ, PAN, Warszawa*, ppp 29, pp. 34–37.
- Baejczyk, Krzysztof (2010). “UTCI-nowy wskanik oceny obcie cieplnych czowieka= UTCI-new index for assessment of heat stress in man”. In: *Przegld geograficzny* 82.1, pp. 49–71.
- Brake, Derrick J and Graham P Bates (2002). “Limiting metabolic rate (thermal work limit) as an index of thermal stress”. In: *Applied occupational and environmental hygiene* 17.3, pp. 176–186.
- Brocherie, Franck, Olivier Girard, and Grégoire P Millet (2015). “Emerging environmental and weather challenges in outdoor sports”. In: *Climate* 3.3, pp. 492–521.
- Brocherie, Franck, Olivier Girard, Alessandro Pezzoli, et al. (2014). “Outdoor exercise performance in ambient heat: Time to overcome challenging factors?” In: *International Journal of Hyperthermia* 30.8, pp. 547–549.
- Brocherie, Franck and Grégoire P Millet (2015). “Is the wet-bulb globe temperature (WBGT) index relevant for exercise in the heat?” In: *Sports Medicine* 45.11, pp. 1619–1621.
- Brotherhood, John R (2008). “Heat stress and strain in exercise and sport”. In: *Journal of Science and Medicine in Sport* 11.1, pp. 6–19.

- Budd, Grahame M (2008). “Wet-bulb globe temperature (WBGT)its history and its limitations”. In: *Journal of science and medicine in sport* 11.1, pp. 20–32.
- Burton, Ian, Kristie L Ebi, and Glenn McGregor (2009). “Biometeorology for adaptation to climate variability and change”. In: *Biometeorology for adaptation to climate variability and change*. Springer, pp. 1–5.
- Cannon, Alex J, Stephen R Sobie, and Trevor Q Murdock (2015). “Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes?” In: *Journal of Climate* 28.17, pp. 6938–6959.
- Carvalho, MJ et al. (2016). “Regionalization of Europe based on a K-Means Cluster Analysis of the climate change of temperatures and precipitation”. In: *Physics and Chemistry of the Earth, Parts A/B/C* 94, pp. 22–28.
- Change, Intergovernmental Panel On Climate (2007). “Climate change 2007: The physical science basis”. In: *Agenda* 6.07, p. 333.
- Change, IPCC Climate (2014). “impact”. In: *Adaptation and vulnerability, IPCC WGII AR5 summary for policymakers*.
- Change, Projected Climate (2018). “Global warming of 1.5ř C”. In: *World Meteorological Organization: Geneva, Switzerland*.
- Chard, Chris and Cheryl Mallen (2012). “Examining the linkages between automobile use and carbon impacts of community-based ice hockey”. In: *Sport Management Review* 15.4, pp. 476–484.
- Chou, Sin Chan et al. (2014). “Assessment of climate change over South America under RCP 4.5 and 8.5 downscaling scenarios”. In: *American Journal of Climate Change* 3.05, p. 512.
- Corkhill, Barney (2023). *European Super Ligue: Which clubs have ruled themselves out of participating*. Accessed on: 28/12/2023. URL: https://www.sportsmole.co.uk/football/man-utd/european-super-league/feature/european-super-league-which-clubs-have-ruled-themselves-out-of-participating_531939.html?t=123987456&s=zero.
- Dawson, Jackie, Daniel Scott, and Mark Havitz (2013). “Skier demand and behavioural adaptation to climate change in the US Northeast”. In: *Leisure/Loisir* 37.2, pp. 127–143.
- Di Napoli, Claudia, Florian Pappenberger, and Hannah L Cloke (2019). “Verification of heat stress thresholds for a health-based heat-wave definition”. In: *Journal of Applied Meteorology and Climatology* 58.6, pp. 1177–1194.
- Diffenbaugh, Noah S and Filippo Giorgi (2012). “Climate change hotspots in the CMIP5 global climate model ensemble”. In: *Climatic change* 114, pp. 813–822.
- Dingle, Greg William and Bob Stewart (2020). “Playing the climate game: Climate change impacts, resilience and adaptation in the climate-dependent sport sector”. In: *Creating and Managing a Sustainable Sporting Future*. Routledge, pp. 39–59.

- Eichinger, Michael (2019). “Transformational change in the Anthropocene epoch”. In: *The Lancet Planetary Health* 3.3, e116–e117.
- Eis, Dieter et al. (2011). “Klimawandel und Gesundheit-Ein Sachstandsbericht”. In:
- Elsasser, Hans and Rolf Bürki (2002). “Climate change as a threat to tourism in the Alps”. In: *Climate research* 20.3, pp. 253–257.
- Europa, Weltgesundheitsorganisation Regionalbüro für (2019). *Gesundheitshinweise zur Prävention hitzebedingter: Gesundheitsschäden neue und aktualisierte Hinweise für unterschiedliche Zielgruppen*. Tech. rep. Weltgesundheitsorganisation. Regionalbüro für Europa.
- Fairley, Sheranne, Lisa Ruhanen, and Hannah Lovegrove (2015). “On frozen ponds: The impact of climate change on hosting pond hockey tournaments”. In: *Sport Management Review* 18.4, pp. 618–626.
- Fiala, Dusan, George Havenith, et al. (2012). “UTCI-Fiala multi-node model of human heat transfer and temperature regulation”. In: *International journal of biometeorology* 56, pp. 429–441.
- Fiala, Dusan, Kevin J Lomas, and Martin Stohrer (2001). “Computer prediction of human thermoregulatory and temperature responses to a wide range of environmental conditions”. In: *International journal of biometeorology* 45, pp. 143–159.
- FIFPRO (2023). *Eleven tips for dealing with hot conditions in professional football*. Accessed on: 02/12/2023. URL: <https://fifpro.org/en/supporting-players/health-and-performance/extreme-weather-and-climate-change/eleven-tips-for-dealing-with-hot-conditions-in-professional-football>.
- (n.d.). *Extreme Weather in Football*. Accessed on: 02/12/2023. URL: <https://fifpro.org/en/supporting-players/health-and-performance/extreme-weather-and-climate-change/extreme-weather-in-football/>.
- Filo, Kevin, Graham Cuskelly, and Pamela Wicker (2015). “Resource utilisation and power relations of community sport clubs in the aftermath of natural disasters”. In: *Sport Management Review* 18.4, pp. 555–569.
- Fischer, Erich M and Christoph Schär (2009). “Future changes in daily summer temperature variability: driving processes and role for temperature extremes”. In: *Climate dynamics* 33.7-8, pp. 917–935.
- Flyvbjerg, Bent and Allison Stewart (2012). “Olympic proportions: Cost and cost overrun at the Olympics 1960-2012”. In:
- Funk, Daniel C and Jeff James (2001). “The psychological continuum model: A conceptual framework for understanding an individual’s psychological connection to sport”. In: *Sport management review* 4.2, pp. 119–150.
- García-Mozo, H, A Mestre, C Galán, et al. (2010). “Phenological trends in southern Spain: A response to climate change”. In: *Agricultural and Forest Meteorology* 150.4, pp. 575–580.

- Gerrett, Nicola et al. (2019). “Ambient conditions prior to Tokyo 2020 Olympic and Paralympic games: considerations for acclimation or acclimatization strategies”. In: *Frontiers in Physiology* 10, p. 414.
- Giorgi, Filippo and William J Gutowski Jr (2015). “Regional dynamical downscaling and the CORDEX initiative”. In: *Annual review of environment and resources* 40, pp. 467–490.
- Giorgi, Filippo, Colin Jones, Ghassem R Asrar, et al. (2009). “Addressing climate information needs at the regional level: the CORDEX framework”. In: *World Meteorological Organization (WMO) Bulletin* 58.3, p. 175.
- Grantham, Justin et al. (2010). “Current knowledge on playing football in hot environments”. In: *Scandinavian journal of medicine & science in sports* 20, pp. 161–167.
- Grundstein, Andrew et al. (2013). “Exceedance of wet bulb globe temperature safety thresholds in sports under a warming climate”. In: *Climate research* 58.2, pp. 183–191.
- Grundstein, Andrew J et al. (2021). “From theory to practice: operationalizing a climate vulnerability for sport organizations framework for heat hazards among US High schools”. In: *Journal of Science and Medicine in Sport* 24.8, pp. 718–722.
- Guttmann, Allen (2004). *From ritual to record: The nature of modern sports*. Columbia university press.
- Gyri, Ferenc and László Balogh (2017). “Rethinking the relationship between sport, recreation and tourism”. In: *Leisure, Health and Well-Being: A Holistic Approach*, pp. 121–133.
- Hanna, Elizabeth G et al. (2011). “Climate change and rising heat: population health implications for working people in Australia”. In: *Asia Pacific Journal of Public Health* 23.2_suppl, 14S–26S.
- Hewitson, Bruce C and Robert George Crane (1996). “Climate downscaling: techniques and application”. In: *Climate Research* 7.2, pp. 85–95.
- Honjo, Tsuyoshi et al. (2018). “Thermal comfort along the marathon course of the 2020 Tokyo Olympics”. In: *International journal of biometeorology* 62, pp. 1407–1419.
- Houghton, John Theodore et al. (2001). *Climate change 2001: the scientific basis*. Vol. 881. 9. Cambridge university press Cambridge.
- IATA (2023). *Industry Statistics: Fact Sheet*. Accessed on: 11/12/2023. URL: <https://www.iata.org/en/iata-repository/pressroom/fact-sheets/industry-statistics/>.
- ISO (2004). *Ergonomics of thermal environments strategy of evaluation of the risk for the prevention of constraints or discomfort under thermal working conditions—ISO 15265 Standard*.
- James, JD, M Walker, and W Kuminka (2009). “Becoming a professional football team fan: Differences based on level of team internalization”. In: *International Journal of Sport Management* 10.1, pp. 14–34.

- Kakamu, Takeyasu et al. (2017). “Preventing heat illness in the anticipated hot climate of the Tokyo 2020 Summer Olympic Games”. In: *Environmental health and preventive medicine* 22, pp. 1–6.
- Kappas, Martin (2009). *Klimatologie: Klimaforschung im 21. Jahrhundert-Herausforderung für Natur-und Sozialwissenschaften*. Springer-Verlag.
- Kay, Joyce and Wray Vamplew (2006). “Under the weather: Combating the climate in British sport”. In: *Sport in Society* 9.1, pp. 94–107.
- Kerr, Zachary Y et al. (2019). “The association between mandated preseason heat acclimatization guidelines and exertional heat illness during preseason high school American football practices”. In: *Environmental health perspectives* 127.4, p. 047003.
- Kiko (2023). *15 Outstanding Sports Industry Revenue Statistics in History*. Accessed on: 11/12/2023. URL: <https://playtoday.co/blog/stats/sports-industry-revenue-statistics/#:~:text=With%20over%20%24512.14%20billion%20in,Soccer%3A%20%2450%20billion%20a%20year>.
- King, Andrew D and David J Karoly (2017). “Climate extremes in Europe at 1.5 and 2 degrees of global warming”. In: *Environmental Research Letters* 12.11, p. 114031.
- Kotlarski, Sven et al. (2014). “Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble”. In: *Geoscientific Model Development* 7.4, pp. 1297–1333.
- Lemke, Bruno and Tord Kjellstrom (2012). “Calculating workplace WBGT from meteorological data: a tool for climate change assessment”. In: *Industrial health* 50.4, pp. 267–278.
- Maughan, RJ et al. (2010). “Living, training and playing in the heat: challenges to the football player and strategies for coping with environmental extremes”. In: *Scandinavian journal of medicine & science in sports* 20, pp. 117–124.
- McCullough, Brian P, Madeleine Orr, and Timothy Kellison (2020). “Sport ecology: Conceptualizing an emerging subdiscipline within sport management”. In: *Journal of Sport Management* 34.6, pp. 509–520.
- Meehl, Gerald A and Claudia Tebaldi (2004). “More intense, more frequent, and longer lasting heat waves in the 21st century”. In: *Science* 305.5686, pp. 994–997.
- Miller, Toby et al. (2001). *Globalization and sport: Playing the world*. Sage.
- Mountjoy, Margo et al. (2012). “Hyperthermic-related challenges in aquatics, athletics, football, tennis and triathlon”. In: *British Journal of Sports Medicine* 46.11, pp. 800–804.
- Mücke, HG and A Matzarakis (2020). *Klimawandel und Gesundheit: Tipps für sommerliche Hitze und Hitzewellen*. Umweltbundesamt.
- Nag, Pranab Kumar, Anjali Nag, and Shirish Pratap Ashtekar (2007). “Thermal limits of men in moderate to heavy work in tropical farming”. In: *Industrial health* 45.1, pp. 107–117.

- Nybo, Lars et al. (2021). *Football facing a future with global warming: perspectives for players health and performance*.
- Olya, Hossein GT (2019). “A call for weather condition revaluation in mega-events management”. In: *Current Issues in Tourism* 22.1, pp. 16–20.
- Orr, Madeleine (2020). “On the potential impacts of climate change on baseball and cross-country skiing”. In: *Managing Sport and Leisure* 25.4, pp. 307–320.
- Orr, Madeleine and Yuhei Inoue (2019). “Sport versus climate: Introducing the climate vulnerability of sport organizations framework”. In: *Sport Management Review* 22.4, pp. 452–463.
- Orr, Madeleine, Yuhei Inoue, et al. (2022). “Impacts of climate change on organized sport: A scoping review”. In: *Wiley Interdisciplinary Reviews: Climate Change* 13.3, e760.
- Parsons, Ken (2006). “Heat stress standard ISO 7243 and its global application”. In: *Industrial health* 44.3, pp. 368–379.
- (2014). *Human thermal environments: the effects of hot, moderate, and cold environments on human health, comfort, and performance*. CRC press.
- Pham, Duc Truong, Stefan S Dimov, and Chi D Nguyen (2005). “Selection of K in K-means clustering”. In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 219.1, pp. 103–119.
- Sawka, Michael N et al. (2007). “American College of Sports Medicine position stand. Exercise and fluid replacement.” In: *Medicine and science in sports and exercise* 39.2, pp. 377–390.
- Schneider, Sven (2021). “Sport and climate change how will climate change affect sport?” In.
- Scott, Daniel, Robert Steiger, Michelle Ruddy, and Yan Fang (2019). “The changing geography of the Winter Olympic and Paralympic Games in a warmer world”. In: *Current Issues in Tourism* 22.11, pp. 1301–1311.
- Scott, Daniel, Robert Steiger, Michelle Ruddy, and Peter Johnson (2015). “The future of the Olympic Winter Games in an era of climate change”. In: *Current Issues in Tourism* 18.10, pp. 913–930.
- Shutaywi, Meshal and Nezamoddin N Kachouie (2021). “Silhouette analysis for performance evaluation in machine learning with applications to clustering”. In: *Entropy* 23.6, p. 759.
- Sofotasiou, Polytimi, Benjamin Richard Hughes, and John Kaiser Calautit (2015). “Qatar 2022: Facing the FIFA World Cup climatic and legacy challenges”. In: *Sustainable cities and society* 14, pp. 16–30.
- Tebaldi, Claudia et al. (2006). “Going to the extremes: an intercomparison of model-simulated historical and future changes in extreme events”. In: *Climatic change* 79.3-4, pp. 185–211.

- Tipton, Mike et al. (2019). “Hit for six: The impact of climate change on cricket”. In: *The British Association for Sustainable Sport*. <https://basis.org.uk/hit-for-six>.
- Townsend, Mardie et al. (2003). “Too hot to trot? Exploring potential links between climate change, physical activity and health”. In: *Journal of Science and Medicine in Sport* 6.3, pp. 260–265.
- Trail, Galen T, Dean F Anderson, and Janet S Fink (2000). “A theoretical model of sport spectator consumption behavior”. In: *International journal of sport management* 1.3, pp. 154–180.
- Trendafilova, Sylvia, Kathy Babiak, and Kathryn Heinze (2013). “Corporate social responsibility and environmental sustainability: Why professional sport is greening the playing field”. In: *Sport management review* 16.3, pp. 298–313.
- Vanos, Jennifer K et al. (2019). “Planning for spectator thermal comfort and health in the face of extreme heat: The Tokyo 2020 Olympic marathons”. In: *Science of The Total Environment* 657, pp. 904–917.
- Vicente S., Lopez de (2023). *Real Madrid dealt early kick-off in searing summer heat*. Accessed on: 02/12/2023. URL: <https://en.as.com/soccer/real-madrid-dealt-early-kick-off-in-searing-summer-heat-n/>.
- Wonisch, Manfred and R Pokan Ledl-Kurkowski (2017). *Kompendium der Sportmedizin*. Vol. 2. Springer.
- Yaglou, CP, D Minaed, et al. (1957). “Control of heat casualties at military training centers.” In: *Arch. Indust. Health* 16.4, pp. 302–16.
- Young, A (2017). *Universal Thermal Climate Index*.
- Zare, Sajad et al. (2018). “Comparing Universal Thermal Climate Index (UTCI) with selected thermal indices/environmental parameters during 12 months of the year”. In: *Weather and climate extremes* 19, pp. 49–57.

A Appendix

Home Team	Timestamp	True_Game	Capacity	UTCI	Stress Category
Valence FC	2022-08-29 18:00:00	0	49500	25.042688	No thermal stress
Cadix	2022-08-29 20:00:00	1	25033	21.906213	No thermal stress
Cadix	2022-08-29 21:00:00	0	25033	21.230341	No thermal stress
Valence FC	2022-08-29 21:00:00	0	49500	23.518121	No thermal stress
Valence FC	2022-08-29 22:00:00	1	49500	24.019098	No thermal stress

Table 4: Dataset overview

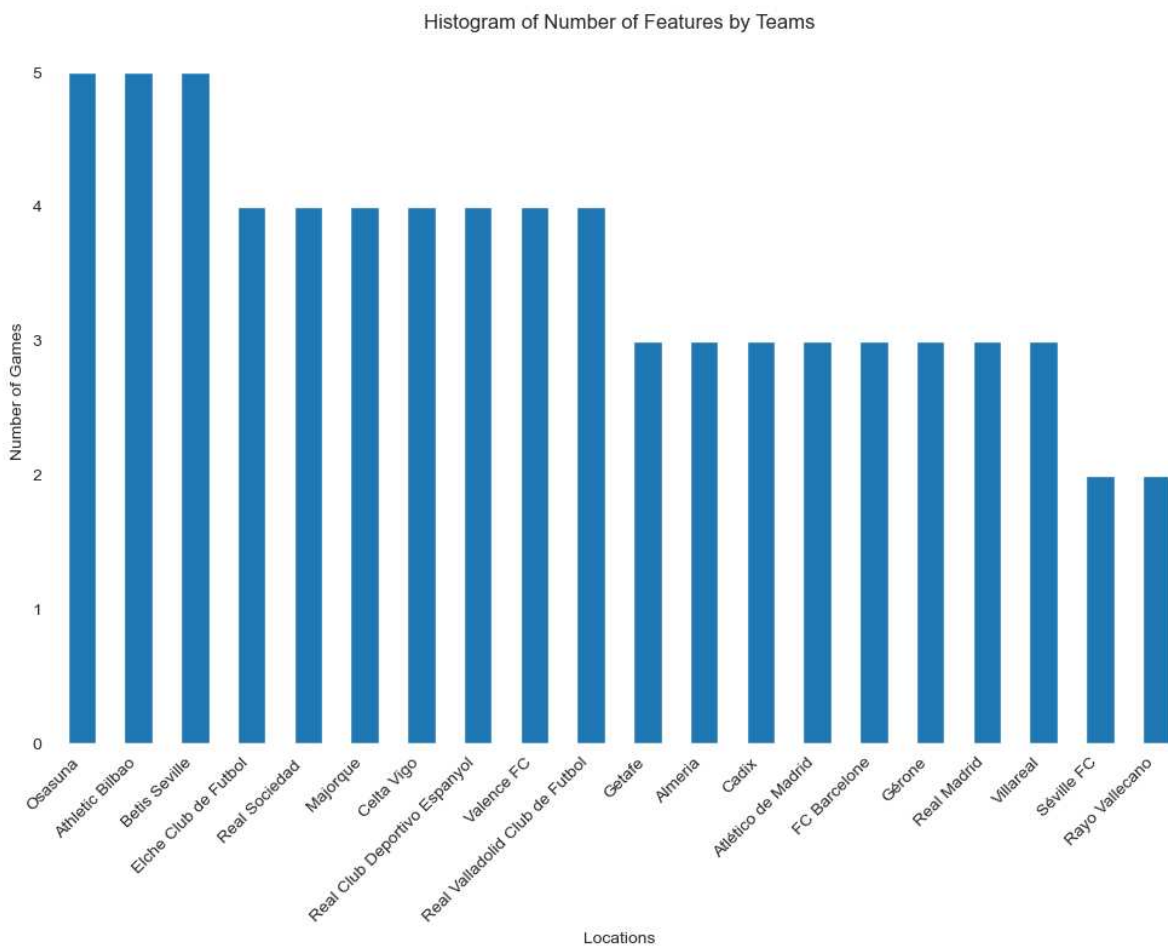


Figure 12: Number of Games by Location

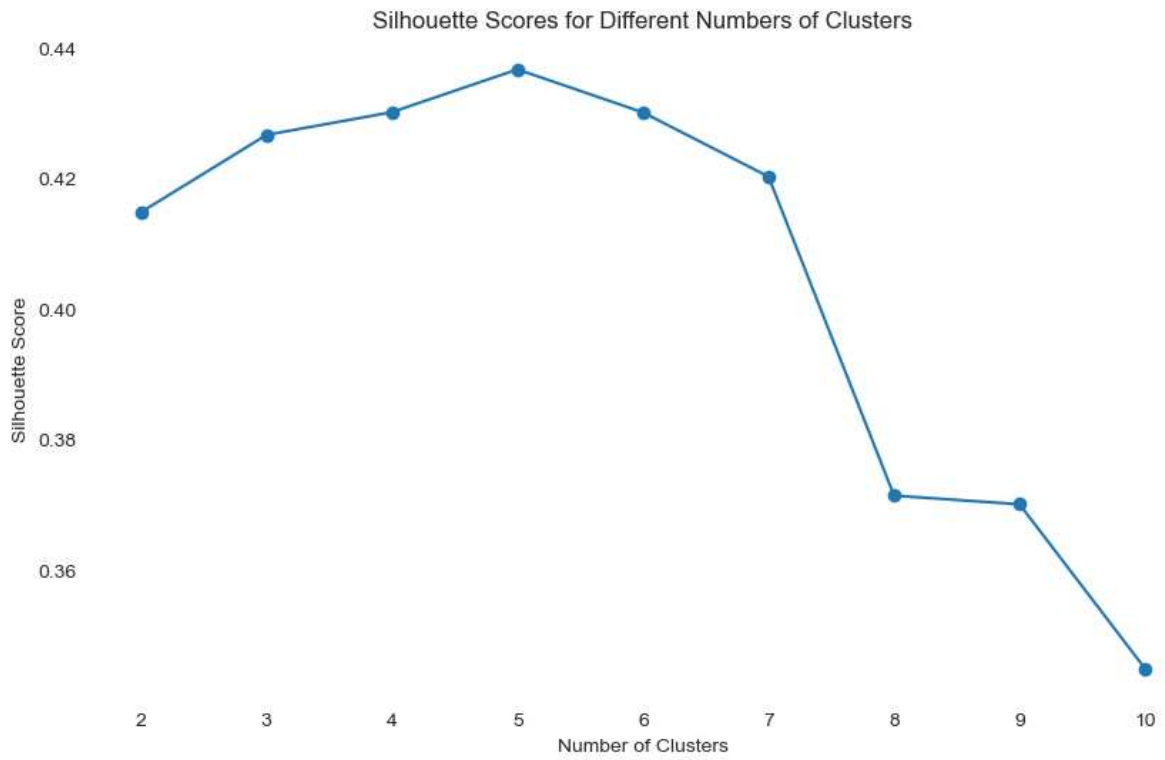


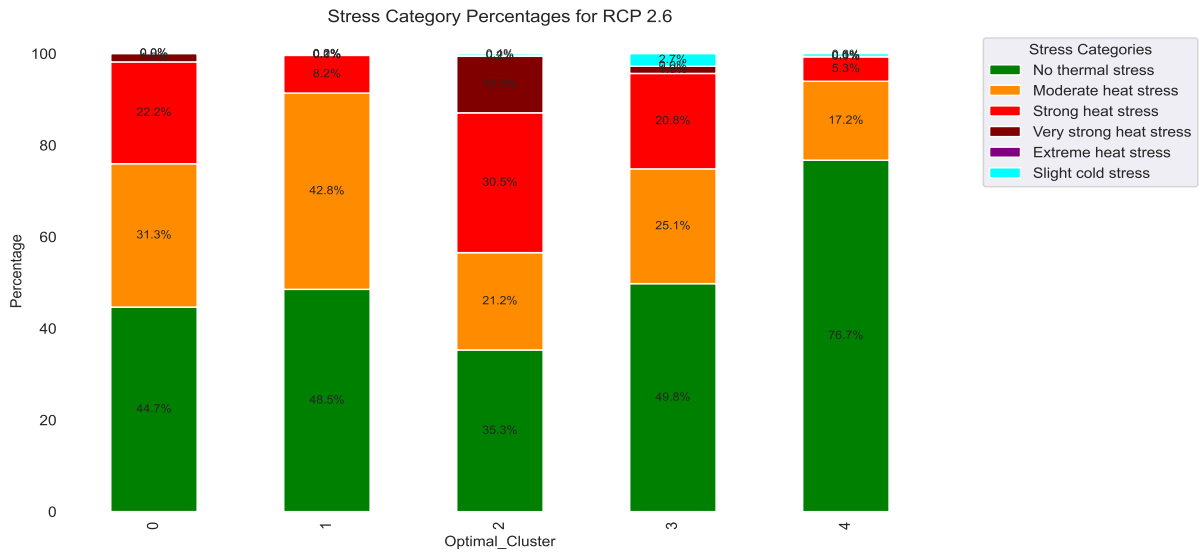
Figure 13: Silhouette Analysis

Time	Number of Games
14.00	6
16.25	6
17.00	1
17.50	6
18.50	7
19.00	11
19.50	6
20.00	4
21.00	14
21.50	1
22.00	9

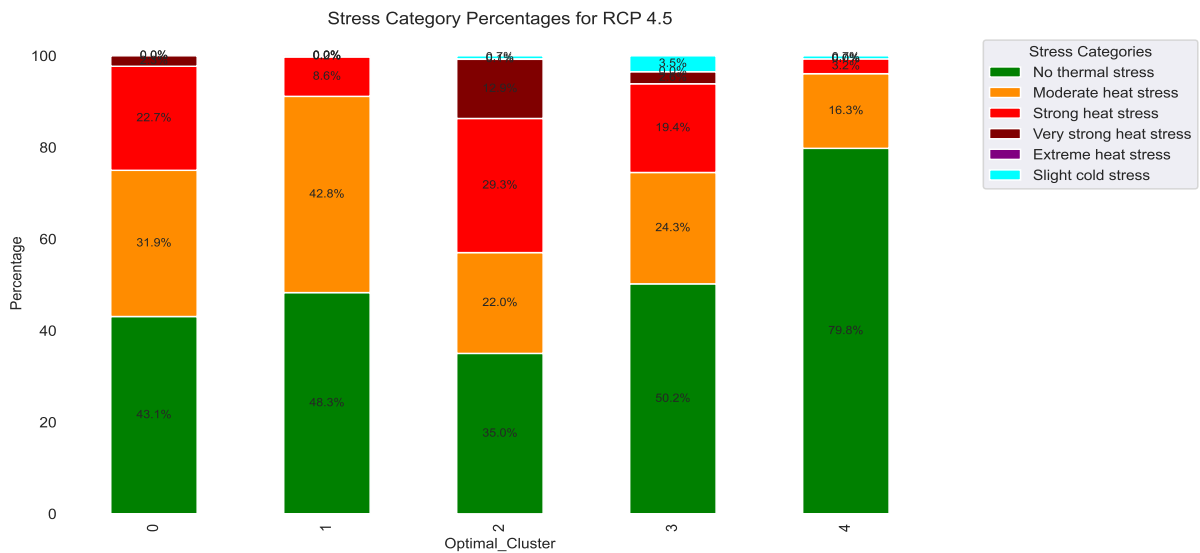
Table 5: Number of Games by Hour

Table 6: Standardized Residuals

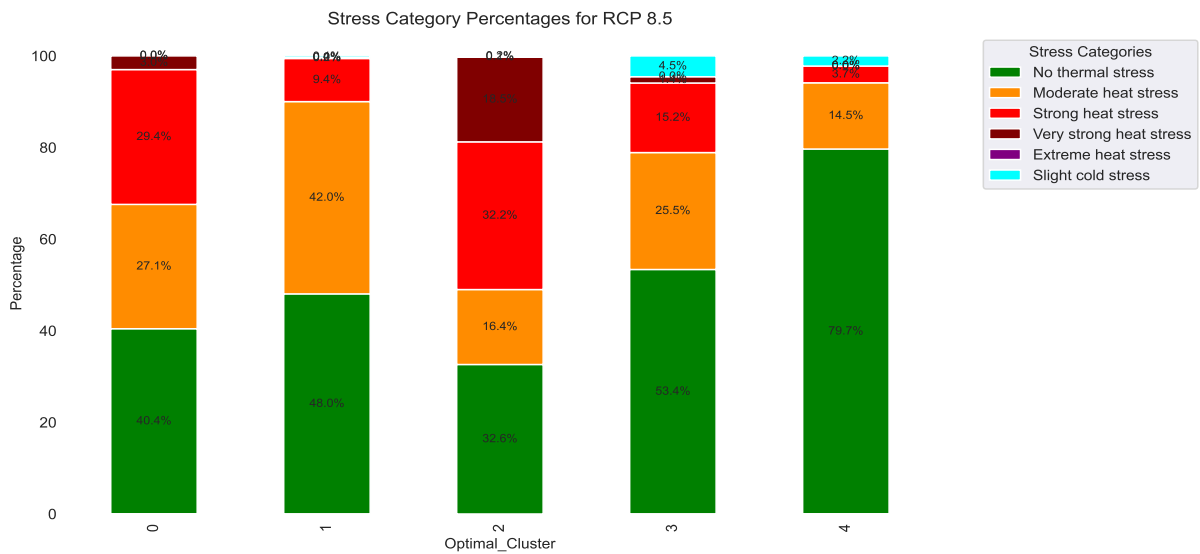
	Stress Category 26	Stress Category 45	Stress Category 85
No thermal stress	0.002210	0.005366	-0.007576
Moderate heat stress	0.029434	0.028864	-0.058299
Strong heat stress	-0.008946	-0.046326	0.055272
Very strong heat stress	-0.165595	-0.049839	0.215434
Extreme heat stress	0.50	-0.25	-0.25
Slight cold stress	-0.259036	-0.132530	0.391566
Moderate cold stress	-1.0	-1.0	2.0



(a) Percentages of Heat Stress Categories across Optimal Clusters under RCP2.6

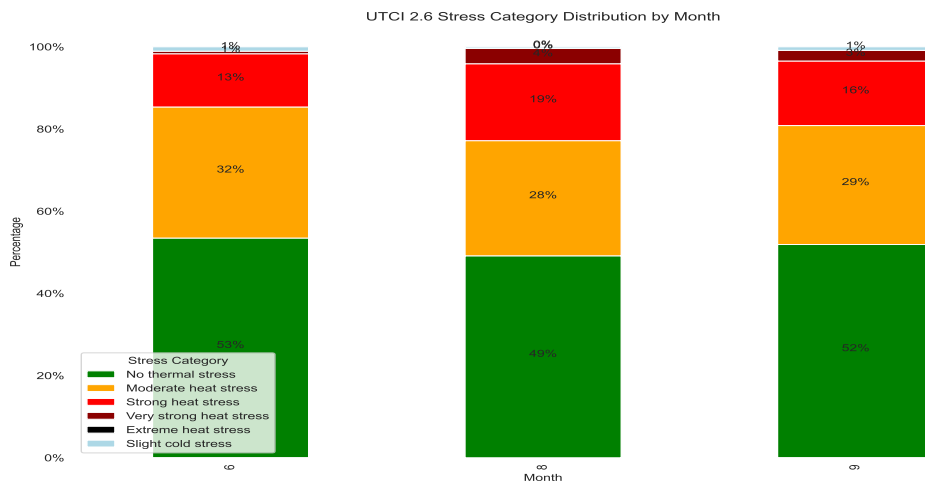


(b) Percentages of Heat Stress Categories across Optimal Clusters under RCP4.5

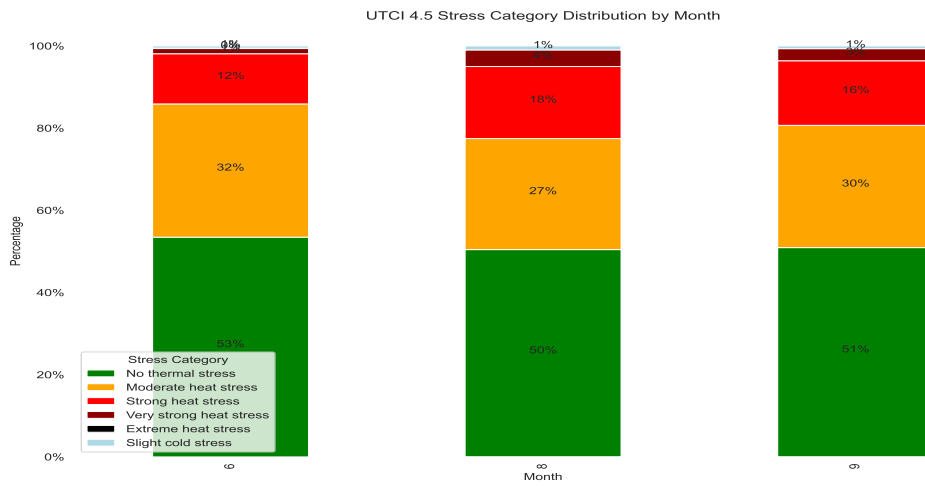


(c) Percentages of Heat Stress Categories across Optimal Clusters under RCP8.5

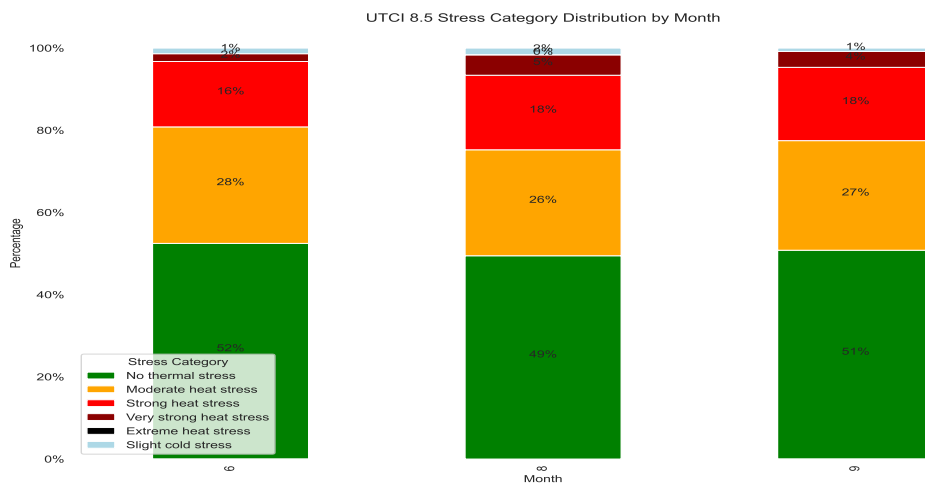
Figure 14: Percentages of Heat Stress Categories across Optimal Clusters under 3 different scenarios



(a) Percentages of Heat Stress Categories across Months under RCP2.6



(b) Percentages of Heat Stress Categories across Months under RCP4.5



(c) Percentages of Heat Stress Categories across Months under RCP8.5

Figure 15: Percentages of Heat Stress Categories across Months under 3 different scenarios

B Code

https://github.com/mathleng/impact_extreme_heat_spain/blob/main/final_code.ipynb