**The Impact of an Athlete's Home Country Weather and Climate on Soccer Performance**

Miguel Montesinos, Zachary De Aguiar, Daniel Dang, Mason Russo, Merzan Aymaz

**Detailed description of the research question**

Does an athlete's acclimation to their home country's weather conditions affect their performance when competing in differing weather environments in soccer matches? Furthermore, can understanding this correlation provide insights for optimization in athletic training and match strategy?

The motivation for this question is to understand how differences/similarities of weather features (temperature, humidity, wind speed, precipitation, altitude, and air quality) from their hometown versus the stadium can affect the performance of professional soccer players, and how coaches and trainers can use this information to prepare their teams for different climatic scenarios.

**A thorough literature survey with appropriate references**

The study from Périard, Racinais, and Sawka explores the topic of heat acclimation and its implications on competitive athletes and sports performance. They traverse various adaptations and mechanisms that cause heat acclimation while also explaining the physiological processes. The paper thoroughly analyzes the applications of heat acclimation on athletes, how it can enhance performance, and reduce the risk of heat-like illnesses. The authors offer a broad synopsis of the field; detailing benefits for heat acclimation, and offer practical considerations for athletes performing sports or physical activities in demanding environmental conditions. From the study, the following were found as conclusions, "Exercise heat acclimation induces physiological adaptations that improve thermoregulation, attenuate physiological strain, reduce the risk of serious heat illness, and improve aerobic performance in warm-hot environments." (Périard et al. 20-38)

From the study conducted by Aughey et al. (2013), the main focus was to examine the effects of a high-altitude environment on soccer players from both sea-level and high-altitude regions. These researchers observed both the physiological and performance adaptations when players were subject to training in reduced oxygen climates. In analyzing their performance, this study breaks down the challenges and adjustments soccer players face when forced to adapt to stress of high-altitude. The results of these findings not only contribute to our understanding of the effects of acclimatization, but offer insight to the specific demands and responses of soccer players in high-altitude environments. From the study, we reach two very important conclusions. "High altitude reduces the distance covered by elite youth soccer players during matches for sea level and high-altitude natives. Neither thirteen days of acclimatization nor lifelong residence at high altitude protects against the detrimental effects of altitude on match activity profile." (Aughey et al. i107-i113)

Valuable information is received from this particular article submission in the 'Science and Medicine In Football' journal. It summarizes the ideas and suggestions from multiple 'real world' studies, in particular, recent studies published within the last three years. For instance, ideas are introduced about mitigation approaches for environmental challenges. We can view an example of this with the FIFA World Cup in Qatar. According to the article, guidelines were introduced to provide players with "cooling" breaks and restricted kick-off times due to the extreme heat. There's multiple mitigation factors brought up in this article that repeat the idea that factors like this are actually taken into account for fixtures. Another example could be read,

“official competition is banned above a specific altitude, in the interest of player health and safety” (Sato, 2007).

In the research by Peñas and Ballesteros, they examined many performance factors to see how the location of a game will affect a team’s performance. They confirm that ”more successful technical and tactical indicators would be performed at home compared to away” (Peñas and Ballesteros 468–69). These indicators would include: goals, shots, passes, crosses, etc. This would mean that many characteristics that determine how well a player did in a game had increased rates playing at home compared to away. This was further confirmed when they found out that in “ the 2008-2009 season of the Spanish League, 61.95% of the games were victories for the home teams” (Peñas and Ballesteros 467). As the probability of winning in home territory is over 60%, it is fair to suggest that there is an advantage of some sort playing in the home stadium. While the research paper examines the increased statistics and probabilities at home, they do not examine what could possibly lead to this, especially nothing related to weather and climate. Thus, we use this paper to compare the probability globally and also to use the same type of statistics for comparisons between home and away.

McSharry’s paper reveals that physiological performance in international football games is significantly impacted by altitude, affecting both aerobic and anaerobic activities. Higher altitudes are correlated with decreased performance levels, marked by alterations in strategy, pacing, and overall physical exertion. This can be analyzed by pulling out an insert from the paper, it states, “The surprising result is that the high altitude teams also had an advantage when playing at low altitude, benefiting from a significant advantage over their low altitude opponents at all locations.” (NLOM). We can expand on the idea that it is feasible to calculate an advantage between two teams based on historical data. Although the paper highlights how altitude influences performance, a detailed exploration of how varying altitude levels impact teams originating from different base altitudes could offer nuanced insights. Furthermore, intertwining the effects of altitude with weather patterns and their collective impact on performance would expand the research framework, correlating it with a multifactorial approach toward understanding performance in different environmental contexts.

**Details of the algorithm(s) to be implemented**

- Multiple Linear Regression: This will be used to model the relationship between an athlete’s adaptation to their home country’s weather conditions and their performance in various weather and climate conditions. We will use this to be able to predict how athletes would perform based on different climate scenarios compared to their homes. The dependent variable would be the athlete’s performance in various weather environments, whereas the independent variable would be the athlete’s adaptation to their home climate.
- ANOVA: We can categorize 3 types of regions for use in ANOVA, such as tropical, temperate, and arctic, where they would have high temperatures and high humidity, moderate temperatures and humidity, and low temperatures and humidity respectively. From there, we can compare the performance of the soccer athletes in different regions to determine whether the adaptation to one region could lead to an advantage or disadvantage in another region or even the same region.
- Random Forest Model: Random Forest will be the primary predictive model. Using an ensemble of decision trees, we aim to capture the non-linear relationships between an athlete's home country weather conditions and their performance in different

environments. It will also help in feature importance to determine which weather conditions have the most significant impact on performance. The model will be evaluated based on accuracy, precision, recall, and the F1 score.

**Expected experiments and analysis to be performed:**

1. Correlation between an athlete's home weather conditions and their performance under different conditions using the Random Forest model.
2. Feature importance analysis using the Random Forest to determine the most influential weather parameters.
3. Comparative analysis of performances among athletes from similar climatic regions.
4. Identify potential outliers
    a. Athletes that defy the general trend and perform consistently regardless of the climate change
5. Evaluating the model's performance and fine-tuning it as necessary.

**List of datasets to be used for the experiments:**

- https://github.com/statsbomb/open-data: The StatsBomb Open Data repository holds a variety of soccer-related datasets. These datasets contain detailed player statistics, team statistics, and event data from soccer matches. Users accessing the sets can expect to find player performance, team performance, player information, match locations, managers, referees, etc.
- https://fbref.com/en/: FBREF is a data website dedicated to tracking soccer players and team from around the world. Created by Sports Reference, users can access data from both mens and womens players, clubs, soccer leagues, cups, and countries. Each of the aforementioned categories go into additional detail that cover all of the player or team's history.
- https://www.whoscored.com/: WhoScored provides “Betting” data to the public which can be used to get a gist on the historical performance of a team at a specific location.
- https://www.wunderground.com/: Weather Underground has provided historical data to the public as a part of a global project since 1993. The database available to the public goes back a number of years with precise data such as temperature, wind speed, precipitation, and hourly updates of weather conditions (“cloudy”, “mostly cloudy”, and “heavy t-storm”).
- https://www.ncei.noaa.gov/cdo-web/: Climate Data Online (CDO) is a free service provided by the National Climatic Data Center (NCDC) that provides access to historical weather and climate data, including temperature, precipitation, wind, and degree days, as well as radar data and 30-year Climate Normals. Data can be searched given a time period and location (with zip code, city, etc.).

**A timeline of significant steps in the project**

- Week 1: Literature review and dataset acquisition
- Weeks 2-3: Data preprocessing and merging athlete and weather datasets
- Week 4: Weather adaptation analysis and initial findings
- Weeks 5-6: Regression model development
- Week 7: Visualization and interpretation of results
- Week 8: Final analysis and revisions

**Bibliography**

Aughey, Robert J., et al. “Soccer Activity Profile of Altitude Versus Sea-level Natives During Acclimatisation to 3600 M (ISA3600).” *British Journal of Sports Medicine*, vol. 47, no. Suppl 1, BMJ, Nov. 2013, pp. i107–13, doi:10.1136/bjsports-2013-092776.

Illmer, Sarah, and Frank Daumann. “The Effects of Weather Factors and Altitude on Physical and Technical Performance in Professional Soccer: A Systematic Review.” *JSAMS Plus*, vol. 1, Elsevier BV, Oct. 2022, p. 100002, doi:10.1016/j.jsampl.2022.100002.

McSharry, Patrick E. “Effect of Altitude on Physiological Performance: A Statistical Analysis Using Results of International Football Games.” *BMJ*, vol. 335, no. 7633, BMJ, Dec. 2007, pp. 1278–81, doi:10.1136/bmj.39393.451516.ad.

Peñas, Carlos Lago, and Joaquín Lago Ballesteros. “Game Location and Team Quality Effects on Performance Profiles in Professional Soccer.” *Journal of Sports Science and Medicine*, vol. 10, no. 3, National Institutes of Health, Feb. 2011, pp. 465–71.

Périard, Julien D., et al. “Adaptations and Mechanisms of Human Heat Acclimation: Applications for Competitive Athletes and Sports.” *Scandinavian Journal of Medicine & Science in Sports*, vol. 25, Wiley-Blackwell, May 2015, pp. 20–38, doi:10.1111/sms.12408.