Modelling the Effect of Adverse Weather Conditions on Quarterback Performance in the National Football League

Atul Venkatesh[a], **Nelson Mendels**[a], **Billy McKittrick**[a], **and Niccolo Campolo**[a]

This manuscript was compiled on November 27, 2024

How much does extreme weather shake the performance of top NFL quarterbacks? We examine 10 years of NFL play-by-play data in order to analyze the effects of temperature, precipitation, and wind on the Expected Points Added (EPA) of quarterbacks, culminating in our own metric called EPA over expected (EPAOE). We find that there is significant effects that wind, temperature, and rain cause on quarterback quality.

NFL | Football |Weather | EPA

The date is December 8, 2013. The 7-5 Detroit Lions are taking on the 7-5 Philadelphia Eagles in a pivotal inter-conference matchup with potential playoff implications. While this would seem like a traditional afternoon game, there was one key difference: a blizzard was attacking the Lincoln Financial Field in Philadelphia. With snow pounding the surface, it was hard enough to see, let alone play the game of football. The game started off sloppy, with fumbled snaps and kickers/punters struggling to plant correctly. It became abundantly clear that the winner would be the team who could handle the conditions. In the end, the Eagles prevailed, winning the game 34 points to 20.

While an above average number of points were scored, the performance of both starting quarterbacks were notably underwhelming. Lions quarterback Matthew Stafford completed 10 of his 25 passes for a measly 151 yards - well below his expectations. Eagles quarterback Nick Foles was not much better, completing 11 of his 22 passes for 179 yards. Taken out of context, these quarterback performances were awful. But was it surprising? With several inches of snow on the ground and even more in the air, it would be reasonable to expect quarterbacks to not play well. But how much worse should they be expected to play? And more importantly, how can we quantify this?

This serves as the central research question to our paper, where we create a model to predict quarterback performance in adverse weather.

Literature review

Weather has been shown to have a direct impact on athletic performance, including American football. This section will begin with a description of the various literature on this phenomenon as it relates to other sports, namely marathon running, and proceed towards an explanation of the existing literature on weather conditions and NFL performance.

Significance Statement

Weather impacts quarterback play but it has rarely been quantitatively analyzed. This paper provides a holistic analysis on player performance at the individual and positional level. Our analysis has the potential to impact strategy and decision-making across the league. **Link to github**

Author affiliations: [a]Department of Quantitative Social Science, Dartmouth College

Effect of weather on performance in marathon running . Elements of weather have been demonstrated to have an observable effect on performance in various sports. Both of the following studies we reviewed suggest a significant positive correlation between marathon performance and air temperature. Firstly, in analyzing the results of six European and six American marathon races from 2001 to 2010, Nour et al. in 2012 attempted to demonstrate the effect of several variables involved in environment—namely, temperature, humidity, dew point, and atmospheric pressure at sea level, as well as the concentration of pollutants—on the performance of long-distance runners (1). Additionally, it was found that temperature was the most important factor in marathon performance, as these were found to be correlated through the quadratic model they used, whereas the other variables were not correlated to a level of significance (1). Similarly, another study attempted to examine the effects that air temperature, relative and specific humidity, wind speed, solar shortwave radiation, thermal longwave radiation, and rain have on marathon performance by taking the results of the Stockholm Marathon from 1980 to 2008, finding through a linear regression analysis that there is a significant correlation between the number of non-finishers and both air temperature and specific humidity (2). Considering the level of endurance called for in success with American football, studies such as these give cause to infer the existence of such an effect in this sport as well, as the following studies suggest.

Effect of weather on performance in American football . In a 2021 NFL game between the New England Patriots and the Buffalo Bills, Patriots quarterback Mac Jones completed a mere three passes, a departure from his yearly single-game averages. This occurred in the presence of extreme weather conditions as a high-wind storm raged on field. Despite this, and despite the high-performing Bills defense, the Patriots narrowly won the game 14-10 (Sporting News, 2022) (3). Such anecdotes motivate research into the effect of extreme weather on player and team performance. Joly and Dik in 2021 examined the effect that weather has on NFL performance for playing at home, using data for winning percentages from 2002 to 2020 and dependent samples paired differences tests, The authors found that there is a significant effect of cold weather on performance for home teams in the NFL. Cold weather during the winter months is associated with a significant advantage, leading them to reject their null hypothesis of "no effect" (4). Another study sought to determine the effect that atmospheric conditions have on both the performance of NFL teams and that of the betting market, the latter of which can likewise provide direct insight on NFL performance (5). Specifically, data on atmospheric conditions such as temperature, humidity, precipitation, wind speed, barometric pressure, and altitude were employed in a regression analysis on data from game outcomes and the betting market, showing that humidity positively affected scoring performance, precipitation and wind speed negatively impacted scoring, and the variables of temperature and barometric pressure had weaker or inconsistent effects (5). It is worth mentioning that this paper, unlike our study, sought to find statistical significance between overall team performance and weather affects, rather than for passing statistics alone. That being said, a paper more closely aligned with the nature of our study sought to find a relationship between the variables of temperature and wind on NFL passing and rushing performance, using ordinary least squares (OLS) regression with heteroskedasticity-robust standard errors to come to the finding that wind speed had a negative impact on passing (specifically, that a 10 mph increase in wind speed reduced passing yards by 6.8% and completion by 2.4%), and a positive impact on rush attempts and yards (6). Additionally, it was found that lower temperatures led to a reduction in pass performance by 1.4% per 10 degree Fahrenheit decrease and an insignificant impact on rush performance. The varied literature on weather parameters and their relationship with NFL performance, including pass performance, serves to motivate our study as we seek to fill the gaps in this literature with our own statistical analyses (6).

Data

To conduct our study, we used a Python library nfl-data-py in order to work with data and imported functions from a variety of sources that gives us play-by-play data, descriptive info on players, and basic statistics for a given play such as passing yards (7). The variable we were most interested in is Expected Points Added (EPA), an advanced statistic that shows how much a player contributes to team success. This is the gold standard for player performance and we concluded that it would be an appropriate way to measure player success, since it was included in the dataset. The dataset was imported to analyze our research question using data from the past ten years (2013-2023). The documentation for this data set can be found here (7): https://pypi.org/project/nfl-data-py/

Key Observation and Variable Descriptions. The data set includes 481,740 observations and 392 unique columns. Each observation represents a specific play conducted in NFL games from the past 10 years, with relevant variables to this study including: **pass_attempt**, a boolean that says whether or not the play that occurred was a pass or not; **passer_player_name**, the name of the player who passes the ball in a passing play, **epa**, and **weather**, which gives description of the weather conditions during the play in the form of: '[Weather Type] : [Temperature]° F, Humidity: [Humidity percent], Wind: [Wind Speed and Direction]'. The descriptions of the weather can be *tokenized* in order to analyze the different effects of the weather's elements on the performance and success of pass attempts.

Data filtering and cleaning.

Pass attempts. Weather particularly has an effect on passing plays in football. Furthermore, such plays are almost entirely led by and attempted by the same player: the quarterback. This means consistent observations can be observed under pass attempts rather than running attempts, which can vary on which player contributes to the run. Because of this and the scope of this research, we filter the data frame, pbp, such that there are only observations of plays that were pass attempts: pbp = pbp[pbp["pass_attempt"] == 1].

Weather tokenization. The first step in processing this weather data involved categorizing the conditions into broader weather categories. This was accomplished by tokenizing the text in the "weather" column, extracting the first word (the weather condition), and creating binary variables to represent whether a particular type of weather condition was present. Specifically, we defined three weather categories:

1. **Sunny Conditions**: Represented by words such as "Sunny," "Clear," or variations thereof.
2. **Cloudy Conditions**: Captured by keywords or misspellings like "Cloudy" or "Cloudly."
3. **Precipitation**: This category was created to encompass any mention of rain, snow, thunderstorms, or scattered showers.

For each of these categories, we created binary flags (1 for the presence of the condition, 0 for its absence). The final dataset includes three new binary columns: is_sunny, is_cloudy, and is_precipitation, representing the weather conditions during each play. Any games that did not fit these three categories were ommitted. **Table 1** shows some basic statistics for all of the variables. Cloudy and sunny weather, to no surprise, are the most prevalent while rainy weather is a lot more rare.

Category	Percent	Count
is_sunny	0.250	50795
is_cloudy	0.380	77060
is_precipitation	0.052	10757
Omitted	0.320	64911

Table 1. Weather categories and their respective percentages and counts

In addition, we were able to extract the temperature and wind speed from our weather variables. Both variables are continuous integers present in every observation.

Pre-model data analysis

As mentioned earlier, EPA will be used as the measurement of quarterback performance. In this section, we will first look at the average EPA per play in each of the weather situations before conducting an OLS regression to confirm the associations between each weather condition and EPA.

Introductory analysis. **Table 2** details the average EPA per play in each of the three weather conditions. This can give us a general idea of what we can expect from our regressions.

Category	EPA
is_sunny	0.030
is_cloudy	0.022
is_precipitation	-0.044

Table 2. Average EPA in each weather condition

As we can see, quarterbacks perform pretty similarly in sunny and cloudy weather. However, in rainy and snowy weather, their performance plummets to a **negative EPA**. This tells us that on average, quarterbacks are likely to perform in a way that **loses** the team expected points than not.

We additionally wanted to see how quarterbacks perform in cold and windy weather. **Table 3** summarizes this. We defined cold weather as below freezing conditions. Windy weather was where winds exceeded 15 miles per hour. It is worth noting that being windy or cold is not mutually exclusive with being sunny/cloudy/rainy.

Category	EPA
is_cold	-0.016
is_windy	-0.018

Table 3. Average EPA in cold and windy weather

Unsurprisingly enough, in both cold and windy conditions, the quarterback performance drops pretty significantly. For our actual model, we will treat temperature and wind as continuous.

Regression. We conducted three separate introductory regressions. One each for player performance in sunny weather, cloudy weather, and rainy weather. There is no need to worry about these variables confounding each other because the weather can only be put in one of the three categories. In each regression, our continuous wind and temperature variables were added as supplementary variables. The outcome variable was a given player's EPA **per game** since that provides a more interpretable statistic. Wind and temperature were simply the average throughout the game. **Table 4** displays the coefficients of the OLS regression in sunny weather.

Variable	Coefficient	Std Error	P value
Intercept	-0.432	0.545	0.428
is_sunny	0.383	0.270	0.157
wind***	-0.083	0.026	0.001
temperature***	0.025	0.008	0.001

Table 4. Sunny regression results. Stars (***) indicate significance at the 0.001 level.

Holding the other two factors constant, the weather being sunny is associated with a 0.383 increase in a quarterback's EPA per game. In other words, a quarterback is expected to add 0.383 points to their team if the weather is sunny. This association is not statistically significant. Holding sunny constant, a one mph increase in wind speed is associated with a 0.083 EPA decrease and a one degree increase in temperature is associated with a 0.025 increase in EPA. Both wind and temperature had P values that siggested there was a significant effect that should be taken into account.

Table 5 displays the results of the cloudy regression.

Variable	Coefficient	Std Error	P value
Intercept	-0.504	0.554	0.363
is_cloudy	0.360	0.264	0.172
wind***	-0.089	0.026	0.001
temperature***	0.026	0.008	0.001

Table 5. Cloudy regression results. Stars (***) indicate significance at the 0.001 level.

These results are pretty similar to the sunny weather regression. Based on these results, it would be reasonable to conclude that there is no real difference in quarterback play if the weather is sunny or cloudy. The coefficients are both similar and have relatively similar errors.

Table 6 visualizes the coefficients of the rainy regression.

Variable	Coefficient	Std Error	P value
Intercept	-0.149	0.543	0.363
is_rainy***	-1.861	0.545	0.001
wind***	-0.084	0.026	0.001
temperature**	0.024	0.008	0.002

Table 6. Rainy regression results. Stars (***) indicate significance at the 0.001 level.

The coefficient associated with rainy weather is both large in magnitude and statistically significant, indicating that rain indeed has a negative effect on player performance. It also makes intuitive sense that wind is negative and temperature is positive. We expect player performance to decrease in windy weather and increase in warmer weather. Interestingly, rain was the only binary variable that showed statistical significance.

The purpose of the regressions was to determine associations between the predictor variables and the outcome. This will provide valuable context heading into the model section of the paper. The key takeaways are that wind, rain, and temperature all play key negaitve factors, while sunny and cloudiness play similar positive factors that aren't as significant statistically.

Model

Since our goal was to predict EPA based on weather, we would need a model that would look at the patterns of certain weather conditions and the resulting EPA in order to make predictions. A gradient boosting algorithm seems like the most appropriate machine learning technique to do this. We settled on conducting an XGBoost - a proper algorithm for predictive modeling in sports. In an XGBoost, weaker predictive models are iterated over each other to create a strong model that minimizes the overall residual error.

Our label variable was the player's EPA per game and our predictor variables were the different weather situations - five binary variables that told us if the weather was sunny, cloudy, rainy, windy, or cold and two continuous variables that measured the temperature and wind speed respectively. We randomly split our model into training and testing with an 80 to 20 split. With that, our model was ready to be processed.

When conducting an XGBoost algorithm, we needed to specify a few parameters in the model. These parameters were specified with the goal of reducing the overall root mean-squared error. The maximum tree depth was three, the learning rate was 0.14, the number of parallel threads to run the XGBoost was four, the number of rounds for boosting was 500, and the early stopping rounds parameter was one. A combination of k-fold cross validation and trial/error was used to fine tune the parameters.

Results

Model Evaluation

Now that the model was conducted, we took a holistic approach in evaluating it. **Figure 2** displays the importance of the key input variables to the model by gain. This represents how much each factor contributes to the model's overall accuracy.

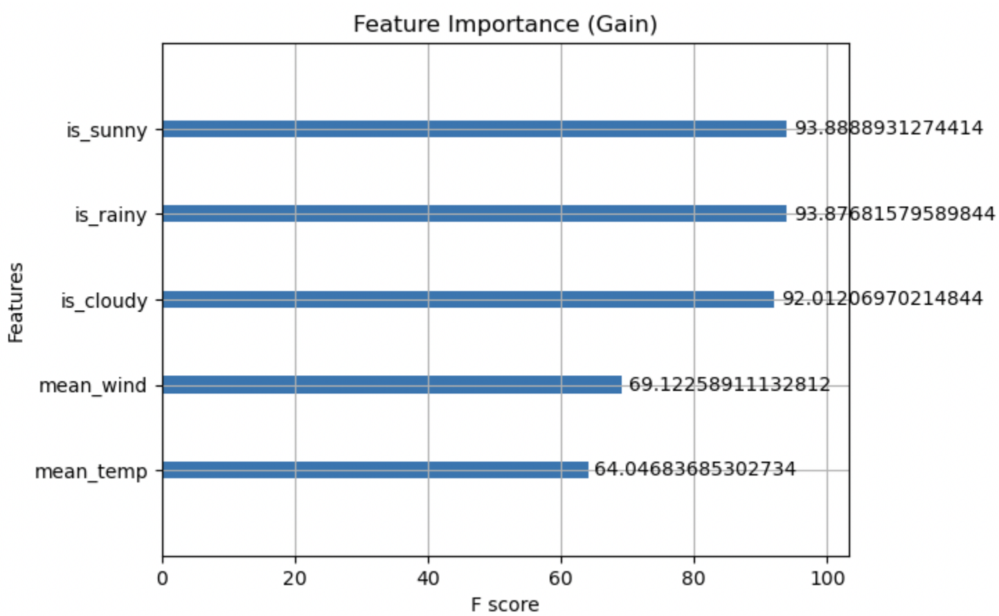

Fig. 1. Importance of Features in XGBoost Model Based on F Score

Since the values are relatively close to each other, no one variable impacted the model more than others. The score is simply the "weight" or the extent to which each feature contributed to the overall accuracy of the model. Binary variables, such as the state of weather, likely played a larger importance because they are the most broad and applicable to the situations derived and branched by the model, whereas minor variations in the continuous variables didn't contribute to a greater marginal addition to the accuracy of the model. It is worth noting that the continuous variables might be used more in the model simply because they appear in every observation. In contrast, a rainy game only appears in five percent of observations.

Figure 1 displays the learning curve of the model given inputted data from the NFL data set compiled and cleaned.

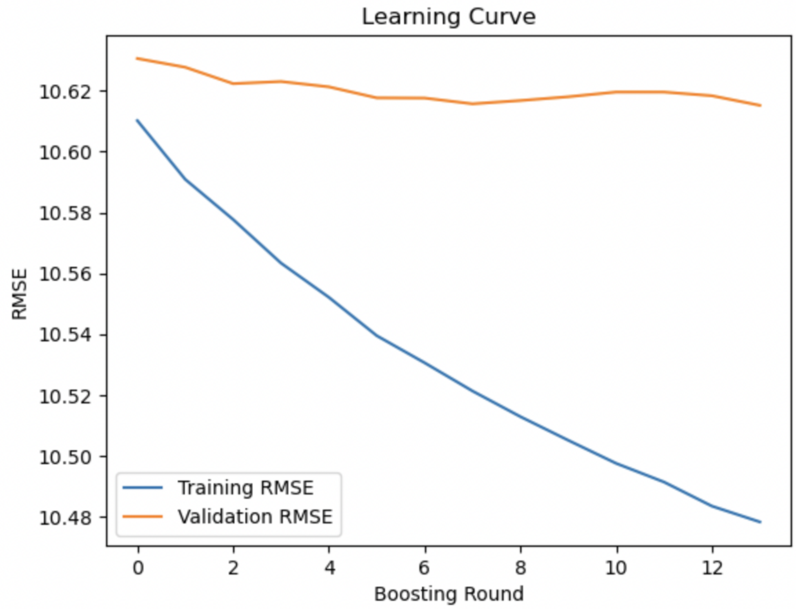

Fig. 2. Learning Curve of Model Given Training and Input Data

Given the training data, for each boost and update of the model there was a clear decrease in the root mean squared error (RMSE) of the models prediction to the actual value. This typically decreases as any model learns, because it gets better at fitting the training data. Comparatively, The RMSE calculated on a validation set (or the testing data) is relatively high, likely due to the fact that the data was less familiar. The model continues to try and generalize its results, and the slight decrease in RMSE is a strong sign that the model outputs generalizable results.

Figure 3 compares the model's predicted EPA with the observed EPA.

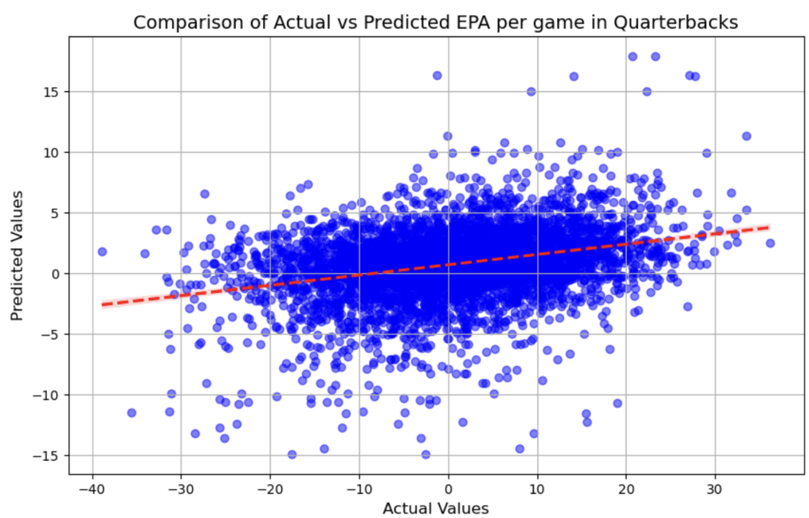

Fig. 3. Visualizing expected EPA vs. actual EPA

While there is a lot of noise, there is a pretty clear positive association, indicating the model has at least some sort of predictive power. The model had an overall mean absolute error of **8.70**, a mean squared error of **121.46**, and a root mean squared error of **11.02**. While these numbers are high, it makes sense in the context of our problem. We are trying to predict quarterback performance based on the weather. There are several factors that go into a quarterbacks performance - offensive situation, defensive strength, playcalling, injury, etc. The purpose of this paper is not to perfectly predict EPA. Rather it is to see if quarterbacks over perform or underperform our estimates, allowing us to create a metric to quantify that.

Analysis

Taking the difference of the observed EPA per game and the predicted EPA per game can do more than simply evaluate our model. If a player has a positive residual (observed value was higher than the prediction), the player outperformed their expected output and had a good game. Since our expectations have now been adjusted for weather, we can take the average residual for every game in a player's career. Ranking this value can give us a **weather-adjusted quarterback ranking**. We called this average residual *EPA Over Expected*, or **EPAOE** for short. **Table 7** shows the top players in terms of the weather-adjusted EPAOE with their average EPA per game included.

Passer Player Name	EPAOE	EPA	Attempts
P. Mahomes	8.054	8.600	109
T. Romo	7.206	8.899	14
B. Purdy	6.760	8.200	27
T. Brady	5.473	6.356	151
A. Rodgers	4.784	5.817	125
D. Brees	4.174	5.921	92
J. Love	3.734	3.831	18
J. Garoppolo	3.694	4.653	63
D. Prescott	3.163	3.477	69
B. Roethlisberger	3.147	4.238	112

Table 7. Top 10 quarterbacks by EPAOE

The EPA rankings are pretty comparable to EPAOE, indicating our metric passes the ”eye test”. By this, it means that from an initial review, the model shows promising ability to predict if a player will over or under perform. What happens if we subset by weather condition? Since we concluded that sunny and cloudy weather weren't major predictors of success, we chose to focus on games with extreme weather conditions. **Table 8** looks at the 10 best quarterbacks by EPAOE under rainy conditions.

Passer Player Name	EPAOE	EPA	Attempts
D. Brees	11.055	13.282	5
T. Brady	7.418	6.041	12
R. Tannehill	4.755	2.792	5
D. Prescott	3.059	-0.762	5
B. Roethlisberger	2.968	0.622	8
A. Rodgers	2.220	2.289	8
R. Wilson	2.209	-0.059	12
J. Allen	1.775	-0.540	12

Table 8. Top 10 quarterbacks by EPAOE - Rainy weather

There seems to be a pretty steep drop off in quality following Drew Brees and Tom Brady. This makes sense since we do not expect a player to outperform their expectations in the rain. It is also worth noting that sample size might be an issue as some players may predominantly play in domes or drier areas. Playing in the rain is quite uncommon, therefore the extent to which these results are precise is suspect. The deviations from the expected value are likely caused by the the minimal sample size in both the model and the data set (only 5.2% of observations were considered having precipitation).

Table 9 visualizes the top 10 quarterbacks in cold weather (below freezing).

Passer Player Name	EPAOE	EPA	Attempts
P. Mahomes	7.976	8.794	95
T. Romo	7.874	9.296	13
B. Purdy	6.760	8.200	27
T. Brady	5.274	6.174	138
A. Rodgers	4.479	5.602	105
D. Brees	4.174	5.921	92
J. Garoppolo	3.761	4.850	62
B. Roethlisberger	3.277	4.484	105
P. Rivers	3.264	4.441	105
T. Tagovailoa	3.133	3.899	49

Table 9. Top 10 quarterbacks by EPAOE - Cold weather

The results are quite similar to the overall top ten quarterbacks, Table 7. Likely it is the case that players will consistently over or under perform in cold weather rather in the whole league rather than have effects on particular players. It is worth noting that there is a general trend that players who are more comfortable or have more experience in cold weather seem to perform better.

Table 10 visualizes the top quarterbacks in windy weather.

Passer Player Name	EPAOE	EPA	Attempts
P. Mahomes	10.328	8.877	15
D. Brees	9.558	14.395	10
J. Goff	6.856	5.984	7
T. Brady	6.478	7.018	13
T. Tagovailoa	6.350	3.194	7
J. Burrow	5.721	4.772	6
L. Jackson	5.298	4.003	7
R. Fitzpatrick	4.554	2.606	13
A. Luck	3.650	4.480	5
J. Allen	3.581	2.132	15

Table 10. Top 10 quarterbacks - Windy weather

It makes a lot of sense to see Patrick Mahomes, one of the best quarterbacks in the 21st century, at the top of the latter two tables. Arrowhead Stadium, the home stadium for the KC Chiefs where he plays half of his games, has its fair share of cold and windy weather. Furthermore, Mahomes is known to perform at an elite level regardless of weather. Can that be said for other players? **Figure 4** visualizes how the top quarterbacks differ by their EPAOE in different situations.

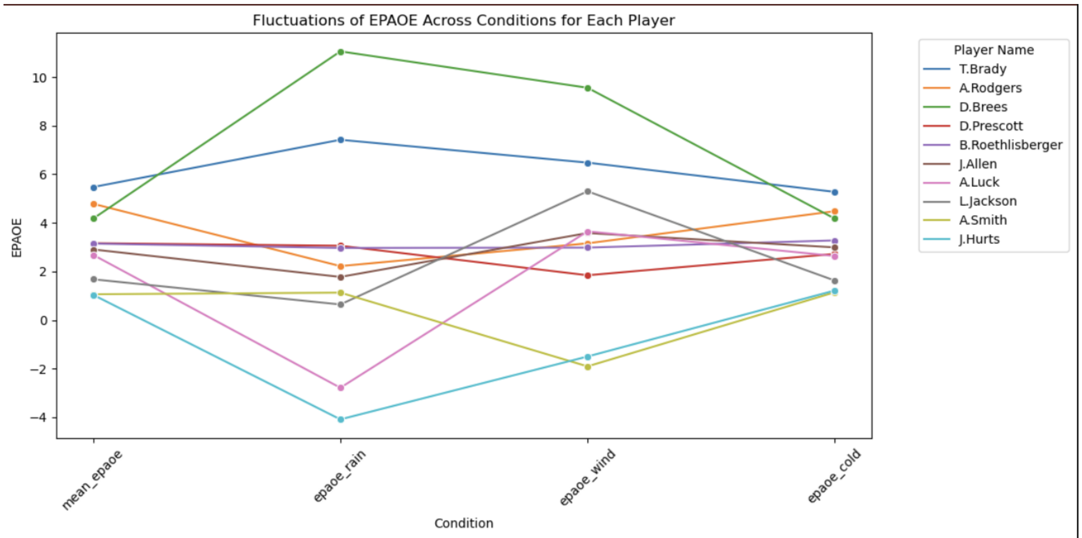

Fig. 4. How EPAOE deviates in different weather conditions for different players: Top 10

Players like Drew Brees and Tom Brady **thrive** in adverse weather circumstances while players like Jalen Hurts and Alex Smith are significantly worse in rainy and windy weather. Questions of sample size might prevent us from generalizing these results, but it is worth noting that negative weather conditions can have a major effect on quarterbacks. And this effect is not the same for every player.

Limitations

While this project has been useful in defining the relationship between weather and QB performance, it is important that we address our model's limitations. First, only 5.2% of games in the dataset had precipitation, significantly limiting the ability to analyze rare weather events like heavy rain, snow, or extreme wind. The underrepresentation of such weather conditions slightly hindered our ability to quantify their true impact on QB performance. Additionally, we chose to exclude player-specific context and other unaccounted variables such as injuries sustained during the game, strength of opposing defense, or prior experience in adverse weather. The lack of these variables in our analysis can introduce confounding effects, making it difficult to accurately isolate the relationship between weather and QB performance. Another limitation is using EPA as a proxy for performance. Although EPA is a standard metric used in football analytics, it is influenced by many team-level factors, such as receiver's ability to catch the QB's pass. The reliance on EPA to entirely capture performance could potentially skew results. Despite these limitations, many of these challenges are inherent to sports performance analysis as a whole, and we believe they don't fully diminish the valuable insights that we still were able to uncover.

Future Research Directions

Future research can further refine our data analysis and expand our understanding of weather's influence on QB performance. By simply expanding the dataset to cover more NFL seasons or combining college football data and data from international leagues with more extreme weather, we can provide greater context for weather's impact. Furthermore, future models could integrate more situational variables as highlighted in the limitations, including time remaining, field position, game score and more. These variables could provide greater context for QB decision making and performance in various situations for example a QB's lower EPA in the fourth quarter could be explained by riskier play calls to close a scoring gap, which would be independent of weather effects. An exciting future application that could greatly benefit individual NFL teams in player evaluation is the use of player-specific models. Such models could predict individual QB's performance trends in specific situations using greater and more specific historical data for each player. There are definitely opportunities to extend this research to other sports (something we already saw in the literature review) such as for baseball (pitching under wind) or soccer (passing and goalkeeping in rain) which could provide a comparative understanding of weather's influence on athletic performance across sports. Finally, this analysis could be used to create tools that leverage weather-based performance trends to offer actionable recommendations for fantasy football lineups, or to help sports bettors make more informed decisions by incorporating weather impacts into their analysis. Overall, exploring these future applications will open the door to innovative applications that can transform how teams, players, and fans engage with the game.

Conclusion

Adverse weather conditions are known to affect quarterback play. However, the extent was relatively unknown. Our paper attempts to quantify the impact of weather on quarterback performance. We understand that predicting player success solely based on weather will lead to error and residuals of high magnitude. However, these residuals can give us insight into how well a player performs when conditions are not ideal.

Mere days before the completion of this paper, the Cleveland Browns and the Pittsburgh Steelers faced each other in a rendition of the original "Blizzard Bowl". With snow transforming Cleveland into a winter wonderland, the two-win Browns, led by Jameis Winston, managed to upset the eight-win Steelers in a fairytale back-and-forth affair.

Did Winston and Steelers quarterback Russell Wilson outperform their expectations? With an average temperature of 36 degrees and an average wind speed of 13 miles per hour, our model projected Winston and Wilson to each have an EPA of **-2.154**. Despite the win, Winston produced a fairly pedestrian statline, completing 18 of 27 passes for 219 yards and an interception. He had an EPA of **-5.81**, underperforming expectations. Russell Wilson, on the other hand, was statistically dominant, posting a final statline of 21 completions, 28 attempts, 275 yards, and one touchdown. His final EPA for the game was **7.81**, meaning he significantly outperformed his expectations. Wilson's performance looks a lot more impressive considering the low outlook on the game.

Despite the significantly worse quarterback performance, the Browns still managed to come out on top. Perhaps this is a cruel twist of fate for the Steelers. Or an early christmas miracle for the Browns. Or maybe this exposes the limitations of EPA. One thing is for certain: when put in unconventional environments, quarterback performances vary significantly - even in a single game. And we believe our model can provide context for each of these performances.

References

1. N El Helou, et al., Impact of environmental parameters on marathon running performance. *PloS one* **7**, e37407 (2012).
2. T Vihma, Effects of weather on the performance of marathon runners. *Int. journal biometeorology* **54**, 297–306 (2010).
3. S News, Mac jones stats vs. bills: Patriots qb struggles with worst outing of 2023 season. *Sporting News* (2022).
4. B Joly, M Dik, Cold weather teams in the national football league and home-field advantage. *Proc. Int. Math. Sci.* **3**, 10–24 (2021).
5. RJ Paul, The impact of atmospheric conditions on actual and expected scoring in the nfl. *Int. J. Sport Finance* **12**, 14 (2017).
6. MA Zipperman, Quantifying the impact of temperature and wind on nfl passing and rushing performance. *Unpubl. Work.* (2014).
7. PyPI, nfl-data-py: A python library for interacting with nfl data (2023).
8. P Eagles, Flashback to snow bowl 2013 (2013).
9. A Sports, Jameis winston's wife sent four-word message during wild snow game vs. steelers (2023).