Temporal projection of natural atmospheric events in the United States Regions

Haejeong Choi, Yanyu Long, Seung Ho Woo

April 2021

1 Introduction

Extreme weather events such as hurricanes, floods, and droughts can often lead to detrimental effects on agricultural production, severe infrastructure and property damage, and loss of life. The U.S. has seen a rapid increase in weather and climate disasters in recent years. In 2020, there were 22 "billion-dollar" extreme weather events (events where losses exceed $1 billion) in the U.S., hitting a record high since 1980 (National Centers for Environmental Information 2021b). The most recent billion-dollar event was the Winter Storm Uri in February 2021, which caused at least 136 deaths and damage of over $195 billion (Wikipedia contributors 2021).

Climate change is expected to increase the frequency and intensity of these events. For example, in dry climates, higher temperature can lead to an increase in surface evaporation, rapid drying of soils, and the occurrence of severe droughts (Briffa, Schrier, and Jones 2009; Cai et al. 2009). In wet climates, warmer weather can lead to abundant atmospheric water vapors (Willett et al. 2008), and the increase in water vapors is found to be a primary cause for the increase in extreme precipitation events (Kunkel et al. 2013).

In this project, we study the relationship between weather conditions and the occurrences of extreme weather events in the U.S. Our goal is to identify the most relevant indicators and the best-performing classification algorithms and build an alarm system of weather and climate disasters based on monthly weather statistics.

2 Data

2.1 Data Preprocessing

The Storm Event Database For the outcome variable, this study uses the Storm Event Database maintained by the National Weather Service (National Centers for Environmental Information 2021a). This dataset documents the occurrence of extreme weather events such as thunderstorm wind, flash flood, drought, etc. at the county level from 1950 to 2020. We will compute the number of episodes observed in each county in each month,[1] and transform it into a boolean variable `event` which equals one when there is at least one extreme weather event, and zero otherwise.

The historical weather dataset For the predictors, it is difficult to obtain weather data aggregated at the county level, so we will instead look at a city level historical weather dataset

[1]Each event in the NOAA database has an `episode_id` and a `event_id`. An episode can correspond to multiple related events. This study uses episodes as the unit of extreme weather events.

from the OpenWeatherMap API (Historical Weather API, OpenWeather 2021) and use the city-level weather data to approximate the county-level conditions. This dataset contains hourly observations of temperature, humidity, pressure, wind speed, and wind direction in 27 U.S. cities (see Table 4) from 2012 to 2017. To reduce the dimensionality and impute the missing values, monthly summary statistics are extracted from the hourly weather data by first computing daily average values of the observed variables, then computing the monthly average of the daily records.

Merging the two datasets Due to the limited coverage of the weather data, we only look at the time frame of November 2012 - December 2017, and focus on the 27 counties we have weather records for. To match the cities in the weather dataset to the counties in the wind storm dataset, we use the city names and locations (latitude and longitude) information from the United States Cities Database (SimpleMaps.com 2021). We specifically match the extreme weather event records to the weather data in the previous month, allowing the model to forecast future events with observed weather measurements.

The final dataset The final dataset we obtain is relatively balanced, with 948 negative samples and 726 positive ones. We randomly draw 25% of the observations to be used as the test set. As the predictors are not on the same scale, we scale each feature variable individually so that they are all in the range of [0,1] before entering models.

Table 1: Descriptions of variables used in the modeling process

Variable	Description
event	outcome variable, equals 1 when num_episodes>0, and 0 otherwise
meantemp_avg	monthly average of daily mean temperature (Kelvins)
difftemp_avg	monthly average of diurnal temperature variation (max(T)-min(T), Kelvins)
humidity_avg	monthly average of daily mean humidity (%)
pressure_avg	monthly average of daily mean atmospheric pressure (hPa)
wind_speed_avg	monthly average of daily mean wind speed (meter/sec)
wind_direction_avg	monthly average of daily mean wind direction (degrees, meteorological)
state	names of the states that the counties are in (categorical)

[a] Observations are uniquely identified by county and year-month.

2.2 Exploratory Data Analysis

To look into the relationship between the predictors and outcome variable, we will perform some descriptive data analysis. From the pairwise scatterplots between the predictors (Figure 1), we do not observe highly correlated features, except that the average diurnal temperature variation (`difftemp_avg`) and the humidity have a slightly worrying but still acceptable correlation coefficient of 0.746. From the boxplots of the weather measurements against the outcome variable, we can see that higher temperature is positively associated with the probability of having extreme weather events. It seems that our research question will likely require a complicated model that allows for nonlinear decision boundaries and/or interactions between

predictors, so we would expect algorithms like nonlinear Support Vector Machines (SVM) and tree-based ensemble methods to work better on this dataset.

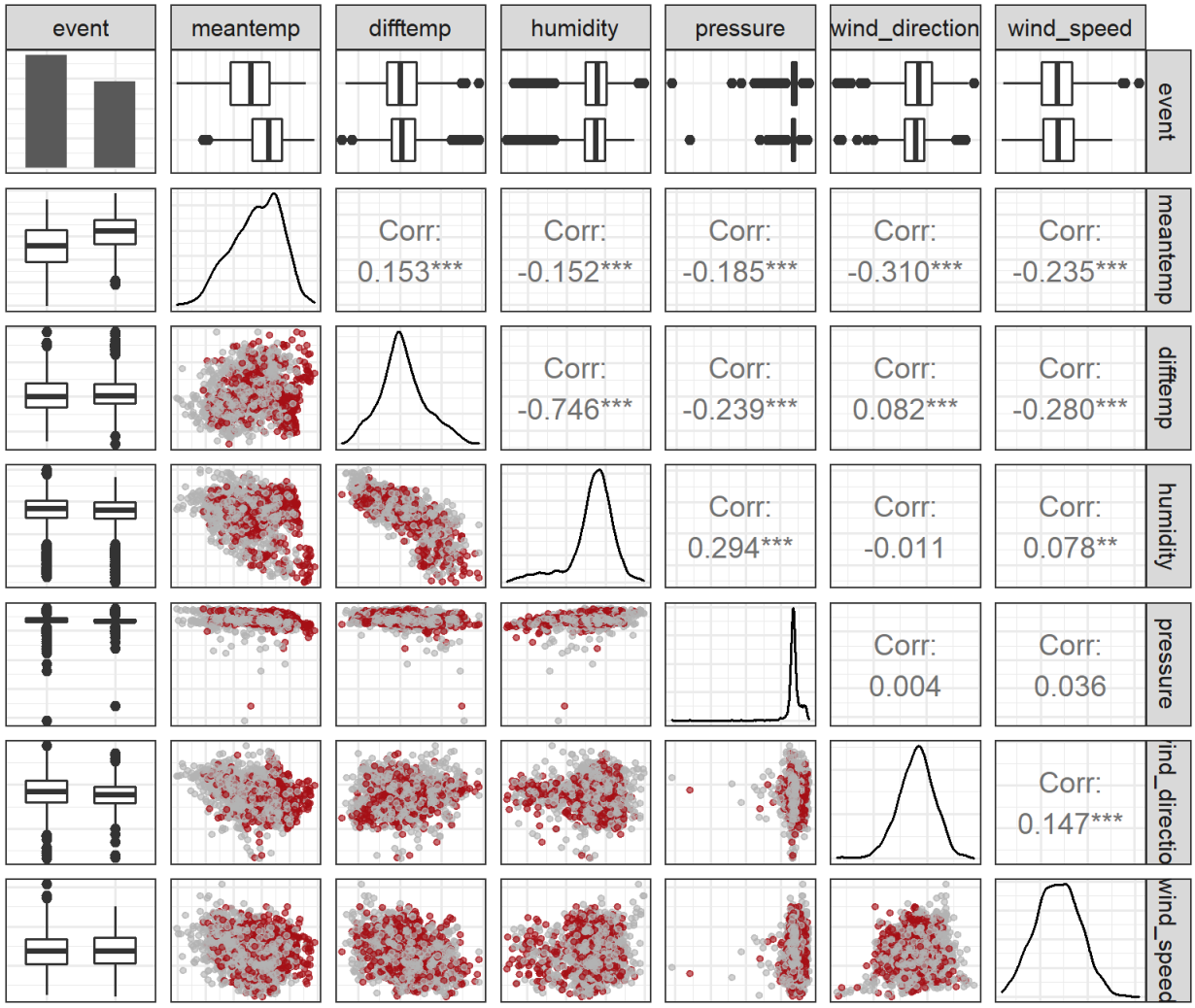

Figure 1: Density plots, pairwise scatterplots, and boxplots of the predictors (Red points: event = 1, grey points: event = 0)

From Figure 2, we can see that the probability of encountering severe weather events are different in each state. Texas had a total of 3,750 episodes from 2012 to 2017 and ranked first among all states. Texas experiences multiple types of weather conditions - the western one-third of the state suffers from cold winters and low humidity, while the eastern two-thirds experiences sub-tropical weather. The state sustains frequent thunderstorms which can lead to damaging hails and flash flooding, and all of Texas is susceptible to drought induced by the intense summer heat and the lack of rain (Blaisdell, Molly 2019). We will introduce the state as a predictor to account for the geographical factors affecting the probability of having extreme weather events other than weather conditions.

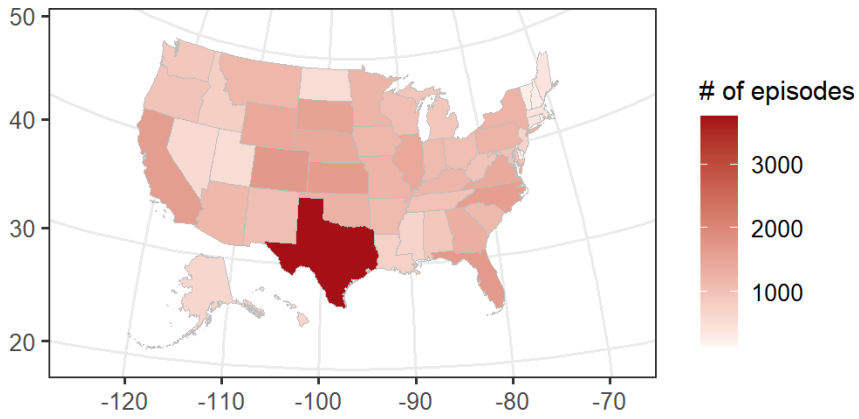

Figure 2: Cumulative number of extreme weather events (episodes) in each state, 01/2012 - 12/2017

3 Methodology

3.1 Research Approach

This project aims to identify the best-performing classification algorithms and the most important indicators that help forecast future extreme weather events. Therefore, we would want to compare multiple classification methods that excel at either interpretability or predictive power. We will consider the following algorithms: Naive Bayes, K-Nearest Neighbors, Decision Tree, AdaBoost, Random Forest, Support Vector Machine (SVM), Neural Networks, and Gaussian Process Classifier. For each algorithm, we tune the parameters using cross validation, and evaluate the performance on the test set using the following metrics: (1) Accuracy $= (TP + TN)/(TP + FP + FN + TN)$; (2) Balanced Error Rate (BER) $= 0.5(FP/(TN + FP) + FN/(FN + TP))$; and (3) the area under the receiver operating characteristic curve (AUC). The accuracy and BER will be used to evaluate the predicted labels, while the AUC can evaluate the predicted probabilities.

3.2 Preliminary Screening

Since there are many available classification methods to be considered, and we have inferred from the EDA that certain types of classifiers might work better on this dataset than others (i.e., those with nonlinear decision boundaries or allow variable interactions), we would not want to tune parameters for all of them. Instead, we will perform a preliminary screening by evaluating the model performance using the untuned parameters, and only tune the parameters for the set of algorithms that look promising. As Table 2 shows, the Gaussian Process Classifier, SVM with a Radial Basis Function (RBF) kernel, and Random Forest have much better accuracy, BER, and AUC than other competing methods. Thus, we will focus on these three classification methods to tune the parameters and evaluate their performance.

Table 2: Accuracy, BER, and AUC of different classification algorithms

Classifier	Accuracy	BER	AUC
Gaussian Process Classifier	68.11%	0.3277	0.7078
Support Vector Machine (Radial Kernel)	66.43%	0.3425	0.7102
Random Forest	65.95%	0.3281	0.7230
Neural Network	64.75%	0.3828	0.6405
K-Nearest Neighbors	64.51%	0.3658	0.6808
Naive Bayes	63.55%	0.4113	0.6266
AdaBoost	63.31%	0.3788	0.6905
Decision Tree	62.35%	0.3619	0.6489

3.3 Hyperparameter optimization

Random Forest The Random Forest is an ensemble algorithm that builds multiple base classifiers (in our case, decision trees) in parallel and combines them together for more stable predictions. By considering a random subset of predictors for each split, the Random Forest de-correlates the trees and at the same time makes sure the final decisions head in the correct direction since the majority of base classifiers would make the right choices even if some are wrong. For a Random Forest classifier, a lot of hyperparameters can potentially influence the model performance, for example the number of base classifiers to employ, the maximum depth and minimum number of nodes in each tree, the number of features to consider in a split, and the impurity measure, etc.

SVM Using the Support Vector Machines, the most important parameter to tune is the kernel function. Some of the commonly used nonlinear kernels are polynomial, radial basis, and Sigmoid function kernels. We would expect a polynomial or radial basis kernel to perform better as they allow more flexible decision boundaries.

Gaussian Process Classifier The Gaussian Process Classifier is based on a Bayesian methodology. It assumes some prior distribution on the underlying probability densities that guarantee some smoothness properties. As a probabilistic method, it allows us to compute empirical confidence intervals and decide if one should refit the prediction in some region of interest. The final classification is then determined as the one that provides a good fit for the observed data, while at the same time guaranteeing certain level of smoothness. To train a GP classifier, we need to specify a kernel function which describes the covariance of the prior distribution.

4 Results

We use the Python Machine Learning Library `scikit-learn` to implement the models and tune the hyperparameters (Pedregosa et al. 2011).

Random Forest With a randomized grid search, we found the the following setting performs the best: training 800 base classifiers with a maximum depth of 40, considering $\sqrt{p}$ features in each split where p is the total number of features, splitting an internal node only when there are at least 15 samples in the current node and at least 2 samples in each of the left and right branches, and using the Gini index as the impurity criterion.

Using the tuned Random Forest classifier, the accuracy on our training data set and test set are 91.87% and 69.06%, respectively. As we can see from the receiver operating characteristic (ROC) curve (Figure 4) and the confusion matrix (Table 3), the AUC reaches 0.7230 and the BER is 0.3281. The Random Forest classifier seems to work well in terms of accuracy and AUC, but it makes a lot of type II errors and thus cannot identify weather conditions that will lead to severe weather events, which is against the goal of our study.

Despite our concerns for the false negative predictions, the Random Forest classifier has better interpretability than the other two models, as it allows us to compare the importance of features. From the relative influence plot (Figure 3), we see that the most helpful indicator is the monthly average temperature, followed by the average diurnal temperature variation. This fits our prior expectation that higher temperatures add to the probability of having extreme climatic events. The rest of the weather measurements have similar importance, while the state variable is the least important one, indicating that the weather measurements alone can account for most of the geographical factors that influence the extreme weather events.

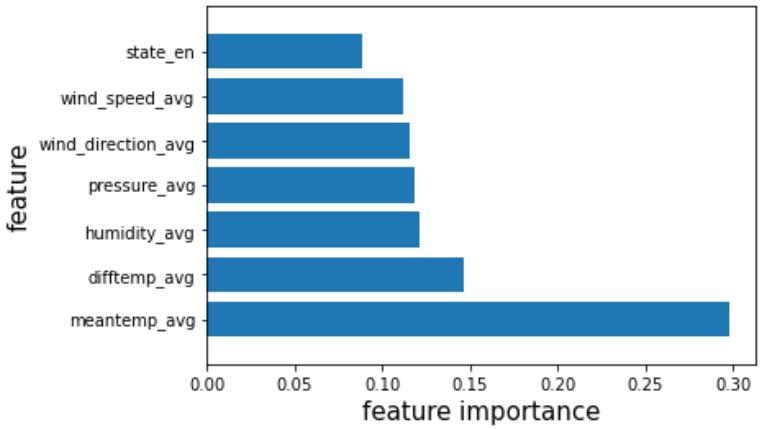

Figure 3: Relative influence plot of the features

SVM The best performance is given by an RBF kernel SVM with $\gamma = 1$ and $C = 5$. The parameter γ defines how far the influence of a single training example reaches, and C is the coefficient for the penalty term, determines how much error is bearable and controls the bias-variance trade-off. We have relatively low γ and C values, which means that we give less weight to the training data and our model will have a simpler decision function with a larger margin.

The tuned SVM model gives a training-set accuracy of 73.92%, a test accuracy of 66.19%, an AUC of 0.7102 and a BER of 0.3425. It is less satisfactory than the Random Forest in terms of accuracy, AUC and BER. While the SVM makes fewer type II errors, it makes the most type I errors out of the three classifiers, which can lead to false alarms.

Gaussian Process Classifier Of the available kernel functions, we find that an RBF kernel works best for this dataset. The rest of the hyperparameters are optimized during fitting. The tuned GP classifier gives a training-set accuracy of 73.68%, test accuracy of 68.11%, AUC of 0.7078 and BER of 0.3277. We choose this model as the optimal classifier as it has fairly good accuracy, AUC, and BER and it classified both true 0 and 1 samples well.

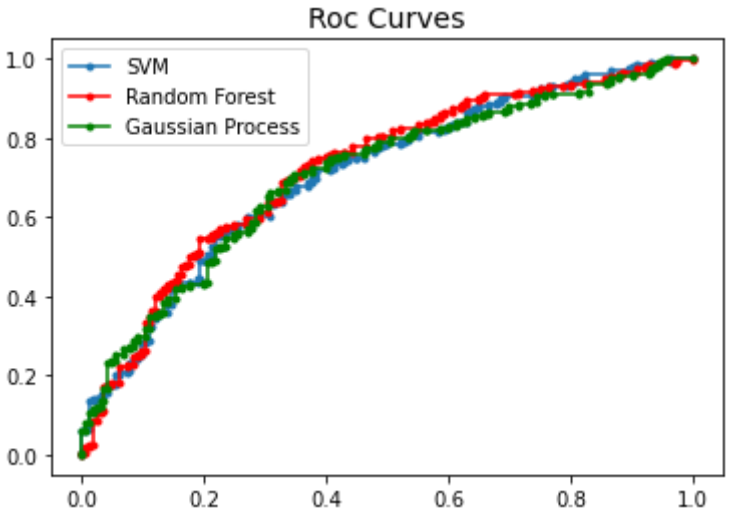

Figure 4: Receiver operating characteristic curves of the three algorithms with tuned hyperparameters

Table 3: Confusion matrix, accuracy, BER, and AUC on the test set

Model	TN	FP	FN	TP	Accuracy	BER	AUC
Random Forest	192	60	69	96	0.6906	0.3281	0.7230
RBF-SVM	171	81	60	105	0.6619	0.3425	0.7102
Gaussian Process	180	72	61	104	0.6811	0.3277	0.7078

5 Conclusions

This project studies the relationship between extreme weather events in the U.S. and historical weather data. We tried different classification algorithms to compare their performance. When it comes to predictive power, the Gaussian Process classifier gives the lowest BER and is our optimal classifier if we deem the correct predicted labels most important. While the Random Forest classifier has the best test accuracy and AUC value, it gives many false negative predictions, which is against our primary goal of identifying weather conditions that can lead to severe weather disasters. In terms of interpretability, the Random Forest classifier allows us to compare feature importance. The monthly mean temperature is the most helpful indicator for forecasting extreme weather events, followed by the monthly average diurnal temperature variation. The categorical variable `state` is the least helpful feature, suggesting that the weather measurements themselves are able to capture most of the geographical factors impacting the outcome variable.

6 Future Work

When constructing the outcome variable, this study aggregates all types of extreme weather events, so we are only able to discuss how weather measurements contribute to the probability that any type of events will happen. This may not be an ideal practice as the relationship between the predictors and the outcome might vary if we look at different types of events. To take this study further, we can divide the events into subcategories or even look at individual event types for better predictive power and more meaningful parameter interpretations.

When it comes to the predictors, this study uses the city-level data as an approximation for counties' weather conditions due to the difficulty of gathering county-level weather data. However, the cities where weather data are available are not guaranteed to represent the counties well. In the future, we can select multiple land surface stations from each county and construct a more representative county-level weather dataset, which might generate more accurate predictions.

References

[1] Blaisdell, Molly. 2019. "What Types of Weather Conditions Does Texas Experience?" https://sciencing.com/types-weather-conditions-texas-experience-8222227.html.

[2] Briffa, Keith R, Gerard van der Schrier, and Philip D Jones. 2009. "Wet and Dry Summers in Europe Since 1750: Evidence of Increasing Drought." *International Journal of Climatology: A Journal of the Royal Meteorological Society* 29 (13). Wiley Online Library: 1894–1905.

[3] Cai, Wenju, Tim Cowan, Peter Briggs, and Michael Raupach. 2009. "Rising Temperature Depletes Soil Moisture and Exacerbates Severe Drought Conditions Across Southeast Australia." *Geophysical Research Letters* 36 (21). Wiley Online Library.

[4] Historical Weather API, OpenWeather. 2021. "Historical Hourly Weather Data 2012-2017." https://www.kaggle.com/selfishgene/historical-hourly-weather-data/.

[5] Kunkel, Kenneth E, Thomas R Karl, Harold Brooks, James Kossin, Jay H Lawrimore, Derek Arndt, Lance Bosart, et al. 2013. "Monitoring and Understanding Trends in Extreme Storms: State of Knowledge." *Bulletin of the American Meteorological Society* 94 (4). American Meteorological Society: 499–514.

[6] National Centers for Environmental Information. 2021a. "Storm Events Database." https://www.ncdc.noaa.gov/stormevents/.

[7] National Centers for Environmental Information. 2021b. "U.S. Billion-Dollar Weather and Climate Disasters." doi:10.25921/stkw-7w73.

[8] Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30.

[9] SimpleMaps.com. 2021. "United States Cities Database." https://simplemaps.com/data/us-cities.

[10] Wikipedia contributors. 2021. "February 13–17, 2021 North American Winter Storm —

Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/w/index.php?title=February_13%E2%80%9317,_2021_North_American_winter_storm&oldid=1019729582.

[11] Willett, Katharine M, Philip D Jones, Nathan P Gillett, and Peter W Thorne. 2008. "Recent Changes in Surface Humidity: Development of the HadCRUH Dataset." *Journal of Climate* 21 (20): 5364–83.

Appendix

Table 4: The 27 U.S. cities where weather data are available

city	county	state	city	county	state
Portland	Multnomah	Oregon	**Saint Louis**	St. Louis	Missouri
San Francisco	San Francisco	California	**Chicago**	Cook	Illinois
Seattle	King	Washington	**Nashville**	Davidson	Tennessee
Los Angeles	Los Angeles	California	**Indianapolis**	Marion	Indiana
San Diego	San Diego	California	**Atlanta**	Fulton	Georgia
Las Vegas	Clark	Nevada	**Detroit**	Wayne	Michigan
Phoenix	Maricopa	Arizona	**Jacksonville**	Duval	Florida
Albuquerque	Bernalillo	New Mexico	**Charlotte**	Mecklenburg	North Carolina
Denver	Denver	Colorado	**Miami**	Miami-Dade	Florida
San Antonio	Bexar	Texas	**Pittsburgh**	Allegheny	Pennsylvania
Dallas	Dallas	Texas	**Philadelphia**	Philadelphia	Pennsylvania
Houston	Harris	Texas	**New York**	New York	New York
Kansas City	Jackson	Missouri	**Boston**	Suffolk	Massachusetts
Minneapolis	Hennepin	Minnesota			