Understanding Nationwide Power Outage and Restoration for Future Prediction

Yasmeen Haleem★
University of North Texas

Isabelle Wagenvoord★
Colorado College

Qianjun Wei
University of North Texas

Ting Xiao
University of North Texas

Tong Shu
University of North Texas

Yuede Ji*
University of North Texas

ABSTRACT

Every year, millions of people in US experience power outages, especially during extreme weather events like hurricanes, extreme cold and icy weather, and wildfires. For example, 6.5 million customers in Florida were out of power due to Hurricane Irma in September 2019 [3], and it took 10 days until power was restored. Similarly, more than 4.5 million customers lost electricity at the worst of Texas freeze in February 2021 for a several days [32]. To understand nationwide power outage and restoration, this project aims to address the challenge questions in "Challenge 7: EAGLE-I Outage Data 2014-2022" of 2023 SMC Data Challenge. In particular, we have studied four major questions and made interesting findings for analyzing restoration times (Question 1), correlation with relevant datasets (Question 2), predicting power outages (Question 5), and identifying patterns over space and time (Question 6).

ACM Reference Format:
Yasmeen Haleem★, Isabelle Wagenvoord★, Qianjun Wei, Ting Xiao, Tong Shu, and Yuede Ji*. 2018. Understanding Nationwide Power Outage and Restoration for Future Prediction. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Reliable electricity is a critical function for the economy and public health. Extreme weather events can lead to prolonged outages, leading to health risks and economic issues [5]. For example, the deep freeze of 2021 in Texas caused power outages to around 4.5 million customers for about 5 days [32]. Therefore, understanding the power outage and restoration patterns is of critical importance for the response and recovery of those in the emergency response sector.

The Department of Energy has an informational platform, called Environment for the Analysis of Geo Located Energy Information (EAGLE-I) [23], which monitors the power outages in nearly every county, accounting for 92% coverage of the US and its territories. The 2023 SMC Data Challenge provides a dataset with eight years of power outage information at the county level from 2014 to 2022 at 15-minute intervals collected by the EAGLE-I program at Oak Ridge National Laboratory (ORNL) [23]. Monitoring power outages at a fine-grained level across the nation is challenging as electricity providers provide varying amounts of information. This dataset provides invaluable knowledge to understand power outages and restoration. A limitation of this dataset that is inherent in tracking outages is that power outage counts can be artificially low [14], but this is relatively rare. This is because large power outages can themselves hinder efforts to keep track of the scale of outages.

We examine these power outages alongside weather events and socioeconomic factors, as these can be important determinants for resource allocation.

Weather is an important predictor of large power outages. In 2018-2020, 62.1% of over eight-hour power outages were found to have taken place alongside extreme weather events [11]. Socioeconomic indicators can provide important insights into how vulnerable counties may have their restoration times impacted.

This project aims to address the challenge questions in "**Challenge 7: EAGLE-I Outage Data 2014-2022**" of 2023 SMC Data Challenge [23]. In particular, we are targeting four major questions: analyzing restoration times (Question 1), correlation with relevant datasets (Question 2), predicting power outages (Question 5), and identifying patterns over space and time (Question 6) as shown in Figure 1.

Our approaches and major findings for the four questions are summarized below[1].

★ Equal contribution.
* Corresponding author: Yuede Ji, University of North Texas.

[1]The source code can be found at https://github.com/SC-Lab-Go/SMCDC-EAGLE-I.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

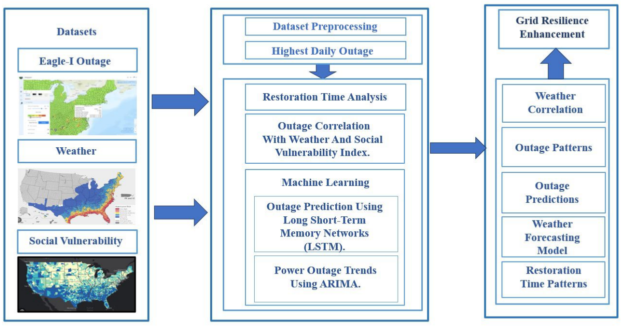

Figure 1: A basic workflow describing the time series analysis for power outages.

Question 1: Analyzing restoration time. To analyze restoration time, we test the efficacy of a variety of anomaly detection methods. These methods are seasonal decomposition, isolation forest, and county-specific outage thresholds. We choose a county-specific threshold to identify power outage periods as the EAGLE-I dataset does not have labels for outage events. In our time series analysis, we extract the trend component using a seasonal decomposition algorithm implemented in the statsmodels Python package with a period of four hours and using a threshold from [29]. The aim is to identify significant outages without significantly reducing the dataset's resolution.

Question 2: Correlation with relevant datasets. We find correlations between power outage scale and durations with national weather data and socioeconomic factors.

We utilize a dataset documenting extreme weather events in the United States from 2016 to 2022 [27]. We find large spikes in the number of power outages in certain counties–particularly during Hurricane Irma in Florida in September 2017 and during a deep freeze in Texas in mid-February 2021.

We also explore warning and advisory datasets provided by the National Weather Service (NWS) [18]. Since EAGLE-I and the NWS datasets are updated in real-time, they provide the potential for monitoring power outages alongside a wide range of documented weather hazards. Due to computational resource and time constraints, we do not have a comprehensive analysis of this particular dataset in relation to outages, but provide a algorithm for preprocessing and combining NWS data with outage data Appendix Section 8.2. Moreover, we provide an example of correlating the outage durations with official flood warnings and alerts .

Finally, we examine outages in relation to social burdens data provided by the United States Council on Environmental Quality.

Question 5: Predicting power outage. We train power outage prediction models based on eight years of historic power outage data. In particular, we use the long short-term memory (LSTM) [19], XGBoost [10], Prophet [30], and CatBoost regressor [12] models to predict the daily outages of San Diego County and Miami-Dade county. The highest daily outages for San Diego County are calculated and a CatBoost regressor outage prediction model is created with a Test MAPE: 3.22 and Test Accuracy: 96.78. We use the power outage data from 2014 to 2020 as the training dataset and the data from 2020 to 2022 as the testing dataset.

Question 6: Identifying patterns over space and time. We perform a detailed time series analysis on the daily outages of San Diego County for the past eight years. We find that there was a increase in power outages during the months of November, December and January in San Diego from 2021 to 2023. More interestingly, the highest weekly power outages tended to be on Fridays and Saturdays, and large-scale outages mainly occurred towards the beginning and end of each year.

2 BACKGROUND AND RELATED WORK

2.1 Definition

Stationarity is a common assumption for many inferencing methods in time series data. A given time series is considered stationary if it has constant statistical properties that are not dependent on the timestamps at which they are observed.

Federal information processing standards (FIPS) codes are standard numeric codes for identifying geographic entities in US, including states, counties, tracts, and territories. Additional definitions have been added in Appendix Section 8.1.

2.2 Related Work

Several research papers have explored the application of machine learning algorithms to predict power outage recovery times [14, 33, 34]. These studies utilize outage data, weather information, and other relevant variables to develop predictive models.

Tabassum *et al.* develops a black-box segmentation algorithm and applies it to EAGLE-I data on areas impacted during the times which Hurricane Harvey or Hurricane Irma took place. It associates phases of the hurricane with outage failures in these areas. Wang *et al.* developed a data-driven predictive model for predicting restoration time using spectral analysis and transfer learning, tested on a six-year dataset of outages in New York. Ericson *et al.* calculated the durations of power outages during extreme weather events in the United States using EAGLE-I data [14]. The algorithm aggregates the data to the maximum number of outages recorded during each 4-hour period, then identifies large power outages during extreme weather events, calculating the outage, peak, and restoration times. Our study aims to identify and analyze outage events across the United States at a 15 minute or daily resolution.

3 DATA PREPROCESSING

The EAGLE-I dataset reports the number of outages every 15 minutes during the entire day for a period of 8 years from 2014 to 2022 for 92% of US counties. We use the EAGLE-I dataset to identify and estimate the durations of outages for each county, as discussed in section 4.2. We then compare the average duration of outages across counties experiencing a variety of social burdens. We find that counties identified as disadvantaged in any category have outages that last longer on average than non-disadvantaged counties.

We also group the outage data by date and FIPS code and identified the highest number of outages for each day, obtaining a single data frame with around 5 million rows such that it can be correlated with the U.S. Weather Events dataset [26] discussed in 5.1. For the outage prediction models, we identify the daily highest outage for 8 years for San Diego County and group it into a data frame with 2932 rows. Similarly, the highest hourly outages for San Diego County are calculated and grouped into a data frame with 7081 rows. San Diego County is used for outage predictions as an example as we found that it had a consistent number of outages between 2014-2022 without any significant gaps. The final grouped dataframes have two columns showing date and highest outage. Further details about NWS dataset have been added in Appendix Section A.2.

4 QUESTION 1: ANALYZING RESTORATION TIME

4.1 Augmented Dickey-Fuller (ADF) Test

An essential step in time series analysis is testing if the time series dataset is stationary. We use the Augmented Dickey-Fuller (ADF) test, which looks for the presence of the unit root. The null hypothesis of the ADF test is that the time series has a unit root, meaning that it's non-stationary. The alternative hypothesis is that the time series has no unit root and is stationary. In contrast to the Dickey-Fuller test, the Augmented Dickey-Fuller test is applicable to a wider range of time series [1]. We perform the ADF test by using the implementation in the statsmodels Python package where the p-values are calculated using regression surface approximation from [24], using the updated tables from [25]. We chose a critical value of 0.01. When we run the ADF test on 10% of the counties in the year 2022, we find that 285 out of 305 counties had a p-value of less than 0.01. A small p-value (typically less than 0.05) in the ADF test indicates that the time series is likely stationary, while a large p-value suggests it is likely non-stationary [16].

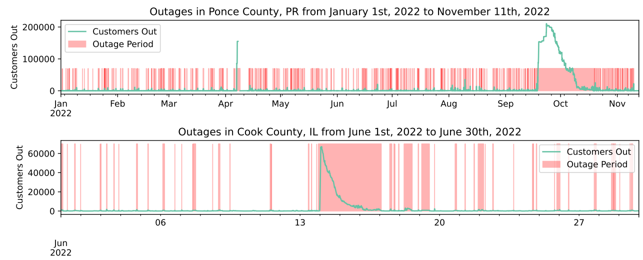

Figure 2: Outage periods identified using a convolution filter and county-specific thresholds.

Table 1 shows the test results for counties with a p-value greater than 0.01. From this, we see that such counties have a small number of observations. We also find that the majority of each of these time series data is stationary, meaning that most of the time the statistical properties of the time series remain constant. This gives confidence to thresholding techniques for detecting spikes in outages since we know that the overall mean and standard deviation of outages for each county will be constant, except when significant outages occur.

Additional anomaly detection methods for outages are explored in Appendix Section A.3.

4.2 Proposed Algorithm for Identifying Outage Periods

Using a threshold to detect outage periods. Current techniques for calculating restoration times in EAGLE-I data focus mainly on outages that occur during extreme weather events [14, 20]. Another method evaluates thresholds proportional to the maximum number of outages [7]. Our method aims to provide an unsupervised and efficient method for identifying outages. Unsupervised spike detection has been studied for high-frequency multi-channel data [8, 22], making relevant techniques strong candidates for identifying significant power outages in large counties where power shutoffs are frequent. To detect significant power outages, we use a threshold that is robust to large spikes and high-frequency data [29]:

$$\sigma_n = med(\frac{x}{0.6745}) \tag{1}$$

$$Thr = 4 \cdot \sigma_n \tag{2}$$

x is the time series signal, σ_n estimates the standard deviation of background noise in the data, and $med(\cdot)$ denotes the median function. For this threshold, we do not take the standard deviation of the entire time series as that can lead to high threshold values in the presence of frequent and large spikes in the time series [29].

Table 1: ADF p-value and the number of observations for each county where the p-value exceeds 0.01. Each FIPS code represents a county. Ideally, outages should be continuously recorded every 15 minutes throughout the year. A typical county should have around 35,040 observations. Ponce, Puerto Rico (FIPS: 72113) experiences an extremely large outage in 2022, making the time series less stationary.

FIPS code	38011	31023	1071	8111	15009	46071	38081	31049	46065	46061	72113	27121	31003
p-value	0.746	0.574	0.490	0.252	0.165	0.109	0.096	0.083	0.055	0.033	0.021	0.019	0.011
Observations	72	305	122	91	282	1,150	4,612	176	175	1,760	28,390	1,533	177

Identifying Outage Periods. Based on these findings, we propose the following algorithm to preprocess and identify outage periods in the EAGLE-I data:

(1) We apply the same convolution filter as the one used seasonal_decompose method in the statsmodels Python package for calculating the trend to remove irregularities. We use a window of 4 hours.
(2) We calculate the threshold from [29] using the trend component.
(3) We identify outage periods based on whether the trend component or original time series passes the threshold.
(4) We obtain the starting and ending timestamps for each outage to calculate the outage durations.

Note that we first apply a convolution filter identical to that used to compute the trend component in the seasonal decomposition algorithm in section 7. We noticed that this helps remove artificially low values that tend to be present in the EAGLE-I dataset [14] without significantly reducing its resolution. Figure 2 presents an example of outage periods identified using this algorithm.

A limitation of this threshold is that counties with low populations and low frequency of outage events will have an extremely low threshold, leading to false identification for outage events. To overcome this, we remove outage events where the maximum customers impacted is 0. A demonstration of comparing outage durations across counties with varying socioeconomic burdens is explored in 5.2. This could also be mitigated by setting a minimum threshold for such counties.

5 QUESTION 2: CORRELATION WITH RELEVANT DATASETS

5.1 Weather Dataset

We obtain a weather dataset from Kaggle [26], which documents 8.6 million weather events in the U.S. from January 2016 to December 2022. Sourced from 2,071 airport-based weather stations nationwide, this dataset records events and their severity in 7 categories: severe-cold, fog, hail, rain, snow, storm, and precipitation. This dataset was developed in [27], and defines each weather event as follows:

Severe-Cold: Extremely cold, where temperature $\leq -23.7°C$.

Table 2: Yearly percentage (%) of different extreme weather events across US.

Type	Cold	Fog	Hail	Storm	Rain	Snow	Precipitation
Duration	0.43	3.18	0.01	0.09	7.41	2.43	0.12

Hail: Solid precipitation including ice pellets and hail.

Rain: Rain, ranging from light to heavy.

Snow: Ranging from light to heavy.

Storm: Extremely windy conditions, where the wind speed is at least 60 km/h.

Precipitation: Any other type of precipitation which cannot be assigned to previously described event types.

The categorizations of weather and thresholds for their severity were determined using the original dataset and K-Means clustering [27].

The percentage of each extreme weather event in different states is shown in Table 2.

We correlate this with the EAGLE-I data by merging the highest daily outage with the type, severity, county columns in the weather dataset [26]. Based on this weather data we implemented a probabilistic Markov weather forecasting model, which is discussed at the end of this section.

Findings. During the period between 2016-2022 we find the top 10 counties with the highest daily outages. Table 3 illustrates that Texas and Florida are among the states in the south with the highest outages, thus they are chosen for further analysis. We also focused on North Carolina and South Carolina as they are affected by hurricanes. A spike in power outages occurs in the month of February 2021, in Dallas, Texas. This is correlated to the Texas deep freeze that happened during the mid-February in 2021. Table 6 shows that around 374,745 outages were recorded for Dallas county on February 15th and even two days later on February 17th, there were still 283,256 outages. It can be observed that along with Dallas, other counties like Harris County, and Travis County had one of the largest power outages that lasted for around 5 days from February 15th that was correlated to extremely cold temperatures of around 5°F with around 4.5 million customers without power in Texas. [32]

Table 3: Top ten counties with highest outages between 2014-2022.

County	Miami-Dade	Broward	Palm Beach	Duval	Brevard	Lee	Pinellas	Harris	Tarrant	Dallas
Outage	1,777,800	1,398,920	1,098,160	528,830	522,957	462,570	417,553	374,745	367,316	358,650

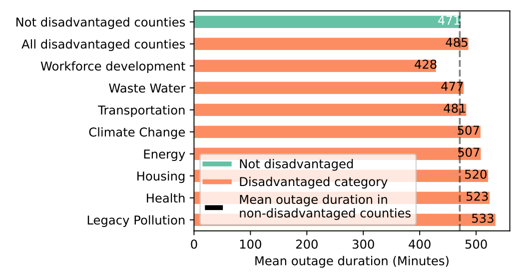

Figure 3: Mean duration of significant outages by category of socioeconomic burden.

We also study the highest daily power outages and extreme weather events between 2016-2022 in Florida. Table 4 shows that Palm Beach County, Florida had one of the highest daily outages, of about 1.2 million, during category 4 Hurricane Irma that made landfall in September 2017. It caused widespread outages all across Florida including in Miami-Dade county. In Table 4, it can be observed that during the peak hurricane season from August to October, October had the highest number of outages in 2016 due to Hurricane Mathew. Moreover, September shows the highest outages from 2017 to 2022, exceeding 1 million outages in Palm Beach County due to extreme wind and rain from Hurricane Irma. Because of this, more than a million customers were without power on September 10th, 2017. The majority of these power outages were gradually restored within five days. Similarly high power outage periods were seen during September 2022 in Florida due to Hurricane Ian. Category 4 hurricanes cause extreme wind speeds of more than 120mph. Such extreme wind speeds and rain can cause grid and infrastructure damage, tree and electric pole damage, and inaccessibility that may lead to prolonged restoration times. As can be seen in Table 4 it can be observed that due to severe storm conditions high power outages exceeding 100,000 customers in Florida all take place during the hurricane season in 2016, 2017, and 2022.

Because of the strong correlation between weather and power outages, we propose that utilities could benefit from a weather forecast model that can identify counties impacted by particular extreme weather events. Examining what specific weather conditions that affect the particular county can thereby provide insight for improving restoration times. A related work [9] discusses a hurricane prediction model

Table 4: Extreme power outages during hurricane season in Florida between 2016-2022

Date	County	Outages	Type	Severity
2016-10-07	Brevard	200,786	Storm	Severe
2016-10-07	Volusia	235,478	Storm	Severe
2017-09-10	Broward	709,360	Storm	Severe
2017-09-10	Lee	227,810	Storm	Severe
2017-09-10	Miami-Dade	898,340	Storm	Severe
2017-09-10	Palm Beach	543,450	Rain	Heavy
2017-09-11	Duval	528,830	Rain	Heavy
2017-09-11	Orange	331,670	Storm	Severe
2017-09-11	Palm Beach	1,098,160	Storm	Severe
2017-09-11	Pinellas	416,237	Rain	Heavy
2017-09-13	Miami-Dade	573,570	Precipitation	UNK
2020-09-16	Escambia	156,882	Storm	Severe
2021-02-13	Marion	117,645	Rain	Moderate
2022-09-28	Lee	456,323	Storm	Severe
2022-09-29	Orange	229,025	Rain	Heavy

which takes uncertainties in dealing with climate into consideration.

We created a basic weather forecast model to forecast different extreme weather events like storm, rain, and snow using a Markov model with a score of 295.228 and we analysed how each county is prone to different extreme weather events as shown in Figure 4.

Moreover, we used the K-Means algorithm to group different states into four different clusters as shown in Figure 4. Compared to other clusters, cluster 3 (purple) has the highest chance of rain, cluster 2 (green) has the highest chance of snow, cluster 1 (red) has the lowest chance of rain while the highest chance of storm; cluster 0 (blue) is similar to cluster 3 while its' chance of rain is slightly lower than cluster 3.

5.2 Climate and Economic Justice Screening Tool Dataset

This 42MB dataset by the United States Council on Environment Quality contains 132 columns on 74,134 census tracts and is used to highlight disadvantaged census tracts nationwide. Each row is uniquely identified by a FIPS code.

In particular, the tool that relies on this dataset uses its own methodology and the data to identify disadvantaged

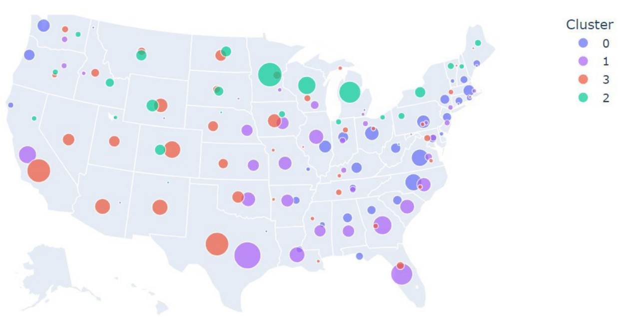

Figure 4: Clusters showing states with different chances of extreme weather events like fog, snow, rain, and hurricane.

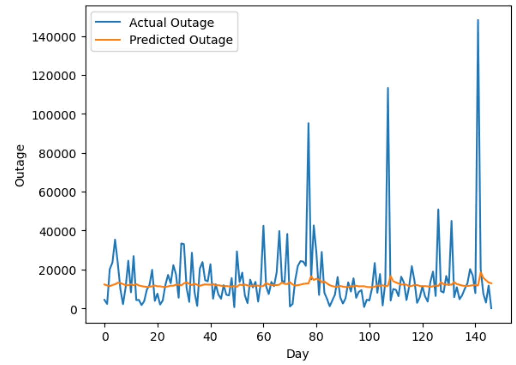

Figure 5: LSTM power outage prediction for San Diego county for the last 5 months of 2022.

communities in 8 categories: climate change, energy, health, housing, legacy pollution, transportation, water and wastewater, and workforce development.

Because the EAGLE-I dataset is at the county resolution, we aggregate this data to the county level and apply the tool's methodology from [28] to identify disadvantaged counties. For a county to be disadvantaged in a particular category, it needs to a) be in at least the 90th percentile in the dataset for at least one of the burdens specified in that category and b) be at least the 65th percentile for proportion of people in low-income households or less than 10 percent high school attainment for the workforce development category.

After identifying counties as disadvantaged in each category, we compare the mean outage durations for each group of counties in Figure 3. On average, a significant outage at any disadvantaged county will last 14 minutes longer than a county that has not been identified as disadvantaged in any category. Counties identified as disadvantaged in the categories of climate change, energy, health, housing, and legacy pollution have outage events that last at least 40 minutes longer on average than counties that are not disadvantaged in any category.

6 QUESTION 5: PREDICTING POWER OUTAGE

6.1 Power Outage Prediction with LSTM

As the EAGLE-I dataset is large with all the county's data, we use the highest daily power outages of San Diego County to implement an LSTM power outage prediction model. We then implement an LSTM model using Pytorch with 10 epochs and a learning rate of 0.001 and a batch size of 16. The test root mean square error is 12398.51 and the training root mean square error is 12596.13. An important issue during training is that many counties like Dallas County had gaps in the power outage information for a few months. This discontinuity led issues in training the prediction models. Noise and gaps in the power outage data can cause issues in creating an accurate outage prediction model. We choose San Diego County as it has a consistent number of power outages. In future work, we would like to include weather severity also as a feature to increase the accuracy of our prediction models.

Figure 5 shows power outages predicted by the LSTM model for 150 days. The y-axis shows the number of outages and the x-axis shows the forecast for the year 2022. It predicts a few spikes in power outages based on the training dataset provided. One of the issues faced while training the model is the unpredictability of certain climatic conditions like storms in San Diego. In future work, we would like to include weather severity also as a feature to train the model to increase the accuracy of the prediction model.

6.2 Power Outage Prediction with CatBoostRegressor

The highest daily outages for San Diego County are calculated and a CatBoost regressor outage prediction model is created with a test mean absolute percentage error (MAPE) of 3.22 and test accuracy of 96.78. It employs an algorithm called symmetric weighted quantile sketch (SWQS) which automatically handles the missing values in the dataset to reduce overfitting and improve the overall performance of the dataset. In Figure 6, the y-axis shows the count of the highest daily outage and as between 2014 to 2022 we have around 3000 days of data, the predicted outage has been calculated for the years 2021-2022 shown as between the days 2500-2900. Few outage peaks are also predicted by the CatBoost regressor.

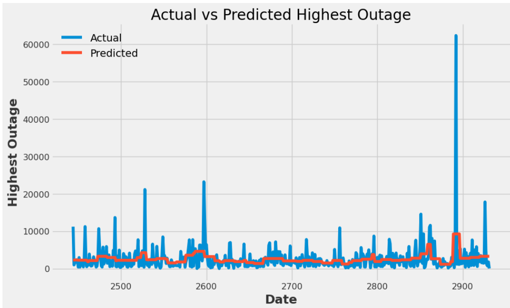

Figure 6: CatBoost Regressor power outage prediction model with a test accuracy of 96.78 for San Diego county for year 2022.

6.3 Power Outage Prediction with XGBoost

Similarly, the highest daily power outages are calculated for San Diego County and then an 80-20 training and testing split is done, by taking the power outages from 2014-2020 as the training dataset and the power outages from 2020-2022 as the testing dataset and a prediction model is created using an XGBoost regressor. The importance of different features like year, hour, day, and month are calculated and it is found that the day of the year had the highest feature importance. The mean absolute error is 643.62. Different graphs are then plotted showing the actual and predicted forecasts. After, we calculate best predicted days and worst predicted days and the error and the absolute error. The best predicted day was September 7th, 2021 as shown in Table 7.

6.4 Power Outage Prediction with Darts and Prophet

The highest daily outages of San Diego County per hour and the highest daily outages of Dallas County per hour are calculated and then the forecasting model is implemented using the Darts library with a train-test with an 80-20 split by choosing the outages from 2014-2020 for training and the outages from 2020-2022 for testing. The prophet forecasting model is used to forecast the power outages for 7 days and 30 days, and to forecast the trend of outages between 2014-2023. The Prophet model obtains a MAPE of 568.8 and the Darts model obtains a MAPE of 643.6.

7 QUESTION 6: PATTERNS OVER SPACE AND TIME

This question asks to analyze how power outage events in the USA have evolved over time and to observe patterns over space and time. To do so, we employ the Python packages Darts, Prophet, and statsmodels.

ARIMA. An ARIMA model implemented using Prophet is used to identify temporal patterns in the power outage data

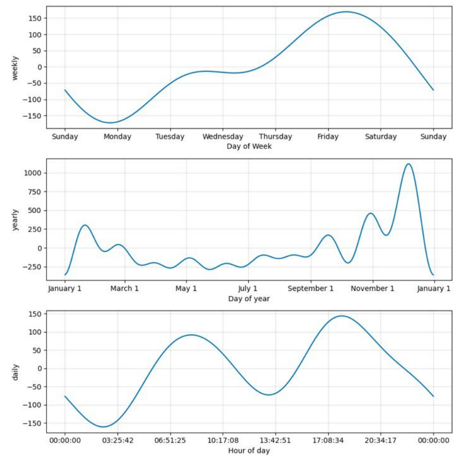

Figure 7: Power outage trends for San Diego county using Prophet library.

and to understand seasonality, trends, autocorrelation, lags and moving averages in the time series. We then use Darts to create a time-series dataframe from the power outage data and perform temporal analysis using Darts' built-in functions. The graphs generated give a clear trend in the power outages for the years between 2014 - 2022 for San Diego County.

Periodic Trends. As the EAGLE-I dataset includes the power outages for all counties, the highest daily outages for San Diego county are calculated for the years between 2014-2022. We then plot the patterns and trends in the outages using the Python packages Darts and Prophet.

Figure 7 shows that in general, the power outages in San Diego county have dropped from previous years, and from a related study it is due to introducing battery power as an addition to power supply [15]. The outages are significantly higher on Friday and Saturday, and during the months of October and January.

Seasonal Decomposition. Seasonal decomposition decomposes time series into three components:

(1) Trend, which is the overall change and direction of the time series;
(2) Seasonality, which is the regular fluctuation over a fixed period of time;
(3) Residual component, that remains after extracting the trend and seasonality component from the original data.

We use a naive decomposition method implemented through the "seasonal_decompose" method in the "statsmodels" Python package that applies a convolution filter to obtain the trend and extracts seasonal and residual components such that all the components can be added together to get the original values [6]. When we perform a seasonal decomposition,

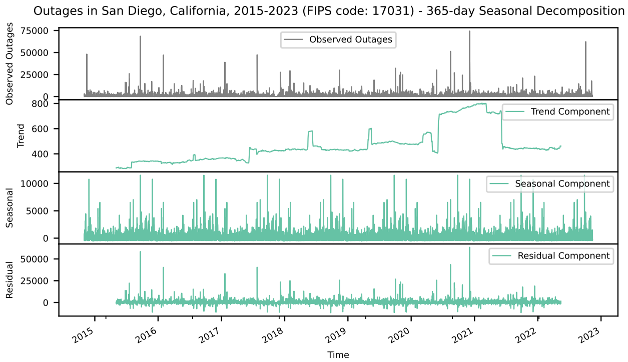

Figure 9: 365-day seasonal decomposition for power outages in San Diego.

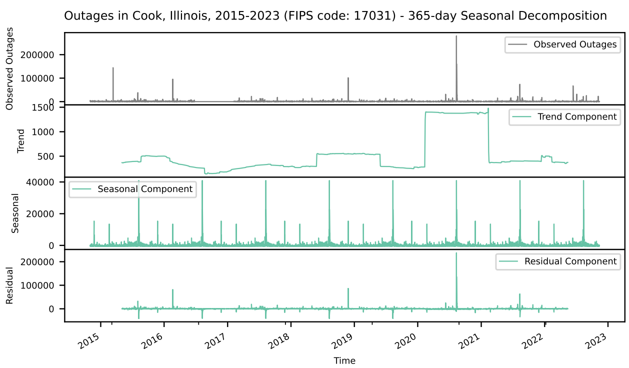

Figure 8: Applying seasonal decomposition on 2014-2023 outage data for Cook County, Illinois with a season of 365 days.

we find in the seasonal component that the largest outages tend to occur towards the end and beginning of each year in Figure 9.

Performing a 365-day seasonal decomposition on a different county, Cook County, Illinois reveals that the largest outages outages typically occur in the fall. Figure 8 demonstrates this in the large spikes in the seasonal component. The trend component also indicates an increase in the magnitude of outages over time until 2022. Applications of seasonal decomposition for anomaly detection are discussed in the appendix.

In the future, we would like to find trends in patterns in major U.S. counties and use an ARIMA model to forecast the expected outage frequency for the next few years and compare it with the historical data. If the forecasted outage frequency is higher than the historical average, it may indicate that there is an increase in outages.

8 CONCLUSION

This paper addresses the challenge questions in "Challenge 7: EAGLE-I Outage Data 2014-2022" of 2023 SMC Data Challenge. In particular, we have studied four major questions and made interesting findings for analyzing restoration times (Question 1), correlation with relevant datasets (Question 2), predicting power outages (Question 5), and identifying patterns over space and time (Question 6).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable suggestions. This work was supported in part by National Science Foundation grants 2319975 and 2331301. The views, opinions, and/or findings expressed in this material are those of the authors and should not be interpreted as representing the official views of the National Science Foundation, or the U.S. Government.

REFERENCES

[1] 2006. *Modeling Financial Time Series with S-PLUS®*. Springer New York. https://doi.org/10.1007/978-0-387-32348-0
[2] 2015. Decoding the Valid Time Event Coding. https://www.weather.gov/bmx/vtec
[3] 2017. Florida power outage: Irma leaves more than 6.5 million in the dark. https://www.cbsnews.com/news/florida-power-outage-irma-leaves-millions-without-power/
[4] 2019. TIGER/Line Shapefiles. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2019.html
[5] 2021. Health Impacts of Extreme Weather. https://www.niehs.nih.gov/research/programs/climatechange/health_impacts/weather_related_morbidity/index.cfm/
[6] 2023. https://www.statsmodels.org/stable/generated/statsmodels.tsa.seasonal.seasonal_decompose.html
[7] Michael Abdelmalak, Sean Ericson, Jordan Cox, Mohammed Ben-Idris, and Eliza Hotchkiss. 2022. A Power Outage Data Informed Resilience Assessment Framework. In *2022 Resilience Week (RWS)*. 1–6. https://doi.org/10.1109/RWS55399.2022.9984016
[8] M. Abeles and M.H. Goldstein. 1977. Multispike train analysis. *Proc. IEEE* 65, 5 (1977), 762–773. https://doi.org/10.1109/PROC.1977.10559
[9] Negin Alemazkoor, Benjamin Rachunok, Daniel R Chavas, Andrea Staid, Arghavan Louhghalam, Roshanak Nateghi, and Mazdak Tootkaboni. 2020. Hurricane-induced power outage risk under climate change is primarily driven by the uncertainty in projections of future hurricane frequency. *Scientific Reports* 10, 1 (Sept. 2020). https://doi.org/10.1038/s41598-020-72207-z
[10] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. *arXiv* (2016). https://doi.org/10.48550/ARXIV.1603.02754
[11] Vivian Do, Heather McBrien, Nina M. Flores, Alexander J. Northrop, Jeffrey Schlegelmilch, Mathew V. Kiang, and Joan A. Casey. 2023. Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA. *Nature Communications* 14 (April 2023), 2470. https://doi.org/10.1038/s41467-023-38084-6
[12] Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin. 2018. CatBoost: gradient boosting with categorical features support. (2018). https://doi.org/10.48550/ARXIV.1810.11363
[13] Romney B. Duffey. 2018. Power Restoration Prediction Following Extreme Events and Disasters. *International Journal of Disaster Risk Science* 10, 1 (Oct. 2018), 134–148. https://doi.org/10.1007/s13753-018-0189-2

[14] Sean Ericson, Jordan Cox, Michael Abdelmalak, Hannah Rabinowitz, and Eliza Hotchkiss. 2022. *Exceedance Probabilities and Recurrence Intervals for Extended Power Outages in the United States.* Technical Report NREL/TP-5R00-83092, 1894479, MainId:83865. NREL/TP–5R00–83092, 1894479, MainId:83865 pages. https://doi.org/10.2172/1894479

[15] Mike Ferry. 2022. Op-ed: California's giant new batteries kept the lights on during the heat wave. https://www.latimes.com/opinion/story/2022-09-13/california-electric-grid-batteries-heat-wave-september-2022

[16] William H. Greene. 1997. *Econometric Analysis, 3rd edition.*

[17] Julien Herzen, Francesco Lässig, Samuele Giuliano Piazzetta, Thomas Neuer, Léo Tafti, Guillaume Raille, Tomas Van Pottelbergh, Marek Pasieka, Andrzej Skrodzki, Nicolas Huguenin, et al. 2022. Darts: User-friendly modern machine learning for time series. *The Journal of Machine Learning Research* 23, 1 (2022), 5442–5447.

[18] Daryl Herzmann. 2019. NWS Valid Time Extent Code (VTEC) Archives. https://mesonet.agron.iastate.edu/info/datasets/vtec.html

[19] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[20] Bandana Kar, John Brewer, Olufemi A. Omitaomu, Benjamin Turner, Samuel Levinson, Yang Chen, Nicholas Roberts, Marija Prica, and Arun Iyengar. 2022. *RePOWERD: Restoration of Power Outage from Wide-area Severe Weather Disruptions.* Technical Report ORNL/TM-2022/2621. Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States). https://doi.org/10.2172/1887668

[21] Pierre Legendre. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74, 6 (1993), 1659–1673.

[22] Michael S Lewicki. 1998. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems* 9, 4 (1998), R53–R78. https://doi.org/10.1088/0954-898X_9_4_001 arXiv:https://doi.org/10.1088/0954-898X$_{940}$01 PMID: 10221571.

[23] Hong Liu and Suzanne T. Parete-Koon. 2023. https://smc-datachallenge.ornl.gov/2023-challenges/

[24] James MacKinnon. 1994. Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests. *Journal of Business Economic Statistics* 12, 2 (1994), 167–76. https://EconPapers.repec.org/RePEc:bes:jnlbes:v:12:y:1994:i:2:p:167-76

[25] James G. MacKinnon. 2010. *Critical Values For Cointegration Tests.* Working Paper 1227. Economics Department, Queen's University. https://ideas.repec.org/p/qed/wpaper/1227.html

[26] Sobhan Moosavi. 2021. US Weather Events (2016 - 2022). https://www.kaggle.com/datasets/sobhanmoosavi/us-weather-events

[27] Sobhan Moosavi, Mohammad Hossein Samavatian, Arnab Nandi, Srinivasan Parthasarathy, and Rajiv Ramnath. 2019. Short and Long-term Pattern Discovery Over Large-Scale Geo-Spatiotemporal Data. *CoRR* abs/1902.06792 (2019). arXiv:1902.06792 http://arxiv.org/abs/1902.06792

[28] Council on Environmental Quality. [n. d.]. Climate and Economic Justice Screening Tool. https://screeningtool.geoplatform.gov/

[29] R Quian Quiroga, Zoltan Nadasdy, and Yoram Ben-Shaul. 2004. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural computation* 16, 8 (2004), 1661–1687.

[30] Christophorus Beneditto Aditya Satrio, William Darmawan, Bellatasya Unrica Nadia, and Novita Hanafiah. 2021. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science* 179 (2021), 524–532.

[31] Skipper Seabold and Josef Perktold. 2010. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, Vol. 57. Austin, TX, 10–25080.

[32] Brian K Sullivan and Naureen S Malik. 2021. Texas power outage: 5 million affected after Winter Storm. https://time.com/5939633/texas-power-outage-blackouts/

[33] Anika Tabassum, Supriya Chinthavali, Varisara Tansakul, and B Aditya Prakash. 2021. Actionable Insights in Urban Multivariate Time-series. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management.* 1774–1783.

[34] Dingwei Wang, Yuxuan Yuan, Rui Cheng, and Zhaoyu Wang. 2023. Data-Driven Outage Restoration Time Prediction via Transfer Learning with Cluster Ensembles. *IEEE Transactions on Power Systems* (2023), 1–14. https://doi.org/10.1109/TPWRS.2023.3236513

APPENDIX

A.1 Time Series Analysis

Darts is a Python library for user-friendly forecasting and anomaly detection on time series [17]. It contains a variety of models useful for time series analysis, including autoregressive integrated moving average (ARIMA) [30], and deep neural networks. The library also makes it easy to backtest models, combine the predictions of several models, and take external data into account. Darts can support both univariate and multivariate time series and models.

Statsmodels is a Python library that provides implementations for many statistical models [31]. We used this library for popular time series techniques including ARIMA, the Dickey-Fuller test, and seasonal decomposition.

ARIMA models are powerful tools for handling time series data with trends and seasonality [30]. By combining autoregression, moving averages, and differencing, these models can effectively capture the patterns and correlations in non-stationary data, making them valuable for forecasting future values and understanding underlying trends.

In time series analysis, lags, moving averages, and autocorrelation are important concepts that help in understanding the model and the temporal patterns and dependencies within the data. Lags refer to the time displacement of a variable's past observations in relation to its current observation. Lags are commonly used in autoregressive models, where the current value of a variable is modeled as a linear function of its past values at various lags [13].

Autocorrelation is a crucial concept in time series analysis, which helps identify patterns and dependencies within the data [21]. Autocorrelation functions (ACF) and partial autocorrelation functions (PACF) are graphical tools used to assess the presence of autocorrelation at various lags. They are essential for determining the appropriate order of autoregressive and moving average terms in autoregressive integrated moving average (ARIMA) models.

Long short-term memory (LSTM) algorithms have been successfully applied in outage prediction due to their ability to handle time-series data with complex patterns and seasonal behavior [19].

Categorical Boosting (CatBoost) is an open-source boosting library [12]. It is a variant of gradient boosting that can handle both categorical and numerical features. It does not require any feature encoding techniques such as one-hot encoding or label encoding to convert categorical features into numerical features. Their usage of weighted preceding data points and nonlinear activation functions makes them well-suited for capturing the correlations between load and weather data patterns.

XGBoost is a convenient algorithm that can be used in classification or regression [10]. Its hyperparameters can be tuned to deal with missing values, down-sample data size, randomly select features to avoid overfitting, and provide the weight of the data to handle an unbalanced dataset.

A.2 National Weather Service Valid Time Extent Code Archives (VTEC) Dataset

Referred as the NWS dataset here for short. This contains the latest metadata and geometries on all of the official VTEC-enabled watch, warning, and advisory events issued by the National Weather Service. These are issued in real time, and are available as ZIP archives for each full year starting from 1986. We collected archives from 2014-2022, matching the temporal extent of the EAGLE-I data. The total storage space of these archives is 9.1GB. Weather events are issued by the NWS in their official VTEC format [2], which documents the event's spatial extent, timeline, severity, weather phenomena, and the latest updates made to the event. [18] We only examine NWS warnings and advisories, as these are for events that actually took place. Due to limitations on RAM and time, we only document the preprocessing algorithm to combine the NWS data with EAGLE-I data and provide an example of correlating the outage durations with official flood warnings and alerts in Figure **??**. The result is that each timestamp and FIPS code in the EAGLE-I data is accompanied by the class of the weather event associated with the weather warning or advisory if there is one present in the county at that time. This pipeline is also available in our GitHub repository. We include this since NWS [2] offers real-time tracking of weather events, creating the possibility of real-time 1:1 monitoring of outages alongside weather hazards at the county level.

To correlate weather events with the outage data, we perform the following steps:

(1) We obtain census tract boundaries from [4] and use their FIPS codes to aggregate the shape files to county boundaries.
(2) We use shapefiles from [4] to overlap the weather event polygons with the county boundaries to get the corresponding FIPS codes.
(3) We filter for events with a significance level of "W" (Warning) and "Y" (Advisory) to get weather hazards that actually took place.
(4) We remove rows where the STATUS is CAN (Cancelled).
(5) We group unique weather events by WFO (Weather Forecasting Office), PHENOM (Phenomena, the type of weather hazard), and ETN (Event Tracking Number, which is only unique to the WFO and Phenomena).
(6) We aggregate each weather event into a single row. Many values in each row are identical and mainly vary by geography. The duration of each weather event is calculated

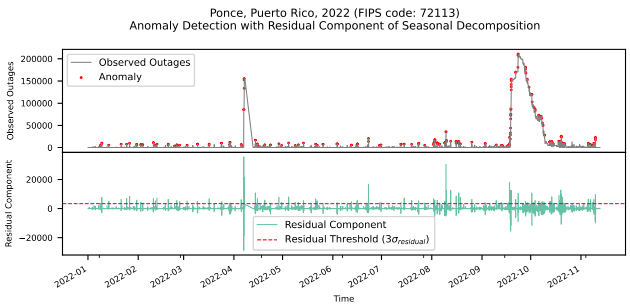

Figure 11: Using a threshold equal to 3 times the standard deviation of the residual component. This only detects irregularities in the time series which have not been accounted for by the trend component.

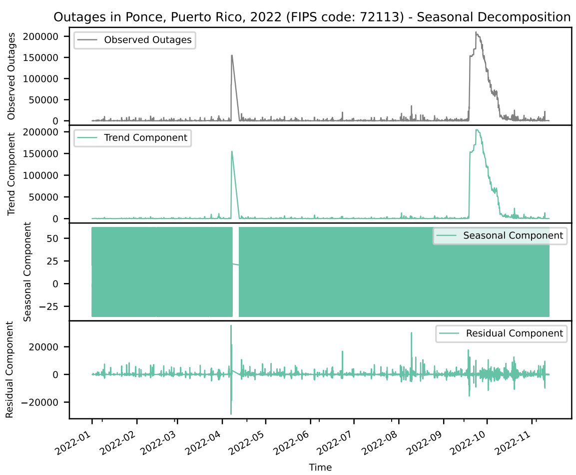

Figure 10: seasonal decomposition on Ponce County, Puerto Rico in 2022 over a period of 4 hours.

using the difference between the latest EXPIRED and earliest INIT_ISS timestamps.

(7) Timestamps are rounded to the nearest 15 minutes and rows are resampled from their start to their end timestamp to align with the timestamps in EAGLE-I.
(8) Duplicate rows resulting from multiple weather warnings for the same hazard in the same county are handled by dropping all duplicates except for those with the latest UPDATED timestamp.
(9) The dataset is merged with the outages and restoration times to find the correlation.

A.3 Anomaly Detection

Analyzing outage restoration times motivates finding a method of identifying and estimating the duration of outage events. We explore spike and anomaly detection techniques to identify power outages and their duration. The use of anomaly detection methods is inspired by the sparse nature of the EAGLE-I data to identify power outages. Unexpected power outages are not typical events, and we wondered if this would be reflected in anomaly detection algorithms. However, advanced anomaly detection requires a concrete definition of what constitutes an anomaly. Due to differences in population and infrastructure, outages vary greatly in range and frequency across counties. Thus, analyses need to be systematically specialized for each county. Therefore, we decided to use a thresholding method for identifying outage periods.

Anomaly detection using seasonal decomposition. We test seasonal decomposition, discussed in section 7, as a method for anomaly detection of outages.

Since the residual component is the original data minus the trend and seasonality components, it represents irregularities in the data. The residual component can be used in many time series datasets to detect anomalies. However, without a small and consistent seasonal period, this seasonal decomposition is not ideal for detecting unusual outage events at a fine-grained level. Seasonal decomposition performed on Ponce, Puerto Rico with a season of 4 hours in Figure 10 leads to a very small season.

Figure 11 demonstrates how anomalies are detected using this method.

However, we find that setting the season period to a longer length provides insight into longer-term patterns in outages, as discussed in section 7.

Anomaly detection using isolation forest. Isolation forest is an efficient ensemble algorithm for identifying anomalies. We apply this to the county-level data using the implementation from the Python package scikit-learn. Most of the time, the number of outages is low. When an outage occurs, there will be a rapid spike, which we hoped the isolation forest would detect. The challenge we faced with this approach is that there are multiple hyperparameters that can be tuned. However, without labels identifying significant power outages for each county, we do not have a good way of optimizing isolation forests reliably for each county. For a handful of counties we observe, isolation forest seems to do well in isolating spikes as consecutive anomalous points, as illustrated by Figure 12.

Table 5 shows the outages correlated with heavy rain and storm conditions from Hurricane Sally in September 2020 and Hurricane Ian in 2022 September caused high daily outages for more than 100,000 customers in North Carolina. Similarly, different extreme weather events like heavy rain and severe storms cause high daily outages in South Carolina. In addition, heavy snow, heavy rain, and storms caused high daily outages impacting more than 100,000 customers in Massachusetts.

Table 7: Power outage prediction for San Diego County by Xgboost regressor model.

Date	Outage	Prediction	Error	Abs_error
(2021, 7, 6)	54 .0	96.1	-41.9	48.0
(2022, 1, 11)	77.0	121.3	-44.0	49.8
(2020, 3, 5)	50.0	96.2	-45.7	52.0
(2020, 4, 14)	49.0	96.4	-46.9	54.1
(2020, 11, 1)	41.0	76.3	-34.7	54.7
(2020, 5, 20)	53.0	96.4	-42.9	54.8
(2022, 6, 26)	90.0	117.6	-27.5	58.1
(2020, 4, 13)	50.0	96.1	-45.3	58.1
(2020, 4, 15)	66.0	96.1	-29.3	58.6
(2021, 9, 7)	116.0	128.2	-11.8	58.9

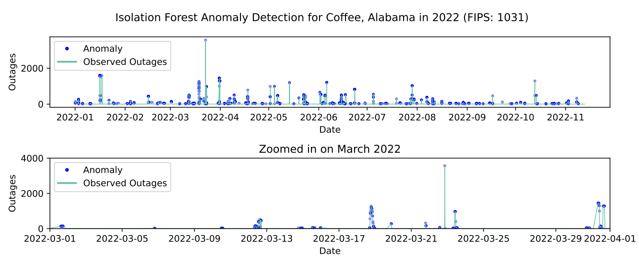

Figure 12: Testing isolation forest on power outages in Coffee, Alabama. The top chart shows anomalous points that were identified over the entire year. The second row zooms into March and demonstrates that some of these anomalous points are consecutive.

Table 5: Extreme power outages for North Carolina during hurricane season between 2016-2022

Date	County	Outages	Type	Severity
2016-10-08	Chatham	141,489	Rain	Heavy
2016-10-08	Wake	117,734	Rain	Heavy
2017-09-12	Lee	208,020	Rain	Moderate
2018-09-14	New Hanover	114,738	Storm	Severe
2021-08-15	Wayne	158,211	Rain	Moderate
2022-09-30	Lee	417,086	Rain	Heavy

Table 6: Extreme power outages during Texas freeze in 2021

Date	County	Outages	Type	Severity
2021-02-15	Bexar	218,382	Cold	Severe
2021-02-15	Dallas	313,948	Cold	Severe
2021-02-15	Denton	108,891	Cold	Severe
2021-02-15	Tarrant	347,719	Cold	Severe
2021-02-15	Travis	237,258	Cold	Severe
2021-02-16	Bexar	329,782	Cold	Severe
2021-02-16	Fort Bend	163,337	Cold	Severe
2021-02-16	Tarrant	367,316	Cold	Severe
2021-02-17	Harris	374,745	Cold	Severe
2021-02-17	Wharton	128,746	Cold	Severe
2021-02-17	Dallas	283,256	Snow	Moderate
2021-02-17	Tarrant	235,554	Snow	Moderate

Table 7 shows the best predicted days by the XGBoost model .