Predicting Power Outage During Extreme Weather Events with EAGLE-I and NWS Datasets

Sangkeun Lee*, Jong Choi*, GS Jung†, Anika Tabassum*, Nils Stenvig‡, Supriya Chinthavali§
*Computer Science and Mathematics Division, Oak Ridge National Laboratory, {*lees4,choij,tabassuma*}*@ornl.gov*
†Computational Sciences and Engineering Division, Oak Ridge National Laboratory, *jungg@ornl.gov*
‡Electrification and Energy Infrastructures Division, Oak Ridge National Laboratory, *stenvignm@ornl.gov*
§Geospatial Science and Human Security Division, Oak Ridge National Laboratory, *chinthavalis@ornl.gov*

***Abstract*—Extreme weather events, such as hurricanes, severe thunderstorms, and floods can significantly disrupt power grid systems, leading to electrical outages that result in inconvenience, economic losses, and life-threatening situations. There is a growing need for a robust and precise predictive model to forecast power outages, which will help prioritize emergency response before, during, and after extreme weather events. In this paper, we introduce machine-learning models that predict power outage risk at the state level during and after extreme weather events. We jointly utilized two publicly available datasets: the U.S. historical power outage data collected by the Environment for Analysis of Geo-Located Energy Information (EAGLE-I™) system, and the National Weather Service historical weather alert data sets. We highlight our initial result and discuss future work aimed at enhancing the model's robustness and accuracy for real-world applications.**

***Index Terms*—Energy resiliency, Severe weather, Power outages, Restoration time, Emergency response**

[1]

I. Introduction

It is essential to comprehend the effects of diverse extreme weather events like hurricanes, thunderstorms, and floods on power outages in different regions [1], as this understanding allows utility companies and emergency responders to react to these incidents more efficiently [2]. In this paper, we formally define a power outage prediction problem of predicting power outage risk at the state level after extreme weather events and benchmark machine learning models including Random Forest, k-NN regression and XGBoost. We jointly utilized two publicly available datasets: the U.S. historical power outage data collected by the Environment for Analysis of Geo-Located Energy Information (EAGLE-I™) system, and the National Weather Service historical weather alert datasets. As a proof-of-concept, the initial result of our tested models is very encouraging and can offer valuable insights into the relationship between weather events and power outages, assisting stakeholders in making informed decisions. Furthermore, we highlighted future research directions to improve the models.

[1]Notice: This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid up, irrevocable, worldwide license to publish or reproduce the published form of the manuscript, or allow others to do so, for U.S. Government purposes. The DOE will provide public access to these results in accordance with the DOE Public Access Plan http://energy.gov/downloads/doe-public-access-plan).

II. Data

Historical power outage data: For the U.S. historical power outage data, we used the data gathered by Oak Ridge National Laboratory (ORNL)'s EAGLE-I™ (Environment for Analysis of Geo-Located Energy Information system) [3]. The EAGLE-I platform has been collecting U.S. electric outage data since 2014. EAGLE-I's electricity outage data includes records of the total number of customers without power within a geographic area at the county level, accompanied by information about the relevant utility company. More precisely, the number of outages represents the number of affected customers, and the outage number snapshot is recorded every 15 minutes.

Weather alert data: For weather information, we utilize data collected by the National Weather Service (NWS), a U.S. government agency that provides weather, water, and climate data, forecasts, warnings, and impact-based decision support services. Specifically, we use the NWS Valid Time Extent Code (VTEC) Archives data processed by Iowa State University's Iowa Environmental Mesonet (IEM) [4]. This dataset contains information about the geography, type (e.g, hurricane, flood, thunderstorm), and advisory level (Watch, Warning, Advisory, etc.) that occurred in the U.S. The dataset and its metadata are publicly available in various formats. To preprocess the dataset, we first excluded entries that were neither warnings nor advisories to focus on extreme weather events. In the second step, we converted all polygon-based data in the NWS data set into county-based data to join with county-level EAGLE-I data set. As polygons can cover multiple counties, we identified all the counties intersecting the polygons and replaced the original data entries with county geometry. Lastly, we removed duplicates and problematic data instances with negative or extremely long event durations (over seven days). Out of a total of 447,266 data rows, 992 rows (0.221%) had negative durations, and 1,503 rows (0.336%) had durations over seven days. We filtered out these problematic cases and retained the remaining 444,661 rows (99.442%). The data size varied per state, as weather event frequency differs among regions. In this paper, we specifically focused on 3 states with varying data sizes (Texas with 15,272 events, Michigan with 4,742 events, and Hawaii with 503 events).

III. Model and Result

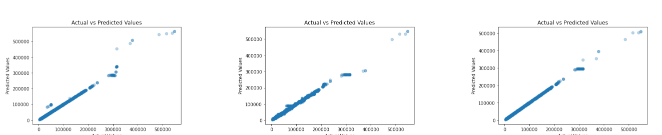
The problem of power outages can be examined from multiple angles. In this study, our main objective is to perform state-level worst-case prediction, by estimating the total summation of the maximum power outages for each county grouping by the state over the next 12 hours when a weather event alert is issued. A state-level risk assessment can provide valuable insights and benefits for various stakeholders, including utility companies, emergency management agencies, and residents, helping them prioritize their actions and make well-informed decisions. We reorganized the EAGLE-I and NWS data to create a training dataset for the selected states (Texas, Michigan, and Hawaii) with varying data availability due to different weather event frequencies. For each past NWS weather alert, we structured the data as follows:

- **Input feature 1**: Maximum outages for each county during the last 12 hours before the weather alert's issuance date and timestamp.
- **Input feature 2**: Type of weather alert (e.g., flood, severe thunderstorm, etc.).
- **Input feature 3**: List of affected counties by a weather alert.
- **Output**: Total summation of the maximum outages from each county grouping by state within the 12 hours following the weather alert's issuance date and timestamp.

With this configuration, we tested 3 machine learning approaches capable of predicting multiple output variables, including Random Forest (RF) [5], k-Nearest Neighbor (k-NN) [6], and Extreme Gradient Boosting (XGBoost [7]). We sorted the data by the timestamp of the NWS weather events and allocated the first 80% of the data for training and the remaining 20% for testing. The number of estimators was selected to be 100 for both RF and XGBoost. In the case of k-NN, the neighborhood size was set to 3.

State	Data Size (# of events)	R^2(RMSE)		
		RF	kNN	XGBoost
Texas	Large (15,272)	0.989 (5922.375)	0.994 (4501.371)	0.998 (2115.354)
Michigan	Medium (4,742)	0.997 (2650.425)	0.984 (6270.681)	0.998 (2092.054)
Hawaii	Small (503)	0.916 (2486.644)	0.930 (2255.152)	0.844 (3385.078)

TABLE I: Evaluation of prediction results of three models on states with different data sizes.



(a) RF (b) k-NN (c) XGBoost

Fig. 1: Comparison of actual and predicted values

Table I shows the summary of the prediction accuracy. We observed R^2 improved with larger data sizes. All three models showed encouraging results for all three states. All three models showed competitive results, but XGBoost showed better RMSE. Fig. 1 shows the visualized result (actual vs. prediction) for the Texas state.

IV. Discussion and Conclusion

In this paper, we demonstrate the feasibility of predicting power outages by training machine learning models using two publicly available datasets. The accuracy of the models is particularly promising for states with more historical data. Specifically, all three models have shown promising results with an R^2 metric of over 0.98 in the case of Texas and Michigan. This is very encouraging because further improvement in accuracy is expected with additional model optimization.

For future work, we will investigate the feasibility of county-level prediction instead of state-level, which will give much more detailed insight to stakeholders. Specifically, we will explore the use of Graph Neural Networks (GNN) [8] to further incorporate the interdependency of power system components. The reliability of power systems during extreme weather events is heavily influenced by the network structure of power grid components, and they can also be affected by other interdependent critical infrastructure components. Thus, modeling the data as a graph will be a natural next step. We observed that the size of the data varies depending on the region. Additionally, some weather events or power outage events may be less common in certain areas, leading to uneven data availability across the nation. Learning from one region with more data and transferring the knowledge for prediction to other regions is a challenge that needs to be addressed. This is important, as drastic climate change has led to rare weather events occurring in unfamiliar regions.

In conclusion, our study serves as a foundation for developing more sophisticated models that can effectively predict power outages during extreme weather events, ultimately enabling better preparedness and response strategies for emergency management and utility companies.

References

[1] E. L. Hotchkiss and A. Dane, "Resilience roadmap: a collaborative approach to multi-jurisdictional resilience planning," tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2019.

[2] F. H. Jufri, V. Widiputra, and J. Jung, "State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies," *Applied energy*, vol. 239, pp. 1049–1065, 2019.

[3] O. R. N. Laboratory, "Environment for analysis of geo-located energy information (eagle-i) system." https://eagle-i.doe.gov/, accessed 2023-05-02.

[4] I. S. University, "Iowa environmental mesonet: Datasets." https://mesonet.agron.iastate.edu/info/datasets, accessed 2023-05-02.

[5] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017.

[6] J. Goldberger, G. E. Hinton, S. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," *Advances in neural information processing systems*, vol. 17, 2004.

[7] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[8] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI open*, vol. 1, pp. 57–81, 2020.