



Juan P. Montoya-Rincon¹

Department of Mechanical Engineering,
The City College of New York,
160 Convent Avenue,
New York, NY 10031
e-mail: jmontoyarincon@ccny.cuny.edu

Jorge E. Gonzalez-Cruz

Department of Atmospheric and Environmental
Sciences,
University at Albany,
1400 Washington Avenue,
Albany, NY 12222
e-mail: jgonzalez-cruz@albany.edu

Michael P. Jensen

Brookhaven National Laboratory,
Upton, NY 11973
e-mail: mjensen@bnl.gov

Evaluation of Power Transmission Lines Hardening Scenarios Using a Machine Learning Approach

The power transmission infrastructure is vulnerable to extreme weather events, particularly hurricanes and tropical storms. A recent example is the damage caused by Hurricane Maria (H-Maria) in the archipelago of Puerto Rico in September 2017, where major failures in the transmission infrastructure led to a total blackout. Numerous studies have been conducted to examine strategies to strengthen the transmission system, including burying the power lines underground or increasing the frequency of tree trimming. However, few studies focus on the direct hardening of the transmission towers to accomplish an increase in resiliency. This machine learning-based study fills this need by analyzing three direct hardening scenarios and determining the effectiveness of these changes in the context of H-Maria. A methodology for estimating transmission tower damage is presented here as well as an analysis of impact of replacing structures with a high failure rate with more resilient ones. We found the steel self-support-pole to be the best replacement option for the towers with high failure rate. Furthermore, the third hardening scenario, where all wooden poles were replaced, exhibited a maximum reduction in damaged towers in a single line of 66% while lowering the mean number of damaged towers per line by 10%. [DOI: 10.1115/1.4063012]

Keywords: grid resilience hardening, power outage prediction, power tower, power transmission, hurricane, machine learning

1 Introduction

The power infrastructure in coastal areas is regularly exposed to hazardous wind and precipitation events, and in particular power transmission system is often catastrophically affected by these weather events. A recent example, Hurricane Maria (H-Maria), damaged more than 55% of Puerto Rico's (PR) transmission towers [1], leaving the island with nearly all of its 2400 miles of transmission and 30,000 miles of distribution lines nonfunctional [2]. Following an event of such massive destruction, the reconstruction of the grid is a major focus. Consequently, studies that guide the reconstruction process and provide guidance for increasing the resiliency of the transmission lines are of particular importance. A number of studies have investigated weather-related damage to the distribution and transmission systems, focusing on the ranking of strategies, and prioritization of techniques for system enhancements. Salman et al. [3] developed fragility curves for the utility poles in the distribution system. These curves were then coupled with a synthetic network model to compare the effectiveness of three different hardening measures, considering cost and critical parts of the distribution system. Other studies created fragility models of pole-wire systems to more accurately account for the damage in the lines [4–6]. These fragility

models were then used to investigate the benefits of burying the distribution lines [7]. Yuan et al. [4] showed how the pole-wire models are used to investigate multiple hardening prioritization options in a distribution line. Ryan et al. [8] used event-based Monte Carlo simulations to study the effect of changing the maintenance strategy for the distribution poles on the infrastructure performance. Furthermore, Hughes et al. [9] demonstrated that replacing aging poles in the distribution lines can decrease wind-induced power outages. Likewise, in other studies, when sufficient data is available, data-driven approaches were demonstrated to correctly estimate the failure of poles in the distribution lines [10–12]. Data-driven methods are also used to investigate how new approaches affect performance under extreme weather events. In particular, the effect of increasing tree trimming frequency and burying the overhead distribution lines on the power outage frequency and duration [13]. Due to the large expense, most transmission line transitions from overhead to underground or relocation are not feasible. On the other hand, the targeted replacement of towers is a more common and cost-effective strategy to increase system resilience. In such a replacement process it is important to know how the newly installed infrastructure will compare with the previous in an extreme weather event. This machine learning-based study addresses this gap by building a model that is capable of predicting hurricane-induced damage to the transmission lines. This model is further used to estimate the damage caused by H-Maria in three different power infrastructure hardening scenarios, where some of the weak structures supporting the lines are replaced by stronger ones.

¹Corresponding author.

Manuscript received March 28, 2023; final manuscript received July 17, 2023; published online August 4, 2023. Assoc. Editor: Cao Wang.

Finally, a comprehensive comparison of all results is conducted to evaluate the effectiveness of the power infrastructure changes and the extent to which each approach can enhance resiliency. Additionally, it is worth noting that recent studies [7] focusing on resilience enhancement can benefit from further improvements by incorporating the hardening options proposed in this paper.

2 Methodology

A data-driven model of transmission structure failure rates is developed using detailed information on transmission structure specifications, damage of the structures, and environmental variables from a numerical weather prediction model.

2.1 Explanatory Variables. The weather variables (i.e., wind speed variables and cumulative rainfall) in this investigation were simulated using a single-layer urban canopy version of the weather research and forecasting (WRF v 3.8.1) model [14], a numerical weather prediction system created by the National Center for Atmospheric Research. The simulation consists of three nested domains, with resolutions of 25 km, 5 km, and 1 km. The third domain, which has a spatial resolution of 1 km, covers the entire island of Puerto Rico (336 points by 156 points). The model comprises 50 vertical levels, with 35 of them being less than 2 kilometers in height. The simulations took place between Sept. 19 and 22, 2017. Pokhrel et al. [15] provide more detailed information on the WRF design and results of H-Maria.

Based on previous studies [16,17], we selected a combination of weather and geographical variables that contributed to the damage to the transmission infrastructure as our explanatory variables.

First, the maximum wind speed during Hurricane Maria was calculated using WRF simulated data from Sept. 19–22. Then, the total duration of high wind speed was calculated by counting the hours of the simulated day with wind speeds higher than 20, 30, and 40 miles per hour (MPH). Moreover, the cumulative rainfall was calculated by adding the one-hour simulated precipitation throughout the storm.

Other important explanatory environmental variables include the land cover type and the land surface elevation. The land cover dataset, with a resolution of 30 m \times 30 m and 12 different land classifications, was retrieved from the National Land Cover database [18]. The land surface elevation was obtained from the United States Geological Survey [19], with a horizontal resolution of 100 m. Table 1 shows a description of all the used variables, including the units, resolution, and range within which each variable is varying.

For this study, PREPA provided damage reports on the transmission lines. The reports included the specific location, material, and tower type of most of the towers in the transmission lines of PR.

To characterize each power tower, the dataset included two additional categorical explanatory variables. The first is the tower's construction type. This variable assigned each tower to a separate category based on its shape and size. The single-pole tower is the most prevalent form of structure in the lines. The material of the tower is the second factor from the reports to consider. According to the reports, wood emerged as the predominant material for towers

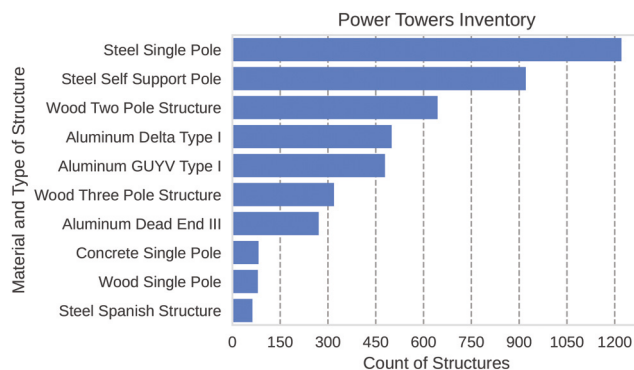


Fig. 1 Inventory of power towers in the transmission lines

along the 115 kV lines, while steel was the prevailing choice for towers along the 230 kV lines. Figure 1 shows the number of structures for each type and material based on the PREPA damage report, for a total of 4647 structures in the dataset.

To build the dataset, the value of each explanatory variable at the position of each power tower was determined using nearest-neighbor interpolation, creating a dataset with the precise value of each explanatory variable in the location of all the towers.

2.2 Response Variable. For the response variable, utility reports were also used. The damage reports provided by PREPA also included information on the type of damage that each of the power towers had sustained, allowing us to identify structures that can withstand wind damage and those that cannot. To incorporate the damage into the dataset, a structure will be categorized as damaged in the model if it requires any repairs after the hurricane.

Given the nature of the failure data, the two classes in the response variable (i.e., damaged, and nondamaged) are heavily unbalanced, with the nondamage category being dominant. This unbalanced pattern is a common problem in natural hazards risk analysis, known as zero-inflation [20]. Two techniques were used to deal with the zero-inflation. First, we randomly reduced the number of samples for the dominant class, balancing the number of samples for the two categories. This technique is also known as under-sampling. Second, we randomly duplicated samples from the minority class to balance both categories. This is known as over-sampling. In Table 2, the number of samples utilized for training in the unbalanced dataset is presented, along with the two employed techniques for reducing zero-inflation.

Table 2 Training dataset size for each sampling method

Model	Training dataset size
Unbalanced	3716
Under-sampled	1478
Over-sampled	5954

Table 1 Explanatory variables

Explanatory variable	Source	Resolution	Units	Range
Maximum wind speed	WRF	1 km	MPH	25–150
Duration of wind speed greater than 20 MPH	WRF	1 km	hours	0–40
Duration of wind speed greater than 30 MPH	WRF	1 km	hours	0–40
Duration of wind speed greater than 40 MPH	WRF	1 km	hours	0–25
Cumulative rainfall	WRF	1 km	inches	0–25
Elevation	United States Geological Survey (USGS)	100 m	Feet	0–1200
Land cover	USGS National Land Cover Database	30 m	Categorical	Categorical
Type of the tower	Puerto Rico Power Authority (PREPA)	Per tower	Categorical	Categorical
Material of the tower	PREPA	Per tower	Categorical	Categorical

2.3 Machine Learning Model. The random forest regression model (RF) [21] was used as a tool to estimate the damage in the power towers given the previously mentioned explanatory variables. To create predictions, RF uses a nonparametric, supervised learning approach that averages the outcomes of an ensemble of decision trees. For training, RF uses the bagging technique, which involves randomly resampling the original dataset with replacement [22]. A limited collection of explanatory variables is chosen at random from the complete set, to do a binary partition of the tree nodes, where each child node is classified as purer than its parent node. This process is used to maximize purity by finding the optimal split candidate [23]. The Gini index is the most used indicator of impurity, so the criteria for dividing the regression tree is based on selecting the input variable with the lowest Gini index [24]

$$I_G(t_{X(x_i)}) = 1 - \sum_j f(t_{X(x_i)}, j)^2 \quad (1)$$

Here, $f(t_{X(x_i)}, j)$ is the percentage of samples with value x_i belonging to leave j as node t [25]. In order to perform regression, RF constructs K $T(x)$ trees and averages the outcomes [22]. As these trees grow, they become the RF regression predictor represented by the following equation:

$$\hat{f}_{rf}^K(x) = \frac{1}{K} \sum_{k=1}^K T(x) \quad (2)$$

Using a random hyperparameter grid search with 300 RF model replicates and a five-fold cross-validation we discovered the best hyperparameters for the RF to be 200 trees, a maximum depth of the tree to be 10, a minimum of five data points placed in a node before the node is split, a maximum of six features considered for splitting a node and default for the remaining.

The interpolated explanatory variables listed in Table 1 were used along with the damage in each tower to construct the dataset. Moreover, using a random split 70% of the data was chosen to be used as the training dataset. The remaining 30% was not included in the training and was utilized to test the model.

2.4 Hardening Scenarios. The state of the transmission system following H-Maria clearly demonstrated the necessity for efforts to minimize high-wind-induced line damage on the island. Using the utility's damage report after H-Maria, we were able to aggregate the percentage of power towers damaged along each

transmission line, with most lines experiencing a percentage of more than 17%, and some as high as 66%, Fig. 2. Additionally, some of the most prevalent types of power towers in Puerto Rico transmission lines are shown in Fig. 3. To increase the resilience of a transmission line to hurricanes, we first must identify the weak factors in the power towers.

Taking the damaged towers from the report and aggregating by material, we found that the power towers made from wood were most likely to be damaged, Fig. 4. To quantify the weakest types of towers, we determined the percentage of power towers that were damaged for each type. Accordingly, the type of tower with the highest failure percent was the two-pole structure, followed by the three-pole structure and the single-pole, Fig. 5.

Based on these findings we selected three hardening scenarios to study. The first one consists of replacing the wooden two-poles with a stronger structure. In the second one, the wooden two-poles and three-poles were replaced. Finally, in the third one, the wooden two-pole, three-pole, and single-poles were all replaced.

To determine which type was the best replacement, a section of a 115 kV line, located on the north-east side of the island was used to analyze different replacement alternatives in the first hardening scenario. As options, we considered the structures and material with the lowest failure frequency on the 115 kV lines: Steel self-support pole; steel single-pole; steel tubular tower; wood self-support pole. Thus, four versions of the transmission line were created, by replacing the wooden two-pole structures with each of the replacement options. Furthermore, we estimated the failure percentages for these distinct versions of the line. The results, ranking, and selection of the ideal option are discussed in the Results section.

After finding an optimal replacement option for the hardening scenarios, we proceeded to expand the study to all the 115 kV transmission lines on the island. The utility damage report included data for most of the transmission lines on the island with the exception of six lines. In order to conduct the island-wide analysis, we created and included a set of synthetic data for these lines. As a reference, we used the U.S. Electric Power Transmission [26] Lines publicly available shapefile, with the approximate location of the lines. Then, following the lines in the shapefile, we constructed points to represent the power towers. The distance (line-span) between the new towers was determined by calculating the median of the distances between towers. These distances were computed based on the information extracted from the damage reports provided by the utility. Furthermore, to maintain a similar balance between the types and materials of structures as observed in the real data, the material and type assignments were randomized with the same percentage of inclusion as found in the actual dataset.

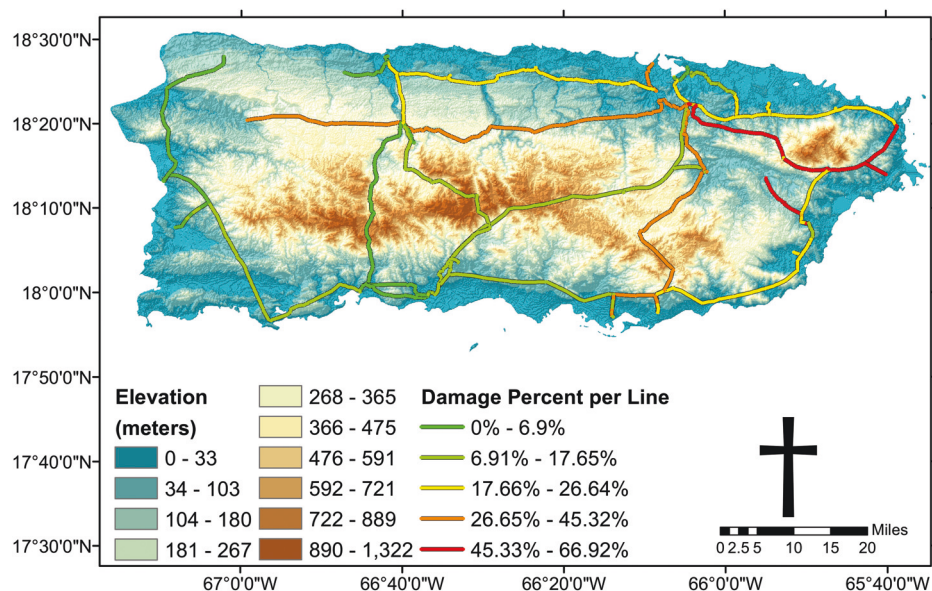


Fig. 2 Percentage of damaged power towers per transmission line

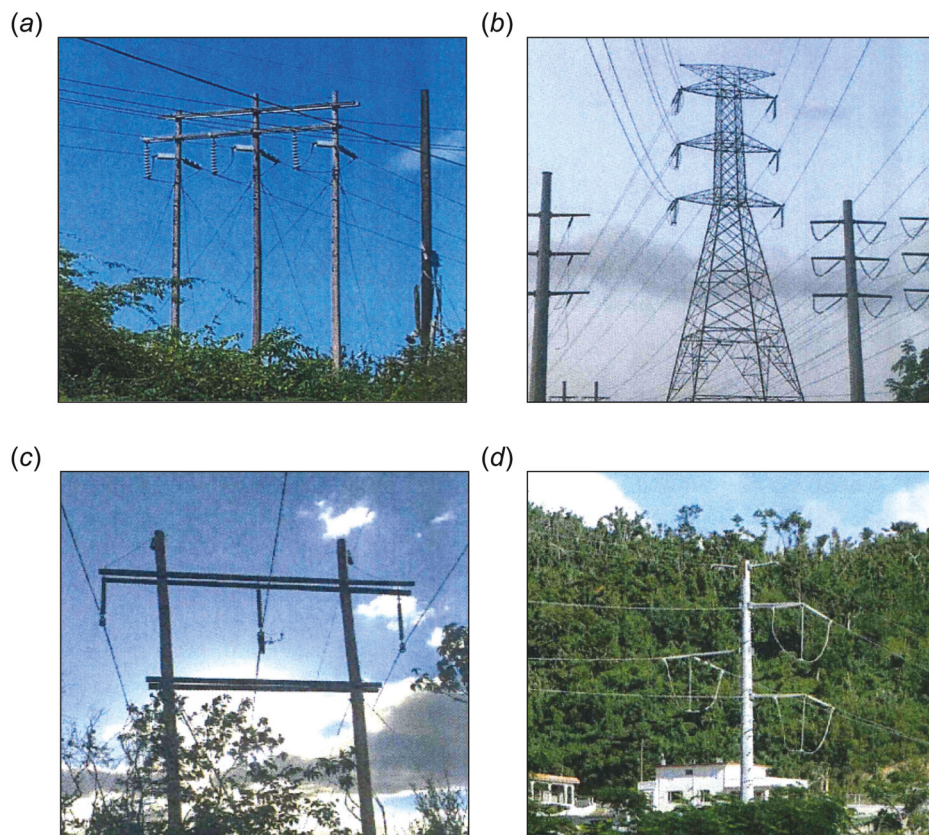


Fig. 3 Type of towers: (a) three-pole structure, (b) self-support tower, (c) two-pole structure, and (d) self-support pole

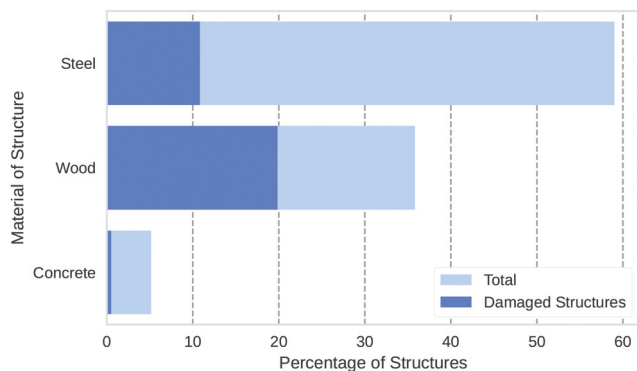


Fig. 4 Damage inventory, grouped by a material of the structure

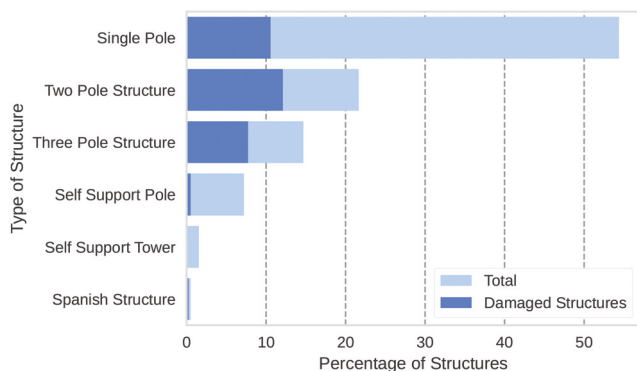


Fig. 5 Damage inventory, grouped by type of structure

3 Results

To test which was the most appropriate sampling technique for the model, we trained and tested the RF with both the over-sampled and under-sampled dataset. In order to evaluate the performance of the various sampling techniques, we utilized the bootstrap method with 100 model replicates. Precision, recall, and F1 score were initially employed as evaluation metrics for each sampling technique. However, considering the imbalanced nature of the dataset, we further incorporated precision gain, recall gain, and F1 Gain as additional evaluation metrics. These gain metrics are particularly suitable for comparing model predictions in skewed datasets, as they undergo transformations that consider the class balance within the dataset [27].

Figure 6 shows a comparison of the performance of both zero-inflation sampling methods with the unbalanced model. The unbalanced model had a high precision with a low recall. This indicates that the model is predicting a high number of false negatives, mostly predicting the dominant class (nondamaged). On the other hand, the under-sampled model showed a low precision with a high recall. This implies a high number of false positives predictions. Moreover, the over-sampled model demonstrated a favorable equilibrium between precision and recall, achieving the highest F1 score and F1 gain score among all the models. Based on this study, over-sampling was determined to be the optimal sampling method for the model, showcasing a commendable balance between precision and recall, and attaining an F1 score and F1 gain score of 0.6 and 0.83, respectively.

As discussed in the methodology section, four replacement alternatives were considered in the study. Table 3 illustrates the proportion of undamaged towers in the transmission line segment for each of the four replacement options that were examined. The steel self-support pole was selected as the best replacement option, followed by the steel single-pole. Replacing the wood two-pole structures with these two options reduced the towers damaged in the

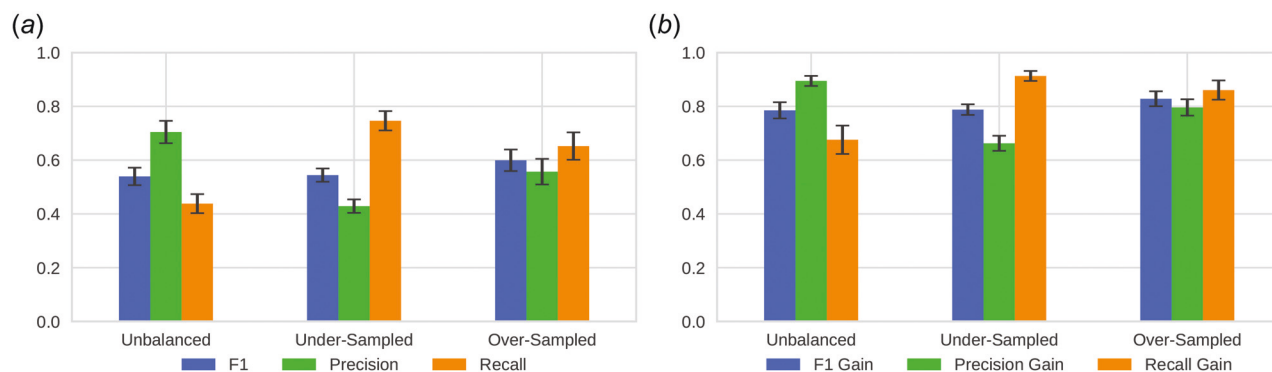


Fig. 6 Evaluation metrics: (a) F1, precision, and recall and (b) F1 gain, precision gain, and recall gain

Table 3 Hardening study results for 115 kV line section

Type of hardening	Undamaged structures
Steel self-support pole	66%
Steel single pole	61%
Tubular	43%
Without replacing	26%
Wood self-support pole	21%

Table 4 Hardening scenarios summary

Hardening Scenarios	Percentage of Damage Towers per Line	
	Mean	Maximum Improvement
Current Infrastructure	22%	NA
Scenario #1	15%	60%
Scenario #2	13%	60%
Scenario #3	12%	66%

section of the line by 40% and 35%, respectively. On the other hand, using the wood self-support pole as a replacement caused an increase in the towers damaged in the section of 5%.

Moreover, the analysis was scaled for all the 115 kV lines in PR using synthetic data for the missing lines. Accordingly, the identified weak structures in each of the three hardening scenarios were replaced by steel self-support poles. Finally, the over-sampled RF model was used to estimate the failure of the power towers in these three hardening scenarios.

The results of the analysis are shown in Fig. 7 and Table 4. The three hardening scenarios show a significant decrease in the damaged structures from the current infrastructure. In addition, the second and the third hardening scenarios have fewer damaged

structures across the lines, with a mean reduction in the damaged structures per line of 9% and 10%, respectively. The difference between the mean improvement of the second and third scenarios is not significant. However, the maximum decrease in damaged structures for a single line improves by 6% in the third hardening scenario, as seen in line 29 in Fig. 7. As a result, the third scenario was chosen as the best configuration of the infrastructure, with decreases in the damaged towers ranging from 1% to 66% for the 115 kV lines.

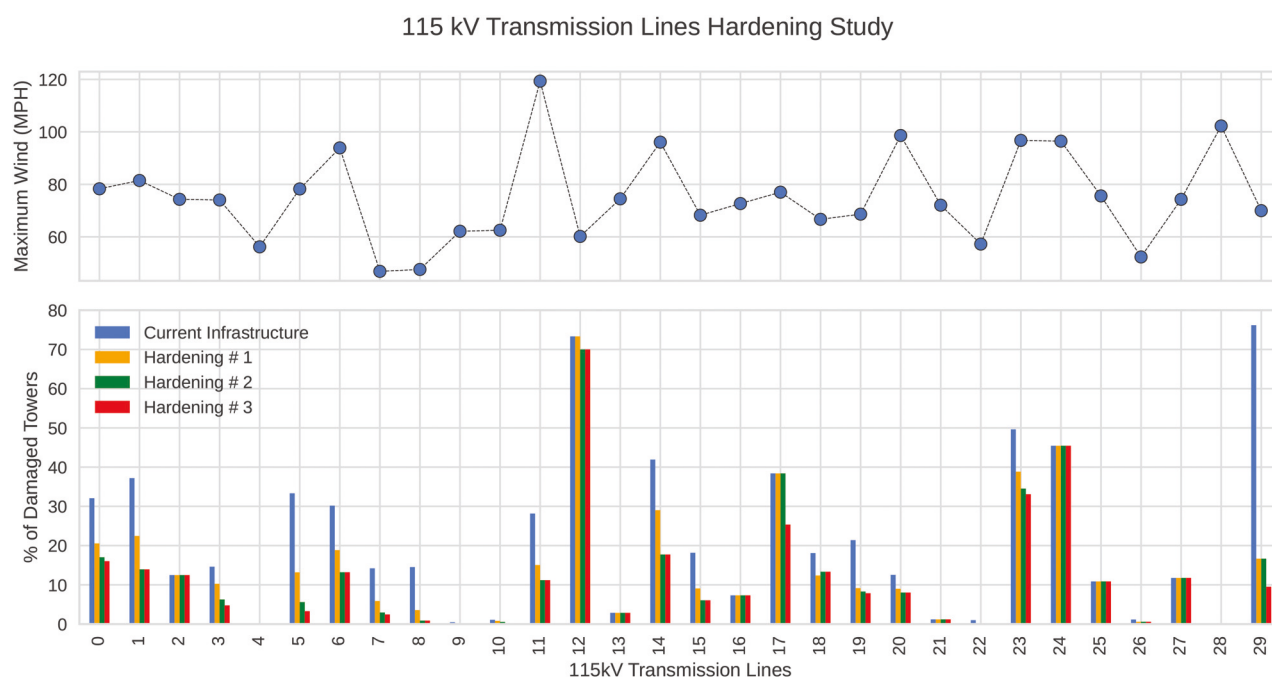


Fig. 7 Island-wide hardening study, each number in the x-axis represents a 115 kV transmission line

4 Conclusions

In this study, we developed a data-driven failure estimation model and a methodology to investigate how the changes in the transmission infrastructure impact the overall resiliency of a transmission line. Puerto Rico was employed as a case study, with damage data from its utility during Hurricane Maria (2017) being used to establish and develop the failure model, as well as rebuild the transmission network topology. Moreover, three different transmission lines hardening scenarios were proposed and explored. The first involves replacing the wooden two-poles with a more durable structure. The wooden two-poles and three-poles were substituted in the second. Finally, the wooden two-pole, three-pole, and single-pole poles were replaced in the third one. Furthermore, four different structures were investigated as replacement options for the hardening scenarios. The ranking of these options was done by quantifying the percentage of damaged structures after replacing each alternative in the first hardening scenario. Looking at the percentage of damaged structures on the north-east 115 kV line section, we concluded that the steel self-support pole was the ideal replacement option, reducing the damaged structures in the section of the line by 40%.

Subsequently, the hardening analysis was scaled to all the lines on the Island. Based on the findings, we conclude that all three hardening scenarios are viable option to increase the resiliency of the lines. However, the third hardening scenario decreased the mean damaged structures per line by 10% and had a maximum decrease in damaged structures in a single line of 66%.

Future extensions of this work may focus on the hardening possibilities for the 230 kV transmission lines in PR. Additionally, other regions of study can be included in the analysis as data become available from the power utilities.

Acknowledgment

The authors gratefully acknowledge the data and support of Puerto Rico Power Authority, Luma Energy, Eng. Alex Echeverria, and Eng. Larry Marini. This support is greatly appreciated.

Funding Data

- The U.S. National Science Foundation (Grant No. CBET-1832678; Funder ID: 10.13039/100000001).
- The National Science Foundation Program for Critical Resilient Interdependent Infrastructure Systems and Processes titled "Integrated Socio-Technical Modeling Framework to Evaluate and Enhance Resiliency in Islanded Communities" (Award No. CBET-1832678; Funder ID: 10.13039/100000001).
- Brookhaven National Laboratory Directed Research and Development (Grant No. 18-020; Funder ID: 10.13039/100006231).

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

Appendix A: Error Metrics

The following error metrics were used in this study:

- Precision

$$\text{precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (\text{A1})$$

- Recall

$$\text{recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (\text{A2})$$

- F1 score

$$\text{F1} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (\text{A3})$$

The precision is describing how accurate the predicted positive values were, by comparing the total predicted positive values with the true positives. This metric is used in models where the cost of high positives is high. The recall evaluates how many of the true positives were correctly predicted by comparing the total actual positives with the true positives. This metric is used to penalize false negatives. Finally, the F1 score is a metric that is used when a balance between precision and recall is wanted but the tested model has an uneven class distribution (unbalanced dataset).

References

- [1] Kwasinski, A., Andrade, F., Castro-Sitiriche, M. J., and O'Neill-Carrillo, E., 2019, "Hurricane Maria Effects on Puerto Rico Electric Power Infrastructure," *IEEE Power Energy Technol. Syst. J.*, **6**(1), pp. 85–94.
- [2] Robles, F., and Bidgood, J., 2017, "Three Months After Maria, Roughly Half of Puerto Ricans Still Without Power," *New York Times*, New York.
- [3] Salman, A. M., Li, Y., and Stewart, M. G., 2015, "Evaluating System Reliability and Targeted Hardening Strategies of Power Distribution Systems Subjected to Hurricanes," *Reliab. Eng. Syst. Saf.*, **144**, pp. 319–333.
- [4] Yuan, H., Zhang, W., Zhu, J., and Bagtzoglou, A. C., 2018, "Resilience Assessment of Overhead Power Distribution Systems Under Strong Winds for Hardening Prioritization," *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.*, **4**(4), p. 04018037.
- [5] Xue, J., Mohammadi, F., Li, X., Sahraei-Ardakani, M., Ou, G., and Pu, Z., 2020, "Impact of Transmission Tower-Line Interaction to the Bulk Power System During Hurricane," *Reliab. Eng. Syst. Saf.*, **203**, p. 107079.
- [6] Lee, S., and Ham, Y., 2021, "Probabilistic Framework for Assessing the Vulnerability of Power Distribution Infrastructures Under Extreme Wind Conditions," *Sustainable Cities Soc.*, **65**, p. 102587.
- [7] Bennett, J. A., Trevisan, C. N., DeCarolis, J. F., Ortiz-García, C., Pérez-Lugo, M., Etienne, B. T., and Clares, A. F., 2021, "Extending Energy System Modelling to Include Extreme Weather Risks and Application to Hurricane Events in Puerto Rico," *Nat. Energy*, **6**(3), pp. 240–249.
- [8] Ryan, P. C., Stewart, M. G., Spencer, N., and Li, Y., 2014, "Reliability Assessment of Power Pole Infrastructure Incorporating Deterioration and Network Maintenance," *Reliab. Eng. Syst. Saf.*, **132**, pp. 261–273.
- [9] Hughes, W., Zhang, W., Bagtzoglou, A. C., Wanik, D., Pensado, O., Yuan, H., and Zhang, J., 2021, "Damage Modeling Framework for Resilience Hardening Strategy for Overhead Power Distribution Systems," *Reliab. Eng. Syst. Saf.*, **207**, p. 107367.
- [10] Guikema, S. D., Quiring, S. M., and Han, S. R., 2010, "Prestorm Estimation of Hurricane Damage to Electric Power Distribution Systems," *Risk Anal.*, **30**(12), pp. 1744–1752.
- [11] Alam, M. M., Zhu, Z., Eren Tokgoz, B., Zhang, J., and Hwang, S., 2020, "Automatic Assessment and Prediction of the Resilience of Utility Poles Using Unmanned Aerial Vehicles and Computer Vision Techniques," *Int. J. Disaster Risk Sci.*, **11**(1), pp. 119–132.
- [12] Hou, H., Yu, S., Wang, H., Huang, Y., Wu, H., Xu, Y., Li, X., and Geng, H., 2019, "Risk Assessment and Its Visualization of Power Tower Under Typhoon Disaster Based on Machine Learning Algorithms," *Energies*, **12**(2), p. 205.
- [13] Nateghi, R., 2018, "Multi-Dimensional Infrastructure Resilience Modeling: An Application to Hurricane-Prone Electric Power Distribution Systems," *IEEE Access*, **6**, pp. 13478–13489.
- [14] Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D., and Duda, M. G., 2008, "A Description of the Advanced Research WRF Version 3," *University Corporation for Atmospheric Research*.
- [15] Pokhrel, R., del Cos, S., Montoya Rincon, J. P., Glenn, E., and González, J. E., 2021, "Observation and Modeling of Hurricane Maria for Damage Assessment," *Weather Clim. Ext.*, **33**, p. 100331.
- [16] Montoya-Rincon, J. P., Azad, S., Pokhrel, R., Ghandehari, M., Jensen, M. P., and Gonzalez, J. E., 2022, "On the Use of Satellite Nightlights for Power Outages Prediction," *IEEE Access*, **10**, pp. 16729–16739.
- [17] Wanik, D. W., Anagnostou, E. N., Astitha, M., Hartman, B. M., Lackmann, G. M., Yang, J., Cerrai, D., He, J., and Frediani, M. E. B., 2018, "A Case Study on Power Outage Impacts From Future Hurricane Sandy Scenarios," *J. Appl. Meteorol. Climatol.*, **57**(1), pp. 51–79.
- [18] Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M., 2004, "Development of a 2001 National Land-Cover Database for the United States," *Photogram. Eng. Remote Sensing*, **70**(7), pp. 829–840.
- [19] National Atlas of the United States, 2012, "100-Meter Resolution Elevation of Puerto Rico and the U.S. Virgin Islands, Albers Projection," *Stanford Digital Repository*, Reston, VA, accessed May 24, 2020, <https://purl.stanford.edu/mbl82my4891>

- [20] Guikema, S., Jun. 2020, "Artificial Intelligence for Natural Hazards Risk Analysis: Potential, Challenges, and Research Needs," *Risk Anal.*, **40**(6), pp. 1117–1123.
- [21] Breiman, L., 2001, "Random Forests," *Machine Learning*, **45**, pp. 5–32.
- [22] Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., and Ribeiro, L., 2014, "Predictive Modeling of Groundwater Nitrate Pollution Using Random Forest and Multisource Variables Related to Intrinsic and Specific Vulnerability: A Case Study in an Agricultural Setting (Southern Spain)," *Sci. Total Environ.*, **476–477**, pp. 189–206.
- [23] Gordon, A. D., Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 1984, "Classification and Regression Trees," *Biometrics*, **40**(3), p. 874.
- [24] Pal, M., 2007, "Random Forest Classifier for Remote Sensing Classification," *Int. J. Remote Sensing*, **26**(1), pp. 217–222.
- [25] Wang, L., Zhou, X., Zhu, X., Dong, Z., and Guo, W., 2016, "Estimation of Biomass in Wheat Using Random Forest Regression Algorithm and Remote Sensing Data," *Crop J.*, **4**(3), pp. 212–219.
- [26] The Homeland Infrastructure Foundation-Level Data, 2021, "U.S. Electric Power Transmission Lines," ArcGIS GeoService, Redlands, CA, accessed June 27, 2023, <https://www.arcgis.com/home/item.html?id=d4090758322c4d32a4cd002ffaa0aa12>
- [27] Flach, P., and Kull, M., 2015, "Precision-Recall-Gain Curves: PR Analysis Done Right," **1**, pp. 838–846.