IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 March 2024, accepted 29 April 2024, date of publication 2 May 2024, date of current version 10 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3396414

RESEARCH ARTICLE

Integrating Structural Vulnerability Analysis and Data-Driven Machine Learning to Evaluate Storm Impacts on the Power Grid

PETER L. WATSON[1,2], (Member, IEEE), WILLIAM HUGHES[2], DIEGO CERRAI[2], WEI ZHANG[2], AMVROSSIOS BAGTZOGLOU[2], AND EMMANOUIL ANAGNOSTOU[2]

[1]Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[2]Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT 06269, USA

Corresponding author: Peter L. Watson (pwatson@lanl.gov)

This work was supported in part by Eversource Energy through the Eversource Energy Center, University of Connecticut, USA.

ABSTRACT The complex interactions between the weather, the environment, and electrical infrastructure that result in power outages are not fully understood, but because of the threat of climate change, the need for models that describe how these factors produce power grid failures is acute. Without them, it remains difficult to understand the amount of weather-related damage we may expect in the future, as well as how changes or upgrades to the infrastructure may mitigate it. To address this problem, a modeling framework is proposed in this article that integrates data derived from structural vulnerability analysis into a machine-learning based weather-related power outage prediction model to create a model that is sensitive both to the weather and the technical configuration of the infrastructure. This Physics Informed Machine Learning (PIML) approach is demonstrated using data from a major power utility operating in the US State of Connecticut, and is compared against a fragility curve modeling approach using some of the same data. The validation of the PIML model shows superior predictive ability, as well as variable sensitivities that follow expected patterns. These results suggest that the model would be able to evaluate the influence that different configurations of the infrastructure would have on the occurrence of power outages caused by severe storms, allowing for the anticipated effects of investments in infrastructural upgrades to be quantified and optimized.

INDEX TERMS Electrical distribution, fragility curves, machine learning, power grid, power outages, reliability.

I. INTRODUCTION

Severe weather events are becoming more frequent and more disruptive to the power grid, and are currently responsible for billions of dollars in economic damage every year [1]. In the United States, this is a particular concern because current climate projections show that the frequency and intensity of thunderstorms are going to significantly increase in North America [2], [3], [4], where they are already a leading cause of power outages [5], [6]. Combined with the effects of aging infrastructure, it will become increasingly difficult for power utilities to deliver reliable electrical power to their customers.

Whereas this a recognized problem, as demonstrated by the recent $10 billion of US Government funding directed to power grid resilience projects [7], how various infrastructural, meteorological, and environmental factors influence the risks of weather-related power outages is still not fully understood. This is not from a lack of trying to understand these processes. To date, there have been many attempts to create predictive models and other analytical frameworks to quantify storm-related infrastructure risk. But because the power grid is so large and complex, and because there are many different factors that potentially influence the risk of failures in the grid (storm characteristics, proximity to trees, drought, vegetation health, infrastructural configuration, etc), attempts to quantitatively describe the

The associate editor coordinating the review of this manuscript and approving it for publication was Hazlie Mokhlis.

processes that contribute to the risk of power outages have involved significant simplification.

II. LITERATURE REVIEW

One established method for weather-related power outage prediction involves fitting an empirical model to data describing various environmental, meteorological and infrastructural factors together with historical outages, often using machine-learning. This method has been applied in various regions of United States [8], [9], [10], [11], using different datasets to describe the weather and the surrounding environment [8], [12], [13], [14], [15], and has been proven to be effective for different types of weather [16], [17], [18]. However, this approach often treats the infrastructural system simply, often ignoring the networked nature of the grid and other aspects of the power system. Infrastructure is accounted for in these models often only by counting the number of utility poles, measuring the length of overhead conductors in an area, or inferring the number of customers. By aggregating information from a large amount of infrastructure, this approach generates models that simplify the complexity of the power system, but have good predictive power in the overall impacts of storms.

These types of weather and environmentally-oriented power outage models can also be effective at quantifying the effects of specific grid hardening measures implemented by power utilities. For example, a recent paper by Taylor et al. [19] describes a weather-related power outage prediction model trained with information that included the application of a new vegetation management standard, called ''Enhanced Tree Trimming'' (ETT) by the local utility. By incorporating this information into a model that is sensitive to various weather patterns, the authors were able to control for the severity of weather over time. The resultant comparisons demonstrated that this new vegetation management standard was able to reduce weather-related power outages by between 25.7 and 42.5% in some cases [19]. However, power utilities have many technical options for upgrading their infrastructure (e.g. new utility poles, stronger power cables, improved crossarms), and it is difficult to evaluate the effect of these types of improvements without more detailed consideration of the infrastructural components and their configuration.

Fortunately there has been considerable work in analyzing the structural vulnerability of power system components to environmental hazards. Specifically, mechanistic modeling of the failure probability of overhead utility pole-wire distribution systems under storm loadings has been particularly widely studied [20]. By treating the loadings and load capacities as random variables in structural simulations, fragility curves or surfaces expressing the pole failure probabilities have been developed as a function of the wind speed [21], [22], [23], [24], [25] as well as combinations of wind and ice [26] or wind and storm surge [27]. However, this physics-based modeling approach is limited because it can only capture certain failure modes, such as pole rupture or foundation failures [27]. Meanwhile, power outages due to nonstructural reasons, such as contact with vegetation, are more challenging to capture with such assessments. This is an important consideration because vegetation-related damages to power infrastructure is very common, with nearly 90% of storm-induced power outages in the US state of Connecticut being related to trees interacting with the overhead lines [11]. Although some recent literature has attempted to use tree fragility curves to calculate the conditional probabilities of trees falling and striking powerlines [28], the diversity in the various tree morphologies and failure mechanisms limit the predictive capacity of such models when applied at scale. To this end, data-driven models have been found to have superior predictive capabilities for in-situ outages [29]. However, such empirical models are limited by the available quantity and quality of training data. This limitation manifests itself particularly for extreme events due to their rarity and relative lack of data. However, physics-based models are not prohibited by these data constraints and predictions at the extremes can be done with more confidence [30], [31]. And in some cases fragility curves are used in more complex, hybrid modeling approaches including Bayesian updating of the fragility curves based on data from observations [32], [33] or using the physics-based simulations of the grid to generate synthetic training data for machine learning models [34].

Because storm damages to the power grid are the result of in-situ environmental and meteorological factors, as well as the structural properties of the infrastructural components, the modeling approaches referenced above all have significant simplifying assumptions that could affect a model's accuracy and limit its potential application. Therefore, the integration of the physics-based modeling with the data-driven predictions could serve to improve the power outage predictions and better capture the complexities of the system.

To address the shortcomings described above, there has been a marked rise in amount of reliability modeling that uses Physics Informed Machine Learning (PIML) in recent years [35], [36]. This term describes a range of methods that combine machine learning and knowledge of physics to create predictive tools or simulations with enhanced capabilities. This can take the form of Machine Learning models that fit data using loss functions informed by physical equations, using physical simulations to generate data for machine learning algorithms, or machine learning based analysis of physical information in complex or engineered systems [36]. Whereas this type of modeling is often used to predict failures in complex engineered systems, it has also been used to inform maintenance schedules, and other types of operational decision-making [36], [37], [38]. PIML approaches have been applied to analyze power systems recently. In Gjorgiev et al, a deep learning model was trained with simulation data from a power systems model to be able to detect failing insulators in the power transmission system [39], and in Varbella et al a physics-based model of power transmission circuits was used to generate training

data for a graph neural network to predict the occurrence of cascading outage events [40].

III. METHODOLOGY

Below, we describe a novel PIML analytical approach that:

- Creates a predictive model for weather-related power outages
- Combines methods from machine-learning weather-related outage prediction models and structural vulnerability analysis
- Sensitive to environmental and meteorological conditions, as well as the age, materials, and configuration of the infrastructural components

We compare this approach with a more typical fragility curve outage model, based on the same structural vulnerability analysis data. Whereas the PIML model is a more complex predictive model, it is able to more completely represent the in-situ risk factors related to weather-related damages to the power grid than previously published models of this type. The motivation for this is not primarily to produce a model with greater predictive accuracy. Because the PIML modeling approach is more comprehensive and includes information derived both the environmental hazards and the infrastructural configuration, it has applications in evaluating theoretical configurations of power infrastructure and the corresponding risks of in-situ weather related outages. The PIML modeling approach has a range of applications, including quantifying the benefits of a variety of technical resilience upgrades or grid hardening measures, very much like Taylor et al. did for specifically for vegetation management [19]. This in-turn would allow power utilities to calculate the expected returns on investments in grid resilience, and to optimize their planned improvements.

The multidisciplinary PIML power outage modeling approach is similar to other machine-learning weather-related power outage prediction models which use a machine-learning algorithm trained on data of various environmental risk factors and the observed outages from historical storms. But in addition to those established methods, we integrate a new set of data derived from fragility curves produced by structural fragility analysis of the power infrastructure in high wind conditions. This information allows us to describe the meteorological and environmental risk factors, along with the mechanical strength of the infrastructure for various ages, materials, and technical configurations. The fragility curves produced can also be used to create a competitive outage model based solely on this information, so that the benefits of our proposed PIML approach can be evaluated.

The PIML model can be considered a Hybrid Physics Informed Machine Learning approach that uses physical simulations to generate datasets [36], but the specific process and aims of this approach are unique in the literature to our knowledge. Whereas PIML has been used in the past to predict failures in the power transmission systems by using physics-based models to generate data [34], [39], [40], we are applying PIML in a way where the physics-based system is used to elucidate one of the risk factors present in a complex system exposed to environmental hazards. Unlike previous works, the ML model is not being applied to replicate the results of the physics-based model, rather the physics-based model is being used to generate information that the ML model can interpret and characterize. Also, the potential applications of this approach are not limited to the prediction of failures or operational decision support. Because the resultant ML outage model will be sensitive to the physical configuration of infrastructural components, it has applications in the design and planning of power systems in order to improve their in-situ resilience to environmental hazards.

Fig. 1 describes the general architecture of this integrated outage prediction model. Numerical Weather Prediction (NWP) simulations are generated and serve as the meteorological inputs to the outage prediction model alongside the vegetation, climatic, and topographic data from various sources, as well detailed infrastructural data provided by a power utility that services about 1.2 million customers in Connecticut, a state in the Northeastern US. That infrastructure data is analyzed to produce the structural failure probabilities of the poles via fragility scores generated from physics-based simulations informed by simulated weather conditions and constrained by the infrastructure's physical characteristics. Finally, all data is aggregated to the power distribution circuit level, where machine learning models are trained to predict the amount of damage in each circuit for given environmental and infrastructural conditions. While this approach is more data intensive than other established methods, and thus more sensitive to data availability, it will produce a model with fewer simplifying assumptions and more fully represents the system it is representing.

A. MODELING DATA

Data for this model came from a wide range of sources including physics-based structural and weather models, utility databases, and national environmental datasets describing the local vegetation, elevation, and climate (e.g., drought) conditions. We aggregated this data for each of the 173 storm events considered in this analysis and for each of the 912 power distribution circuits in the Eversource Connecticut service territory, which produced a database of 31 variables and 157,776 entries that describe the conditions during storms, and the amount of resultant infrastructural damage. All data processing was programmed in R with the `sf` and `terra` libraries supporting for geospatial processing tasks [41], [42], [43].

1) OUTAGE DATA

The power outage models described here predict a specific outage metric: the number of Damage Locations (aka Trouble Spots [TS]), which is any location where the infrastructure gets damaged, causing a power outage and requiring a service

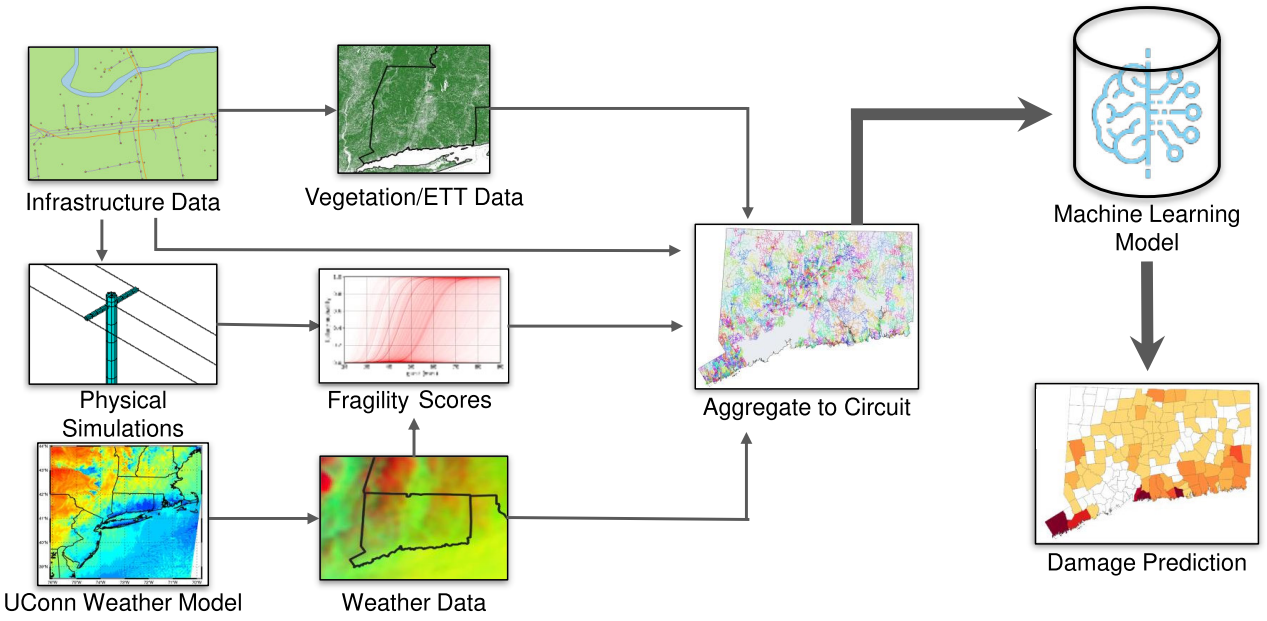


FIGURE 1. Architecture of statistical-mechanistic outage prediction modeling approach, demonstrating how data is processed to produce damage predictions.

crew to repair the damage. This metric is particularly salient to power utilities because the number of Damage Locations is proportional to the number of service crews required to repair the damage promptly, and can strongly influence restoration times. Information for each Damage Location is tracked by the utility's Outage Management System (OMS), which is a computer based system that tracks the damages and is used to dispatch repair crews. For each historical storm used in this analysis, we extracted the total number of damages from the OMS database by assigning each outage that started during the storm to the nearest power circuit, and counting the number of outages associated with each circuit. Then we applied a log transformation to the outages to reduce the numerical range of the trouble spot counts. This shrinks the relative magnitude of extreme values and helps the machine learning algorithm fit the long tailed distribution of the outages [44].

2) INFRASTRUCTURE DATA

Detailed information describing the location and various technical characteristics of the infrastructure is often maintained by major electrical utility companies for asset management purposes. For this project, we used data from the Eversource Connecticut utility service territory that describes the age, class, and location of utility poles; and the location, material, size, and insulation of overhead primary conductors.

The geospatial information of conductors consisted of a collection of line segments which are one or more overhead line spans with similar characteristics. The raw utility data had several flaws in the geometries. We corrected them with a procedural cleaning process, which combined line segments shorter than a full span between two utility poles with neighboring line segments. Using these completed line segments, we were able to associate them spatially with nearby utility poles and calculate the number of poles, average span length, and average age of the poles in each segment. This granular information about the characteristics of the infrastructural components (e.g. age, class, etc) was used to inform the parameters of the structural vulnerability analysis and the resultant fragility scores for each line segment, and was also used to directly derive some variables for the ML outage model. More specifically, to create the infrastructure variables for ML outage model, we aggregated the information about utility poles, overhead conductors, and reclosers for each circuit considered in the model to generate the total number of utility poles, the total length of overhead primary and secondary conductors, and the number of reclosers in each power circuit.

3) WEATHER DATA & STORM SELECTION

173 extratropical and tropical weather events from 2005 to 2020 were included in this analysis, with impacts across the domain that range from about 50 to 20,000 damage locations. The event selection process was guided by an analysis of outage records and weather observations, to identify when and where severe weather coincided with power outages. However, the collection of storms used is not comprehensive. Winter and thunderstorms were not considered for this study because of their meteorological complexity. While related ML outage models have been developed [16], [18], uncertainly in precipitation type, and the chaotic nature of how convective storms develop make fitting accurate outage models challenging. Some events were also removed from consideration because they could not be clearly delineated

from other sources of power outages (e.g. multiple storms in succession, occurrence during heatwaves), or because of missing data for initial and boundary conditions used by the Numerical Weather Prediction (NWP) simulations. The NWP simulations were generated with the Weather Research Forecasting (WRF) system [45], initialized with North American Mesoscale (NAM) forecast system analysis data [46]. These simulations generated meteorological information about each of the 173 storms considered, including descriptions of Winds & Gusts, Planetary Boundary Layer (PBL) height, Temperature, Humidity, and Precipitation over the course of each 48 hour weather event. This is the same weather simulation configuration used in other published works where the technical details of the model configuration are more comprehensively described [10], [13], [19], [44]. Because these simulations are based on a weather analysis product, which are refined with observations of the historical weather conditions, it can be considered to be a best estimate of the weather conditions throughout the simulation domain. The grid cells of the NWP simulation tended to overlap the power circuits, with most power circuits spanning several cells. The weather information was summarized for each power circuit by taking an average of the values of the grid cells that the power circuit overlaps with, weighted by the percentage of overhead lines in each grid cell. This produced an hourly time series of weather conditions for each power circuit for each storm, and summary statistics (mean, max, etc.) were calculated to use as variables for outage modeling, so that total impact of the storm could be evaluated by overall storm characteristics.

4) ENVIRONMENTAL DATA

Several additional variables were developed to describe different environmental risk factors. Specifically variables describing land cover including the tree canopy coverage were developed by sampling and summarizing the National Land Cover Database (NLCD) Canopy Cover [47] and the NLCD Land Cover pixels [48] within 60m of primary overhead conductors for each power circuit. Both of the tree canopy coverage and the land cover datasets are available spatially referenced at 30 meter resolution, and summary statistics of values near overhead lines were calculated via geospatial processing to produce modeling variables.

As with previously published models, we also included a variable derived from the climatology of the leaf area index (LAI) [10], [13], [44]. This variable varies seasonally, and is a reasonable estimate of the amount of leaf surface area presented by vegetation at any given time. This information was available at coarse 10km resolution and is best considered a seasonal indicator of the leaf status of the trees that are present in an area. Because of spatial correlations associated with seasonal changes of leaves, the relatively coarse resolution of this information was consider sufficient to capture the seasonal variability of vegetation, as seasonal change is regional process that affects areas at the scale of hundreds or thousands of kilometers. This data was also used to calculate `KEprox`, a hybrid variable which is the product of the LAI and the square of the maximum wind gust speed from the NWP simulations to approximate the kinetic energy of the wind stress exerted on tree branches.

Variables describing additional environmental conditions were also included in the outage model. Maximum and mean soil moisture levels were included from the NWP simulations to describe soil conditions, which can affect the mechanical properties of the soils and vegetation. To understand the potential longer-term effects of drought on the health of vegetation, we also included the Standardized Precipitation Index (SPI), a drought index, for 1, 3, 12 and 24 month periods by taking an average of each value found in the area of each power circuit [49], producing average 1, 3, 12, and 24 SPIs for each circuit and storm. A variable that describes the elevation of the infrastructure, which can have an influence on localized weather conditions was also included by sampling and summarizing the values from a USGS dataset in the same way as we processed land cover and tree canopy values [50].

A variable describing recent vegetation management activity was also used to condition the risk that vegetation presents to the infrastructure. More specifically, it is a value that describes the proportion of the infrastructure treated with an aggressive vegetation management standard, Enhanced Tree Trimming (ETT), which clears all vegetation that overhangs power lines within the power company's right-of-way around the power distribution infrastructure. This is the same variable used in other published works, whereas mentioned in the Introduction, it has also been used as a variable in machine-learning based analysis quantifying the effectiveness of this method of vegetation management [19], [51], [52]. The processing of this data involved measuring the length of overhead conductors in treated areas and calculating the percentage of each circuit treated with this type of vegetation management.

B. STRUCTURAL VULNERABILITY ANALYSIS

Structural vulnerability analysis was conducted to predict the conditional probability of structural failure under a given set of infrastructure and weather conditions while accounting for variations in the loadings and material properties. Such analyses are based off the fundamental underlying engineering and mechanics principles, such as stress and strain. A limit state function is defined to determine the failure definition, such as when the stress exceeds the material strength or when a certain deflection limit has been exceeded. In the case of comparing the structural resistance R to the stress S, the probability of failure P_f is expressed as:

$$P_f = P(R - S \leq 0) \tag{1}$$

The failure probability can be evaluated through different methods, including simulation-based approaches such as Monte Carlo simulation, as described in a collection of previous literature [21], [26], [27], [31]. Fragility curves can

TABLE 1. Inventory of missing infrastructure data.

Conductor Size	Pole Age	Span Length	Pole Class
4.24%	13.05%	14.96%	56.11%

then be developed, which express the failure probability as a function of some engineering demand parameter, such as earthquake peak ground acceleration or water depth.

For the application to the power distribution system, the limit state of the structural simulations was defined as pole rupture at the groundline, when the bending stress exceeds the pole's flexural strength [21]. The structural model was developed similar to that in [21], where the details of the distributions of the material properties, wind load coefficients, and resistances can be found in detail. The simulations were carried out using the commercial structural analysis software ANSYS. The fragility curves were developed using the wind gust V as the demand parameter by varying the wind speeds and calculating the failure probability at each speed using Monte Carlo simulations. The wind pressures P_w (N/m^2) on the poles and conductors were calculated as:

$$P_w = 0.613V^2 k_z G C_f \tag{2}$$

where V is the 3-second wind gust at 10 m (m/s), k_z is the velocity pressure exposure coefficient, G is the gust response factor, and C_f is the shape factor or force coefficient. The poles are assumed to be Southern yellow pine (SYP) as also done in related work [21], [26], [31]. As the pole thickness increases accordingly with pole height, the height was assumed to be a noncritical parameter and was taken as 13.72 m (45 ft). As the wind forces on the conductors are directly proportional to their projected areas, the conductor projected area was considered for various configurations of the span lengths and conductor size and number [53]. Additionally, the pole age is a critical factor in the pole's bending strength. Presently, the age degradation model proposed by [54] was utilized to model the percentage of strength loss as a function of pole age.

For various combinations of pole class, age, and conductor area, the fragility curves were developed as a function of the wind speed, and lognormal cumulative distribution functions (CDFs) were fit. The effect of the various pole classes, ages, and conductor areas were then investigated through sensitivity analysis, and the mean of the lognormal distributions were conditioned to consider the effects of the various parameter combinations.

To calculate the fragility curve of each line segment, data on the configurations and conditions of the poles and overhead lines were extracted from the utility database. This infrastructure data was made available for each subcircuit line segment, as described above. Due to missing data in some cases, for each line segment, the conductor size, average pole age, average span length, and typical pole class from the available data were calculated and taken as constant throughout the line segment. An inventory of how much missing data was present is shown on Table 1. For the rare cases where no records of characteristics for an infrastructural component were available, the mean values across the system were assumed. For each storm event, the pole ages are first updated based on the event date, and the fragility score was calculated based on the maximum gust values produced by the NWP model and applied to all poles in the line. The fragilities were then compiled to the circuit level by taking the mean of the fragility scores of the lines segments in each circuit. This provided us an average fragility score for each circuit for each event, which is used in both the fragility curve and the PIML outage models.

C. OUTAGE MODELING

Two outage modeling approaches were employed to evaluate the PIML approach over an approach that only uses fragility curves based on the structural analysis. Both modeling methodologies are described below.

1) FRAGILITY CURVE APPROACH

An outage modeling approach that uses the mean pole fragility scores for each circuit, the number of overhead line spans in each circuit, the cumulative density function (CDF) of the binomial distribution, and quantile mapping was developed to evaluate the predictive power of the fragility scores alone over the occurrence of weather related power outages.The binomial distribution was chosen as it is a good representation of the system. The infrastructure has two potential states: damaged or not, and the fragility curves describe the probability of damage. The CDF of the binomial distribution used is defined as:

$$F(x; p, n) = \sum_{i=0}^{x} \binom{n}{i} (p)^i (1-p)^{n-i} \tag{3}$$

where p is the probability of success, and n is the number of draws or attempts made, which in this case was the mean fragility score of a circuit, and the number of overhead wire spans in that circuit respectively. This produced a function for the probability of different levels of damage in each circuit for each storm event. For this modeling method to produce deterministic predictions of the number of damage locations, a specific probability to sample the CDF needed to be determined. This allowed us to effectively tune the sensitivity of the deterministic damage predictions based on the event probability used. Many event probabilities were considered when developing this model, including some very low probabilities to compensate for the typically very low fragility scores and improve the sensitivity to storm events with low and moderate wind gust speeds.

Also due to the bias expected because of the limitations of the fragility curves described in the Introduction, quantile mapping was applied to the deterministic fragility curve outage predictions produced via the method described above. More specifically, the `qmap` library was used to fit parametric

transformations of the predicted outage to more closely match the statistical distribution of the actual outages [55]. The quantile mapping was applied in a leave-one-storm-out manner, to prevent overfitting and produce results comparable to the cross-validation results of the PIML outage model.

2) MACHINE LEARNING APPROACH

In the course of developing the PIML model, many variables were considered for inclusion, but the full set of candidate predictor variables was reduced down to the final set of 31, as described in Table 2, by a multi-stage variable selection process in an effort to improve the simplicity and interpretability of the model. Variables were primarily removed from use for weak variable importance (as described below), or being strongly correlated to other predictor variables. This was particularly problematic if there were correlations with the amount of infrastructure in each circuit because of the strong influence the amount of infrastructure has on amount of outages. Whereas all correlates were not removed from the final list of predictor variables, the accuracy of the algorithms used for modeling is not adversely affected by their inclusion. However, minimizing cross-correlations improves how well we can interpret model sensitives and demonstrate that the model produces consistent and reasonable results for various infrastructural and environmental scenarios. Table 2 also describes the sources of the data for each variable used, and states whether each variable is "Static," with values associated with a particular power circuit which does not vary in time, or "Dynamic," consisting of values that vary from storm to storm.

For outage modeling, we used a Gradient Boosted Tree Model (GBM) configured with a set of tuned hyperparameters. Other algorithms common in ML outage prediction were assessed including Random Forest (RF) and Bayesian Additive Regression Trees (BART), but GBM was selected due to its optimal combination of accuracy and computational efficiency. GBM is a popular tree-based machine learning algorithm that uses a series of simple decision trees to optimize the weights of a larger tree-based model. It does this iteratively, adjusting the model weights based on each subsequent decision tree created, in a way that is very similar to common optimization algorithms [56]. The training process, and thus the fit of the model, is primarily controlled by three hyperparameters: the number of trees used in training, the depth of those trees, and the learning rate used to calculate the optimized weights. Optimal hyperparameters were determined via a search process performed by a differential evolution optimization algorithm [57], as well as an evaluation of model sensitivities to ensure it was consistently fitting the variable space. The model implementation used was from the `gbm` library for R [58] and the hyperparameter search was performed with `DEoptim` [59]. The results from the search are shown in Table 3.

Because of the distortion of the log transformation previously applied to the outages, the predictions from the machine learning model were re-transformed into linear space, and then adjusted using a conditional bias correction. The bias correction consists of a scaling factor applied to storm predictions with a predicted combined impact larger than an optimized threshold. The scaling factors were calculated separately for each storm, only considering the predictions of other storms, such that each bias-correction factor was determined using only out-of-sample information, simulating the accuracy of what a forecast would be for an unknown storm. This is the same type of bias correction used in related work [44], and serves specifically to correct the expected under-prediction of the model for extreme events, which results from the log transformation applied to the outage variable. The log transformation helps manage the extreme distribution of outages by reducing the distance between the most extreme events and the most common events, which improves how the model predicts moderate cases. However, this usually comes at the cost of underestimating the extreme events. In this specific case, scaling factors were only applied to the strongest seven out of the 173 storms considered, and the calculated scaling factors averaged around 4.15, with a relatively tight standard deviation of 0.44.

D. EVALUATION METHODS

Several methods were used to evaluate the developed outage models. Firstly, to evaluate the PIML model's fit to the data and its overall predictive ability, we performed cross-validation and evaluated its predictions. Specifically, we performed a leave-one-storm-out cross validation (LOSO CV), where the data from one storm is removed from the training data, the model is fit on the remaining data, and then the trained model is evaluated by predicting the outcome of the storm that was removed. This process is then repeated for all storms in the training data. The results closely simulate the performance of what we would expect from an operational outage prediction system because it issues predictions for events that the model has not been trained on, and comprehensively evaluates the model on all available storms. It has the additional benefit of preventing data leakage due to the spatial correlations within each storm [13], which is not done in random k-fold cross-validation. These cross-validation results were compared against the quantile mapped predictions of the fragility curve model, which were developed in a similar leave-one-storm-out fashion.

Secondly, to establish the importance of the new predictor variables, we performed an evaluation of the variable importance for the model. This is performed with a model that is trained on all available data. Then for each variable, the information of that variable is destroyed by randomizing the order of the entries, and in-sample model prediction performance is measured with an error metric generically called drop-out loss, but in this case we used the Root Mean Squared Logarithmic Error (RMSLE). Any observed decrease in performance can be attributed to that variable's individual contribution to the predictive skill in the model, and variables with higher dropout loss are interpreted as

TABLE 2. Description and source of information for variables used for outage prediction model.

Variable	Description	Source	Characteristic	Type
poleCounts	Number of Utility Poles	Utility	Static	Integer
pOHlength	Primary Overhead Line Length	Utility	Static	Continuous
sOHlength	Secondary Overhead Line Length	Utility	Static	Continuous
rclrCount	Number of Reclosers	Utility	Static	Integer
MAXWind10m	Maximum 10m Wind Speed	WRF [45]	Dynamic	Continuous
MEANWind10m	Mean 10m Wind Speed	WRF [45]	Dynamic	Continuous
MAXGust	Maximum Wind Gust Speed	WRF [45]	Dynamic	Continuous
MEANGust	Mean Wind Gust Speed	WRF [45]	Dynamic	Continuous
wgt5	Hours of Winds above 5m/s	WRF [45]	Dynamic	Continuous
wgt9	Hours of Winds above 9m/s	WRF [45]	Dynamic	Continuous
Cowgt5	Cont. Hours of Wind above 5m/s	WRF [45]	Dynamic	Continuous
Cowgt9	Cont. Hours of Wind above 9m/s	WRF [45]	Dynamic	Continuous
MEANPBLH	Mean PBL Height	WRF [45]	Dynamic	Continuous
MAXTemp	Maximum 2m Air Temperature	WRF [45]	Dynamic	Continuous
MEANTemp	Mean 2m Air Temperature	WRF [45]	Dynamic	Continuous
MAXSpecHum	Maximum Specific Humidity	WRF [45]	Dynamic	Continuous
MAXPreRate	Maximum Precipitation Rate	WRF [45]	Dynamic	Continuous
MAXTotPrec	Total Precipitation	WRF [45]	Dynamic	Continuous
MAXSoilMst	Maximum Soil Moisture	WRF [45]	Dynamic	Continuous
MEANSoilMst	Mean Soil Moisture	WRF [45]	Dynamic	Continuous
KEprox	Mean Leaf Stress	WRF, Cerrai [13], [45]	Dynamic	Continuous
spi1	1 Month SPI	WWDT [49]	Dynamic	Continuous
spi3	3 Month SPI	WWDT [49]	Dynamic	Continuous
spi12	12 Month SPI	WWDT [49]	Dynamic	Continuous
spi24	24 Month SPI	WWDT [49]	Dynamic	Continuous
avgDEM	Mean Elevation	NED [50]	Static	Continuous
avgCanopy	Mean of Tree Canopy Coverage	NLCD Canopy [47]	Static	Continuous
percLand21	Percent Developed, Open Space	NLCD [48]	Static	Continuous
LAI	Climatological Leaf Area Index	Cerrai [13]	Dynamic	Continuous
percETT	Percent Application of ETT	Utility	Dynamic	Continuous
fragilitymu	Mean Fragility Score	Fragility Curve	Dynamic	Continuous

TABLE 3. Tuned hyperparameter values for GBM algorithm in PIML model.

Tree Number	Tree Depth	Shrinkage
468	5	0.086412

being more important. This process, and the theory behind it is described in Robnik-Sikonja and Kononenko [60]. Because the gradient boosted tree model training process has a randomized component, the outage model was trained 100 times, and an average dropout loss and corresponding 95% confidence intervals were calculated for each variable. These variable importances were calculated using the DALEX library for R [61].

Thirdly, to evaluate the PIML model's sensitivity to different variables, we developed a series of one dimensional partial dependence profiles. Partial dependence profiles (PDPs), also known as partial dependence plots, are used to map out the model's sensitivity to different variable values. PDPs are effectively an overall average of the dynamics of a particular variable based on all observations in a dataset. The other model input variables are effectively fixed, and the individual variable sensitivity is tested by forcing the model with a range of values of the variable being tested. This method was originally developed by Jerome Friedman, and has since become quite popular [56], [62]. Because a PDP is aggregated from all observations in a dataset, the results do not represent interactions between variables particularly well. To quantify the range and distribution of possible model fits, the partial dependence profiles were calculated separately for 100 different model trainings. All PDPs were generated in R using the pdp library [63].

E. ERROR METRICS

For evaluating model accuracy for various analyses, we used several different error metrics: Absolute Percent Error (APE), Root Mean Squared Logarithmic Error (RMSLE), Nash-Suttcliffe Efficiency (NSE), and the Coefficient of Determination (R^2). Equations for these metrics are included below, where actual outages (A) are evaluated against predicted values (P). APE was calculated for first, second, and third quartiles, as well as the mean (Q1 APE, Q2 APE, Q3 APE, MAPE). Also, it should be noted that RMSLE has several beneficial proprieties for evaluating this type of model: firstly, it accepts zero values in both the actual and predicted values, making it appropriate for evaluating circuit level accuracy, which is zero-inflated; and secondly, it is less sensitive to the influence of the extreme events present in the data. NSE is also a useful metric for quantifying the predictive power of a model. It shares many properties with R^2, but is generally considered a more sensitive metric [64]. Possible values range from negative infinity to 1, with higher values being better, and 0 being produced when predictions are comparable to the accuracy of a naive average of

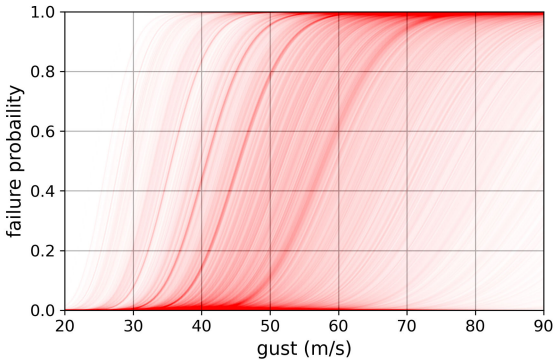


FIGURE 2. Example fragility curve density plot of several thousand sample line segments, showing that the probability of damages in distribution circuits is only significant when wind gusts are very high according to the structural vulnerability analysis conducted.

actual outages.

$$APE = \frac{|P-A|}{A} \times 100 \tag{4}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|P-A|}{A} \times 100 \tag{5}$$

$$\text{RMSLE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(log(P_i+1) - log(A_i+1))^2} \tag{6}$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N}(P_i - A_i)^2}{\sum_{i=1}^{N}(P_i - \bar{A})^2} \tag{7}$$

$$\text{R}^2 = \left(\frac{\sum_{i=1}^{N}(P_i - \bar{P})(A_i - \bar{A})}{\sqrt{\sum_{i=1}^{N}(P_i - \bar{P})^2 \sum_{i=1}^{N}(A_i - \bar{A})^2}}\right)^2 \tag{8}$$

IV. RESULTS

A. FRAGILITY CURVES

Fig. 2 shows a density plot of the variation in the fragility curves for different sample line segments. The bulk of the fragility curves see failure probabilities begin to increase rapidly around wind gusts of 40 to 50 m/s. However, due to combinations of aging infrastructure and less reliable designs, some segments are more fragile and could see high failure probabilities with gusts well below 40 m/s. These results suggest that the fragility curves are able to identify segments of the power distribution system that are particularly vulnerable to structural damage, which would be novel information for the power outage prediction model.

B. MODEL PREDICTIONS

Based on the metrics in Table 4 and Fig. 3, the predictive accuracy for both outage prediction models can be evaluated. Results from the best performing fragility curve model, after developing deterministic predictions at various probabilities of occurrence, are presented here. Using the number of damage locations that had probability of 0.1% of occurring produced predictions with the best balance of NSE and MAPE. This low event probability was required to increase the sensitivity of fragility curve model, and prevent severe under-prediction of events. Even lower probabilities were also considered (down to $1 \cdot 10^{-10}\%$), which continued to reduce under-prediction, but this also resulted in more over-predicted cases, and higher MAPE. The quantile mapping process proved to very effective at improving the raw deterministic outage predictions of this model, and allowed us produce outage predictions with positive NSE values at much higher event probabilities than otherwise. The results from the fragility curve outage model would likely be better if the fragility curve model was more sensitive to lower impact events, or our analysis focused on high impact events. As shown in Figure IV-A this lack of sensitivity to certain events in the fragility curve model is associated with impactful events with lower wind gust speeds. This is intuitive because fragility scores are a function of wind gust speeds, and if a storm was not associated with elevated wind guts the fragility model will not predict outages.

The results of the PIML model are superior to the fragility curve outage model by every metric considered, and are comparable with that of similar ML-based outage models. As shown in Table 4, this model has better MAPE and Q2 APE than all of the models evaluated in Watson et al. [10], similar Q2 APE as the model presented in Wanik et al. [11], and similar MAPE and Q2 APE scores as the model from Yang et al. [37]. Whereas this model has many similarities to the one presented in Taylor et al. [19], it demonstrates better performance than their leave-one-storm-out cross-validated model. Also, it is important to note that the PIML model has the largest dynamic range of any of these models.

It should also be noted that optimizing model accuracy is not necessarily the primary goal for the PIML model. To be applicable for estimating the outcomes of theoretical scenarios, it is also important that this model fits the variables in an explainable way, and its various sensitivities can be shown to be reasonable. This is often challenging for machine-learning models, because it is not possible to directly control how the machine learning algorithm fits the data, but is important for the proposed applications of this model to clearly represent how the various dimensions of power outage risk combine and interact to produce different outcomes.

C. VARIABLE IMPORTANCE

Individual variable importance in the PIML model is shown in Fig. 5. These results are consistent with other outage prediction models where infrastructure, wind, humidity, and leaf stress are evaluated as important [44]. Also, it should be noted that the variables most closely associated with resilience upgrades that power utilities might implement (`percETT` & `fragilitymu`) are also moderately important. Because the training data encompasses examples of very strong and very weak storms, as well as large, expansive power circuits and small ones, it makes sense that the variables that best describe the severity of the storm, as well as the amount of

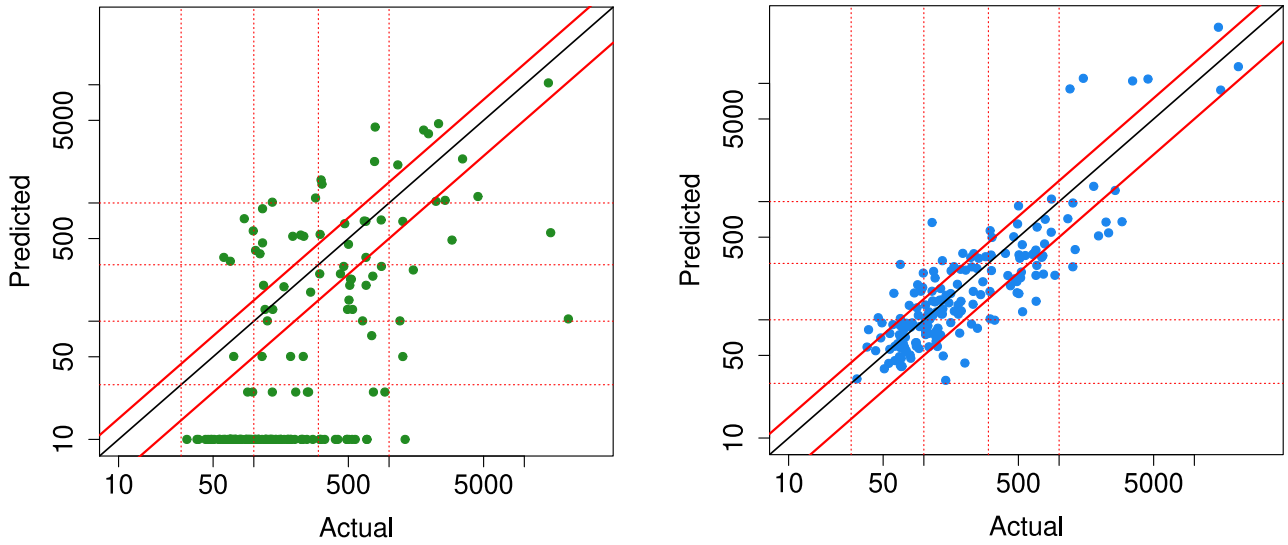


FIGURE 3. Scatterplots of the number of actual damage locations plotted against the predictions of the best performing fragility curve outage model (Green) and the cross-validation predictions of the PIML outage model (Blue). The black diagonal line indicates perfect accuracy, and the solid red diagonal lines indicate an interval of ± 50% error.

TABLE 4. Event-level error metrics from fragility score model with best performing probability of occurrence and PIML cross-validation results. Also includes reported error metrics from related ML models from literature. Asterisks (*) indicate values approximated from figures.

Model	Storms	Outage Range	CV Method	Q1 APE	Q2 APE	Q3 APE	MAPE	RMSLE	R^2	NSE
Fragility Score	173	32 – 21,151	LOSO	90%	100%	100%	115%	3.8458	0.20	0.18
PIML	173	32 – 21,151	LOSO	20%	37%	63%	51%	0.6549	0.64	0.63
Wanik et al [11]	89	20 – 15,000	10 fold	15%*	30%*	70%*	85%*	NA	NA	NA
Watson et al [10]	100	5 – 2,983	LOSO	NA	39%	NA	57%	NA	NA	0.47
Yang et al [37]	100	30* – 4,626	LOSO	19%	38%	60%	46%	NA	0.79	0.79
Talyor et al [19]	165	40* – 5,000*	LOSO	NA	NA	NA	NA	NA	NA	0.55

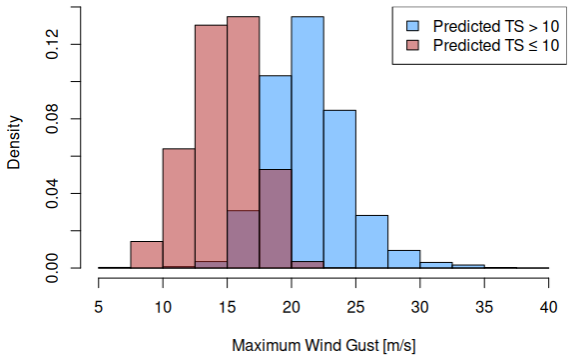


FIGURE 4. Overlapping histograms of maximum wind gust speeds separated by magnitude of fragility model predictions by event.

infrastructure present, are the most influential. However, the `percETT` and `fragilitymu` variables have comparable importance to that of `avgCanopy`, which is the only variable present that describes the risks associated with proximity to vegetation, so the influence of these structural and resilience related variables appears significant.

However, the reader should note that straightforward interpretation of these results can be difficult when predictor variables have correlations between themselves, as is the case for some variables in this model. This is because the information within those variables is not completely destroyed in the importance evaluation process because some other unaltered correlated variable holds similar information. For example, because `MEANGust`, `MAXGust`, `MEANWind10m`, `MAXWind10m`, and the other wind-related variables are significantly correlated, the influence of each of these variable on the model is diffuse, and it is very likely that the overall influence of winds and gusts on model predictions is greater than what Fig. 5 may imply.

Similarly, this consideration can extended to evaluating the importance of the `fragilitymu` variable. The observed dropout loss of `fragilitymu` can be interpreted as the importance of the unique information that it provides to the model, and fragility scores are a function of wind gust speed and the structural characteristics of the infrastructure. Because other sources of information related to magnitude and duration of winds are present, we can surmise that the measured importance of `fragilitymu` is primarily related to its unique contribution to the model, which would be related to the structural fragility of the infrastructure.

D. PARTIAL DEPENDENCE

As shown in Fig. 6, the individual variable partial dependence profiles (PDP) indicate that the PIML model fits the variable space in a largely intuitive way. Because the X axes are in quantiles, it may be difficult to associate the patterns

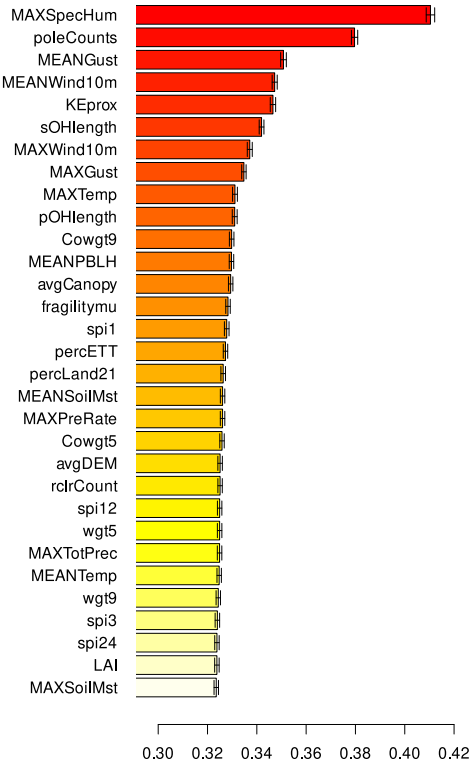


FIGURE 5. Dropout loss indicative of variable importance to the PIML outage prediction model. 95% Confidence intervals are shown based on 100 trainings of the model.

shown PDPs with particular values of each variable, but each plot is comparable to others despite differences in the statistical distributions of each variable. Of the included PDPs, `MAXSpecHum`, `MAXGust`, `KEprox`, `poleCounts`, `pOHlength`, `avgCanopy`, and `fragilitymu` all seem to have a similar dynamic pattern, where the influence of the variable is negligible until about the second quartile, after which impacts increase exponentially. Also the machine learning algorithm appears to very consistently fit the variables, and dynamic patterns appear very consistent across all 100 model trainings. Some extreme values, however are more uncertain: for `MAXGust`, the model fits in divergent ways for the highest values. This effect may be related to interactions with `MEANGust`, which is significantly correlated, or to the relative rarity of extreme winds.

It is interesting to note how much more sensitive the machine learning outage model is to wind speeds than the structural models used to develop the fragility curves. As shown in Fig. 2, the probability of damages in the structural models are very low until high wind gust speeds, 40 m/s as previously mentioned, which are very rare. So rare that values of wind gusts that high are not present in the storms used in this outage model, with the maximum wind gust value in our dataset is 37.7 m/s. Based on the PDPs in Fig. 6, the machine learning outage model demonstrates a greater sensitivity to wind gust speed. The Max Wind Gust speed shows increasing outages starting around the 80th percentile, which corresponds to gust of only about 21 m/s. This discrepancy between the ML outage model and the structural vulnerability analysis can be attributed to the additional in-situ risk factors facing operational power infrastructure that are not accounted for in the idealized simulations used in the structural vulnerability analysis.

V. DISCUSSION

Despite the additional complexity, data processing, and simulation required to develop information about the structural characteristics of power distribution infrastructure using structural vulnerability analysis, when it is incorporated into a machine-learning based outage prediction model, the results demonstrate that we have been able to create a PIML outage model that is sensitive to various infrastructural configurations, and has superior accuracy over a model that only uses information from fragility curves produced from structural analysis. The variable sensitivities of the PIML model appear to be consistent, have intuitive dynamics, and have a significant influence on outage predictions. This model has realistic sensitives to variables like the average fragility score, the amount of Enhanced Tree Trimming, the percentage of tree canopy coverage, the length of primary overhead conductor, the length of secondary overhead conductor, and the number of utility poles. With these variables we can describe a range of different infrastructural characteristics in power infrastructure: utility poles of different ages and classes, various vegetation management practices, or different amounts of overhead infrastructure.

However, it must be noted that the empirical approach used for this model is both a strength and a weakness. Instead of relying on idealized or theoretical frameworks to estimate the occurrence of power outages, this approach uses observed in-situ failures to estimate the risk of outages for various meteorological and environmental conditions as well as infrastructural configurations. By learning from the diversity of conditions that already exists in the distribution grid, this model is able to produce estimates of outage risk that are grounded in reality.

This means the PIML model is also limited by the available data and can only predict with confidence when conditions are represented in the original training data. Extrapolation of findings from this model should be done cautiously. For example, because the training data for this model is limited to Connecticut, and does not include examples of winter storms, one should not make confident conclusions about outage risk in different places or meteorological conditions. This does not mean that the methodology presented would only work under these conditions or generalizability of the presented findings are limited. As discussed in the Introduction, data-driven impact models have been applied successfully under various conditions. It should also be noted that where as Connecticut is a small state, it includes urban, suburban, and rural areas which have a range of different vegetation conditions and infrastructural densities, as shown in the histograms of Figure 7. This means that conclusions derived from this model about the processes that contribute to

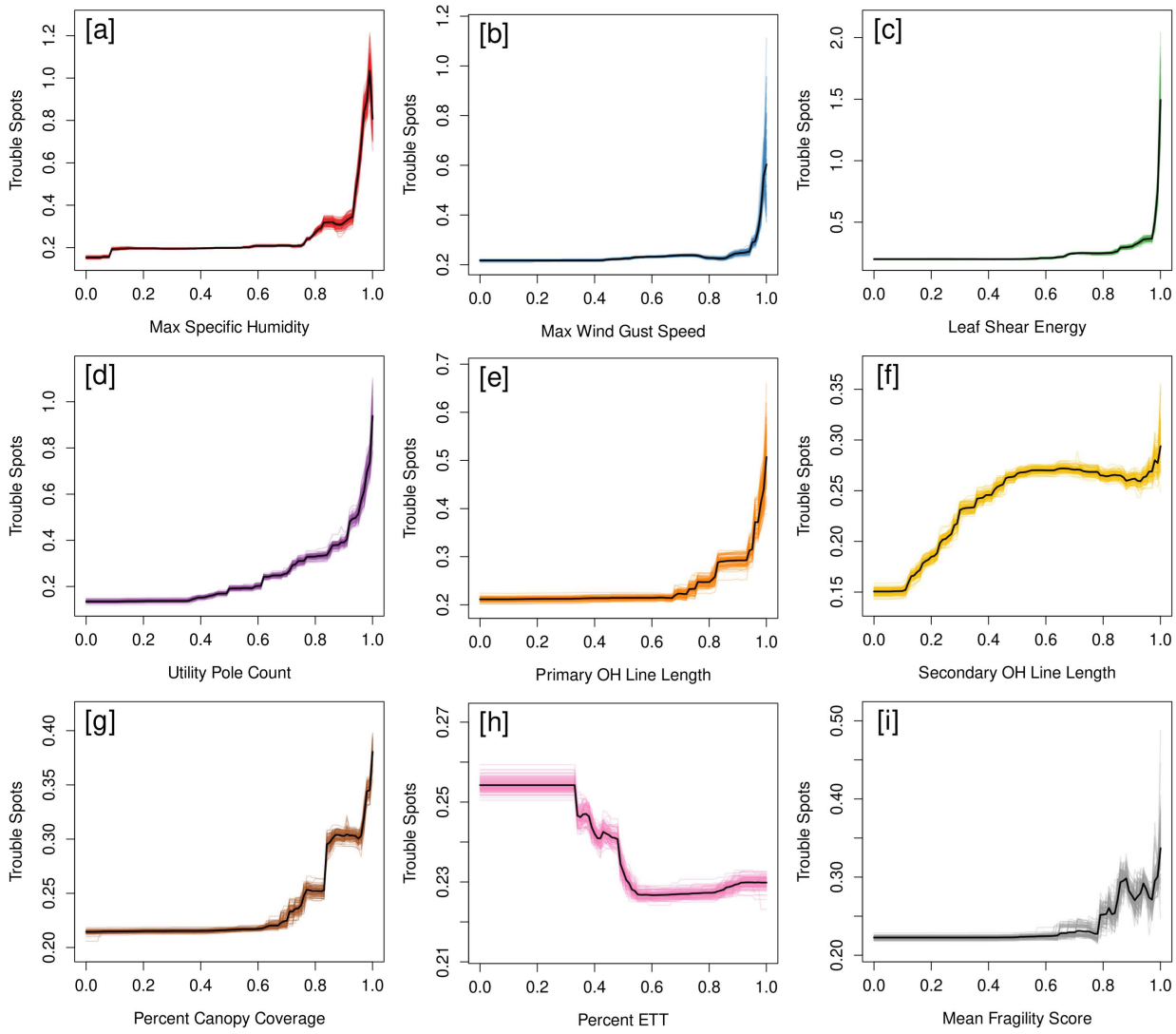


FIGURE 6. Partial dependence profiles for select variable quantiles. Individual model trainings are shown as colored lines, and average values for all 100 trainings are shown in black. The ML algorithm appears to fit the data in a consistent and intuitive manner.

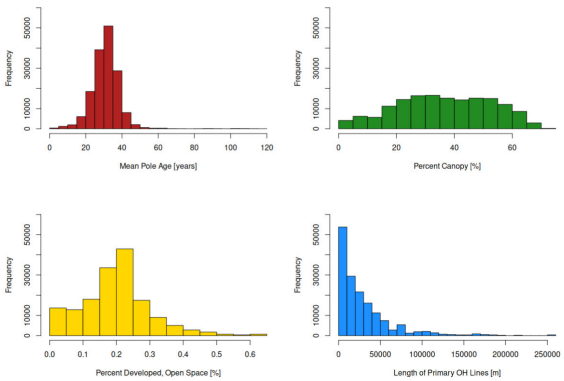


FIGURE 7. Histograms of select environmental and infrastructural variables from the training dataset, including: Mean pole age (Red), percent canopy (Green), percent developed, open space (Yellow), length of primary OH lines (Blue).

weather-related power outages could have some degree of generalizability to similar areas with similar infrastructure, but further investigation and validation is required.

The availability of a wide range of training examples is also a concern for some of the predictor variables used. `percETT` has limited examples of high levels of trimming, especially on circuits with large amounts of infrastructure [19], [52]. This could be why the outage reduction associated with this variable appears to level off around the 55th quantile as seen in Figure 6, and its possible that higher levels of ETT could produce further improvements in grid reliability. However without more examples of high ETT levels in larger circuits we cannot be conclusive. But this is not a problem for all variables. For example, because it is a function of both the structural characteristics of the infrastructure and the maximum wind gusts for each storm, the range of `fragilitymu` values are quite well represented for circuits of all sizes. This allows for confident quantitative predictions of different infrastructural configurations where `fragilitymu` is varied, as long as it remains within the bounds of the maximum and minimum values.

Whereas the generalizability of ML outage models to different types of storms and various places has been demonstrated [10], [16], [17], expansion of this particular PIML modeling approach is limited to service territories that have detailed records of the locations and technical specifications of their infrastructural assets. However, as more and more of asset management practices are digitized, the empirical nature of this analysis will become less and less of a barrier. Indeed, one potential side-benefit of the proposed US government spending on updating the power distribution grid is the generation of additional data about the grid that could be used to validate this data-driven approach. If high-quality and detailed information about federally-funded infrastructural improvements are maintained, it is likely that this methodology can be applied more widely and more comprehensively validated.

As previously stated, because of the interpretability and intuitive sensitivities of this model, it can be used to better understand the processes that contribute to weather-related power outages, as well as quantify outage risk for a range of different meteorological and infrastructural scenarios. This has particular applications in evaluating the effectiveness of various grid hardening measures that power utilities may plan on implementing. Because of the complexities of the system, it is particularly difficult to evaluate how changes or upgrades to the infrastructure influence outage risk. This model could be used for such evaluations, and could be a foundation for cost-benefit analysis and optimization of such measures. For example, if we test the model presented above with a scenario where 10% of the existing utilities poles are upgraded to a larger class, the predicted number outages are reduced by more than 30% in some of the largest storms. This type of information could be extremely important information for electrical utilities planning to upgrade their infrastructure, but are not sure how or where would be most effective. Models like the one presented above have the potential to be the foundation of a decision making support tool that enables the optimization of reliability upgrades to the power grid, which would help us ensure that our infrastructure will remain reliable even as climate change affects the frequency and severity of severe weather events.

APPENDIX
VARIABLES EVALUATED FOR MODELING

TABLE 5. Description of variables evaluated for inclusion in PIML model.

Variable	Description	Source	Group	Status
latLength	Length of Lateral Overhead Lines	Utility	Infrastructure	Removed
boneLength	Length of Backbone Overhead Lines	Utility	Infrastructure	Removed
coveredLength	Length of Covered Overhead Lines	Utility	Infrastructure	Removed
bareLength	Length of Bare Overhead Lines	Utility	Infrastructure	Removed
percCovered	Percent of Covered Overhead Lines	Utility	Infrastructure	Removed
percLat	Percent of Lateral Overhead Lines	Utility	Infrastructure	Removed
fuseCount	Number of Fuses	Utility	Infrastructure	Removed
rclrCount	Number of Reclosers	Utility	Infrastructure	Kept
swtcCount	Number of Switches	Utility	Infrastructure	Removed
poleCounts	Number of Utility Poles	Utility	Infrastructure	Kept
transCounts	Number of Transformers	Utility	Infrastructure	Removed
pOHlength	Primary Overhead Line Length	Utility	Infrastructure	Kept
pUGlength	Primary Underground Line Length	Utility	Infrastructure	Removed
sOHlength	Secondary Overhead Line Length	Utility	Infrastructure	Kept
sUGlength	Secondary Underground Line Length	Utility	Infrastructure	Removed
pUGperc	Percent of Undergroud Primary Conductors	Utility	Infrastructure	Removed
sUGperc	Percent of Undergroud Secondary Conductors	Utility	Infrastructure	Removed
meanPoleAge	Mean Age of Utility Poles	Utility	Infrastructure	Removed
sdPoleAge	Std. Dev. Of Age of Utility Poles	Utility	Infrastructure	Removed
avgCanopy	Mean Tree Canopy Coverage	NLCD Canopy [47]	Vegetation	Kept
stdCanopy	Std. Dev. Of Tree Canopy Coverage	NLCD Canopy [47]	Vegetation	Removed
avgVegHgt	Mean Vegetation Height	GEDI [65]	Vegetation	Removed
stdVegHgt	Std. Dev. Of Veg. Height	GEDI [65]	Vegetation	Removed
avgHardBA	Average Hardwood Basal Area	ITSP [66]	Vegetation	Removed
stdHardBA	Std. Dev. Of Hardwood Basal Area	ITSP [66]	Vegetation	Removed
avgHardSDI	Average Hardwood Stand Density Index	ITSP [66]	Vegetation	Removed
stdHardSDI	Std. Dev. Of Hardwood Stand Density Index	ITSP [66]	Vegetation	Removed
avgSoftBA	Average Softwood Basal Area	ITSP [66]	Vegetation	Removed
stdSoftBA	Std. Dev. Of Softwood Basal Area	ITSP [66]	Vegetation	Removed
avgSoftSDI	Average Softwood Stand Density Index	ITSP [66]	Vegetation	Removed
stdSoftSDI	Std. Dev. Of Softwood Stand Density Index	ITSP [66]	Vegetation	Removed
avgBA	Average Total Basal Area	ITSP [66]	Vegetation	Removed
stdBA	Std. Dev. Of Total Basal Area	ITSP [66]	Vegetation	Removed
avgSDI	Average Total Stand Density Index	ITSP [66]	Vegetation	Removed
stdSDI	Std. Dev. Of Stand Density Index	ITSP [66]	Vegetation	Removed
avgDQ	Average Quadratic Mean Diameter	ITSP [66]	Vegetation	Removed
stdDQ	Std. Dev. Of Quadratic Mean Diameter	ITSP [66]	Vegetation	Removed
avgTF	Average Total Frequency	ITSP [66]	Vegetation	Removed
stdTF	Std. Dev. Of Total Frequency	ITSP [66]	Vegetation	Removed
avgTPA	Average Trees Per Acre	ITSP [66]	Vegetation	Removed
stdTPA	Std. Dev. Of Tree Per Acre	ITSP [66]	Vegetation	Removed
LAI	Climatological Leaf Area Index	Cerrai [13]	Vegetation	Kept

TABLE 5. *(Continued.)* **Description of variables evaluated for inclusion in PIML model.**

percPriProx00	Percent of Primary Conductors within 0m of Veg.	Utility, NLCD Canopy [47]	Vegetation	Removed
percPriProx30	Percent of Primary Conductors within 30m of Veg.	Utility, NLCD Canopy [47]	Vegetation	Removed
percPriProx60	Percent of Primary Conductors within 60m of Veg.	Utility, NLCD Canopy [47]	Vegetation	Removed
percSecProx00	Percent of Secondary Conductors within 0m of Veg.	Utility, NLCD Canopy [47]	Vegetation	Removed
percSecProx30	Percent of Secondary Conductors within 30m of Veg.	Utility, NLCD Canopy [47]	Vegetation	Removed
percSecProx60	Percent of Secondary Conductors within 60m of Veg.	Utility, NLCD Canopy [47]	Vegetation	Removed
percETT	Percent ETT Application to all Lines	Utility, NLCD Canopy [47]	Vegetation	Kept
percAppETT00	Percent ETT to Lines within 0m of Veg.	Utility, NLCD Canopy [47]	Vegetation	Removed
percAppETT30	Percent ETT to Lines within 30m of Veg.	Utility, NLCD Canopy [47]	Vegetation	Removed
percAppETT60	Percent ETT to Lines within 60m of Veg.	Utility, NLCD Canopy [47]	Vegetation	Removed
avgDEM	Mean Elevation	NED [50]	Elevation	Kept
stdDEM	Std. Dev. Of Elevation	NED [50]	Elevation	Removed
spi1	1 Month SPI	WWDT [49]	Drought	Kept
spi3	3 Month SPI	WWDT [49]	Drought	Kept
spi12	12 Month SPI	WWDT [49]	Drought	Kept
spi24	24 Month SPI	WWDT [49]	Drought	Kept
spi36	36 Month SPI	WWDT [49]	Drought	Removed
percLand11	Percent Open Water	NLCD [48]	Land Cover	Removed
percLand21	Percent Developed, Open Space	NLCD [48]	Land Cover	Kept
percLand22	Percent Developed, Low Intensity	NLCD [48]	Land Cover	Removed
percLand23	Percent Developed, Med. Intensity	NLCD [48]	Land Cover	Removed
percLand24	Percent Developed, High Intensity	NLCD [48]	Land Cover	Removed
percLand31	Percent Barren Land	NLCD [48]	Land Cover	Removed
percLand41	Percent Deciduous Forest	NLCD [48]	Land Cover	Removed
percLand43	Percent Evergreen Forest	NLCD [48]	Land Cover	Removed
percLand52	Percent Shrub	NLCD [48]	Land Cover	Removed
percLand71	Percent Grassland	NLCD [48]	Land Cover	Removed
percLand81	Percent Pasture	NLCD [48]	Land Cover	Removed
percLand82	Percent Cropland	NLCD [48]	Land Cover	Removed
percLand90	Percent Woody Wetland	NLCD [48]	Land Cover	Removed
percLand95	Percent Herbaceous Wetlands	NLCD [48]	Land Cover	Removed
wgt5	Hours of Winds above 5m/s	WRF [45]	Weather	Kept
wgt9	Hours of Winds above 9m/s	WRF [45]	Weather	Kept
wgt13	Hours of Winds above 13m/s	WRF [45]	Weather	Removed
Cowgt5	Cont. Hours of Wind above 5m/s	WRF [45]	Weather	Kept
Cowgt9	Cont. Hours of Wind above 9m/s	WRF [45]	Weather	Kept
Cowgt13	Cont. Hours of Wind above 13m/s	WRF [45]	Weather	Removed
ggt13	Hours of Wind Gusts above 13m/s	WRF [45]	Weather	Removed
MEANPBLH	Mean PBL Height	WRF [45]	Weather	Kept
MAXGust	Maximum Wind Gust Speed	WRF [45]	Weather	Kept
MAXWind10m	Maximum 10m Wind Speed	WRF [45]	Weather	Kept
MEANGust	Mean Wind Gust Speed	WRF [45]	Weather	Kept
MEANWind10m	Mean 10m Wind Speed	WRF [45]	Weather	Kept
MAXTotPrec	Total Precipitation	WRF [45]	Weather	Kept
MAXSoilMst	Maximum Soil Moisture	WRF [45]	Weather	Kept
MEANSoilMst	Mean Soil Moisture	WRF [45]	Weather	Kept
MAXPreRate	Maximum Precipitation Rate	WRF [45]	Weather	Kept
MAXSpecHum	Maximum Specific Humidity	WRF [45]	Weather	Kept
MAXTemp	Maximum 2m Air Temperature	WRF [45]	Weather	Kept
MEANTemp	Mean 2m Air Temperature	WRF [45]	Weather	Kept
fragilitymu	Mean Fragility Score	Fragility Curve	Fragility	Kept
fragilitysd	Std. Dev. Of Fragility Score	Fragility Curve	Fragility	Removed
WS_50	Critical Wind Speed for 50% Chance of Failure	Fragility Curve	Fragility	Removed
WS_1	Critical Wind Speed for 1% Chance of Failure	Fragility Curve	Fragility	Removed
WS_50_SD	Std. Dev. Of Critical Wind Speed, 1% Chance of Failure	Fragility Curve	Fragility	Removed
windDiff_01	Difference between Max Gust and 1% Critical Wind Speed	Fragility Curve, WRF [45]	Fragility	Removed
windDiff_50	Difference between Max Gust and 50% Critical Wind Speed	Fragility Curve, WRF [45]	Fragility	Removed
KEprox	Mean Leaf Stress	WRF, Cerrai [13], [45]	Hybrid	Kept

ACKNOWLEDGMENT

The authors have full access to all of the data in this study and they take complete responsibility for the integrity of the data and the accuracy of the analysis. This publication uses classified electric utility data.

AUTHOR CONTRIBUTIONS

- Peter L. Watson: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing - original draft
- William Hughes: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing - original draft
- Diego Cerrai: Conceptualization, Project administration, Supervision, Writing - review & editing
- Wei Zhang: Conceptualization, Project administration, Supervision, Writing - review & editing
- Amvrossios Bagtzoglou: Conceptualization, Project administration, Supervision, Writing - review & editing
- Emmanouil Anagnostou: Conceptualization, Funding acquisition, Supervision, Writing - review & editing

REFERENCES

[1] *Economic Benefits of Increasing Electric Grid Resilience to Weather Outages*, President's Council Econ. Advisers U.S. Dept. Energy's Office Electr. Del. Energy Rel., Washington, DC, USA, 2013.

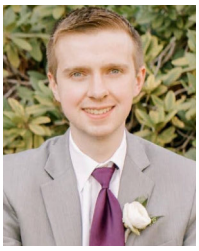
[2] L. Scaff, A. F. Prein, Y. Li, C. Liu, R. Rasmussen, and K. Ikeda, "Simulating the convective precipitation diurnal cycle in North America's current and future climate," *Climate Dyn.*, vol. 55, nos. 1–2, pp. 369–382, Jul. 2020.

[3] N. S. Diffenbaugh, M. Scherer, and R. J. Trapp, "Robust increases in severe thunderstorm environments in response to greenhouse forcing," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 41, pp. 16361–16366, Oct. 2013.

[4] T. R. Moore, H. D. Matthews, C. Simmons, and M. Leduc, "Quantifying changes in extreme weather events in response to warmer global temperature," *Atmosphere-Ocean*, vol. 53, no. 4, pp. 412–425, Aug. 2015.
[5] K. L. Hall, "Out of sight, out of mind," Edison Electr. Inst., Washington, DC, USA, Tech. Rep., 2013. [Online]. Available: https://www.eei.org/-/media/Project/EEI/Documents/Issues-and-Policy/Reliability-and-Emergency-Response/UndergroundReport.pdf
[6] S. Mukherjee, R. Nateghi, and M. Hastak, "A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S." *Rel. Eng. Syst. Saf.*, vol. 175, pp. 283–305, Jul. 2018.
[7] *Infrastructure Investment and Jobs Act*, document Public Law 117-58, 117th U.S. Congr., 2021.
[8] S. M. Quiring, L. Zhu, and S. D. Guikema, "Importance of soil and elevation characteristics for modeling hurricane-induced power outages," *Natural Hazards*, vol. 58, no. 1, pp. 365–390, Jul. 2011.
[9] S. D. Guikema, R. Nateghi, S. M. Quiring, A. Staid, A. C. Reilly, and M. Gao, "Predicting Hurricane power outages to support storm response planning," *IEEE Access*, vol. 2, pp. 1364–1373, 2014.
[10] P. L. Watson, D. Cerrai, M. Koukoula, D. W. Wanik, and E. Anagnostou, "Weather-related power outage model with a growing domain: Structure, performance, and generalisability," *J. Eng.*, vol. 2020, no. 10, pp. 817–826, Oct. 2020.
[11] D. W. Wanik, E. N. Anagnostou, B. M. Hartman, M. E. B. Frediani, and M. Astitha, "Storm outage modeling for an electric distribution network in Northeastern USA," *Natural Hazards*, vol. 79, no. 2, pp. 1359–1384, Nov. 2015.
[12] D. B. McRoberts, S. M. Quiring, and S. D. Guikema, "Improving Hurricane power outage prediction models through the inclusion of local environmental factors," *Risk Anal.*, vol. 38, no. 12, pp. 2722–2737, Dec. 2018.
[13] D. Cerrai, D. W. Wanik, M. A. E. Bhuiyan, X. Zhang, J. Yang, M. E. B. Frediani, and E. N. Anagnostou, "Predicting storm outages through new representations of weather and vegetation," *IEEE Access*, vol. 7, pp. 29639–29654, 2019.
[14] D. F. D'Amico, S. M. Quiring, C. M. Maderia, and D. B. McRoberts, "Improving the Hurricane outage prediction model by including tree species," *Climate Risk Manage.*, vol. 25, 2019, Art. no. 100193.
[15] H. Hou, C. Liu, R. Wei, H. He, L. Wang, and W. Li, "Outage duration prediction under typhoon disaster with stacking ensemble learning," *Rel. Eng. Syst. Saf.*, vol. 237, Sep. 2023, Art. no. 109398.
[16] D. Cerrai, M. Koukoula, P. Watson, and E. N. Anagnostou, "Outage prediction models for snow and ice storms," *Sustain. Energy, Grids Netw.*, vol. 21, Mar. 2020, Art. no. 100294.
[17] B. A. Alpay, D. Wanik, P. Watson, D. Cerrai, G. Liang, and E. Anagnostou, "Dynamic modeling of power outages caused by thunderstorms," *Forecasting*, vol. 2, no. 2, pp. 151–162, May 2020.
[18] P. L. Watson, M. Koukoula, and E. Anagnostou, "Influence of the characteristics of weather information in a thunderstorm-related power outage prediction system," *Forecasting*, vol. 3, no. 3, pp. 541–560, Aug. 2021.
[19] W. O. Taylor, P. L. Watson, D. Cerrai, and E. N. Anagnostou, "Dynamic modeling of the effects of vegetation management on weather-related power outages," *Electr. Power Syst. Res.*, vol. 207, Jun. 2022, Art. no. 107840.
[20] A. Serrano-Fontova, H. Li, Z. Liao, M. R. Jamieson, R. Serrano, A. Parisio, and M. Panteli, "A comprehensive review and comparison of the fragility curves used for resilience assessments in power systems," *IEEE Access*, vol. 11, pp. 108050–108067, 2023.
[21] H. Yuan, W. Zhang, J. Zhu, and A. C. Bagtzoglou, "Resilience assessment of overhead power distribution systems under strong winds for hardening prioritization," *ASCE-ASME J. Risk Uncertainty Eng. Syst. A, Civil Eng.*, vol. 4, no. 4, Dec. 2018, Art. no. 04018037.
[22] M. Panteli and P. Mancarella, "Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies," *Electr. Power Syst. Res.*, vol. 127, pp. 259–270, Oct. 2015.
[23] M. Panteli and P. Mancarella, "Modeling and evaluating the resilience of critical electrical power infrastructure to extreme weather events," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1733–1742, Sep. 2017.
[24] Q. Lu, W. Zhang, and A. C. Bagtzoglou, "Physics-based reliability assessment of community-based power distribution system using synthetic hurricanes," *ASCE-ASME J. Risk Uncertainty Eng. Syst. A, Civil Eng.*, vol. 8, no. 1, Mar. 2022, Art. no. 04021088.
[25] Q. Lu and W. Zhang, "Integrating dynamic Bayesian network and physics-based modeling for risk analysis of a time-dependent power distribution system during hurricanes," *Rel. Eng. Syst. Saf.*, vol. 220, Apr. 2022, Art. no. 108290.
[26] Y. Ma, Q. Dai, and W. Pang, "Reliability assessment of electrical grids subjected to wind hazards and ice accretion with concurrent wind," *J. Struct. Eng.*, vol. 146, no. 7, Jul. 2020, Art. no. 04020134.
[27] Y. Darestani, J. Padgett, and A. Shafieezadeh, "Parametrized wind–surge–wave fragility functions for wood utility poles," *J. Structural Eng.*, vol. 148, no. 6, Jun. 2022, Art. no. 04022057.
[28] Q. Lu and W. Zhang, "An integrated damage modeling and assessment framework for overhead power distribution systems considering tree-failure risks," *Struct. Infrastruct. Eng.*, vol. 19, no. 12, pp. 1745–1760, Mar. 2022.
[29] C. Zhai, T. Y.-J. Chen, A. G. White, and S. D. Guikema, "Power outage prediction for natural hazards using synthetic power distribution systems," *Rel. Eng. Syst. Saf.*, vol. 208, Apr. 2021, Art. no. 107348.
[30] W. Hughes, W. Zhang, A. C. Bagtzoglou, D. Wanik, O. Pensado, H. Yuan, and J. Zhang, "Damage modeling framework for resilience hardening strategy for overhead power distribution systems," *Rel. Eng. Syst. Saf.*, vol. 207, Mar. 2021, Art. no. 107367.
[31] W. Hughes, W. Zhang, D. Cerrai, A. Bagtzoglou, D. Wanik, and E. Anagnostou, "A hybrid physics-based and data-driven model for power distribution system infrastructure hardening and outage simulation," *Rel. Eng. Syst. Saf.*, vol. 225, Sep. 2022, Art. no. 108628.
[32] S.-R. Han, D. Rosowsky, and S. Guikema, "Integrating models and data to estimate the structural reliability of utility poles during hurricanes: Integrating models and data to estimate structural reliability," *Risk Anal.*, vol. 34, no. 6, pp. 1079–1094, Jun. 2014.
[33] A. F. Mensah and L. Dueñas-Osorio, "Outage predictions of electric power systems under Hurricane winds by Bayesian networks," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, Durham, U.K., Jul. 2014, pp. 1–6.
[34] M. Noebels, R. Preece, and M. Panteli, "A machine learning approach for real-time selection of preventive actions improving power network resilience," *IET Gener., Transmiss. Distrib.*, vol. 16, no. 1, pp. 181–192, Jan. 2022.
[35] Z. Xu and J. H. Saleh, "Machine learning for reliability engineering and safety applications: Review of current status and future opportunities," *Rel. Eng. Syst. Saf.*, vol. 211, Jul. 2021, Art. no. 107530.
[36] Y. Xu, S. Kohtz, J. Boakye, P. Gardoni, and P. Wang, "Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges," *Rel. Eng. Syst. Saf.*, vol. 230, Feb. 2023, Art. no. 108900.
[37] F. Yang, P. Watson, M. Koukoula, and E. N. Anagnostou, "Enhancing weather-related power outage prediction by event severity classification," *IEEE Access*, vol. 8, pp. 60029–60042, 2020.
[38] N. Chahrour, M. Nasr, J.-M. Tacnet, and C. Bérenguer, "Deterioration modeling and maintenance assessment using physics-informed stochastic Petri nets: Application to torrent protection structures," *Rel. Eng. Syst. Saf.*, vol. 210, Jun. 2021, Art. no. 107524.
[39] B. Gjorgiev, L. Das, S. Merkel, M. Rohrer, E. Auger, and G. Sansavini, "Simulation-driven deep learning for locating faulty insulators in a power line," *Rel. Eng. Syst. Saf.*, vol. 231, Mar. 2023, Art. no. 108989.
[40] A. Varbella, B. Gjorgiev, and G. Sansavini, "Geometric deep learning for online prediction of cascading failures in power grids," *Rel. Eng. Syst. Saf.*, vol. 237, Sep. 2023, Art. no. 109341.
[41] *R: A Language and Environment for Statistical Computing*, R Found. Stat. Comput., R Core Team, Vienna, Austria, 2022.
[42] E. Pebesma, "Simple features for R: Standardized support for spatial vector data," *R J.*, vol. 10, no. 1, pp. 439–446, 2018.
[43] R. J. Hijmans, *Terra: Spatial Data Analysis*, document Version 1.5-34, R Package, 2022.
[44] P. L. Watson, A. Spaulding, M. Koukoula, and E. Anagnostou, "Improved quantitative prediction of power outages caused by extreme weather events," *Weather Climate Extremes*, vol. 37, Sep. 2022, Art. no. 100487.
[45] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X. Y. Huang, W. Wang, and J. G. Powers, "A description of the advanced research WRF version 3," Nat. Center Atmos. Res. (NCAR), Boulder, CO, USA, NCAR Tech. Rep. TN-475+STR, 2008.
[46] *NCEP North American Mesoscale (NAM) 12 km Analysis*, Environ. Model. Center, Nat. Centers Environ. Predict., Nat. Weather Service, NOAA, U.S. Dept. Commerce, Washington, DC, USA, 2015.
[47] J. W. Coulston, G. G. Moisen, B. T. Wilson, M. V. Finco, W. B. Cohen, and C. K. Brewer, "Modeling percent tree canopy cover: A pilot study," *Photogrammetric Eng. Remote Sens.*, vol. 78, no. 7, pp. 715–727, Jul. 2012.
[48] S. Jin, C. Homer, L. Yang, P. Danielson, J. Dewitz, C. Li, Z. Zhu, G. Xian, and D. Howard, "Overall methodology design for the United States national land cover database 2016 products," *Remote Sens.*, vol. 11, no. 24, p. 2971, Dec. 2019.

[49] J. T. Abatzoglou, D. J. McEvoy, and K. T. Redmond, "The west wide drought tracker: Drought monitoring at fine spatial scales," *Bull. Amer. Meteorol. Soc.*, vol. 98, no. 9, pp. 1815–1820, Sep. 2017.
[50] D. B. Gesch, G. A. Evans, M. J. Oimoen, and S. Arundel, "The national elevation dataset," in *The DEM Users Manual*, 3rd ed. Baton Rouge, LA, USA: American Society for Photogrammetry and Remote Sensing, 2018, pp. 83–110.
[51] D. Cerrai, P. Watson, and E. N. Anagnostou, "Assessing the effects of a vegetation management standard on distribution grid outage rates," *Electr. Power Syst. Res.*, vol. 175, Oct. 2019, Art. no. 105909.
[52] W. O. Taylor, P. L. Watson, D. Cerrai, and E. Anagnostou, "A statistical framework for evaluating the effectiveness of vegetation management in reducing power outages caused during storms in distribution networks," *Sustainability*, vol. 14, no. 2, p. 904, Jan. 2022.
[53] Y. Mohammadi Darestani and A. Shafieezadeh, "Multi-dimensional wind fragility functions for wood utility poles," *Eng. Struct.*, vol. 183, pp. 937–948, Mar. 2019.
[54] A. Shafieezadeh, U. P. Onyewuchi, M. M. Begovic, and R. DesRoches, "Age-dependent fragility models of utility wood poles in power distribution networks against extreme wind hazards," *IEEE Trans. Power Del.*, vol. 29, no. 1, pp. 131–139, Feb. 2014.
[55] L. Gudmundsson, J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen, "Technical note: Downscaling RCM precipitation to the station scale using statistical transformations—A comparison of methods," *Hydrol. Earth Syst. Sci.*, vol. 16, no. 9, pp. 3383–3390, Sep. 2012.
[56] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[57] D. Ardia, K. Boudt, P. Carl, K. M. Mullen, and B. G. Peterson, "Differential Evolution with DEoptim: An application to non-convex portfolio optimization," *R J.*, vol. 3, no. 1, pp. 27–34, 2011.
[58] B. Greenwell, B. Boehmke, and J. Cunningham, *GBM: Generalized Boosted Regression Models*, document Version 2.1.8.1, R Package, GBM Developers, 2022.
[59] K. Mullen, D. Ardia, D. Gil, D. Windover, and J. Cline, "DEoptim: An R package for global optimization by differential evolution," *J. Stat. Softw.*, vol. 40, no. 6, pp. 1–26, 2011.
[60] M. Robnik-Sikonja and I. Kononenko, "Explaining classifications for individual instances," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 5, pp. 589–600, May 2008.
[61] P. Biecek, "DALEX: Explainers for complex predictive models in R," *J. Mach. Learn. Res.*, vol. 19, no. 84, pp. 1–5, 2018.
[62] P. Biecek and T. Burzykowski, *Explanatory Model Analysis*. Boca Raton, FL, USA: CRC Press, 2021.
[63] B. Greenwell, "PDP: An R package for constructing partial dependence plots," *R J.*, vol. 9, no. 1, pp. 421–436, 2017.
[64] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models Part I—A discussion of principles," *J. Hydrol.*, vol. 10, no. 3, pp. 282–290, Apr. 1970.
[65] P. Potapov, X. Li, A. Hernandez-Serna, A. Tyukavina, M. C. Hansen, A. Kommareddy, A. Pickens, S. Turubanova, H. Tang, C. E. Silva, J. Armston, R. Dubayah, J. B. Blair, and M. Hofton, "Mapping global forest canopy height through integration of GEDI and Landsat data," *Remote Sens. Environ.*, vol. 253, Feb. 2021, Art. no. 112165.
[66] J. R. Ellenwood, F. J. Krist Jr., and S. A. Romero, "National individual tree species atlas," United States Dept. Agriculture, Washington, DC, USA, Tech. Rep. FHTET-15-01, 2015. [Online]. Available: https://www.fs.usda.gov/foresthealth/technology/pdfs/FHTET_15_01_National_Individual_Tree_Species_Atlas.pdf

PETER L. WATSON (Member, IEEE) received the bachelor's degree in environmental studies from The University of Chicago, IL, USA, and the Ph.D. degree in environmental engineering from the University of Connecticut, in 2023. He is currently an Infrastructure Scientist with Los Alamos National Laboratory, Los Alamos, NM, USA, where he investigates the intersection of the environmental hazards and infrastructural systems.



WILLIAM HUGHES is currently pursuing the Ph.D. degree with the Civil and Environmental Engineering Department, University of Connecticut. His research interests include assessing the resilience of infrastructure systems and communities against natural hazards.

DIEGO CERRAI received the B.S. degree in physics from the University of Pisa, Italy, in 2012, the M.S. degree in physics of the earth system from the University of Bologna, Italy, in 2015, and the Ph.D. degree in environmental engineering from the University of Connecticut, in 2019. He is currently an Assistant Professor with the Department of Civil and Environmental Engineering, University of Connecticut, where he is also the Associate Director of the Eversource Energy Center. His research interests include understanding the relationships between weather patterns, environmental conditions, and damage to the natural and built environment, with a particular focus on infrastructural damage. Moreover, he is investigating innovative methodologies for improving wintry precipitation measurements.

WEI ZHANG received the Ph.D. degree from Louisiana State University, in 2012. He is currently an Associate Professor with the Department of Civil and Environmental Engineering, University of Connecticut. His research interests include structural reliability, wind engineering, resilience of civil infrastructure and systems, and community resilience.

AMVROSSIOS (ROSS) BAGTZOGLOU received the Diploma degree in civil engineering from the Aristotle University of Thessaloniki, Greece, in 1985, the M.S. degree in hydrology and water resources engineering from Florida Institute of Technology, in 1987, and the Ph.D. degree in water resources and environmental engineering from the University of California at Irvine, in 1990. He is currently a Professor in civil and environmental engineering with the University of Connecticut, where he teaches water resources and environmental engineering courses and specializes in numerical modeling of environmental and hydrologic processes. He is an elected fellow of American Society of Civil Engineers, in 2012, the Institution of Civil Engineers, in 2012, American Water Resources Association, in 2014, and National Groundwater Association, in 2016.

EMMANOUIL ANAGNOSTOU received the Diploma degree in civil and environmental engineering from the National Technical University of Athens, in 1990, and the M.Sc. and Ph.D. degrees in hydrometeorology from the University of Iowa, in 1995 and 1997, respectively. He is currently a Board of Trustees Distinguished Professor in civil and environmental engineering and the Eversource Energy Endowed Chair of Environmental Engineering. He is also the Interim Executive Director of UConn Tech Park and the Director of the Eversource Energy Center. His research interests include remote sensing applications in water resources, with a focus on hydro-meteorological extremes and the synthesis of sociological and engineering methods to create people-centered solutions to the nexus problem of water, food, and energy insecurities and risks.