IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 28, 2021, accepted January 13, 2022, date of publication February 7, 2022, date of current version February 15, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3149485

On the Use of Satellite Nightlights for Power Outages Prediction

JUAN P. MONTOYA-RINCON[1], SHAMS AZAD[2], RABINDRA POKHREL[3], MASOUD GHANDEHARI[2], MICHAEL P. JENSEN[4], AND JORGE E. GONZALEZ[1,4]

[1]Department of Mechanical Engineering, The City College of New York, New York, NY 10031, USA
[2]Department of Civil and Urban Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11201, USA
[3]Department of Mechanical Engineering, Kathmandu University, Dhulikhel 45200, Nepal
[4]Brookhaven National Laboratory, Upton, NY 11973, USA

Corresponding author: Juan P. Montoya-Rincon (jmontoyarincon@ccny.cuny.edu)

This work was supported in part by the U.S. National Science Foundation under Grant CBET-1832678, in part by 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" under Award 1832678, and in part by the Brookhaven National Laboratory Directed Research and Development under Grant 18-020.

ABSTRACT Hurricanes are a dominant disaster in the Caribbean, always causing serious power outages throughout the islands. Hurricane Maria was a prime example, causing unimaginable destruction of the power infrastructure of Puerto Rico (PR). Consequently, one month after the hurricane landfall, approximately 80% of the population was still without power. After an event of such massive destruction, the electric power restoration process progresses very slowly. This timeline can be improved using power outage (PO) forecast models that help identify the vulnerable places before the hurricane landfall. Generally, these models are trained with historical power outages records, associated data on weather conditions, and additional information about the natural and built environments. However, PO records are often difficult to acquire, and, in many instances, the power utility companies may not record them. This study utilizes a satellite-based Visible Infrared Imaging Radiometer Suite (VIIRS) night light data product as a surrogate for the power delivery to predict hurricane-induced PO in areas having limited to nonexistent historical data records. The processed satellite data is then used along with geographic variables, and simulated weather data to formulate machine learning-based algorithms to predict PO for future hurricane events. These models are applied and validated in the context of the PR catastrophic storm, Hurricane Maria.

INDEX TERMS Power outage prediction, machine learning, weather simulation, extreme events, night-time light.

I. INTRODUCTION

Hurricanes cause significant damage to the electrical power infrastructure. A recent example is the passage of Hurricane Maria (H-Maria) in the Caribbean during September 2017, which reached a peak intensity of maximum sustained wind speeds of 175 mph (280 km/h), making it the tenth-most intense Atlantic hurricane on record. H-Maria made landfall in Puerto Rico on September 20th, causing unimaginable destruction. Almost all of the 2,400 miles of transmission lines, 30,000 miles of distribution lines, and 342 substations were damaged by the storm [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale De Meo.

The recovery process of the Puerto Rico power grid was slow due to its near-complete destruction. After one month, less than 20% of the total power capacity had been restored [2]. The preparedness for such events can be improved by anticipating the likely location and timing of storm-induced damage to the power grid. Primarily, this will help the utility companies and emergency managers to direct restoration plans, allowing for a more efficient repair and recovery process after the extreme weather event.

Several studies have investigated the impact of extreme weather events on the power infrastructure. For instance, Reed *et al.* [3] used data on actual power outages experienced during Hurricane Rita along the Gulf of Mexico coast to study how they related to the observed wind gusts. They found a direct relationship between power outages and wind speed;

however, other meteorological and environmental variables were not considered in their study. Bjarnadottir *et al.* [4] performed a risk assessment of the utility poles for the area of Miami and Florida, similarly considering the wind speed as the only explanatory weather variable. Yue *et al.* [5] developed a failure rate model with a Bayesian update for specific, different power grid components (such as transmission lines, transformers, substations), using high-resolution weather data from two National Weather Service Next-Generations Radars (NEXRADs). However, forecasted weather data was not considered in their study. Reed *et al.* [6] studied the energy system performance during Hurricane Katrina and concluded that damage to the utility infrastructure could also occur at relatively lower wind speeds, mainly by heavy inland rainfall. Zhu *et al.* [7] developed a power outage forecast model, making short-term predictions of one and five hours ahead of a storm. McRoberts *et al.* [8] developed a two-step procedure for power outage prediction using classification trees to help reduce the inflation of non-outage cases in the training data, also called the zero-inflation problem, leading to more accurate predictions. Yang *et al.* [9] studied power outage (PO) models for different storm magnitudes finding that the prediction accuracy increased dramatically when a greater number of storms were included in the training dataset, and when selectively training the model with a subset of events representative of the event type of interest.

As a result of anthropogenic climate change, future tropical storms are expected to have stronger winds and heavier rainfall [10]. These storms present a substantial risk to the power infrastructure. Wanik *et al.* [11] studied the impact of a future-climate Superstorm Sandy (the year 2100) on the Connecticut power grid, concluding that a storm of this magnitude will increase the outages by 42%-64%, compared to the already significant impact of superstorm sandy in 2012.

The aim of this study is the development of a PO forecast model that does not rely on the PO records provided by the utility. The experience of H-Maria along with Hurricane Irma (H-Irma) (2017) is used as learning and training events. We used a machine learning based approach to model the complex non-linearity present in the extreme weather-induced PO.

One challenge that is often faced in the development of PO prediction models is the availability of reported PO records for the desired utility area. This data is often incomplete, difficult to acquire, proprietary, or may even be non-existent. Multiple techniques have been developed using NASA's black marble nighttime lights (NTL) product suite (VNP46) [12], developed with the Visible Infrared Imaging Radiometer Suite (VIIRS) day/night band (DNB) onboard the SUOMI national polar-orbiting platform (SNPP). Román *et al.* [13] showed that this dataset can be used as a measure of the PO restoration process, and validated the results with actual restoration data from the utility. Azad and Ghandehari [14] used the NTL data product to estimate outage duration in Puerto Rico following H-Maria with high spatial granularity. Cole *et al.* [15] used the NTL along with population estimates, count of grid assets, and PO reports to train a neural network that was able to predict the count of power outages in extreme weather events. However, this technique fully relies on the availability of power outages reports for the training of the model. Wang *et al.* [16] found that the change in NTL after a natural disaster can be used to monitor PO approximate location and duration.

Developing new approaches that do not require actual PO records is relevant to the current state of the field. In this study, we aim to address this challenge, by developing a prediction model that relies on satellite-based NTL observations as a proxy for PO. Thereby, the model does not require any data from the utility, making it ideal for isolated regions or regions with limited PO records.

Researchers from different domains have used NTL for quantitative analysis of different complicated phenomena. These include economic activity [17], [18], population change [19], [20], power systems and disruptions [21], [22], and urban expansion [23], [24]. The defense meteorological satellite program's operational line scanner (DMSP/OLS) was the first satellite system providing night-time light data since 1992. Six DMSP satellites (f10, f12, f14, f15, f16, and f18) equipped with OLS were of great value, despite the lack of full calibration, posing challenges to perform quantitative and time series analysis of radiance. Researchers often utilized the product by performing different inter-calibration methods, such as assuming constant light intensity throughout the time of interest [25]. DMSP was followed by the SNPP satellite with the VIIRS sensor in 2011. Many of the shortcomings of the DMSP night-time light products were addressed using an onboard solar diffuser calibration system [26], [27]. One of the VIIRS sensor's bands is the DNB which detects light in a range of wavelengths from green to near-infrared. Using these bands, VIIRS is able to deliver high-quality information on night-time lights.

This paper is structured as follows: section II provides a description of the data along with the different methods used to process the data; section III contains details on the power outage prediction model and its initialization hyperparameters; section IV contains the results and a discussion on the sensitivity of the model to different spatial resolutions; limitations are given in section V, and our conclusions are presented in section VI.

II. DATA AND METHODOLOGY

Two storms were considered for the development of the PO prediction model: Hurricanes Irma and Maria. Multiple weather explanatory variables (Independent Variables) were used in the model to describe the destructive capabilities of a hurricane. Moreover, additional non-weather-related variables were also considered. These variables describe potential contributing risks, such as trees near the overhead lines, or provide information on the energy infrastructure. All the selected explanatory variables are explained in detail in section A. In this study satellite-based NTL observations are incorporated as a proxy for PO, which gives the possibility

of developing prediction models for utility areas where PO records are limited or non-existent. The use of NTL as a response variable (Dependent Variable) will be discussed in section B.

A. EXPLANATORY VARIABLES

This study employs a single-layer urban canopy version of the Weather Research and Forecasting (WRF v 3.8.1) model [28], a numerical weather prediction system developed by the National Center for Atmospheric Research (NCAR) to simulate the meteorological variables used in this study. For domain configuration, three two-way nested domains were employed. The Mesoamerican and Caribbean regions are covered under the parent domain at a spatial resolution of 25 km (144 points by 100 points). The Caribbean Sea, Dominican Republic, and the island of Puerto Rico are included in the second domain, which has a spatial resolution of 5km (306 points by 191 points), while the entire island of Puerto Rico is included in the third domain, which has a spatial resolution of 1km (336 points by 156 points). This nested domain configuration is illustrated in the top panel of Figure 1, with the surface elevation contours within the smallest domain shown in the bottom panel. The center of the island contains the Cordillera Central mountain range with elevations as high as 1300 meters. For the 1 km domain, the cumulus parameterization was disabled because WRF can explicitly resolve convective processes at this resolution. The model has 50 vertical levels, 35 of which are below 2 km in height. Two simulations were conducted, from September 4th to 9th, and from September 19th to 22nd, 2017 that covered both H-Irma and H-Maria, respectively.

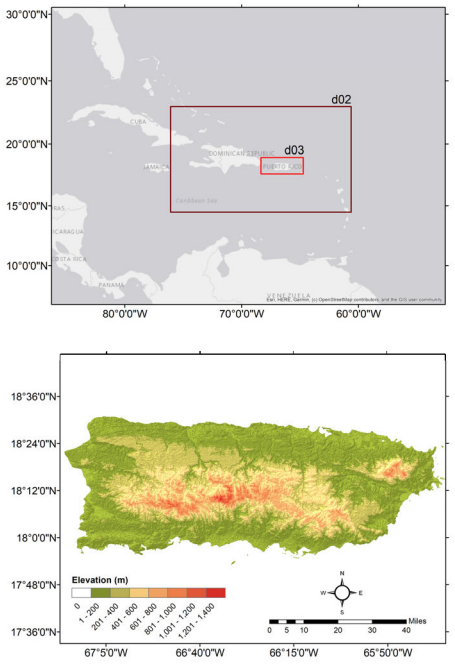

FIGURE 1. WRF domain (top) and topography (bottom).

As part of this analysis, an ensemble of model simulations of H-Maria was considered that included variation in the resolution of the boundary and initial conditions, the PBL schemes and the cumulus parameterizations. We use the explanatory variables output by the ensemble member that best reproduced the observed storm track. H-Irma results were validated with ground station data from Luis Muñoz Marin International Airport (TJSJ) and Roosevelt Roads Naval Station (TJNR). More details of the WRF configuration for both storms, and H-Maria results are given in Pokhrel *et al.* [29].

For H-Irma, data from September 6th and 7th was used, with a resolution of 1km $\times$ 1km. The simulation provides the wind in its U and V components. Moreover, we determined the maximum wind speed magnitude in each grid cell over time. The center and northeast part of the island experienced the greatest maximum wind speeds during H-Irma, where the highest power loss occurred, Figure 2. Furthermore, the cumulative precipitation for each event is calculated as the sum of the hourly precipitation at each location over the lifecycle of the storm. Figure 3 shows that for H-Irma, highest rainfall totals occurred in the same regions as the greatest maximum wind speeds.

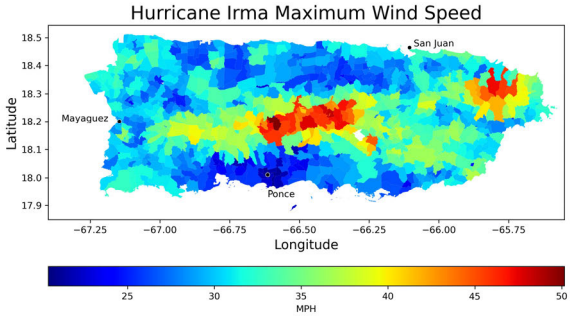

FIGURE 2. H-Irma WRF-simulated maximum wind speed, towns subdivisions spatial resolution.

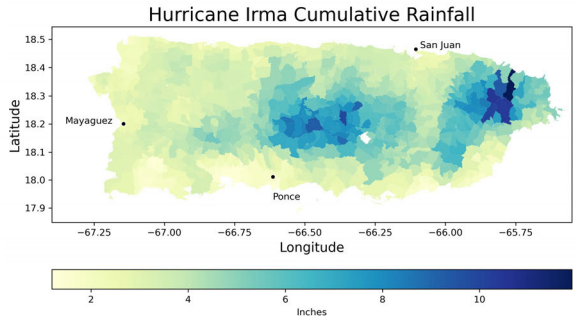

FIGURE 3. H-Irma WRF-simulated cumulative rainfall, towns subdivisions spatial resolution.

For H-Maria, we used similar processing to that applied to the H-Irma wind data, finding the maximum value in each

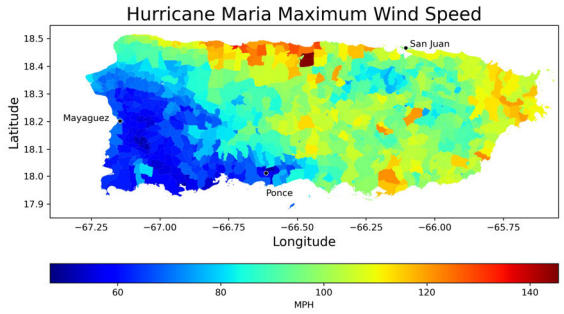

FIGURE 4. H-Maria WRF-simulated maximum wind speed, towns subdivisions spatial resolution.

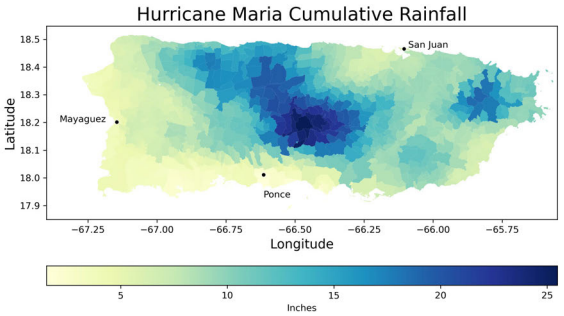

FIGURE 5. H-Maria WRF-simulated cumulative rainfall, towns subdivisions spatial resolution.

grid cell throughout the whole event. As shown in Figure 4, the wind speed in H-Maria was significantly higher than H-Irma, with speeds as high as 145 MPH. Furthermore, we determined the duration of high winds in the service area from the WRF simulated wind speed. Specifically, the duration of wind speed greater than 20, 30, and 40 MPH, resulting in a total of 4 wind-related variables in the training dataset. For H-Maria, model outputs from September 20th and 21st were used. The greatest precipitation in H-Maria was located around the center of the island, with a maximum value of 25 inches, Figure 5.

Besides the weather data, land surface elevation, population, and land cover were added as static geographic variables in the model. The land surface elevation was obtained from the United States Geological Survey [30]. The dataset has a horizontal resolution of 100m $\times$ 100m. The population data was obtained from the US Census, providing an estimation of the population by town. The land cover dataset was downloaded from the National Land Cover database [31], with a resolution of 30m $\times$ 30m, including 12 different land classes. Most of the island is covered by evergreen forest, which presents a significant risk to the overhead transmission and distribution lines, Figure 6.

After processing each variable individually, all the explanatory variables (e.g. weather, elevation) were interpolated to a common spatial resolution of 500m $\times$ 500m to better match the satellite NTL resolution. Additionally, two

TABLE 1. Explanatory variables.

Explanatory Variable	Source	Resolution	Units	Aggregation Method
Maximum Wind Speed. (WS)	WRF	1km x 1km	MPH	Maximum
Duration of Wind Speed greater than 20 MPH. (WS_20)	WRF	1km x 1km	hours	Maximum
Duration of Wind Speed greater than 30 MPH. (WS_30)	WRF	1km x 1km	hours	Maximum
Duration of Wind Speed greater than 40 MPH. (WS_40)	WRF	1km x 1km	hours	Maximum
Cumulative Rainfall. (CR)	WRF	1km x 1km	inches	Maximum
Population by Municipios. (POP)	US Census	Municipios	count	Maximum
Elevation. (EL)	USGS	100m x 100m	feet	Mean
Land Cover. (LC)	USGS NLCD	30m x 30m	categ orical	Median
Pre-Hurricane NTL intensity map. (NTL_Base)	NASA VIIRS	500m x 500m	radian ce	Mean

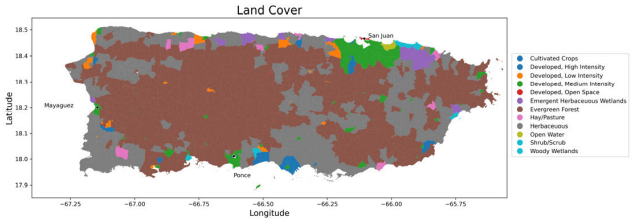

FIGURE 6. Puerto Rico type of land cover, towns subdivisions spatial resolution.

different datasets were created, one where all the variables were aggregated using the census tract into towns and the other where the variables were aggregated into towns subdivisions, using the most appropriate statistical method for each variable. Here, a town is the political boundary, and a town subdivision is a sub-region within the town also referred to as *barrio*. The selected aggregation method for each variable is listed in Table 1. Consequently, three training datasets were created by changing the spatial resolution of the variables (500m $\times$ 500m, Towns, and Towns Subdivisions).

B. RESPONSE VARIABLE

Previous studies have already demonstrated the high efficacy of the NTL product to represent hurricane-induced power loss and recovery [13], [14]. The VIIRS satellite sensor is capable of capturing the upwelling visible and infrared radiance from the Earth at 500m $\times$ 500m resolution. In this study, we used the top-of-atmosphere, at-sensor nighttime radiance product (VNP46A1). First, we examined the cloud-mask layer of the VNP46A1 product to determine the cloud coverage. To quantify the pre-H-Irma and Maria baseline NTL distribution, we removed the pixels with clouds and aggregated

the NTL data between August 20th and August 24th, 2017 to a complete, clear-sky mapping of the NTL over Puerto Rico. Since significant cloud cover is associated with hurricanes, it is not always possible to capture the immediate nightlight radiance following landfall. We aggregated images between September 8th and September 11th to quantify the H-Irma induced power loss. We found the power was completely recovered by September 17th. The cloud cover remained longer for H-Maria with no cloud-free imagery in the first four days following landfall. To create the post-Maria NTL data, we aggregated the cloud-free part of the island captured in images between September 25th and September 30th, to construct a cloud-free image for the entire island. Due to this need for cloud-free observations of the NTL, the estimates of power loss will be impacted by power restoration during the time between outage occurrence and cloud-free observations. This will result in some underestimation of the total power outages from the derived algorithm. Figure 7, shows a box plot of log-transformed pixel-level NTL radiance for the entire island. The median log transformed NTL intensity before H-Irma, between 20 and 24 August, was 0.6 which dropped to 0.09 after H-Irma landfall. Again, between 17 and 19 September, the median radiance became 0.6 which is equal to the intensity prior to H-Irma. This indicates the power infrastructure of the Island completely recovered from the loss caused by H-Irma before the landfall of H-Maria. Therefore, using radiance values between August 20th and 24th as a baseline for both events would give an unbiased estimation of power loss.

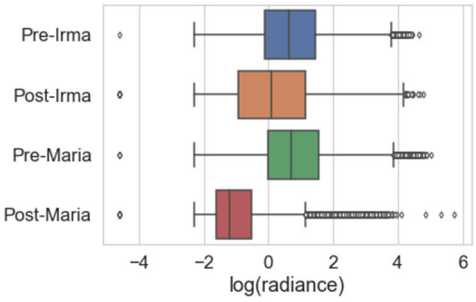

FIGURE 7. Box plot showing log-transformed pixel-level NTL radiance for the Island. Radiance distribution before H-Irma is demonstrated by 20 Aug – 24 Aug, (Pre-Irma) where 8 Sep – 11 Sep (Post-Irma) shows the radiance immediately after the landfall H-Irma. Between 17 and 19 Sep (Pre-Maria), the power was fully recovered from the loss caused by H-Irma. 25 Sep – 30 Sep (Post-Maria) shows the distribution after the landfall of H-Maria.

The loss in power infrastructure can be formulized as,

$$Power\ Loss = \frac{NL_{Base} - NL_{After}}{NL_{Base}} * 100 \tag{1}$$

where NL_{Base} is the night light radiance before and NL_{After} is the radiance after the hurricane. The NL_{Base} was used by itself as an independent variable. Note that in this context Power Loss (PL) represents the change of night light radiance and not the actual electricity power loss. Moreover, the PL metric could be interpreted as the probability of

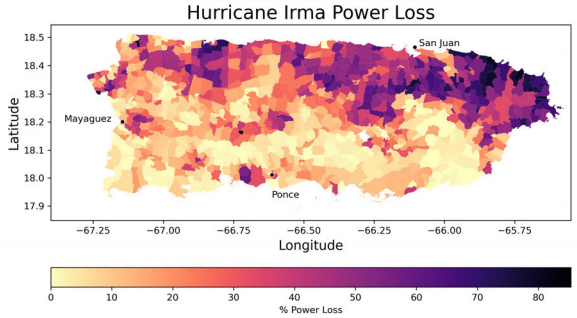

FIGURE 8. H-Irma power loss, towns subdivisions spatial resolution.

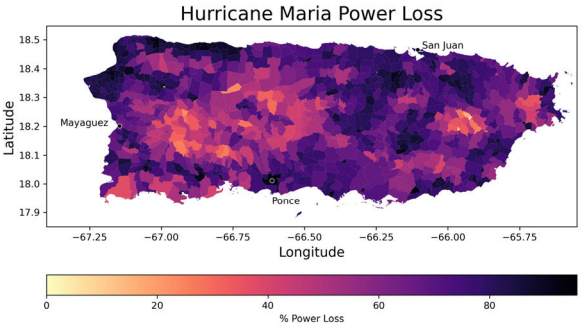

FIGURE 9. H-Maria power loss, towns subdivisions spatial resolution.

power outage (PO) within a given spatial boundary (i.e., 500m, Towns, and Towns Subdivisions). As such, we are relating PL to PO in this contribution. As shown in Figure 8, H-Irma had a notable impact on the power infrastructure, leaving a major power loss on the northeastern side of the island. In contrast, H-Maria severely damaged the power infrastructure, leaving major power loss throughout the island, Figure 9.

III. POWER OUTAGE MODEL

The model development is based upon machine learning tools widely used for power outage forecasting, specifically Bayesian Additive Regression Trees (BART), Random Forest (RF), and Extreme Gradient Boosting (XGBoost).

BART is a data mining, fully Bayesian probability model, with a prior and likelihood. The model is constructed with an ensemble of decision trees. The predictions are made by adding the resulting outputs from each tree together, helping to avoid overfitting in the model. The model can be described with the following equation [32],

$$Y = \sum_{j=1}^{m} g\left(x, T_j, M_j\right) + \epsilon, \quad \epsilon \sim (0, \sigma^2) \tag{2}$$

Here, T_j is a binary regression tree where $M_j = \{\mu_{1j}, \mu_{2j} \ldots \mu_{bj}\}$ is its terminal node parameters. $g\left(x, T_j, M_j\right)$ function assigns $\mu_{ij} \in M_j$ to x. The expected value equals the sum of all the terminal node assigned to x. ϵ is the variance component, assumed to follow normal distribution with zero mean.

BART has been successfully used in different approaches to risk analysis and damage prediction in extreme weather events. Nateghi *et al.* [33] compared the BART model with survival models by predicting power outage duration in Hurricane Ivan (2004). BART was found to give better results than the traditional survival models. Guikema *et al.* [34] compared multiple models including generalized additive models, BART, generalized linear models, and classification and regression trees (CART), for the estimation of damage in the distribution poles during hurricane events. The authors concluded that nonparametric models perform better than parametric models for outage prediction in hurricanes. He *et al.* [35] compared two nonparametric tree-based models, BART, and quantile regression forest, concluding that BART was better for predicting the magnitude and spatial variation of outages. Moreover, BART was also found to perform better when the data was aggregated into larger service areas (e.g., Towns Subdivisions).

The RF [36] regression model is a non-parametric, supervised learning algorithm that averages over the outputs of an ensemble of decision trees to make the predictions. RF follows the bagging technique for training data creation by randomly resampling the original dataset with replacement [37]. From the total set, a small set of input variables is randomly selected for binary partitioning the nodes of a tree. The splitting of the non-terminal node of a regression tree is based on choosing the input variable with the lowest Gini Index.

$$I_G\left(t_{X(x_i)}\right) = 1 - \sum_j^m f\left(t_{X(x_i)}, j\right)^2 \tag{3}$$

Here, $f(t_{X(x_i)}, j)$ is the proportion of samples with value x_i belonging to leave j as node t [38]. The final prediction of the model is done by averaging all trees.

Mukherjee *et al.* [39] developed a two-stage hybrid model to predict PO in the United States, by training a support vector machine model and a RF with publicly available data. Wanik *et al.* [40] created an ensemble model for storm-induced PO by combining RF, boosted gradient tree, and a decision tree. Li *et al.* [41] found that an RF-based model outperformed linear regression, support vector regression, and decision tree regression models for PO quantity predictions in typhoon events. Nateghi *et al.* [42] found that RF outperformed BART in the estimation of PO duration, by increasing the accuracy in predictions by 87%.

XGBoost is a scalable end-to-end tree boosting system that follows the principle of greedy function approximation of a gradient boosting algorithm [43]. The model utilizes additional regularized-model reinforcement to regulate overfitting to enhance the model performance [44]. XGBoost uses a tree ensemble technique which refers to the utilization of a set of CART, and the final prediction is the sum of each CART's score [45]. For prediction, the model minimizes the following regularized objective function [46],

$$L(\phi) = \sum_i l(\hat{y}, y_i) + \sum_k \Omega(f_k) \tag{4}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda |\,|\omega|\,|^2 \tag{5}$$

Here, l is a convex loss function that measures the difference between predicted ($\hat{y}$) and true value (y_i). Moreover, Ω is the regularization parameter that penalizes the complexity of the model to avoid overfitting, where T represents the number of leaves and $|\,|\omega|\,|^2$ is the L2 norm of all leaf scores. The parameters γ and λ controls the degree of conservatism when searching the tree [47]. The efficacy of this algorithm has been recognized widely, which made it a popular choice to model the failures in the power network [48], [49].

To implement the BART model we selected the R library "BartMachine" [50]. We choose this library over the BayesTree R package [32], mainly for its capability to run in parallel, giving higher efficiency in the training process. For the BART model, a 5 – fold cross-validation was used and a total of 50 trees were selected, the other hyperparameters were set to default, as recommended by Chipman *et al.* [32] in his model development. In the training process, 250 burn-in iterations were performed and discarded. Finally, another 1000 iterations were made to build the regression trees. Furthermore, using a random hyperparameter grid search with 150 replicates of the model and a 5 – fold cross-validation we found the optimal hyperparameters for the RF to be 100 trees, a maximum depth of 126 for each tree, a maximum of 4 features considered for splitting a node, a minimum of 5 data points placed in a node before the node is split and default for the remaining. Similarly, a 5 – fold cross validation random hyperparameter grid search with 150 replicates of the model was used for XGBoost. The selected hyperparameters were gbtree as the booster, a total of 100 decision trees, a maximum depth of the tree of 10, a learning rate of 0.3, and a minimum weight of 1 to create a new node in the tree.

PL was analyzed in each storm independently. The behavior of the PL was very different in each storm, Figure 10. H-Maria had very high winds and precipitation. As a result, it caused more severe damage throughout the island, leaving most of the island with 70% to 100% power loss. On the other hand, H-Irma was less destructive, leaving most of the island with minimum PL, Figure. 9. Consequently, both storms are used as training events, allowing the model to be sensitive to both types of events. To build the training dataset, 70% of the data were randomly selected from both H-Irma and H-Maria. The remaining 30% of both storms data was left out of the training process and used to test the model. Explanatory variables in Table 1 were used in conjunction with the PL as inputs in the training process of the model.

IV. RESULTS

To test the sensitivity of the models at different spatial granularities, we formulated them at three different spatial levels: (1) 500m × 500m, (2) towns subdivisions level, and (3) towns level. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-Squared (R2) were used to compare the

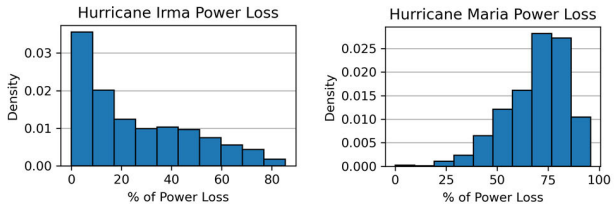

FIGURE 10. Power loss histograms, towns subdivisions spatial resolution.

prediction capabilities of the model at different resolutions. Moreover, a mean-only model was used as a benchmark for BART, RF, and XGBoost.

Table 2 reveals that the RF and XGBoost models have higher explained variance (R2) for the 500m × 500m resolution and the towns subdivisions aggregation. Mainly because the training dataset size was significantly reduced due to the larger areas of aggregation (towns). Pixel resolution, on the other hand, offers the model with a vast dataset to train on. Furthermore, the RMSE shows that the 500m resolution has errors of greater magnitude in all models. Owing to the pixel level daily NTL dataset being noisier and skewed. Most importantly, combining the pixels into a larger spatial resolution minimizes noise and aids in the removal of the skewed response variable distribution.

The towns subdivision aggregation had the smallest prediction error in most of the models as the training dataset remained large enough for a reliable training process. Furthermore, the models had a small variance in the predictions with minimal large residuals in all the resolutions, as indicated by the closeness of the RMSE to the MAE value.

All three models (BART, RF, and XGBoost) performed similarly well on the test dataset (Figure 11). However, The RF model at towns subdivisions resolution was chosen as the best configuration because it had fewer large residuals in the predictions and the explained variance outperformed the other models by a small margin.

After selecting the optimal configuration of the model, we proceeded to study the importance of each variable in the

TABLE 2. Comparison of model resolutions performance, test dataset.

Resolution	Metrics	Mean Only	BART	RF	XGBoost
500m x 500m	RMSE	31.81	18.46	13.16	15.16
	MAE	27.8	14.48	9.10	11.16
	R2	NA	0.67	0.82	0.77
Towns	RMSE	23.86	13.05	13.59	13.65
	MAE	19.65	9.71	10.45	10.27
	R2	NA	0.70	0.66	0.66
Towns Subdivisions	RMSE	29.32	13.76	12.51	12.84
	MAE	25.80	10.49	9.42	9.66
	R2	NA	0.79	0.82	0.81

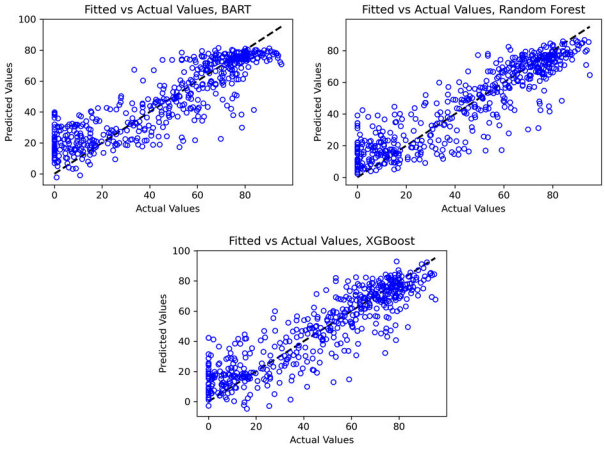

FIGURE 11. Fitted versus actual values for test dataset, towns subdivision spatial resolution.

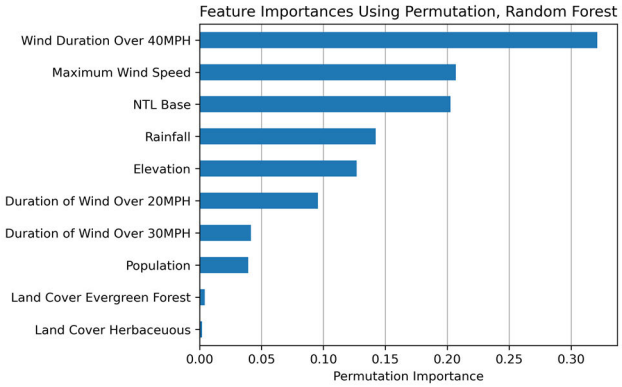

FIGURE 12. Ranking of variable importance (definitions are given in Table 1).

model as a predictor. In order to get a stable study in the test dataset, we used permutation features importance with 100 replicates of RF to generate variable inclusion proportions, Figure 12.

As expected, the first three variables with the most influence in the prediction are weather-related variables that quantify the magnitude of the hurricane. Moreover, the duration of winds over 40 MPH had a higher inclusion proportion than the wind speed magnitude, implying that longer times of high wind exposure can be more critical than maximum wind gusts for PL estimation. Among land cover types, the evergreen forest is detected as an important predictor for power outages. That is plausible as this land type has a high risk for overhead transmission and distribution lines due to falling trees.

To further investigate the influence of the explanatory variables with highest inclusion proportion, we created partial dependence plots (PDP; Figure 13). The PDP were created using 50 bootstrap resamples and a confidence interval of 95%. The PDP shows that a higher duration of wind over 40 MPH strongly influences the PL. Similarly, the maximum

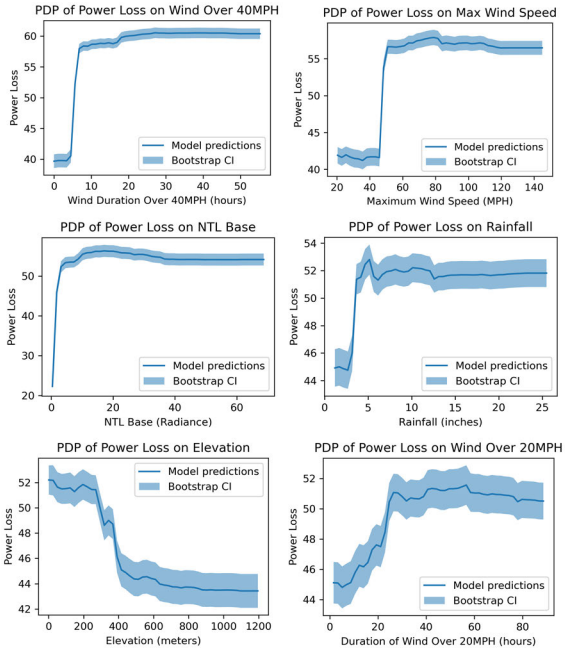

FIGURE 13. Partial dependence plot of power loss in variables with more influence. 95% confidence interval with 50 bootstraps resamples.

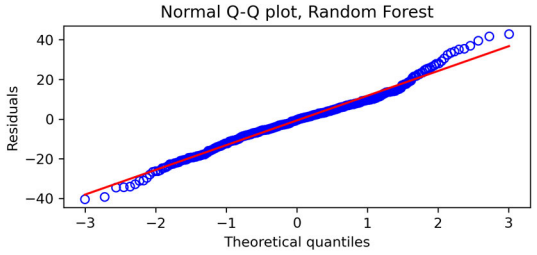

FIGURE 14. QQ-plot for the residuals of the RF.

wind speed and rainfall influence the PL as they increase. However, the influence plateaus when the duration of wind over 40 MPH, maximum wind speed, and rainfall reaches 25 hours, 80 MPH, and 13 inches, respectively. Additionally, we can see an increase in the influence on PL when the NTL_Base increases from 0 to 5. This shows how the NTL_Base helped the model achieve a better distribution of the PL over the island, by giving information on service areas with low NTL radiance, such as rural areas with a small customer count. Finally, looking at the quantile-quantile plot (QQ-plot; Figure 14) we see that most of the residuals fall along the 45-degree red line, which indicates that the residuals follow a normal distribution. This shows that the RF model can capture the variability in our dataset.

V. LIMITATIONS

Prediction of the hurricane-induced PL with NTL data is particularly valuable where the traditional outage information cannot be retained. However, clouds associated with hurricanes can remain for several hours to days, obscuring the NTL and making it impossible for the satellite to capture the actual damage right after the hurricane landfall. The model developed in this study is not applicable if the hurricane-induced outages are restored before nighttime clear sky conditions occur. Again, the satellite captures the image once every night. It would not be possible to detect PL if the damage occurred during the daytime and the power gets restored before the satellite passes over the affected area. In this regard, the model is only appropriate to use if the hurricane-induced POs sustain until the satellite captures a clear NTL image of the affected area. Most importantly, the model is able to capture the PL in our area of study, Puerto Rico, where hurricane-induced PO usually lasts more than one day after the storm landfall.

VI. CONCLUSION

This study presented a new technique for PO prediction in service areas with limited or non-existent direct PO historical records, using the NTL PL as a proxy of PO. The model was built based on simulated data from H-Irma and Maria in the Caribbean during September 2017. Then, the out-of-sample prediction capability was validated with data from both tropical cyclones. Moreover, the three tested models prove to perform well on the test dataset. However, RF was found to outperform BART and XGBoost in PL predicting capabilities. Duration of wind speed greater than 40 MPH and maximum wind speed were found to highly influence the PL in both storms. Similarly, the rainfall had a major influence on the PL in the range of 0 to 13 inches. These variables are directly correlated with the magnitude of the hurricane and the damage it can cause to the infrastructure. Among the non-weather-related variables, elevation was detected as a significant factor. This is mainly because most of the population in Puerto Rico is located near the coast, in low elevations, giving this a higher influence in PL in those areas.

The model was tested using three different spatial granularities (500m, towns subdivisions, and towns) for the same set of explanatory variables. The 500m and towns subdivisions aggregation outperformed the towns aggregation in the prediction capabilities for RF and XGBoost. The towns aggregation was found to be a very coarse domain, decreasing the training dataset size dramatically. Consequently, the towns resolutions models were not able to correctly explain the variability in the data. Moreover, the RMSE indicates that all the 500m resolution models have errors of larger magnitude. Finally, the RF with towns subdivisions resolution was chosen as the optimal model, giving a good balance between the explained variability (R2 0.82) and the reduction of outliers in the predictions (RMSE of 12.51%).

A future expansion of this work may include outages duration estimation using NTL. Which allow us to better understand the behavior of the power grid in extreme weather conditions. Moreover, a study of the power loss in future H-Maria like scenarios will provide us with tools to properly design future grids in complex terrains, focusing on improving resilience.

APPENDIX
ERROR METRICS

The error metrics used for this study are listed below:

- The mean absolute error (MAE)

$$MAE = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n} \tag{6}$$

- The root mean square error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n}} \tag{7}$$

The MAE is used to determine how accurate the predictions are. Furthermore, the RMSE is used to penalize undesired large errors in the prediction. The variance in the sample's prediction errors is then estimated using the difference between MAE and RMSE. Where a small difference between the metrics means there is small variation in the magnitude of errors.

REFERENCES

[1] F. Robles and J. Bidgood, *Three Months After Maria, Roughly Half of Puerto Ricans Still Without Power*. New-York, NY, USA: The New York Times, 2017.

[2] D. Lu and C. Alcantara, *After Hurricane Maria, Puerto Rico was in the Dark for 1037 Days, 14 Hours and 37 Minutes*. Washington, DC, USA: The Washington Post, 2018.

[3] D. A. Reed, M. D. Powell, and J. M. Westerman, "Energy infrastructure damage analysis for hurricane rita," *Nat. Hazards Rev.*, vol. 11, no. 13, pp. 102–109, 2010, doi: 10.1061/(ASCE)NH.1527-6996.0000012.

[4] S. Bjarnadottir, Y. Li, and M. G. Stewart, "Hurricane risk assessment of power distribution poles considering impacts of a changing climate," *J. Infrastruct. Syst.*, vol. 19, no. 1, pp. 12–24, 2013, doi: 10.1061/(ASCE)IS.1943-555X.0000108.

[5] M. Yue, T. Toto, M. P. Jensen, S. E. Giangrande, and R. Lofaro, "A Bayesian approach-based outage prediction in electric utility systems using radar measurement data," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6149–6159, Nov. 2018, doi: 10.1109/TSG.2017.2704288.

[6] D. A. Reed, M. D. Powell, and J. M. Westerman, "Energy supply system performance for Hurricane Katrina," *J. Energy Eng.*, vol. 136, no. 4, pp. 95–102, 2010, doi: 10.1061/(ASCE)EY.1943-7897.0000028.

[7] D. Zhu, D. Cheng, R. P. Broadwater, and C. Scirbona, "Storm modeling for prediction of power distribution system outages," *Electr. Power Syst. Res.*, vol. 77, no. 8, pp. 973–979, Jun. 2007, doi: 10.1016/j.epsr.2006.08.020.

[8] 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, doi: 10.1111/risa.12728.

[9] F. Yang, D. W. Wanik, D. Cerrai, M. A. E. Bhuiyan, and E. N. Anagnostou, "Quantifying uncertainty in machine learning-based power outage prediction model training: A tool for sustainable storm restoration," *Sustainability*, vol. 12, no. 4, p. 1525, Feb. 2020, doi: 10.3390/su12041525.

[10] T. Knutson, S. J. Camargo, J. C. L. Chan, K. Emanuel, C.-H. Ho, J. Kossin, M. Mohapatra, M. Satoh, M. Sugi, K. Walsh, and L. Wu, "Tropical cyclones and climate change assessment: Part I: Detection and attribution," *Bull. Amer. Meteorol. Soc.*, vol. 100, no. 10, pp. 1987–2007, Oct. 2019, doi: 10.1175/BAMS-D-18-0189.1.

[11] D. W. Wanik, E. N. Anagnostou, M. Astitha, B. M. Hartman, G. M. Lackmann, J. Yang, D. Cerrai, J. He, and M. E. B. Frediani, "A case study on power outage impacts from future hurricane sandy scenarios," *J. Appl. Meteorol. Climatol.*, vol. 57, no. 1, pp. 51–79, Jan. 2018, doi: 10.1175/JAMC-D-16-0408.1.

[12] M. Román, Z. Wang, Q. Sun, V. Kalb, and S. Miller, "NASA's black marble nighttime lights product suite," *Remote Sens. Environ.*, vol. 210, pp. 113–143, Jun. 2018, doi: 10.1016/j.rse.2018.03.017.

[13] M. O. Román, E. C. Stokes, R. Shrestha, Z. Wang, L. Schultz, E. A. S. Carlo, Q. Sun, J. Bell, A. Molthan, V. Kalb, C. Ji, K. C. Seto, S. N. McClain, and M. Enenkel, "Satellite-based assessment of electricity restoration efforts in Puerto rico after hurricane Maria," *PLoS ONE*, vol. 14, no. 6, Jun. 2019, Art. no. e0218883, doi: 10.1371/journal.pone.0218883.

[14] S. Azad and M. Ghandehari, "A study on the association of socioeconomic and physical cofactors contributing to power restoration after hurricane Maria," *IEEE Access*, vol. 9, pp. 98654–98664, 2021, doi: 10.1109/ACCESS.2021.3093547.

[15] T. Cole, D. Wanik, A. Molthan, M. Román, and R. Griffin, "Synergistic use of nighttime satellite data, electric utility infrastructure, and ambient population to improve power outage detections in urban areas," *Remote Sens.*, vol. 9, no. 3, p. 286, Mar. 2017, doi: 10.3390/rs9030286.

[16] Z. Wang, M. O. Román, Q. Sun, A. L. Molthan, L. A. Schultz, and V. L. Kalb, "Monitoring disaster-related power outages using nasa black marble nighttime light product," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 1853–1856, Apr. 2018, doi: 10.5194/isprs-archives-XLII-3-1853-2018.

[17] C. N. H. Doll, J.-P. M'uller, and J. G. Morley, "Mapping regional economic activity from night-time light satellite imagery," *Ecol. Econ.*, vol. 57, no. 1, pp. 75–92, Apr. 2006, doi: 10.1016/j.ecolecon.2005.03.007.

[18] T. Ghosh, S. Anderson, R. Powell, P. Sutton, and C. Elvidge, "Estimation of Mexico's informal economy and remittances using nighttime imagery," *Remote Sens.*, vol. 1, no. 3, pp. 418–444, Aug. 2009, doi: 10.3390/rs1030418.

[19] P. Sutton, "Modeling population density with night-time satellite imagery and GIS," *Comput. Environ. Urban Syst.*, vol. 21, no. 34, pp. 227–244, 1997, doi: 10.1016/S0198-9715(97)01005-3.

[20] P. Sutton, D. Roberts, C. Elvidge, and K. Baugh, "Census from heaven: An estimate of the global human population using night-time satellite imagery," *Int. J. Remote Sens.*, vol. 22, no. 16, pp. 3061–3076, Jan. 2001, doi: 10.1080/01431160010007015.

[21] C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis, "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption," *Int. J. Remote Sens.*, vol. 18, no. 6, pp. 1373–1379, Apr. 1997, doi: 10.1080/014311697218485.

[22] T. R. K. Chand, K. V. S. Badarinath, C. D. Elvidge, and B. T. Tuttle, "Spatial characterization of electrical power consumption patterns over India using temporal DMSP-OLS night-time satellite data," *Int. J. Remote Sens.*, vol. 30, no. 3, pp. 647–661, Feb. 2009, doi: 10.1080/01431160802345685.

[23] M. Imhoff, W. Lawrence, C. Elvidge, and T. Paul, "Using nighttime DMSP/OLS images of city lights to estimate the impact of urban land use on soil resources in the United States," *Remote Sens. Environ.*, vol. 59, no. 1, pp. 105–117, 1997, doi: 10.1016/S0034-4257(96)00110-1.

[24] M. Henderson, E. T. Yeh, P. Gong, C. Elvidge, and K. Baugh, "Validation of urban boundaries derived from global night-time satellite imagery," *Int. J. Remote Sens.*, vol. 24, no. 3, pp. 595–609, Jan. 2003, doi: 10.1080/01431160304982.

[25] Q. Huang, X. Yang, B. Gao, Y. Yang, and Y. Zhao, "Application of DMSP/OLS nighttime light images: A meta-analysis and a systematic literature review," *Remote Sens.*, vol. 6, no. 8, pp. 6844–6866, Jul. 2014, doi: 10.3390/rs6086844.

[26] C. D. Elvidge, K. E. Baugh, M. Zhizhin, and F.-C. Hsu, "Why VIIRS data are superior to DMSP for mapping nighttime lights," *Asia–Pacific Adv. Netw.*, vol. 35, p. 62, Jun. 2013, doi: 10.7125/apan.35.7.

[27] C. Cao, X. Shao, and S. Uprety, "Detecting light outages after severe storms using the S-NPP/VIIRS Day/Night band radiances," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1582–1586, Nov. 2013, doi: 10.1109/LGRS.2013.2262258.

[28] W. C. Skamarock *et al.*, "A description of the advanced research WRF version 3," Univ. Corp. Atmos. Res., Boulder, CO, USA, Tech. Rep. NCAR/TN-475+STR, 2008, doi: 10.5065/D68S4MVH.

[29] R. Pokhrel, S. D. Cos, J. P. Montoya Rincon, E. Glenn, and J. E. González, "Observation and modeling of hurricane Maria for damage assessment," *Weather Climate Extremes*, vol. 33, Sep. 2021, Art. no. 100331, doi: 10.1016/j.wace.2021.100331.

[30] 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*. Accessed: May 24, 2020. [Online]. Available: https://purl.stanford.edu/mb182my4891
[31] C. Homer, C. Huang, L. Yang, B. Wylie, and M. Coan, "Development of a 2001 national land-cover database for the united states," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 7, pp. 829–840, Jul. 2004, doi: 10.14358/PERS.70.7.829.
[32] H. A. Chipman, E. I. George, and R. E. McCulloch, "BART: Bayesian additive regression trees," *Ann. Appl. Statist.*, vol. 4, no. 1, pp. 266–298, Mar. 2010, doi: 10.1214/09-AOAS285.
[33] R. Nateghi, S. D. Guikema, and S. M. Quiring, "Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes," *Risk Anal.*, vol. 31, no. 12, pp. 1897–1906, Dec. 2011, doi: 10.1111/j.1539-6924.2011.01618.x.
[34] S. D. Guikema, S. M. Quiring, and S.-R. Han, "Prestorm estimation of hurricane damage to electric power distribution systems," *Risk Anal.*, vol. 30, no. 12, pp. 1744–1752, Dec. 2010, doi: 10.1111/j.1539-6924.2010.01510.x.
[35] J. He, D. W. Wanik, B. M. Hartman, E. N. Anagnostou, M. Astitha, and M. E. B. Frediani, "Nonparametric tree-based predictive modeling of storm outages on an electric distribution network," *Risk Anal.*, vol. 37, no. 3, pp. 441–458, Mar. 2017, doi: 10.1111/risa.12652.
[36] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[37] V. Rodriguez-Galiano, M. P. Mendes, M. J. Garcia-Soldado, M. Chica-Olmo, and L. Ribeiro, "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.*, vols. 476–477, pp. 189–206, Apr. 2014, doi: 10.1016/j.scitotenv.2014.01.001.
[38] L. Wang, X. Zhou, X. Zhu, Z. Dong, and W. Guo, "Estimation of biomass in wheat using random forest regression algorithm and remote sensing data," *Crop J.*, vol. 4, no. 3, pp. 212–219, Jun. 2016, doi: 10.1016/j.cj.2016.01.008.
[39] 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, doi: 10.1016/j.ress.2018.03.015.
[40] 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, doi: 10.1007/s11069-015-1908-2.
[41] M. Li, H. Hou, J. Yu, H. Geng, L. Zhu, Y. Huang, and X. Li, "Prediction of power outage quantity of distribution network users under typhoon disaster based on random forest and important variables," *Math. Problems Eng.*, vol. 2021, pp. 1–14, Jan. 2021, doi: 10.1155/2021/6682242.
[42] R. Nateghi, S. D. Guikema, and S. M. Quiring, "Forecasting hurricane-induced power outage durations," *Natural Hazards*, vol. 74, no. 3, pp. 1795–1811, Dec. 2014, doi: 10.1007/s11069-014-1270-9.
[43] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, Oct. 2001, doi: 10.1214/aos/1013203451.
[44] R. Jain and A. Nayyar, "Predicting employee attrition using xgboost machine learning approach," in *Proc. Int. Conf. Syst. Modeling Advancement Res. Trends*, Nov. 2018, pp. 113–120, doi: 10.1109/SYSMART.2018.8746940.
[45] J. Nobre and R. F. Neves, "Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets," *Expert Syst. Appl.*, vol. 125, pp. 181–194, Jul. 2019, doi: 10.1016/j.eswa.2019.01.083.
[46] C. CHEN and T. GUESTRIN, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, 2016, pp. 113–120.
[47] X. Zhang, C. Yan, C. Gao, B. A. Malin, and Y. Chen, "Predicting missing values in medical data via XGBoost regression," *J. Healthcare Informat. Res.*, vol. 4, no. 4, pp. 383–394, Dec. 2020, doi: 10.1007/s41666-020-00077-1.
[48] J. Cai, Y. Cai, H. Cai, S. Shi, Y. Lin, and M. Xie, "Feeder fault warning of distribution network based on XGBoost," *J. Phys., Conf. Ser.*, vol. 1639, no. 1, Oct. 2020, Art. no. 012037, doi: 10.1088/1742-6596/1639/1/012037.
[49] H. Hou, X. Chen, M. Li, L. Zhu, Y. Huang, and J. Yu, "Prediction of user outage under typhoon disaster based on multi-algorithm stacking integration," *Int. J. Electr. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 107123, doi: 10.1016/j.ijepes.2021.107123.
[50] A. Kapelner and J. Bleich, "BartMachine: Machine learning with Bayesian additive regression trees," *J. Stat. Softw.*, vol. 70, no. 4, pp. 1–40, 2016, doi: 10.18637/jss.v070.i04.

JUAN P. MONTOYA-RINCON received the B.S. degree in mechanical engineering from EAFIT University, Medellin, Colombia, in 2018. He is currently pursuing the Ph.D. degree in mechanical engineering with The City College of New York, New York, NY, USA. His research interests include the extreme weather-related power outages prediction using statistical models, power grid resiliency analysis, and risk assessment of the transmission systems.

SHAMS AZAD received the B.S. degree from the Bangladesh University of Engineering and Technology and the M.S. degree from the Stuttgart University of Applied Sciences, Germany. He is currently a Ph.D. Fellow with the Department of Civil and Urban Engineering, New York University. His research interests include performing innovative research to solve the emerging challenges in cities and promote the transformation of existing cities into more livable, resilient, and sustainable environments.

RABINDRA POKHREL received the bachelor's degree in mechanical engineering from Kathmandu University, Nepal, in 2003, the master's degree from Santa Clara University, California, in 2009, and the Ph.D. degree from The City College of New York (CCNY), in 2021. Since 2011, he has been teaching with Kathmandu University as an Assistant Professor focusing mostly on alternative energy and recently in environmental dynamics and urban planning. His Ph.D. research prepared him on understanding Caribbean climatology, extreme weather events, and especially on urban energy sustainability. He has developed tools to incorporate in urban climate model to study the impacts of building integrated technologies on mitigating peak energy usage for historic and future extreme heat events. His career objective is to use modeling/observation data to understand, implement, and promote measures to adapt to climate change.

MASOUD GHANDEHARI received the B.S. degree from Columbia University, the M.S. degree from McGill University, and the Ph.D. degree from Northwestern University. He is currently with the Faculty of Civil and Urban Engineering, NYU Tandon School of Engineering. His research interests include applications of advanced instrumentation and modeling targeting the health of urban systems, where he studies the interaction of physical, natural, and human systems.

MICHAEL P. JENSEN received the B.S. degree in atmospheric sciences from the State University of New York at Stony Brook, in 1990, and the M.S. and Ph.D. degrees in meteorology from The Pennsylvania State University, in 1993 and 2000, respectively.

He was a Postdoctoral Research Associate with Columbia University and the NASA Goddard Institute of Space Studies, New York, NY, USA, from 2000 to 2004. His accomplishments include the observational study of atmospheric processes that drive cloud lifecycle, particularly applied to marine boundary layer cloud and deep convective cloud systems. As a Meteorologist at the Brookhaven National Laboratory (BNL), Upton, NY, USA his research activities focus on the use of observations from the Department of Energy's Atmospheric Radiation Measurement Facility where he has served as the Principal Investigator for the 2011 Midlatitude Continental Convective Clouds Experiment (MC3E) and the 2021–2022 TRacking Aerosol Convection interactions ExpeRiment (TRACER). He currently leads the Cloud Processes Group, BNL, and works as the Deputy Chair of the Environmental and Climate Sciences Department. He is a member of the American Geophysical Union and the American Meteorological Society.

JORGE E. GONZALEZ received the bachelor's degree in mechanical engineering from the Georgia Institute of Technology, in 1988, and the Ph.D. degree in mechanical engineering from the University of Puerto Rico-Mayagüez, in 1994. He was appointed in 2015 by the Mayor of the City as Member of the Climate Change Panel for the City of New York, and more recently as a Senior Visiting Scientist of the Beijing Institute of Urban Meteorology and the Brookhaven National Laboratory. He is currently the Director of The City College of New York (CCNY) Initiative to Promote Academic Success in STEM (CiPASS), a Lead Scientist of the Coastal-Urban Environmental Research Group (CUERG), and The CCNY Presidential Professor of mechanical engineering at the CCNY. He teaches and conducts research in urban energy sustainability, urban weather and climate, urban remote sensing, and regional climate modeling and analysis. He holds several patents in solar energy equipment, aerosol detection, and energy forecasting for buildings, and was recognized as a Prominent Young Researcher by the National Science Foundation with a Prestigious CAREER Award. He has authored or coauthored more than 100 peer-reviewed publications, has delivered hundreds of conference presentations, and his research has attracted more than $40M in external funding. He is a Fellow Member of the American Society of Mechanical Engineering (ASME), and the Former Vice-Chairperson of the American Meteorological Society Board on the Urban Environment. He is the co-editor of the *ASME Handbook of Integrated and Sustainable Buildings Equipment and Systems*, and was named this year 2019 as the Founding Editor of the newest *ASME Journal of Engineering for Sustainable Buildings and Cities*.