21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

CIGRE US National Committee
2023 Grid of the Future Symposium

Prescriptive Vegetation Management Framework Using High-Resolution Satellite NDVI Imagery for Resilience Planning: A Case Study in Puerto Rico

Shuaiang RONG, Andrija SADIKOVIC*
Quanta Technology, LLC
USA

SUMMARY

Vegetation, particularly trees falling or growing into overhead power lines, is the leading cause of electric power outages in the United States. It significantly impacts power reliability, especially in regions with dense forests like Puerto Rico. To address the issue of vegetation-related outages, the paper presents a prescriptive vegetation management framework to mitigate vegetation-related power outages in distribution systems. The study aims to enhance power reliability and resilience by leveraging historical utility outage data, high-resolution NDVI (Normalized Difference Vegetation Index) satellite imagery, and advanced analytics. Specifically, this paper uses NDVI values, which indicate vegetation density, to develop a vegetation proxy index for each line segment. The index accounts for distribution line segment neighbouring area's impact on power lines and provides detailed information on vegetation exposure along the line segments. The study demonstrated ability to predict outage events and affected customer numbers based on variables - such as line length, customer numbers, and the newly introduced vegetation proxy index – using Poisson log-linear regression. The regression models' performance was evaluated using AICc and BIC criteria, with the results showing that the model parameters contribute to the model's accuracy with statistical significance. To mitigate potential risks of vegetation-related outages, the study recommended tree trimming in specific areas based on the vegetation proxy index. Two tree trimming approaches, narrow and wide, were considered, and their impact on reducing outage events and affected customers was predicted. This creates an opportunity to optimally plan tree trimming, assuring the highest return on invested capital in terms of SAIDI/SAIFI reduction. The results validated the expectation that wide trimming had a more substantial effect on reducing outages and the number of affected customers. The presented work demonstrated the potential of using lower-cost high-resolution satellite imagery and advanced analytics to predict vegetation-related power outages and develop prescriptive vegetation management strategies. With a developed predictive vegetation-related outage model, distribution system operator in Puerto Rico can make informed decisions on tree trimming – where, how deep, and when to trim trees - to enhance the system's reliability and resilience at the lowest cost to the ratepayers.

KEYWORDS

Vegetation-related power outages, satellite NDVI imagery, Distribution system reliability and resilience, Poisson log-linear regression models, Prescriptive vegetation management

asadikovic@quanta-technology.com

1. INTRODUCTION

According to the Federal Energy Regulatory Commission (FERC), vegetation, especially trees that grow or fall into overhead power lines, is the primary cause of electric power outages in the United States. This issue significantly impacts power reliability, particularly in tropical regions like Puerto Rico, where over 59% of the area is covered by forests [1]. In 2018, a single fallen tree on a power line resulted in a massive power outage affecting approximately 900,000 customers in Puerto Rico, leading to significant economic costs [2]. Data from LUMA, the utility company, reveals that vegetation-related outages in Puerto Rico's distribution networks accounted for a substantial 19% of the total outages between 2017 and 2022 due to the high exposure to trees.

LUMA has engaged Quanta Technology to help design cost-effective solutions that improve power reliability and resilience in Puerto Rico. With a deep understanding of the challenges LUMA is facing, Quanta Technology developed a data-driven, prescriptive vegetation management framework by leveraging historical data and advanced analytics, aiming to prevent outages and build resilient distribution systems. In light of the significant impact of vegetation on power outages, it is crucial that the next phase of resilience planning addresses the urgent challenges related to vegetation by better assessing, quantifying potential risks, and prescribing solutions to vegetation-related vulnerabilities.

To mitigate potential risks of vegetation before they cause outages, utilities traditionally conduct periodic tree trimming that involves regularly scheduled maintenance where utility crews proactively trim trees in predetermined areas along power lines. However, determining the trimming regions typically relies on manual patrols or helicopter inspections, which is extremely exhaustive and time-consuming [3]. New sensor technologies, such as LiDAR (Light Detection and Ranging) [4] and vision sensors [5] has been deployed to vegetation monitoring system for strategic defense of power infrastructures. Vegetation height and the encroachment to power lines are estimated in [5] using image data from vision sensors mounted on power towers deploying stereovision and deep learning techniques. While proven effective in transmission systems, such methodology has not been successfully applied at a granular level on a distribution system in a cost-effective fashion.

Airborne LiDAR has been used in [6] to collect vegetation height and GIS (Geographic Information System) data for predicting vegetation-related outages. However, in distribution systems, the use of drones for scanning along the extensive network of distribution lines is time-consuming and costly. Over the past few years, there has been a noticeable decline in the cost of launching satellites, leading to a proliferation of satellites and mini-satellites that are equipped with advanced sensors. Therefore, the cost of commercial satellite imagery has significantly decreased, enabling access to high-resolution images and extensive global coverage. A recent study in [7] use high-resolution satellite imagery to monitor the vegetation proximity to the power lines based on image processing and neural network techniques. However, this study only discusses monitoring system and does not provide an outage prediction methodology. Several studies have been conducted to predict vegetation-related outages using analytical and data-driven approaches [8-11]. These studies have focused on modeling the impact of tree trimming on vegetation-related failures and predicting feeder failure rates. Nevertheless, they either neglect to utilize vegetation indices [8] or employ them directly as measured [9-11], lacking an analytical interpretation of vegetation image data to accurately quantify the vegetation exposure of power lines at high resolutions and its impact on outages for detailed and accurate analysis of outage risks at the feeder level. Additionally, these studies do not include predictions of the number of affected customers during distribution system outages.

In this paper, a comprehensive case study in Puerto Rico is conducted to predict vegetation-related outages in distribution systems using lower-cost, high-resolution satellite NDVI imagery and historical outage data. The study includes the development of a Gaussian kernel-based vegetation proxy index, derived from NDVI values, to accurately quantify the vegetation exposure of power lines. The impact of various indicators – such as distribution line characteristics (e.g., line length, number of phases in a line section), number of customers, and the newly developed vegetation proxy index - on vegetation-related outages and the number of customers affected are investigated. Statistical models using Poisson log-linear regression are generated to predict the number of outages and affected customers due to vegetation. By employing the established models, distribution system operators can estimate the impact of vegetation control measures, such as tree trimming, on outage and customer loss reduction. This predictive capability enables them to make informed decisions on prioritizing tree-trimming activities across the regions and in specific areas, ultimately leading to a more effective capital spent on a resilient and reliable distribution system with minimized vegetation-related outages and customer disruptions.

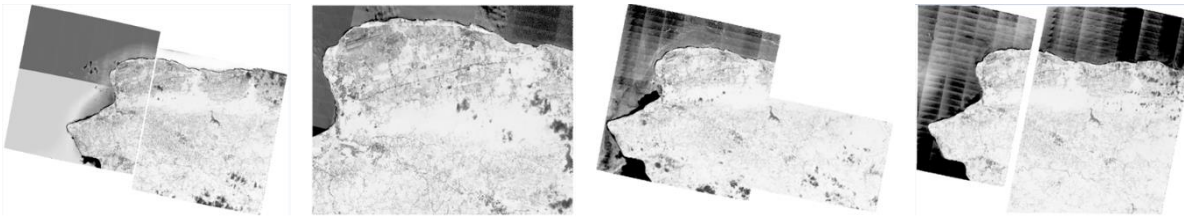
2. USE CASE AND DATA DESCRIPTION

The study is conducted in partnership with LUMA, the power distribution system operator in Puerto Rico. It focuses on the Northwest area of the distribution network, which encompasses both urban business areas and rural regions, including agriculture fields, forests, and water streams, as shown in Fig. 1. Various datasets were collected, including satellite imagery, distribution system infrastructure data, and outage records specific to the studied area, as described below.

Fig. 1 Studied region in Puerto Rico (images from google earth)

2.1 Satellite NDVI imagery

This study uses four high-resolution satellite quarterly NDVI images captured in March, June, September, and December in 2022, as shown in Fig. 2.



(a) Mar. 2022 (b) Jun 2022 (c) Sep. 2022 (d) Dec. 2022

Fig. 2 Satellite NDVI images captured in 2022

These images have a consistent pixel size of 3 meters per pixel and encompass NDVI values ranging from -1 to 1. The NDVI values commonly used to detect vegetation in remote sensing and is computed by measuring the difference between the near-infrared light (NIR),

which is strongly reflected by vegetation, and the red light (RED), which is absorbed by vegetation, as illustrated in (1) [7].

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

The high-resolution NDVI maps provide substantial vegetation data that allows detailed investigation and quantification study of vegetation impact on distribution feeders.

2.2 Power infrastructure and outage data

The GIS data of the line segments reveals that there are 181 feeders within the designated map region. These feeders account for approximately 10% of the total 1698 feeders operated by LUMA in Puerto Rico. They include distribution lines with voltage levels ranging from 2 kV to 13 kV, operating in 1, 2, or 3 phases. The dataset focuses on power outages related to vegetation from 2017 through 2022. In addition, it is disaggregated by individual feeders including line length data of overhead and underground lines in 1, 2, or 3 phases for each feeder, the number of customers served by that feeder, and the count of tree-related incidents and affected customers on that feeder, as described below.

Nomenclature	**Description**
Outage_Event_Num	Total number of outage events in each feeder
Affected_Customer_Num	Total number of customers whose power are interrupted in each feeder
Length 1phase_OH	Total length of overhead 1 phase power line in each feeder
Length 2phase_OH	Total length of overhead 2 phase power line in each feeder
Length 3phase_OH	Total length of overhead 3 phase power line in each feeder
Length_1phase_UG	Total length of underground 1 phase power line in each feeder
Length _2phase_UG	Total length of underground 2 phase power line in each feeder
Length _3phase_UG	Total length of underground 3 phase power line in each feeder
Customer_Num	Total number of the customers served by each feeder

The current dataset lacks information on vegetation, which will be generated based on the satellite NDVI image processing in the subsequent section.

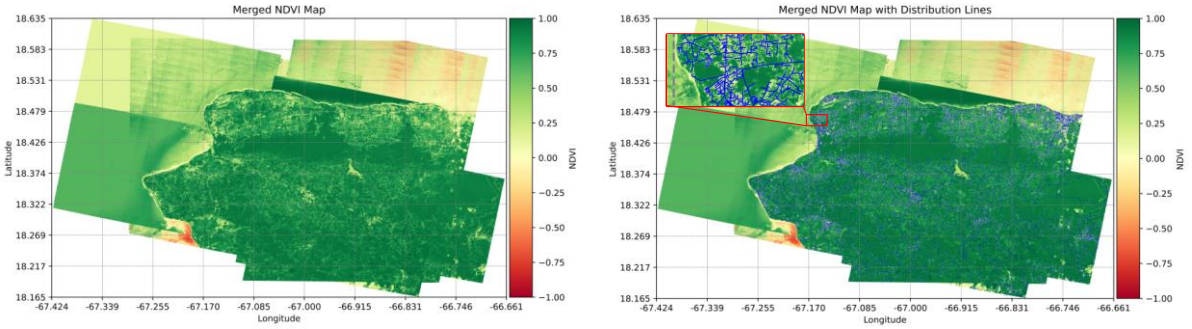
3. SATELLITE IMAGE PROCESSING AND VEGETATION PROXY INDEX

The four seasons' satellite images are merged into a single NDVI image. This resulting NDVI image overlapped with the distribution network based on GIS data by aligning the image coordinate system with the GIS coordinate system. The overlapping provides a spatial correspondence between the NDVI pixels and the distribution network, which is utilized to generate a vegetation proxy index for every distribution line segment at a high resolution that can consist of as little as a few line spans.

3.1 Satellite image data processing

While the four satellite image data focus on the same part of Puerto Rico, they still vary in size and covered regions. To create a comprehensive NDVI map for 2022 and extract as much valuable information as possible from these images, they are merged into a single map using the maximum method. Specifically, the four images are pre-processed to filter out the pixel data outside the NDVI range (-1 to 1). Subsequently, the GIS coordinates of pixels in the four maps are aligned to generate a new merged image. In this merged image, the maximum NDVI value is selected for pixels with the same coordinates. By employing this approach, the final merged NDVI map accurately records the most severe vegetation conditions observed throughout 2022, which have the potential to affect power lines and cause outages in the studied region. Fig. 3 (a) shows the merged NDVI map, with the NDVI values illustrated in color for better visualization, in contrast to the greyscale in Fig. 2. Further, the distribution lines are mapped

on the merged NDVI image based on the GIS data of the starting and ending points of each line segments, as shown in Fig.3 (b).



(a) Merged NDVI map (b) Merged map with distribution lines

Fig. 3 Merged satellite NDVI map covered with distribution lines

3.2 Vegetation proxy index

After mapping the distribution lines with the NDVI map, the vegetation proxy index is generated to quantify the vegetation exposure level of each line segment. Fig. 4 uses an example to illustrate the vegetation proxy index quantification process based on a Gaussian convolution, which is a common operation and building block for algorithms in signal and image processing [12].

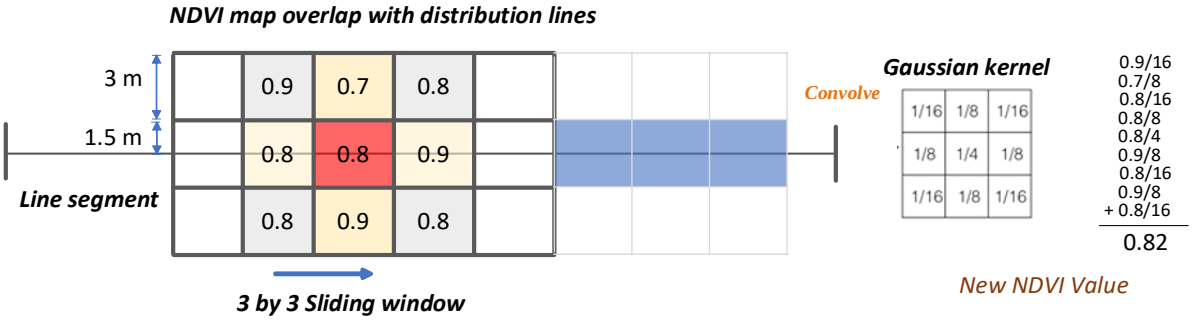


Fig. 4 Vegetation proxy index for line segment

To account for the impact of neighboring areas on the power line, a 3 by 3 sliding window covering a 9m×9m area around the line is created. This window is convolved with a normalized Gaussian kernel to differentiate the levels of vegetation exposure in the covered areas. The resulting vegetation proxy index, denoted as $V_{i,j,m}$, represents the quantification of the vegetation exposure on the i_{th} pixel of the j_{th} line segment ($L_{j,m}$) in the m_{th} feeder (F_m), as shown in (2).

$$V_{i,j,m} = \underbrace{\begin{bmatrix} x_{i-1,i-1} & x_{i-1,i} & x_{i-1,i+1} \\ x_{i-1,i} & x_{i,i} & x_{i,i+1} \\ x_{i-1,i+1} & x_{i+1,i} & x_{i+1,i+1} \end{bmatrix}_{j,m}}_{A_{j,m}} * \underbrace{\begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix}}_{G} \tag{2}$$

To elaborate, matrix $A_{j,m}$ contains the NDVI values in the 3 by 3 sliding window, with $x_{i,i}$ being the NDVI value in the i_{th} pixel on the line, and the rest represent NDVI values in neighboring pixels. Matrix Grepresents the normalized Gaussian kernel, with specific values to weigh the impact of adjacent pixels appropriately. Specifically, $g_{11}, g_{13}, g_{31}, g_{33}$ has the same value of $\frac{1}{16}$, $g_{12}, g_{21}, g_{32}, g_{23}$ has the same value of $\frac{1}{8}$, and g_{22} is $\frac{1}{4}$.

By sliding through the line segment pixel by pixel and convolving with the Gaussian kernel, a new set of NDVI values is generated, forming a vegetation proxy index that indicates vegetation exposure at the pixel level on the line. This index captures vegetation exposure across a wider range in the line corridors, providing detailed information for accurate power line analysis. It is important to note that this vegetation proxy index remains applicable even when line segments do not align parallelly with the pixels like in Fig. 4 as the approach effectively considers neighboring pixels that cover adjacent areas, and provides valuable information for analyzing power lines under varying scenarios.

After generating the vegetation proxy index for each line segment, the next step is to aggregate this data to obtain the feeder-level vegetation proxy index. This aggregation process combines the vegetation proxy index values of all the line segments within a feeder to provide an overall measure of vegetation exposure for that feeder.

4. STATISTICAL MODELING

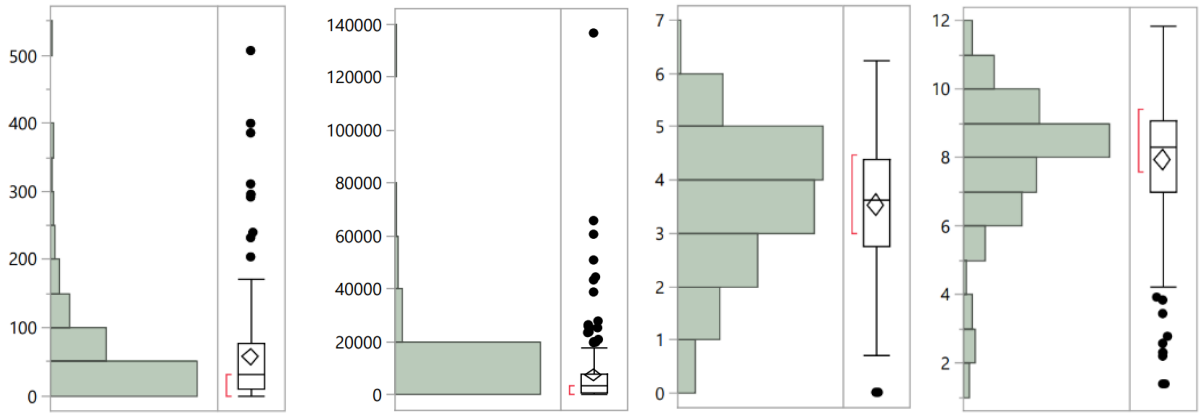
4.1 Poisson log-linear regression model

Two statistical models are developed to analyze the outage characteristics of the feeders, utilizing two different outputs: Outage_Event_Num (Y_1) and Affected_Customer_Num (Y_2) in the feeder. Poisson regression is usually used for modeling when the response variables are counts [13]. As shown in Fig. 5 (a) and (b), both Y_1 and Y_2 display a skewed Poisson distribution, deviating from a normal pattern. Given this non-normality, traditional Poisson linear regression is not applicable for effective modelling. To address this issue, log-linear regression is chosen for this paper. This method involves transforming the outputs into their logarithmic form, as illustrated in Fig. 5 (c) and (d), resulting in a more normal distribution. The transformed outputs are then utilized as the new response variables for the log-linear regression model. The Poisson log-linear models with response variables of Y_1 and Y_2 are represented in (5) and (6).

$$\boldsymbol{y_1} = \exp(\alpha_1 + \beta_{1,1}\boldsymbol{x_{1,1}} + \cdots + \beta_{1,k}\boldsymbol{x_{1,k}}) \tag{5}$$

$$\boldsymbol{y_2} = \exp(\alpha_2 + \beta_{2,1}\boldsymbol{x_{1,1}} + \cdots + \beta_{2,k}\boldsymbol{x_{2,k}}) \tag{6}$$

where $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ indicate the vectors of the Outage_Event_Num and Affected_Customer_Num on the selected feeders, respectively. The models consist of intercepts (α_1 and α_2) and variables $\{\boldsymbol{x_{1,1}}, \ldots \boldsymbol{x_{1,k}}\}$ and $\{\boldsymbol{x_{2,1}}, \ldots \boldsymbol{x_{2,k}}\}$ for Y_1 and Y_2 responses, respectively. The weights for these variables are denoted as $\{\beta_{1,1}, \ldots \beta_{1,k}\}$ and $\{\beta_{2,1}, \ldots \beta_{2,k}\}$.



(a) Event count (Y_1) (b) Customer number (Y_2) (c) log (Y_1) (d) log (Y_2)

Fig. 5 Data distribution of event count and number of affected customers from 2017 to 2022

To select the appropriate variables for the models, correlation analysis using the REML (Restricted Maximum Likelihood) method is performed. This method is commonly used to estimate the correlation structure in linear mixed models [14]. Variables that exhibit significant correlations with the response variables will be considered candidates for building the log-linear models. However, correlation analysis only captures variables with evident correlations with the responses. To ensure the best model fit, further screening is performed by comparing regression models with different variables. By doing so, the best-fitting model is identified, and its corresponding variables are selected as the final determinants for the log-linear models.

4.2 Goodness-of-fit criterion and model comparison

Two different methods are used in this paper to compare the regression models for the data set, i.e., akaike information criteria (AICc), and Bayesian information criterion (BIC). The AICc is an advanced method derived from the AIC method, which is usually used to evaluate the model quality by estimating the prediction error [15]. The AIC is defined as (7).

$$AIC = -In(\hat{L}) + 2k \tag{7}$$

where k is the number of independent parameters in the model, and $\hat{L}$ represents the maximized value of the likelihood function of this model. However, when the sample size is small, AIC has a tendency to select models with too many parameters, leading to overfitting. To address this, AICc was introduced to incorporate a correction for small sample sizes. It adds an extra penalty term for the number of parameters, as represented in (8).

$$AICc = AIC + \frac{2k^2 + 2k}{m - k - 1} \tag{8}$$

where m denotes the sample size, i.e., the number of studied feeders in the dataset. The Bayesian Information Criterion (BIC) is similar to AIC but imposes a more substantial penalty for the number of parameters [16]. It uses the natural logarithm of the sample size (m) to ensure that fewer parameters are favored for a simpler model, as shown in (9).

$$BIC = -In(\hat{L}) + In(m)k \tag{9}$$

In model comparison, smaller values of AICc and BIC are preferred, as they indicate better fit and appropriate parameters for the dataset. The paper uses both AICc and BIC to identify the most suitable regression model with minimal risk of overfitting.

5. RESULTS

The Poisson log-linear modeling is applied to the dataset, which combines the utility data and the feeder-level vegetation proxy index data. This statistical approach is applied to analyze the correlation between the variables for predicting two crucial factors: Outage_Event_Num (number of outages) and Affected_Customer_Num (number of affected customers). To create these models, correlation analysis is conducted to identify meaningful relationships between the given variables. Additionally, goodness-of-fit evaluation metrics such as AICc and BIC are utilized to assess the models' performance and choose the best-fit models. The generated models are utilized to forecast the potential reduction in outage numbers and affected customer numbers. This is achieved by simulating the effects of tree

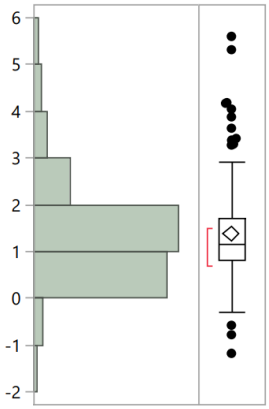


Fig. 6 Data distribution of $\log(\frac{Y_1}{line\ length})$

trimming in the recommended vegetation management areas, which result in decreased NDVI values.

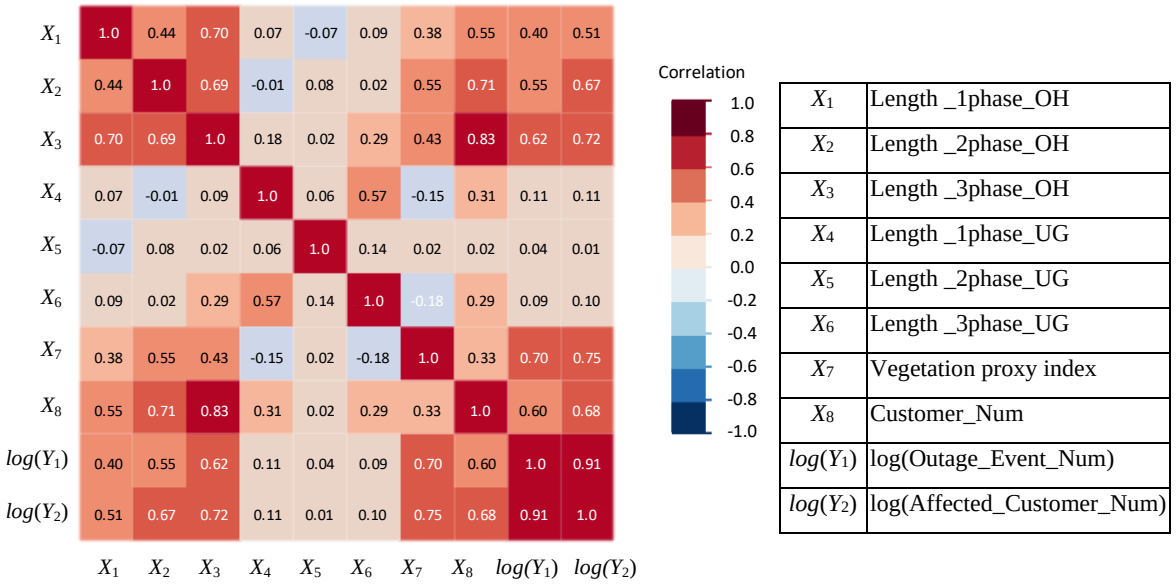
5.1 Outlier removal

To process the data, the initial step is to remove outliers to ensure data quality and accuracy. From Fig. 5, the data distribution also reveals the presence of outliers for both Y_1 and Y_2. However, removing outliers based solely on the distributions may not be feasible since the outage event counts and affected customer counts are inherently related to the overhead bare wire length of the feeders. Shorter feeders with fewer customers may naturally have lower event outage counts and customer impacts, while longer feeders may exhibit higher counts of events and affected customers.

To address this issue, a more nuanced approach is taken. Instead of directly removing outliers, the logarithmic form of the outage event count per mile is utilized as an indicator, represented by $log(\frac{Y_1}{line\ length})$, as shown in Fig. 6. By employing this transformed representation, outliers in the data can be better identified and analyzed without arbitrarily removing feeders solely based on their event and customer counts. This allows for a more robust analysis of the data and ensures that the influence of feeder length on outage characteristics is appropriately considered. The analysis revealed that 14 feeders as outliers, and they have been excluded from the dataset. These outlier feeders may have experienced configurational changes over the past five years, resulting in excessively low or high vegetation-related events considering the overhead line exposure and feeder customer count today. As a result, the remaining 167 feeders are used for further analysis and modeling.

5.2 Correlation analysis and variable selection

The correlation between the indicators in the given utility data and the vegetation proxy index with the model outputs log (Y_1) and log (Y_2) are analyzed based on the REML (Restricted Maximum Likelihood) method. Fig. 7 shows the estimation results in a heatmap.



X_1	Length _1phase_OH
X_2	Length _2phase_OH
X_3	Length _3phase_OH
X_4	Length _1phase_UG
X_5	Length _2phase_UG
X_6	Length _3phase_UG
X_7	Vegetation proxy index
X_8	Customer_Num
$log(Y_1)$	log(Outage_Event_Num)
$log(Y_2)$	log(Affected_Customer_Num)

Fig. 7 Heatmap of correlations

The heatmap analysis reveals noteworthy patterns in correlation between various indicators and vegetation-related outage events (Y_1), as well as their impact on affected customers (Y_2). The line length of overhead lines in 1/2/3 phases (X_1, X_2, and X_3) shows relatively higher positive correlations (0.4~0.7) with vegetation-related outages and affected

customers. This finding aligns with expectations, as feeders with longer overhead lines have greater vegetation exposure potential and are more susceptible to nearby tree contact, leading to potential outages. Essentially, an increased length of overhead lines amplifies a circuit's exposure to tree-related issues. Conversely, distribution circuits with underground line sections in 1/2/3 phases (X_4, X_5, and X_6) exhibit low correlation with the outputs (Y_1 and Y_2). This result is expected since underground lines are not exposed to vegetation, thus not contributing to vegetation-related outages.

Additionally, the vegetation proxy index (X_7) demonstrates a high correlation with the feeder's vegetation-related outages, validating the effectiveness of the index generation method and the significance of considering the vegetation proxy index in this study. This index proves to be a useful tool in predicting and addressing vegetation-related challenges. Moreover, the analysis indicates that an increase in customer base (X_8) significantly contributes to the heightened exposure to tree-related issues. As the customer base grows, so does the demand for power, necessitating an expansion of the distribution network, including longer overhead lines that may encounter vegetation-related challenges. As a result, five indicators with the highest correlations including the length of 1/2/3 phase overhead line (X_1, X_2, X_3), developed vegetation proxy index (X_7), and the served customers (X_8) are selected as variables for building the log-linear regression models.

5.3 Log-linear regression models and model comparison

TABLE I and II shows the estimated parameters of Poisson log-linear models of outage prediction and affected customer prediction and their p-values. In addition, R^2, AICc, and BIC are also shown for models with different variables combinations. From TABLE I, it can be seen that variables X_7 (vegetation proxy index) and X_8 (served customer) consistently exhibit low p-values, indicating their statistical significance in the model. Models 1 to 4 have similar R^2 values above 0.7 and are not highly statistically distinguishable. Among them, Model 3, which retains only three variables (X_3, X_7, and X_8), demonstrates the smallest AICc and BIC values, indicating the best fit for the dataset. Reducing the number of variables further, especially by removing the vegetation proxy index in Model 5, leads to a substantial decline in R^2 and a rise in AICc and BIC, indicating poorer fitting performance. This highlights the importance and effectiveness of the vegetation proxy index, which has a significant impact on outage events.

TABLE I Estimated parameters, p-values, R^2, AICs, BIC of outage prediction

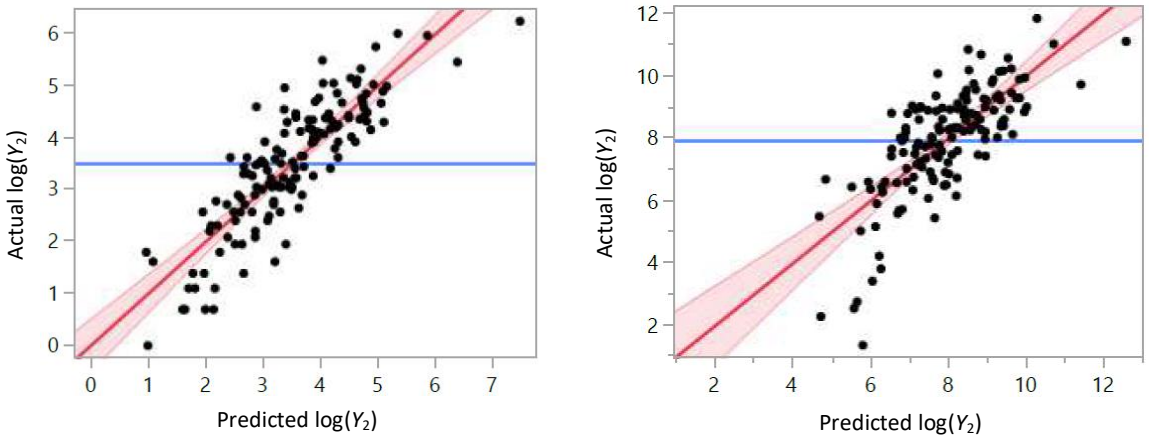
Model	$\beta_{1,1}$ (X_1)	$\beta_{1,2}$ (X_2)	$\beta_{1,3}$ (X_3)	$\beta_{1,4}$ (X_7)	$\beta_{1,5}$ (X_8)	Intercept (α_1)	R^2	AICc	BIC
1	-1.734e-6 0.3930	-1.185e-6 0.6784	1.0453e-5 0.0022	3.5625 <0.0001	3.75e-4 <0.0001	0.8755 <0.0001	0.7376	286.17	306.14
2	N/A	-7.009e-7 0.8022	9.0664e-6 0.0023	3.4676 <0.0001	3.732e-4 <0.0001	0.9207 <0.0001	0.7362	284.73	301.94
3	**N/A**	**N/A**	**9.0035e-6 0.0023**	**3.4139 <0.0001**	**3.643e-4 <0.0001**	**0.9406 <0.0001**	**0.7360**	**282.61**	**297.03**
4	N/A	N/A	N/A	3.7535 <0.0001	5.556e-4 <0.0001	0.8188 <0.0001	0.7179	290.04	301.63
5	N/A	N/A	N/A	N/A	6.961e-4 <0.0001	2.4893 <0.0001	0.4473	383.74	392.49

TABLE II Estimated parameters, p-values, R^2, AICs, BIC of affected customer prediction

Model	$\beta_{2,1}$ (X_1)	$\beta_{2,2}$ (X_2)	$\beta_{2,3}$ (X_3)	$\beta_{2,4}$ (X_7)	$\beta_{2,5}$ (X_8)	Intercept (α_2)	R^2	AICc	BIC
1	-7.348e-6 0.0474	-1.018e-5 0.0511	1.4687e-5 0.0168	5.5810 <0.0001	6.15e-4 0.0001	4.2273 <0.0001	0.5788	457.84	477.80
2	N/A	-8.133e-6 0.1149	8.8102e-6 0.1030	5.1791 <0.0001	6.071e-4 0.0002	4.4185 <0.0001	0.5666	459.74	476.95

3	**N/A**	**N/A**	**8.0806e-6 0.1352**	**4.5565 <0.0001**	**5.05e-4 0.0006**	**4.6496 <0.0001**	**0.5588**	**460.15**	**474.56**
4	N/A	N/A	N/A	4.8613 <0.0001	6.757e-4 <0.0001	4.5403 <0.0001	0.5516	460.30	471.89
5	N/A	N/A	N/A	N/A	1.02e-4 <0.0001	0.1882 <0.0001	0.3337	515.23	523.97

Similarly, Table II reveals a similar pattern, with Models 1 and 4 exhibiting no significant statistical difference, while Model 5 performs significantly worse after removing the vegetation proxy index indicator. However, it is noted that in models 1 to 4, AICc gradually increases, while BIC decreases with the reduction of variables. Considering both trends, we confidently conclude that Model 3 provides the best fit, with relatively low AICc and BIC values. Furthermore, Fig. 8 (a) and (b) illustrate the comparison between the predicted values from the generated models and the actual values of outage event count (Y_1) and affected customer number (Y_2), respectively.



(a) Outage event count (b) Affected customer number

Fig. 8 Comparison of predicted and actual value

5.4 Prescriptive tree trimming

The process of recommending trimming areas is conducted at the pixel level, enabling detailed and precise trimming suggestions to achieve desired (post-trimming) benefits. Based on observations from the map, it has been noted that areas covered by forests typically exhibit NDVI values greater than 0.7, while grasslands or bushes have NDVI values lower than 0.6. Using this information, the recommended trimming areas are identified as regions covered by line segments and pixels with NDVI values equal to or greater than 0.7. To calculate the length of the trimmed area (L_{trim}) along the line (L_{total}), the proportion of pixels to be trimmed (N_{trim}) to the total pixels on the line (N_{total}) is used, as shown in equation (10).

$$L_{trim} = L_{total} \times N_{trim}/N_{total} \tag{10}$$

This study considers two trimming patterns: narrow and wide trimming. Narrow trimming targets only the 3m x 3m area on the power line, while wide trimming covers a 9m x 9m region within the power line corridor. NDVI values in the trimmed pixels are reduced to 0.6. Trimming lengths along 1/2/3 phases of overhead lines are recorded for both trimming methods, as shown in Fig. 9. The recommended trimming areas along power lines are depicted in grey, with white and blue line representing power lines with vegetation proxy index below and above 0.7, respectively. Predictions for outage events and affected customers are made based on the new NDVI values and compared with the counts before trimming. Results in TABLE IV show the percentage reduction in outages and customer losses for each trimming approach.

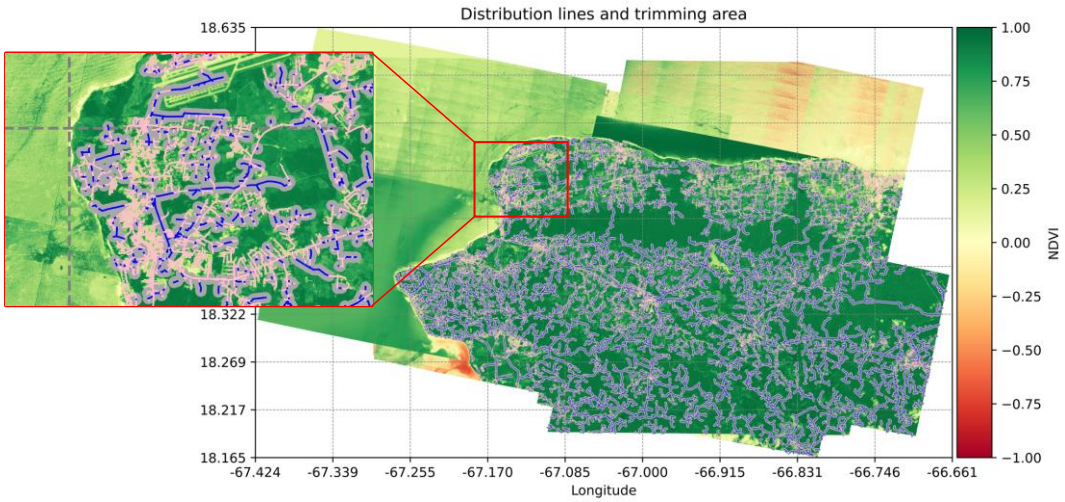


Fig. 9 Distribution lines with vegetation proxy index and recommended trimming area

TABLE III Trimmed length (mile) of narrow trimming and wide trimming areas

Trimmed line type	Total Length	Narrow Trimmed Length	Wide trimmed Length
1phase_OH	619.0	349.0	373.2
2phase_OH	850.7	433.9	474.0
3phase_OH	985.5	378.9	423.9

TABLE IV Estimated reduction of outage events and number of affected customers

Variables	Total 5-year count (before trimming)	Narrow trimming		Wide trimming	
		Prediction	Reduction	Prediction	Reduction
Outage_Event_Num	9369	8,688	7.1%	7,058	24.8 %
Affected_Customer_Num	1,082,445	979,791	9.2%	746,156	31.2%

The results indicate that the trimmed length along the distribution power lines is approximately half of the total length for 1/2/3 phase overhead lines. Narrow tree trimming leads to a 7.1% reduction in outage events and a 9.2% decrease in affected customers. On the other hand, wide tree trimming has a more significant impact, resulting in a 24.8% reduction in outage events and a 31.2% decrease in affected customers, but comes at a higher tree trimming cost. This prescriptive analysis is instrumental in decision-making processes for vegetation management. It provides valuable insights into the potential benefits of tree trimming in the recommended areas, allowing distribution system operators to assess the effectiveness of these actions in reducing the risk of outages and minimizing customer impact.

6. CONCLUSION

This paper presents a prescriptive and proactive vegetation management framework using lower-cost high-resolution satellite NDVI imagery for resilience planning of distribution systems. The results of the case study in Puerto Rico show that the vegetation exposure level of the power lines is effectively quantified using the proposed vegetation proxy index and can be effectively used to construct prediction models, such as the Poisson log-linear regression model, to predict the vegetation related outage risks and prescribe the most cost-effective measures to improve reliability and resilience of the system. The initial results highlight the promise of integrating vegetation factor into the outage risk prediction. In the future, our work will focus on improving the model accuracy using deep learning techniques to fit in a broader data-driven outage prediction framework to enhance electric grid resilience in Puerto Rico.

BIBLIOGRAPHY

[1] D. L. Ocasio, “State and Private Forestry Fact Sheet--- Puerto Rico 2023”, US Forest Service, International Institute of Tropical Forestry, San Juan, Puerto Rico, U.S., Jan. 31, 2023.

[2] Nicole Acevedo, “Puerto Rico: Single fallen tree on power line leaves 900K without power”, NBC News, U.S., April 12, 2018.

[3] L. F. Luque-Vega et al., “Power line inspection via an unmanned aerial system based on the quadrotor helicopter,” in *Proc. IEEE Mediterranean Electrotech. Conf.*, 2014, pp. 393–397.

[4] K. Sikorska-Łukasiewicz, “Methods of automatic vegetation encroachment detection for high voltage power lines,” in *Proc. SPIE Radioelectron. syst. Conf., Jachranka*, Poland, Feb. 2020.

[5] S. Rong, L. He, L. Du, Z. Li and S. Yu, "Intelligent Detection of Vegetation Encroachment of Power Lines With Advanced Stereovision," *IEEE Trans. Power Delivery*, vol. 36, no. 6, pp. 3477-3485, Dec. 2021.

[6] D. Wanik, J. Parent, E. Anagnostou, and B. Hartman, "Using vegetation management and LiDAR-derived tree height data to improve outage predictions for electric utilities," *Electric Power Systems Research*, vol. 146, pp. 236-245, 2017.

[7] M. Gazzea, M. Pacevicius, D. O. Dammann, A. Sapronova, T. M. Lunde and R. Arghandeh, "Automated Power Lines Vegetation Monitoring Using High-Resolution Satellite Imagery," *IEEE Trans. Power Delivery*, vol. 37, no. 1, pp. 308-316, Feb. 2022.

[8] D. T. Radmer, P. A. Kuntz, R. D. Christie, S. S. Venkata, and R. H. Fletcher, “Predicting vegetation-related failure rates for overhead distribution feeders,” *IEEE Trans. Power Del.*, vol. 17, no. 4, pp. 1170–1174, Oct. 2002.

[9] T. Dokic and M. Kezunovic, "Predictive risk management for dynamic tree trimming scheduling for distribution networks," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4776-4785, 2018.

[10] A. U. Melagoda, T. D. L. P. Karunarathna, G. Nisaharan, P. A. G. M. Amarasinghe and S. K. Abeygunawardane, “Application of Machine Learning Algorithms for Predicting Vegetation Related Outages in Power Distribution Systems,” *2021 3rd International Conference on Electrical Engineering (EECon)*, Colombo, Sri Lanka, 2021, pp. 25-30.

[11] S. D. Guikema, R. A. Davidson, and H. Liu, “Statistical models of the effects of tree trimming on power system outages,” *IEEE Trans. Power Delivery*, vol. 21, no. 3, pp. 1549-1557, 2006.

[12] P. Getreuer, “A Survey of Gaussian Convolution Algorithms”, *Image Processing On Line*, Dec. 17, 2013.

[13] A. C. Cameron and P. K. Trivedi, “Regression Analysis of Count Data,” *Econometric Soc. Monographs*, no. 30, Cambridge Univ. Press, Cambridge, U.K., 1998.

[14] M. S. Bartlett, “Properties of Sufficiency and Statistical Tests”, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 160, no. 901, pp. 268-282 May 18, 1937.

[15] P. Stoica and Y. Selen, “Model-order selection: a review of information criterion rules,” *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36-47, July 2004.

[16] K. P. Burnham, D. R. Anderson, “Multimodel inference: understanding AIC and BIC in Model Selection”, *Sociological Methods & Research*, 33: 261–304, 2004.