Electric Power Systems Research 175 (2019) 105909

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Assessing the effects of a vegetation management standard on distribution grid outage rates

Diego Cerrai*, Peter Watson, Emmanouil N. Anagnostou

Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT, USA

ARTICLE INFO

Keywords:
Power grid resilience
Vegetation management
Outage prediction model

ABSTRACT

Enhanced Tree Trimming (ETT) is a vegetation management standard consisting in the trimming or removal of trees in proximity of overhead lines. We quantify the effects of this practice, designed to improve the resiliency of the overhead electrical power grid, using two independent methodologies. The first approach is a statistical study of the change of frequency of outage-free locations. The second approach uses an Outage Prediction Model (OPM) as a vulnerability assessment tool to evaluate the change in the number of outages before and after ETT. The OPM is a machine learning based framework that relates weather, soil, vegetation and electric grid characteristics to storm-related power outages. The two methods introduced in this work are compared against a straightforward assessment of the ETT impact on the number of outages. Applying both methods, the occurrence of power outages is studied for varying tree trimming amounts, for 144 storms occurred between 2005 and 2017 in the Northeastern United States. From the statistical approach we find a reduction of the outage-free grid cells during storms between 49% and 65% due to tree trimming. The OPM based analysis suggests that the number of outages during storms are reduced between 16% and 48% after performing ETT.

1. Introduction

Electric utilities apply tree trimming and other vegetation management standards along power transmission [1] and distribution [2,3] lines to minimize damages caused by trees during storms [4] and maintain system reliability [5]. Trees are among the top causes of outages in electric distribution systems [6,7], and vegetation management is a key component for improving the electric grid resilience to weather-related power outages [8].

In the absence of vegetation maintenance, trees can grow undisturbed and eventually fall on power lines [9], while cyclic tree trimming helps preventing trees and brush from interfering with the lines [10]. This cyclic vegetation management technique, which Eversource Energy-Connecticut performs every four to five years, is called Scheduled Maintenance Trimming (SMT), and consists in the removal of limbs within 8 feet (2.5 m) to the side, 10 feet (3 m) below, and 15 feet (4.5 m) above the wires of the distribution lines [11]. Beyond SMT, Enhanced Tree Trimming (ETT) is also performed, and consists in the complete removal of trees and brush within 8 feet (2.5 m) to the side of power lines [11].

Beyond the utility's perspective, tree trimming and removal in urban areas have climatological and societal dimensions: urban forests change microclimate conditions, by decreasing the temperature during hot summer days [12–14] and improving air quality [15–17], and also provide a more pleasant and comfortable living environment [18,19]. Matching utility benefits with people' expectations in urban areas is challenging, because trees and overhead distribution lines share the same space [20]. Collaboration and coordination between utilities and municipalities is therefore needed to guarantee the reliability of the electric system consequently people' well-being [20], especially where public concerns about tree removal has been raised [21]. Evaluation of the impact of ETT on power grid resilience may provide a vital piece of information for understanding the effects of this management practice.

A first attempt to evaluate the impact of tree trimming on failure rates for overhead (OH) distribution feeders can be found in [5]. According to that study, in a 8-year vegetation management cycle, there is a trend of increase of failure rates between 3 and 7 years after tree trimming is performed. However, the high variability around the mean makes the results of this study quantitatively difficult to interpret. A reason for the lack of statistically significant results can be attributed to the absence of parameters describing the annual variability of storms impacts.

In nonstorm conditions, the effects of tree trimming on power outages were evaluated in [22], for an electric company in the Southeastern US. From the study it emerged, with statistical significance, that electric power system reliability is improved by tree trimming. Under

* Corresponding author at: Department of Civil & Environmental Engineering, School of Engineering, University of Connecticut, Storrs, CT, USA.
E-mail address: diego.cerrai@uconn.edu (D. Cerrai).

https://doi.org/10.1016/j.epsr.2019.105909
Received 15 October 2018; Received in revised form 5 April 2019; Accepted 7 June 2019
Available online 19 June 2019
0378-7796/ © 2019 Elsevier B.V. All rights reserved.

storm conditions, a quantification of the reduction of the number of outages due to ETT has never been evaluated. However, it was demonstrated that tree trimming and other vegetation management techniques improve storm-related outage prediction [23].

This study aims at presenting methodologies for assessing of the ETT effects on the resilience of the electric grid. Specifically, we evaluate the impact of ETT on the rate of outages through two different approaches. The first is a purely statistical analysis of outage data to evaluate the trend of outage reduction to increasing ETT. The second is a resiliency evaluation carried out using an Outage Prediction Model (OPM) that predicts the power outages during storm events. The second approach allows to take into account the variability in storm severity through the use of OPM, which provides a reference for the evaluation. Results from the two evaluation frameworks will be compared against a straightforward assessment, which quantifies the change of outages for different tree trimming amounts without considering any other factor affecting power outage occurrence.

The next section provides a characterization of the dataset and a description of the model setup. Section 3 describes the methodology used for assessing the impact of ETT on electric grid resilience. The results obtained by following the two methods proposed in this paper are presented and discussed in Section 4. Finally, a summary of the principal findings of the study, together with their implications on future studies and the State economy is presented in Section 5.

2. Model and data

2.1. Variable description and model setup

This study is based on 100 extratropical storms exhibiting rain and wind conditions and 44 thunderstorms, occurred in the period 2005–2017 across Connecticut. The storms considered are an extension of the storm events dataset used in [24–26] for OPM evaluation. The storm outage model was based on the combination of the two machine learning models that in [26] exhibited the best performance for outage prediction purposes: Random Forest (RF, [27]) and Bayesian Additive Regression Trees (BART, [28]).

Two hundred decisions trees [29] constitute our RF. Each decision tree is formed by a series of nodes and branches that split the dataset through decision rules, by minimizing the sum of square error. Each tree uses a random subset of predictor variables and is trained on a random subset of training data.

The BART model is a sum of tree models, that can be expressed as [28]:

$$Y = \sum_{j=1}^{m} g_j(x, T_j, M_j) + \epsilon. \tag{1}$$

where $m = 30$ is the number of binary trees T_j with associated sets of parameters M_j; x is the vector of predictors, ϵ is assumed to be normal, with standard deviation σ. The contribution $g_j(x;T_j, M_j)$ to the Bayesian sum provided by each tree is computed through ten thousand Bayesian iterations of the Markov Chain Monte Carlo (MCMC [30]) algorithm, starting from prior specifications for T_j, M_j and σ. Four thousand iterations are used after convergence is reached to obtain the predictions. Number of iterations and trees are the same as [26], and are constrained by computational power.

The two models were trained on the storm datasets and used a number of input variables (described in A) including weather, land cover, soil, vegetation, electric grid characteristics, and historical outages, similarly to [23–25,31], and [26]. Weather and soil parameters were obtained from Weather Research and Forecasting (WRF v3.7) model [32,33], operating in hindcast mode. The WRF model was initialized with Global Forecasting System (GFS) weather analyses and the dynamical downscaling occured through the use of three nested domains, at 18, 6 and 2 km grid spacing. Land cover variables weare obtained from the National Land Cover Database (NLCD) at 30 m resolution, provided by the Multi-Resolution Land Characteristics Consortium (MRLC, [34]). Vegetation characteristics were represented through a Leaf Area Index (LAI) dataset, produced in [26]. This dataset was obtained through post-processing of a global LAI dataset [35] derived from MODIS [36] to compute, for every 8-day period of the year and for every grid cell, a climatological value of the LAI. Electric grid information, provided by Eversource Energy, included OH lines and annual tree trimming data. Storm outage data were provided by the electric utility: each outage is defined as a location where a two-man crew is needed to restore power, and the number of outages on a 2 km grid was the object of the models predictions.

2.2. Aggregation methods

Our analysis was performed on the 2 km inner WRF grid. Weather- and soil-related variables, obtained from WRF output are already at the 2 km grid, but other datasets had to be interpolated to this grid. Land cover parameters, computed in the immediate proximity (60 m) of the OH lines, were aggregated for each 2 km grid cell [31]. A similar procedure was employed for the computation of the OH trimmed and non-trimmed lines: the WRF grid was superimposed to the OH lines map, and the length of the lines in each grid cell was computed, together with the length of trimmed lines for each year (Fig. 1). Consequently, for each cell and for each year, we computed the percentage of lines treated with ETT and the cumulative percentage of ETT from the beginning of the treatment.

To account for vegetation density characteristics we post-processed the LAI data by computing, for each grid cell and for each week of the year, the climatological value of the LAI through the use of a gaussian filter and an autoregressive model applied to the original, global dataset. All the outages measured during or in the immediate hours after each storm were attributed to the grid cell in which they were reported.

For analyzing the dataset at the town level, each grid cell is assigned to the town with the largest areal coverage in the grid cell. Each town is assigned to a division on the basis of the geographic location (Eastern, Western, Southern, Central) within the utility service territory in Connecticut.

2.3. Dataset characteristics

The knowledge of the principal dataset features guided the development of the methods at the basis of the analysis performed in this work. In the following paragraphs we are presenting the characteristics of ETT and outage data, their mutual relationships, and the relationships with other variables of the dataset.

Between 2005 and 2008 ETT was not systematically performed, while between 2009 and 2016, ETT was performed on 21% of the total OH lines length in Connecticut, at a rate varying between 1.4% to 4.1% every year (Fig. 2). The decision as to where to trim each year is affected by considerations, such as budgets and past reliability [37]. Between 2009 and 2017, 80% of the grid cells in the study area received some ETT in portions of the transmission or distribution lines they contain.

Cumulative tree trimming amount presents a low variability across the 4 divisions, varying between 18% and 23%, and across the towns, since no town exceeds 50% of cumulative ETT (Fig. 3a). Moreover, tree trimming is performed in both mostly forested and mostly developed areas (Fig. 3a, b), hence no significant impact on the results of the analysis is expected from relating ETT to land cover variables.

The total number of outages for the considered extratropical storms and thunderstorm events is 58,236. Outages are primarily concentrated in South-Western Connecticut (Fig. 3d), due to the high population and assets density. The outage temporal distribution is uneven across the years (Fig. 3c) and the seasons. In particular, most of the dataset describes storms occurring when leaves are on trees, since the interaction

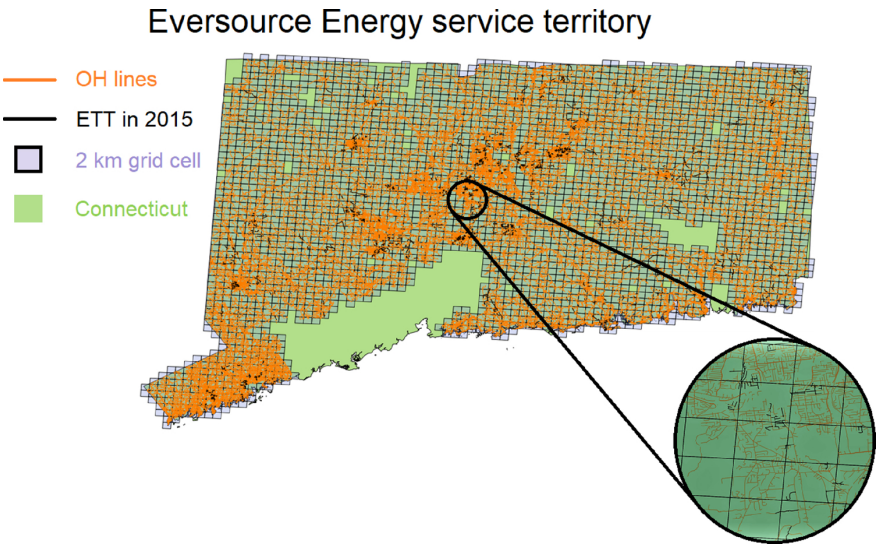

Fig. 1. Map of overhead lines (orange) and ETTed lines (black) during 2015 in the State of Connecticut (green), with WRF model grid overlaid (squares).

between solid or mixed precipitation and the electric grid is not object of this study. Two major hurricanes (Irene, 2011; Sandy, 2012) have been excluded from the study, because their outage predictions require a different OPM, whose error characteristics are difficult to quantify due to the limited dataset. The strongest storm considered in our analysis, occurred on October 2017, produced 4430 outages.

The outage distribution is zero- and one-inflated: on average 91% of the grid cells in each storm do not contain any outage, 7% contain 1 outage, while less than 2% have a number of outages grater or equal than 2. For this reason, the study of the frequency of zeros versus non-zeros is an appropriate metric to describe the most important features of the dataset.

It is important to note that in each grid cell the total number of outages depends on the length of the OH lines (Fig. 4a). The probability of having outages in a given cell is proportional to the line length in that cell. Moreover, also the cumulative ETT percentage depends on the line length, and the dependency of this quantity on the OH line length is shown for the year 2017 in Fig. 4b. All cells with total OH line length over 25 km are associated with ETT ranging between 5% and 55%, while all cells with ETT above 65% are associated with OH line length below 15 km. The first can be explained by the fact that if a grid cell has a high amount of OH lines it is more likely that a part of the lines was selected for the ETT program, but it is also less likely that all the vegetation next to the power lines was trimmed. On the contrary, it is more likely for a grid cell with a limited OH line amount to be completely trimmed.

Based on the above, one can argue that this relationship between ETT, outages, and line length can influence the results of this analysis. In the next section, we describe a methodology that, by taking into account these dependencies, introduces corrections to the raw results. This approach is necessary for the correct interpretation of the ETT impact on the reduction of outages occurrence, and to avoid an overestimation of this impact.

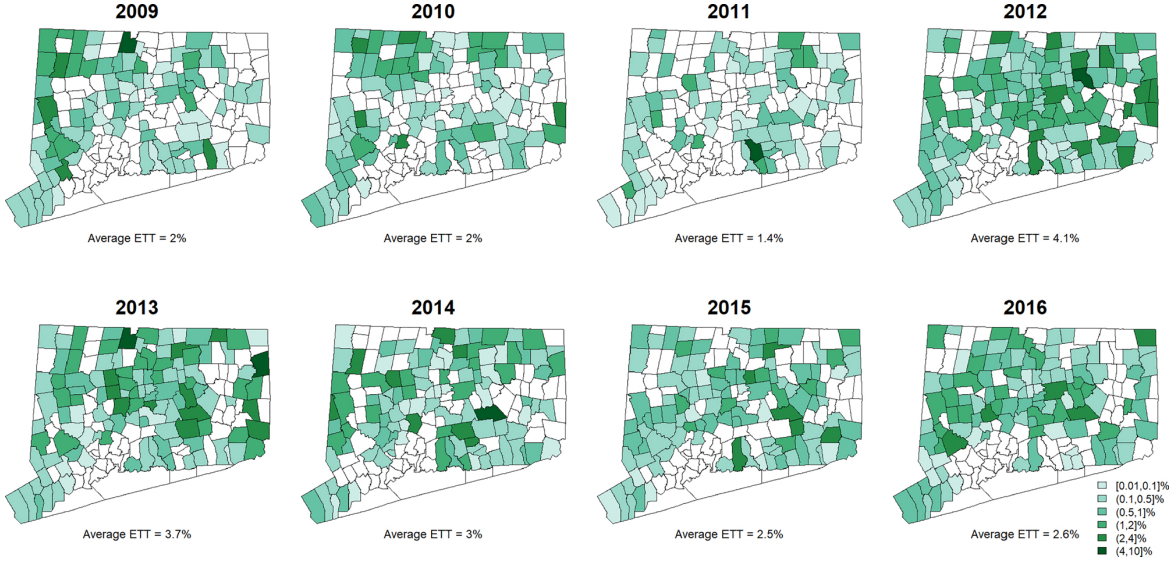

Fig. 2. Percentage of ETTed lines by town in the State of Connecticut between 2009 and 2016.

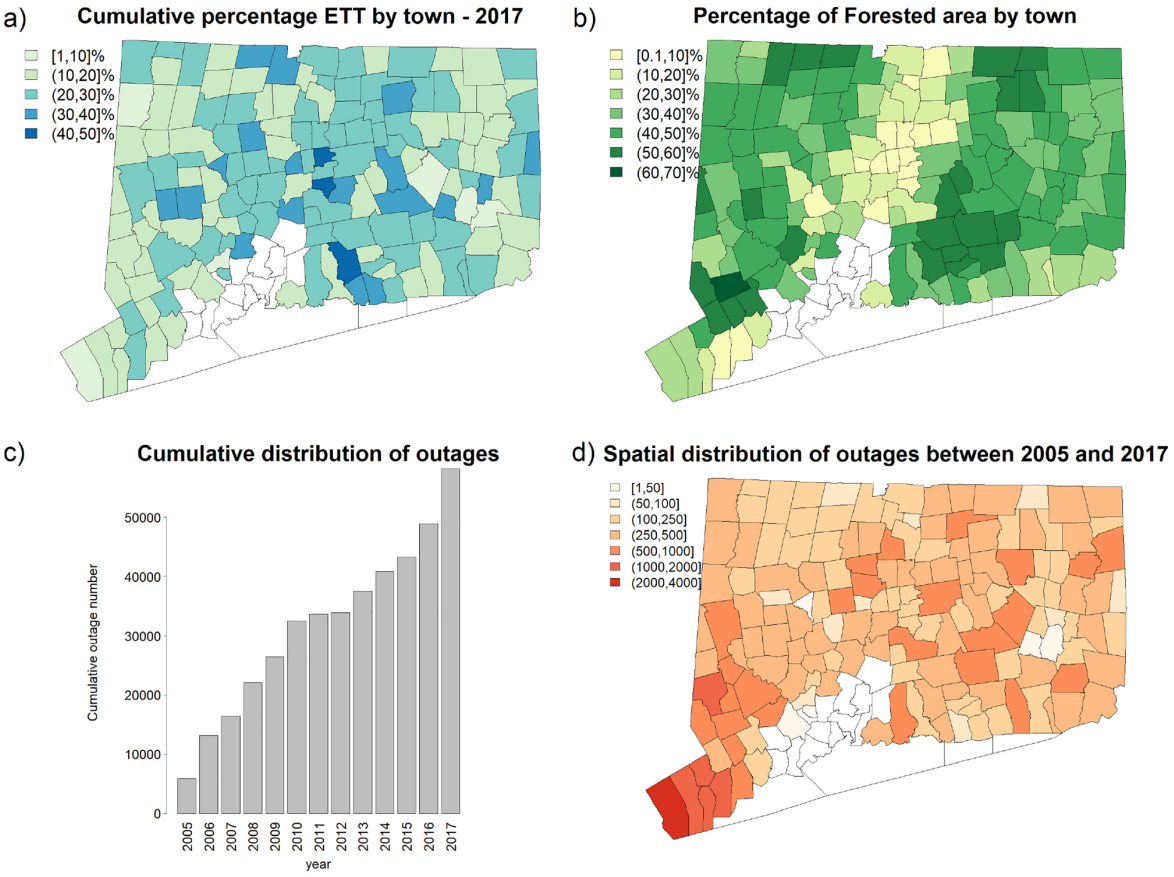

Fig. 3. a) Cumulative ETT percentage as of 2017; b) percentage of forested area; c) cumulative and d) spatial distribution of outages between 2005 and 2017.

3. Methodology

As discussed above, two separate methodologies were used to evaluate the role of ETT on the electric grid resilience:

- a statistical method, describing the relationship between the change of the number of outage-free grid cells and the amount of tree trimming.
- a model based analysis, which uses the OPM for establishing whether the impact of tree trimming on the amount of storm related power outages can be more accurately estimated by taking into account variations of storm frequency and intensity..

Both approaches were applied separately on the two storm types of this study: extratropical and summer thunderstorms. These methods are compared to a straightforward statistical assessment (hereafter named baseline), which measures the percentage of outage free grid cells for different tree trimming amounts.

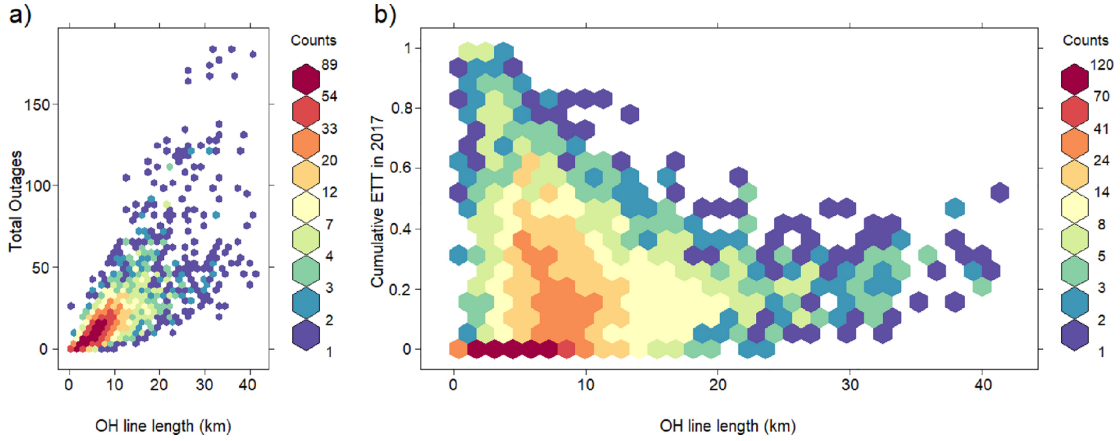

Fig. 4. a) density scatter plot between total outage and OH line length; b) and between cumulative ETT fraction in 2017 and OH line length.

3.1. Baseline assessment

The first step for the analysis of ETT impact on power outages occurrence during storms was the selection of subsets of the original dataset with similar vegetation management characteristics. We set seven thresholds equally spaced between 0 and 100% of cumulative ETT per grid cell to create eight equal intervals. These intervals allowed us to partition the dataset, and to study the confidence intervals [LI, UI] for the proportions between the presence and absence of outages for each subset using the Wilson score interval with continuity correction [38,39]:

$$\mathrm{LI} = \max\left\{0, \frac{2n\hat{p} + z^2 - [z\sqrt{z^2 - 1/n + 4n\hat{p}(1-\hat{p}) + (4\hat{p} - 2)} + 1]}{2(n + z^2)}\right\} \tag{2}$$

$$\mathrm{UI} = \min\left\{1, \frac{2n\hat{p} + z^2 + [z\sqrt{z^2 - 1/n + 4n\hat{p}(1-\hat{p}) - (4\hat{p} - 2)} + 1]}{2(n + z^2)}\right\} \tag{3}$$

where $z = 1.96$ for the 95% confidence level chosen, and $\hat{p} = n_z/n$ is the percentage of zeros, which is the ratio between the number of zeros n_z and the total number of entries n.

This baseline evaluation of the ETT change among the subsets, produces both deceiving and inaccurate results, by misrepresenting the impact of ETT, due to the lack of normalization for OH line length and storm intensity. We will use the baseline results as a starting point of our analysis, and we will compare the results of the herein proposed methods with the findings of this baseline evaluation.

3.2. Statistical data analysis method

In order to provide a more accurate evaluation of the ETT impact on electric grid resilience, we investigated some important factors affecting the occurrence of power outages through a statistical approach.

The change of the percentage of zeros between the subsets is only partially explained by a different tree trimming amount. The principal reason for the change of this percentage is given by the dependency between ETT and OH lines, already introduced in Fig. 4. The boxplots summarizing the principal characteristics of the OH line length distribution computed for all the storms across the subsets (Fig. 5) show a strong decrease of OH line length per grid cell for increasing ETT, if tree trimming is performed. Since the expected number of outages is proportional to the OH line length, the strong decrease of OH line length for increasing ETT suggests, per se, a modification in the expected number of outages. For this reason, it was necessary to perform the following steps:

1. To control our experiment we created new subsets for each tree-trimming interval using the outage history of the grid cells. For each member of the partition studied in the baseline assessment, the new subsets contain the history of weather, outages and trimming for the grid cells of this member. Since all the grid cells had a starting value, before 2009, of 0% ETT, the first subset, corresponding to [0%,12.5%] ETT, contained the entire dataset, and was not considered in this part of the analysis. For each other subset, the ETT histogram was composed of two peaks, that can be identified as before and after ETT. For each peak, we computed the confidence interval on $\hat{p}$ using the Wilson score. Multiple sessions of ETT were identifiable as background signal in the histogram.
2. We computed the quantity F of outage-free grid cells per standard OH line length:

$$F = \frac{\hat{p}^* m}{l} \tag{4}$$

where $m = 9.4$ km is the mean OH line length for all the grid cells, and l is the mean length for the considered subset. This normalization allowed us to take into account of the relationships between the number of outages and the length of the OH lines.
3. We computed the coefficients a and b, and their respective standard errors, both before and after ETT was performed. For this purpose, we used a weighted linear least square regression model [40] for the dependency of F on the ETT intervals:

$$F = a + b \cdot \mathrm{ETT} + \epsilon \tag{5}$$

where a is the intercept, b is the slope and ϵ the vector of errors. Values different than zero for the slope correspond to the presence of a trend in the number of outage-free grid cells for varying tree trimming. The intercept a can be associated to a typical value for the number of outage-free grid cells, therefore changes in a correspond to changes in reliability between before and after tree trimming.
4. We performed a z test for the slope of the regression, following [41] and [42]:

$$Z = \frac{b_1 - b_2}{\sqrt{\mathrm{SE}_{b_1}^2 + \mathrm{SE}_{b_2}^2}} \tag{6}$$

where b_1 and b_2 are the slopes of the regression models respectively before and after ETT, and SE_{b_1}, SE_{b_2} their standard errors. The test

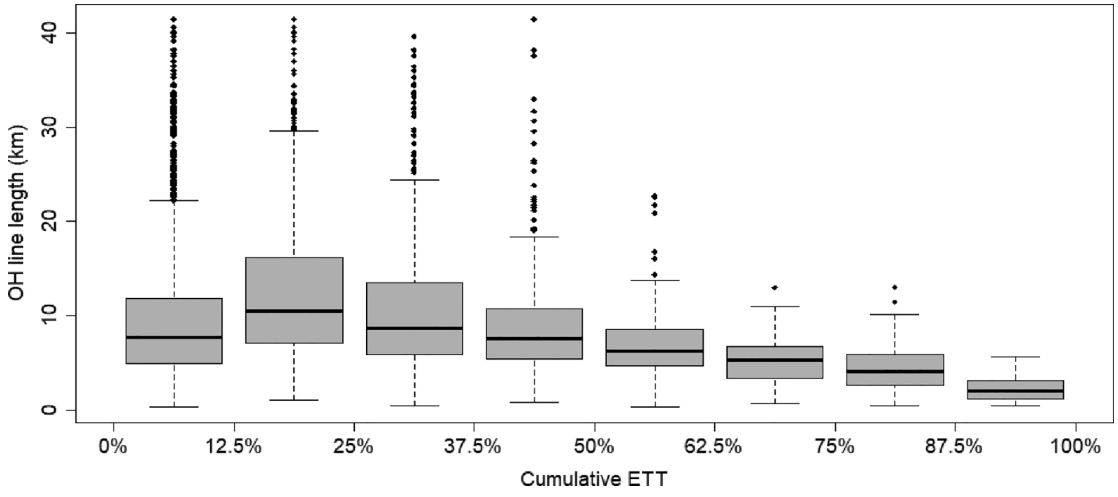

Fig. 5. Boxplots of OH line length for each subset based on ETT amount.

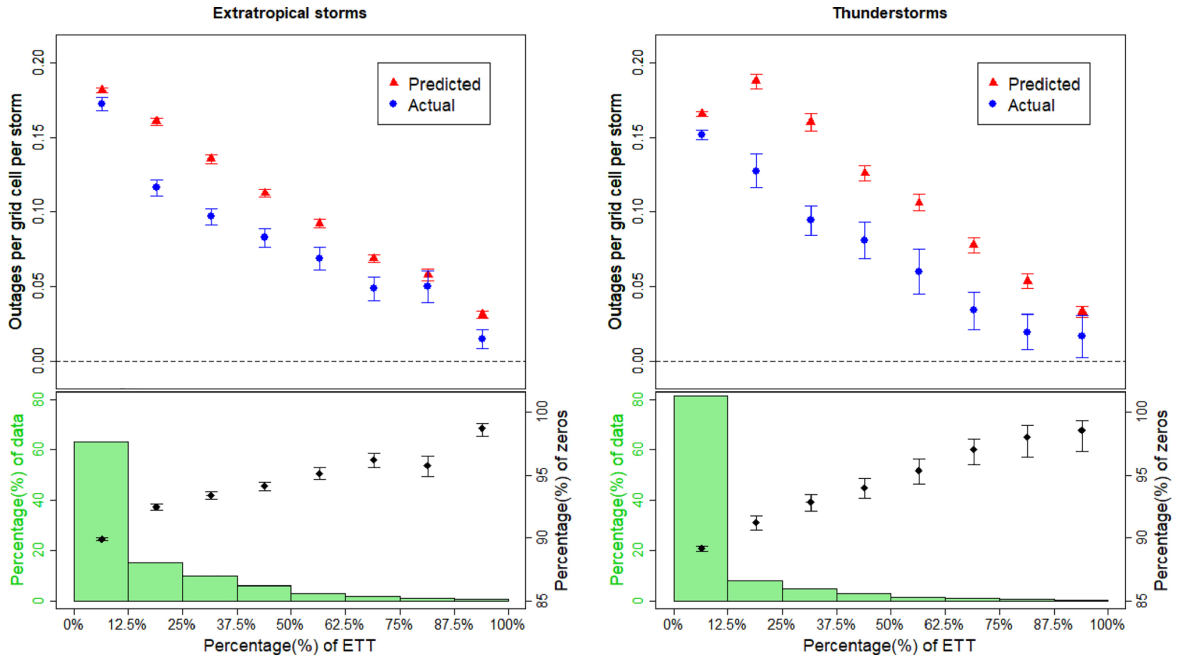

Fig. 6. Trend of predicted (red) and actual (blue) trouble spots versus the percentage of tree trimming (top); percentage of data and percentage of zeros for varying tree trimming (bottom), for extratropical storms (left), and thunderstorms (right).

for the slope allowed us to draw conclusions about the presence of possible trends of outage reduction for increasing tree trimming.

The statistical approach described above is used to evaluate the impact of ETT on the number of outage-free locations. However, since the number of power outages depends on storm intensity, the statistical analysis of the outage dataset alone is not sufficient due to the variability of number and severity of storms each year and the impact this variability has on the results. Therefore, this analysis was extended using the OPM model, that allows to normalize the outages based on the expected number for each storm. This method is discussed in the next section.

3.3. OPM model based method

The trends and the associated uncertainties in the number of outages, and their dependencies on storm intensity, OH line length, and ETT can be simultaneously considered by using the OPM. Specifically, the OPM ability to predict the intensity of storm impacts based on weather, land cover, soil moisture, vegetation variables, and electric grid characteristics allows to take into account of the storm severity variability by studying the relationships between predicted and actual outages.

The evaluation process of ETT impact on outage reduction through the use of the OPM consists of the following stages:

1. For each subset of the partition introduced in Section 3.2, we computed the mean number of OPM predicted and actual outages per grid cell per storm. Following [43], the 95% confidence intervals for the mean were calculated as:

$$\mathrm{CI} = \bar{x} \pm t_{\alpha/2}(s/\sqrt{n}) \tag{7}$$

where $\bar{x}$ is the mean, $t_{\alpha/2}$ is the critical value of the t-distribution, s is the standard deviation of the data, and n the number of samples.From the analysis of the trends of actual and predicted outages, it was possible to understand the model behavior for the different storm types.

2. Following the procedure performed in Section 3.2, we analyzed the family of subsets for an evaluation of the change of outages in locations where ETT was performed. For each subset, the Outage Overestimation Factor (OOF), was computed as the ratio between the predicted and the actual outages per grid cell. This quantity is invariant for both storm intensity and OH line length, and is close to 1 (unbiased) when computed on any sufficiently large dataset. However, ETT is not a variable used in the OPM, hence the OOF provides information on the outage reduction where ETT was performed.
Since the number of outages is much smaller than the number of entries in the dataset, and since in only 2% of the locations the number of outages is greater than one, a good approximation for the confidence intervals for the OOF is given by the confidence interval for the risk ratio. The lower and upper boundaries can be computed, following [44], as:

$$[\mathrm{LI}, \mathrm{UI}] = \mathrm{OOF}\cdot\exp\left(\pm z\sqrt{\frac{1}{x_1} - \frac{1}{n_1} - \frac{1}{x_2} - \frac{1}{n_2}}\right) \tag{8}$$

where x_1 and x_2 are the numbers of predicted and actual outages in each subset, that can be approximated as the number of grid cells with outages, and $n_1 = n_2$ is the dimension of each subset.
Since the results obtained for the absence of ETT are statistically unbiased, there is no need to proceed to further post-processing.

4. Results

Using the above-mentioned methodology, in this section we will quantify the impact of ETT on storm related outage frequency. We will compare results from the statistical and modeling approaches, for both thunderstorms and extratropical storms, in order to explain similarities and differences between storm types and approaches.

The first step of our analysis consisted in the assessment of the change of the number of outages per grid cell per storm and of the

percentage of zeros for increasing ETT. For this purpose, we selected the subsets of the partition introduced in Section 3.2, and we computed means and confidence intervals of the quantities of interest. For both extratropical storms (Fig. 6, left) and thunderstorms (Fig. 6, right), we determined, with statistical significance, a decrease of the number of actual outages per grid cell per storm (Fig. 6, top), and an increase of the percentage of outage-free grid cells in the dataset (Fig. 6, bottom) for increasing ETT. These results allowed us to estimate a 10 fold (87% to 91%) decrease of the number of power outages between the first and the last subset. It is important to note here that the OPM was able to predict the decrease of the number of outages for increasing ETT (Fig. 6, blue markers), although this quantity was not used as model predictor. This finding suggests that the predictability of the decrease of outage frequency for increasing ETT can be mostly explained by the decrease of OH line length for increasing ETT (Fig. 5), and by the relationships between outages and OH line length (Fig. 4a,b).

The OPM predictions, however, present an ETT-dependent bias, which manifests as an increasing overestimation of the average predicted outages with respect to the average actual values. In order to extract information from the bias change, we investigated the change of the OOF, already introduced in Section 3.3, between before and after ETT being performed. It is noted that, for all the considered subsets, the value of the OOF was not significantly different than the target value of one (unbiased) for the data corresponding to storms hitting areas of the territory before ETT was performed. For the same locations, however, the OOF was significantly different than one after trees were trimmed, for all the subsets, except for (75%, 87.5%] ETT, as shown by the blue markers in Fig. 7. These results allowed us to estimate that the impact of ETT on outage occurrence could be quantified as a 16% to 48% reduction in outage frequency (36%–63% for thunderstorms, 13%–52% for extratropical storms, not shown). We did not find, however, any statistically significant trend (not shown) of increase of OOF for increasing ETT, despite the group with the highest ETT amount presenting the highest OOF value.

For the same subsets, we also studied the change of outage-free grid cells before and after tree trimming. By contrasting the percentage of zeros within each subplot of Fig. 7, we found a significant increase of outage-free grid cells for all the groups after ETT was performed. Moreover, from a comparison across the different subplots of Fig. 7, we noticed that a strong increase was also present across different groups.

Through the information brought about by the differences across the groups, we quantified the average ETT impact, and the dependency of this impact on the amount of ETT. To achieve this, we computed the quantity F (equation 3), which allowed us to normalize the percentage of zeros across the ETT groups, and to evaluate whether the differences within the groups varied across the groups. It is important to note that the larger widths of the confidence intervals associated with high ETT values is due to the smaller sample size of these classes (green histograms in Fig. 6).

From this analysis we found that:

- for both extratropical storms and thunderstorms, a significant increase of the number of outage-free grid cells was measured after ETT was performed. This increase, according to Fig. 8, corresponded to an average decrease of grid cells with outages, ranging between 43% and 69% for thunderstorms, 50% and 67% for extratropical storms, and 49% and 65% for the combination of the two.
- the areas that received tree trimming were associated with a lower normalized percentage of zeros before ETT. This means that ETT was correctly performed in the most vulnerable areas;
- for both extratropical storms and thunderstorms, the slopes of the weighted linear regression models before and after ETT differed each other at $\alpha = 0.05$ confidence level. The Z value for the difference in slopes for the dataset obtained by the combination of extratropical storms and thunderstorm was $Z = -2.25$. This finding implied that, under the assumptions valid for this statistical analysis, the higher the cumulative ETT is, the higher its impact on outage reduction will be.

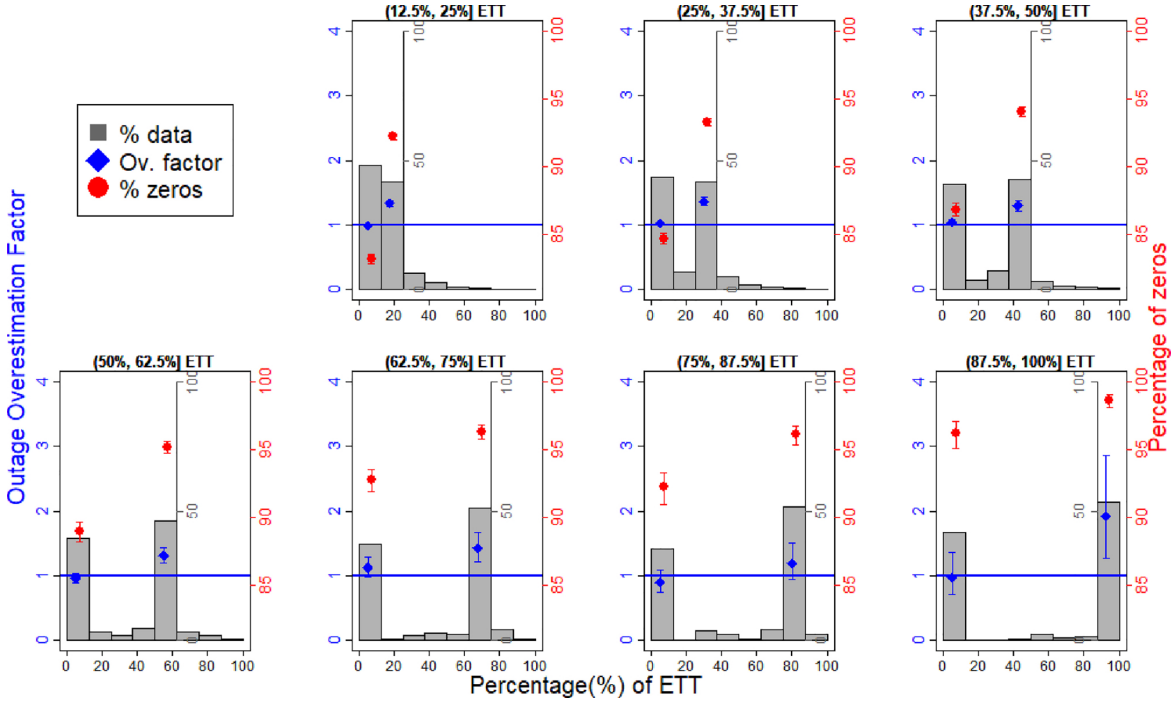

Fig. 7. Change of power outage frequency between before and after ETT is performed, for varying ETT amounts. The blue squares represent the change in the OOF, the red circles represent the change of the number of outage free grid cells. Both quantities are computed before (left peak of each histogram), and after (right peak) ETT is performed. The background noise of the histogram corresponds to multiple ETT in the same grid cell during the years.

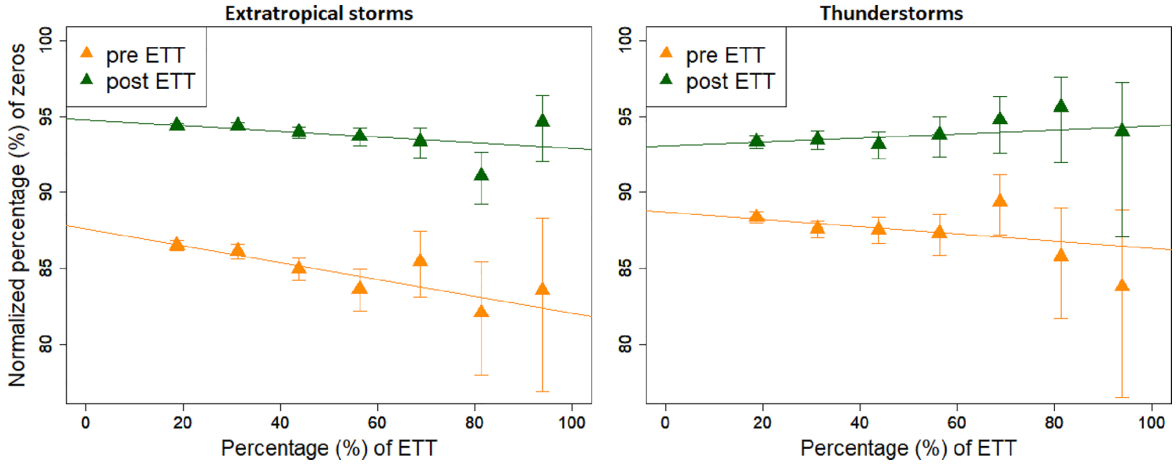

Fig. 8. Normalized percentage of zeros for varying ETT before (orange markers) and after (green markers) ETT was performed, for extratropical storms (left) and thunderstorms (right), with weighted linear least square regression lines overlaid.

This last result based on the count of outage-free grid cells contradicts the absence of trend of OOF for increasing ETT. The difference between these two findings could be explained by the different assumptions used and by the different quantities involved, and more in detail:

- despite the vast majority of the grid cells had zero or one outage per storm, 2% of the cells had more than one outage. A reduction of outages due to tree trimming in highly impacted grid cells did not affect the statistics for presence/absence of outages, but did affect the OOF;
- we assumed that the use of the OPM produced a normalization for storm intensity, and was able to remove temporal inconsistencies in the dataset, due to a varying storm selection across the years, or a trend of storm intensity. This assumption was not valid for the statistical method;
- the Z value (equation 5) of -2.25 was very close to the threshold of rejection of -1.96. This means that the conclusions on statistical significance could be reversed by assuming a 3-sigma limit instead of a 2-sigma.

Moreover, the lack of a clear trend in the increase of reliability for increasing tree trimming may be explained by the fact that not all power lines are placed next to trimmable trees. Consequently, similar increases are found (not shown in this paper) for locations where different amounts of ETT were performed, but corresponding to the same percentage of trimmed overhead lines.

Similarities in the results between the two approaches consist in the statistical significance of the average impact of tree trimming on power outage. By combining the results from the two methods we estimate, under different assumptions, that ETT produced a 16% to 65% reduction of the number of power outages.

5. Concluding remarks

The comparison performed in this work between the statistical analysis of the number of outage-free grid cells and of the OOF provided an improved understanding of the relationships between ETT (a vegetation management standard in Connecticut that represents removal of trees within 8 feet of power lines), outages, line length and storm severity.

The analysis of the data started with a direct evaluation of the reduction of the number of outages and of the increase of outage-free grid cells due to tree trimming. Using such a simplistic approach, we estimated a 10 fold decrease of the number of outges in extensively trimmed areas, with a significant trend of outage reduction for increasing ETT.

However, highly trimmed areas corresponded to areas with a lower OH line length, where a lower value of power outages was expected. By taking into account the OH line length in the study, we found that most of the outage reduction variability was explained by this quantity, and tree trimming accounted for a 49% to 65% reduction of grid cells with outages, with a statistically significant outage reduction trend for increasing ETT.

This approach did not take into account the variability of storm intensity. For this reason, we used an OPM to estimate the change in the OOF for different ETT ranges. The ETT impact on outage reduction based on this method was estimated between 16% and 48%. However, the model overestimation did not show any statistically significant trend for varying ETT (results are summarized in Table 1).

Further investigation is needed to explore the high variability in the relationship of ETT and outage rates. In a future study we will use a LIDAR-derived product describing the proximity of trees to power lines [23] as proxy to trimmed and trimmable areas, and use this dataset to study the change of reliability for varying trimmed areas.

The framework presented in this work will inform utilities as well as regulators and town officials about the efficiency of vegetation management programs in terms of improving the reliability of the system, and consequential financial benefits from reduction of outage rates. The continuation of this study will focus on outage reductions to economic benefits continuation of this work, and a step forward for the definition of optimal vegetation management and urban planning strategies for maximizing the benefits for utilities and its ratepayers.

Moreover, the OPM can benefit from an extension of this study through:

- the implementation of the expected reduction of outages due to ETT,

Table 1
Summary of the principal findings of this work.

Method:	ETT only:	Statistical:	Modeling:
Outage reduction: Thunderstorms	87%–89%	43%–69%	36%–63%
Outage reduction: Extratropical	87%–91%	50%–67%	13%–52%
Outage reduction: Combined	87%–91%	49%–65%	16%–48%
ETT amount (slope): Thunderstorms	Signif.	Signif.	Not signif.
ETT amount (slope): Extratropical	Signif.	Signif.	Not signif.
ETT amount (slope): Combined	Signif.	Signif.	Not signif.

which may provide corrections to the model in areas of significant trimming;
- the use of ETT as a predictor, in order to quantify outage variability due to tree trimming.

Declaration of interest

The authors of this publication have research support from Eversource Energy and DTN, and also hold stock in ACW Analytics. This publication partially uses classified datasets of the electric grid. We have full access to all of the data in this study and we take complete responsibility for the integrity of the data and the accuracy of the data analysis.

Appendix A. Predictor variables

Table 2

Table 2
Variables used as input to the OPM.

Variable	Description	Units
PercConif	Percentage of coniferous forest	%
PercDecid	Percentage of deciduous forest	%
PercDevel	Percentage of developed area	%
Wgt5	Duration of wind at 10 m above 5 m/s	hr
Wgt9	Duration of wind at 10 m above 9 m/s	hr
Wgt13	Duration of wind at 10 m above 13 m/s	hr
MaxW10m	Maximum wind at 10 m	$m\ s^{-1}$
Cowgt5	Continuous duration of wind at 10 m above 5 m/s	hr
Cowgt9	Continuous duration of wind at 10 m above 9 m/s	hr
Cowgt13	Continuous duration of wind at 10 m above 13 m/s	hr
MeanW10m	Mean wind at 10 m	$m\ s^{-1}$
Ggt13	Duration of wind gusts at 10 m above 13 m/s	hr
Ggt17	Duration of wind gusts at 10 m above 17 m/s	hr
MaxGust	Maximum wind gusts at 10 m	$m\ s^{-1}$
MeanGust	Mean wind gusts at 10 m	$m\ s^{-1}$
MaxTotPrec	Total precipitation	mm
MaxPreRate	Maximum precipitation rate	$mm\ h^{-1}$
MeanPreRate	Mean precipitation rate	$mm\ h^{-1}$
MaxSoilMst	Maximum Soil Moisture	$m^3\ m^{-3}$
MeanSoilMst	Mean Soil Moisture	$m^3\ m^{-3}$
MaxSpHum	Maximum Specific Humidity	$g\ g^{-1}$
MaxTemp	Maximum Temperature	K
MeanTemp	Mean Temperature	K
LAI	Leaf area index	$m^2\ m^{-2}$

References

[1] R.H. Yahner, R.J. Hutnik, Integrated vegetation management on an electric transmission right-of-way in southeastern Pennsylvania, US, J. Arboricult.
[2] P.A. Kuntz, R.D. Christie, S.S. Venkata, Optimal vegetation maintenance scheduling of overhead electric power distribution systems, IEEE Trans. Power Del. 17 (4) (2002) 1164–1169.
[3] P. Simpson, R. Van Bossuyt, Tree-caused electric outages, J. Arboricult. 22 (1996) 117–121.
[4] R.J. Campbell, Weather-related power outages and electric system resiliency, Congressional Research Service, Library of Congress, Washington, DC, 2012.
[5] D.T. Radmer, P.A. Kuntz, R.D. Christie, S.S. Venkata, R.H. Fletcher, Predicting vegetation-related failure rates for overhead distribution feeders, IEEE Trans. Power Del. 17 (4) (2002) 1170–1175.
[6] F.E.R. Commission, et al., Utility vegetation management final report, (2004 march) Link.(commissioned to support the investigation of the 2003 Northeast blackout.).
[7] P.J. Appelt, J.W. Goodfellow, Research on how trees cause interruptions-applications to vegetation management, Rural Electric Power Conference, 2004, IEEE, 2004 C6-1.
[8] W. House, Economic benefits of increasing electric grid resilience to weather outages, Washington, DC: Executive Office of the President.
[9] G. Paula, Right-of-way maintenance: Changing options, Elect. World T&D Special Rep 203 (9) (1989) S5–S15.
[10] W. Lovelace, Vegetation management on distribution line right-of-way are you getting top value for your money? Rural Electric Power Conference, 1996. Papers Presented at the 39th Annual Conference, IEEE, 1996, p. B5.
[11] Understanding vegetation management, (2018) https://www.eversource.com/content/docs/default-source/nh–pdfs/eversource-veg-mgmt-guide_final_web.pdf, accessed: 2018-04-10.
[12] T.R. Oke, The micrometeorology of the urban forest, Philos. Trans. Roy. Soc. London. B, Biol. Sci. 324 (1223) (1989) 335–349.
[13] N.J. Georgi, K. Zafiriadis, The impact of park trees on microclimate in urban areas, Urban Ecosyst. 9 (3) (2006) 195–209.
[14] E.G. McPherson, D.J. Nowak, R.A. Rowntree, Chicago's urban forest ecosystem: results of the Chicago Urban Forest Climate Project, vol. 186, US Department of Agriculture, Forest Service, Northeastern Forest Experiment., 1994.
[15] D.J. Nowak, et al., The Effects of Urban Trees on Air Quality, USDA Forest Service, 2002, pp. 96–102.
[16] D.J. Nowak, D.E. Crane, J.C. Stevens, Air pollution removal by urban trees and shrubs in the united states, Urban Forest. Urban Green. 4 (3-4) (2006) 115–123.
[17] J. Yang, J. McBride, J. Zhou, Z. Sun, The urban forest in beijing and its role in air pollution reduction, Urban Forest. Urban Green. 3 (2) (2005) 65–78.
[18] J.F. Dwyer, E.G. McPherson, H.W. Schroeder, R.A. Rowntree, Assessing the benefits and costs of the urban forest, J. Arboricult. 18 (1992) 227–227.
[19] D.J. Nowak, J.F. Dwyer, Understanding the benefits and costs of urban forest ecosystems, Urban and Community Forestry in the Northeast, Springer, 2007, pp. 25–46.
[20] W.B. Most, S. Weissman, Trees and power lines: minimizing conflicts between electric power infrastructure and the urban forest.
[21] Pura calls for suspension of enhanced tree trimming, https://www.ct.gov/pura/cwp/view.asp?Q=541516, accessed: 2019-04-03 (2014).
[22] S.D. Guikema, R.A. Davidson, H. Liu, Statistical models of the effects of tree trimming on power system outages, IEEE Trans. Power Del. 21 (3) (2006) 1549–1557.
[23] D. Wanik, J. Parent, E. Anagnostou, B. Hartman, Using vegetation management and lidar-derived tree height data to improve outage predictions for electric utilities, Electr. Power Syst. Res. 146 (2017) 236–245.
[24] D.W. Wanik, E.N. Anagnostou, B.M. Hartman, M.E.B. Frediani, M. Astitha, Storm outage modeling for an electric distribution network in northeastern USA, Natural Hazards 79 (2) (2015) 1359–1384.
[25] J. He, D.W. Wanik, B.M. Hartman, E.N. Anagnostou, M. Astitha, M.E.B. Frediani, Nonparametric tree-based predictive modeling of storm outages on an electric distribution network, Risk Anal. 37 (3) (2017) 441–458.
[26] D. Cerrai, D.W. Wanik, M.A.E. Bhuiyan, X. Zhang, J. Yang, M.E. Frediani, E.N. Anagnostou, Predicting storm outages through new representations of weather and vegetation, IEEE Access 7 (2019) 29639–29654.

[27] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[28] H.A. Chipman, E.I. George, R.E. McCulloch, et al., BART: Bayesian additive regression trees, Ann. Appl. Stat. 4 (1) (2010) 266–298.
[29] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and Regression Trees, CRC Press, 1984.
[30] T. Hastie, R. Tibshirani, et al., Bayesian backfitting (with comments and a rejoinder by the authors, Stat. Sci. 15 (3) (2000) 196–223.
[31] D. Wanik, E. Anagnostou, M. Astitha, B. Hartman, G. Lackmann, J. Yang, D. Cerrai, J. He, M. Frediani, A case study on power outage impacts from future hurricane sandy scenarios, J. Appl. Meteorol. Climatol. 57 (1) (2018) 51–79.
[32] G. Powers, X.-y. Huang, B. Klemp, C. Skamarock, J. Dudhia, O. Gill, G. Duda, D. Barker, W. Wang, A Description of the Advanced Research, WRF version 3, 2008.
[33] W. Wang, C. Bruyère, M. Duda, J. Dudhia, D. Gill, H.-c. Lin, ARW version 3 modelling system user's guide, (2009).
[34] National land cover database, (2018) https://www.mrlc.gov/, accessed: 2018-07-01.
[35] Leaf Area Index (1 Month - Terra/MODIS), https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD15A2_M_LAI, accessed: 2017-12-17 (2017).
[36] V.V. Salomonson, W. Barnes, P.W. Maymon, H.E. Montgomery, H. Ostrow, Modis: Advanced facility instrument for studies of the earth as a system, IEEE Trans. Geosci. Remote Sens. 27 (2) (1989) 145–153.
[37] C. Light, Power, Transmission and distribution reliability performance 608 report, (2014) Tech. rep..
[38] E.B. Wilson, Probable inference, the law of succession, and statistical inference, J. Am. Stat. Assoc. 22 (158) (1927) 209–212.
[39] R.G. Newcombe, Two-sided confidence intervals for the single proportion: comparison of seven methods, Stat. Med. 17 (8) (1998) 857–872.
[40] T. Strutz, Data fitting and uncertainty: A practical introduction to weighted least squares and beyond, Vieweg and Teubner, 2010.
[41] C.C. Clogg, E. Petkova, A. Haritou, Statistical methods for comparing regression coefficients between models, Am. J. Sociol. 100 (5) (1995) 1261–1293.
[42] R. Paternoster, R. Brame, P. Mazerolle, A. Piquero, Using the correct statistical test for the equality of regression coefficients, Criminology 36 (4) (1998) 859–866.
[43] Student, The probable error of a mean, Biometrika (1908) 1–25.
[44] J.L. Fleiss, B. Levin, M.C. Paik, Statistical Methods for Rates and Proportions, John Wiley & Sons, 2013.