This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2868457, IEEE Transactions on Smart Grid

Predictive Risk Management for Dynamic Tree Trimming Scheduling for Distribution Networks

Tatjana Dokic, *Student Member, IEEE*, and Mladen Kezunovic, *Life Fellow, IEEE*

***Abstract*-- This paper introduces a predictive method for distribution feeder vegetation management based on a risk framework. The state of risk is calculated for each feeder section using a variety of factors extracted from network parameters and historical outage data, historical weather data and weather forecasts, and a variety of vegetation indices. The framework implements the spatiotemporal correlation of all the collected data. The prediction model used is the Gaussian Conditional Random Field, which takes into account spatial interdependencies between different feeder sections. This enables better prediction accuracy, and also offers the capability to deal with missing and bad data. Based on the calculated risk, the dynamic optimal tree trimming schedule, which minimizes the overall risk for the system under a given predetermined budget, is developed. Results obtained on a real utility network show that optimal tree trimming based on the developed risk framework for vegetation management could significantly decrease the overall risk of the feeder outages without increasing the budget.**

***Index Terms*—Asset management, big data, data mining, geographic information system, meteorology, prediction methods, power distribution, risk analysis, smart grid, vegetation mapping.**

I. Introduction

The most common cause of outages in electric power systems is a combination of vegetation activity and severe weather impacts [1]. Thus, vegetation management is of the upmost importance for assuring high levels of network resilience. In addition, good vegetation management practices ensure safety for field workers and the public. Utilities spend millions of dollars on vegetation management every year [2], which makes it one of the highest costs in distribution asset management [3]. Every year several billions of dollars are spent on vegetation management in the U.S.A. [5]. Efficient automated vegetation management could significantly decrease the costs associated with tree trimming [4].

Efforts to automate vegetation management have employed multiple techniques in the last few decades. Work in [6] used a Markov model to find the optimal inspection frequency while finding a compromise between the reliability of the system and the cost of distribution feeder inspection. In [7] the optimal tree trimming schedule was developed based on a hybrid genetic algorithm consisting of simulated annealing, genetic algorithms, and tabu search. Vegetation-related failure rates were predicted using four different algorithms in [3]: linear regression, exponential regression, linear multivariable regression, and an artificial neural network. The developed predictors used historical outage data and some of the weather parameters, but vegetation indices were not considered. In the listed literature, when the weather impact was considered, only a few variables of interest were included and their impact was averaged over time. Two models, negative binomial generalized linear model and a Poisson generalized linear mixed model, were used in [8] to evaluate the impact of tree trimming on the rate of vegetation caused outages in distribution. The data used in this study were limited to the utility collected data, without insight into weather and vegetation indices. In [9],[10] satellite imagery was used to identify dangerous trees around the transmission lines. While the use of high resolution imagery did show the potential in transmission vegetation management, its use in distribution was not discussed. Work in [4], [11] developed a reliability-centered vegetation management while looking closely into the electrical characteristics of vegetation-related outages. The work in [12] demonstrated the potential of spatial correlation of big data for improvements in distribution vegetation management but did not provide related data analytics.

This work provides several contributions:

1) To improve risk predictions, a variety of data sources are used: the historical weather and weather forecast data, various vegetation indices and high resolution imagery data, and historical utility records about outages and maintenance. Their integration and correlation is novel.
2) A spatiotemporal model for correlating a variety of data in time and space is developed, which provides real-time generation of predictive risk maps for assessment of the vegetation around the distribution feeders.
3) Analytical approach is introduced for vegetation risk management based on a Gaussian Conditional Random Field (GCRF), which takes into account both the spatial and the temporal configuration of the network and past events to improve the prediction performances.
4) An optimized, cost-effective dynamic tree trimming scheduler is developed to minimize the overall risk of the network while maintaining the economic investment in periodic tree trimming. The unique benefits of this approach are demonstrated on an actual utility distribution network.

The background about vegetation management is provided

T. Dokic and M. Kezunovic are with Texas A&M University, College Station, TX, USA (e-mail: tatjana.djokic@tamu.edu, kezunov@ece.tamu.edu).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2868457, IEEE Transactions on Smart Grid

in Sec. II. Spatiotemporal correlation of big data is described in Sec. III. Sec. IV defines the vegetation risk management, while the optimal tree-trimming schedule is introduced in Sec. V. The results are presented in Sec. VI, and conclusions in Sec. VII.

II. Background

This section describes the mechanisms of weather and vegetation impacts on vegetation caused outages, and current vegetation management practices implemented by the utilities. As presented in Fig. 1, there are two major classes of vegetation-related feeder outages in distribution systems. They are differentiated by the tree coming in contact with feeders due to 1) overgrowing the feeder height, and 2) being forced into a contact with the feeder due to wind or some other similar weather impacts.

A. Vegetation Impact

Starting from the most recent tree trimming performed, the vegetation-caused failure probability is constantly increasing [6]. For predicting the potential of vegetation to cause faults subsequent to the last tree trimming, the most important factor is the vegetation canopy growth rate. There are two types of models for estimating the canopy growth dynamics [3]: 1) process-based models that aim at defining the processes that cause tree growth [13], and 2) empirical data-based models [14]. The maximum tree crown spread represents the maximum width of the tree crown (branches, leaves) along any axis. It is affected by the tree's age, last tree trimming date, application of herbicides or growth regulators, and weather impacts (primarily temperature and precipitation) [3].

The measured electrical behavior and physical processes and effects surrounding the vegetation-related faults were described in detail in [4], [11]. It was concluded from the experimental results that while the initial current during the tree contact can be quite low (~1A), after a complete carbonization path in the tree branch is formed, the current magnitude quickly increases to a high level.

B. Weather Impact

The weather parameters that can affect vegetation-related outages are wind speed, direction, and gusts, precipitation, temperature, humidity, pressure, and lightning, as listed in Fig. 1. The impact of high speed wind and heavy precipitation may cause trees to come into contact with distribution feeders due to the following reasons: a) branches break off and fly into lines, and b) complete trees topple when moved by wind [12]. The temperature, precipitation level, and humidity have impacts on the tree growth rate. In combination with the type of soil, they are the main factors dictating a tree's growth rate.

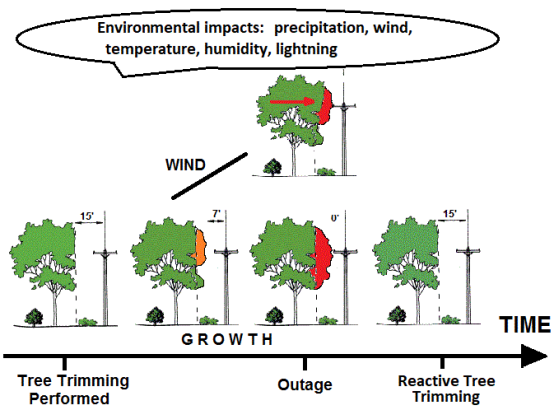

Fig. 1. Environmental impact on vegetation management [15]

C. Vegetation Maintenance

Vegetation maintenance staff are in charge of maintaining the feeder clearance to the surrounding vegetation. This includes trimming and the removal of trees around the distribution poles and lines. Distribution lines are often placed near the surrounding vegetation due to relaxed right-of-way requirements. Due to the high expenses of trimming large areas populated by many distribution feeders, it is not economical to have all trees securely trimmed at all times, so a more economical trimming schedule is used.

In most cases, the process of tree trimming is applied by utilities based on a predetermined periodic schedule. Each feeder section is given a tree trimming frequency, e.g. three or five years, based on the operating voltage and required clearance, leading to the standard fixed interval schedule [7]. The only other occasion when the schedule would be changed is as a reactive measure to a vegetation-caused outage, shown in Fig. 1. There are two types of reactive measures that can be distinguished: 1) only the faulted area is maintained, and 2) the entire tree trimming zone is trimmed. In addition to tree trimming, some utilities inject growth-retarding chemicals into trees (tree-growth regulators) or apply herbicides [7].

The current maintenance practice relies on a visual inspection by helicopters, airplanes, ground vehicles, or people walking up to the lines [16]. Because of the high cost of this practice, it is of economic benefit to develop visual inspection methods that can provide automatic identification of dangerous zones, as it will be described in Sec. III B.

III. Spatiotemporal Correlation of Data

This section describes the data processing that starts from the raw data and prepares the processed inputs for the predictive risk analysis and optimal tree trimming scheduler described in the next two sections respectively. All of the data has to be spatiotemporally correlated. All of the spatial processing of the data is done using ESRI ArcGIS [17]. Temporal data processing is done using Python [18] *datetime* library [19].

A. Data Preprocessing

Raw data are processed to remove unused components. All the data that has a geographical reference is placed into a geodatabase during the preprocessing. Table I lists all the extracted parameters needed for the prediction model, and the associated temporal and spatial references.

Data come with different spatial and temporal resolutions. Historical weather data from ASOS land stations [20] has the highest temporal resolution (up to 1 min); however, the spatial resolution of data is low, including only a few weather stations in the network service area. Vegetation data has a low temporal resolution (collected once per year or two years) but has a high spatial resolution (up to 50 cm). The rate of data collection varies not only between different data sets, but also it can vary

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2868457, IEEE Transactions on Smart Grid

TABLE I
Parameters Extracted in Preprocessing

	Historical Outage Data	Periodic Tree Trimming	Reactive Tree Trimming	Poles	Lines	Vegetation	Weather
Spatial	Point shapefile	Polyline shapefile	Polyline shapefile	Point shapefile	Polyline shapefile	• Raster • Polygon shapefiles	• Points • Polygon shapefiles
Temporal	Start and end time	Year quarter	Date	Static	Static	Year	1 min to 3 hours
Other parameters	• Num. of customers • Cause code	• Trim period • Num. of customers • Cost	• Cost	• Material/class • Height	• Conductor size • Conductor count • Conductor material • Nominal voltage	• Imagery • Vegetation classes	• Wind (speed, gust, direction) • Temperature • Precipitation • Humidity • Pressure • Forecast indices

within a single data set. For example, weather data is collected by land-based weather stations with a maximum rate of one data point per minute, but the rate can go down to one data point per hour. In some rare cases, the rate may go as low as one point within several hours. After preprocessing, the dataset is still not ready for the input into the predictive risk model. All the parameters need to be spatially and temporally correlated, as is described in Sec. III.C and III.D.

B. Vegetation Data Processing

Image data is used to extract the location of vegetation surrounding the network. The imagery is collected from the Texas Natural Resources Information System (TNRIS) database [21]. The following orthoimagery datasets are used in the study:

- National Agriculture Imagery Program 1m NC\CIR for years 2010, 2012, 2014, and 2016;
- Texas Orthoimagery Program 50cm NC\CIR for 2015.

The datasets are loaded into the geodatabase as raster files. First, to reduce the amount of data for processing, imagery raster files are clipped to a 20 m buffer around the distribution lines. Then unsupervised image classification [22] is applied. The iso-cluster is set to 40 classes in all datasets. In the next step the classes are reclassified to "vegetation class" and "non-vegetation class", and converted into a polygon shapefile. The vegetation class is transferred to the next step (spatial correlation), and the rest is discarded. In Fig. 2 we provide examples of the unsupervised classification (a), and the resulting map after reclassifying (c). Map (b) is in Fig. 2 for visual reference.

The result of image processing is a set of historical maps with vegetation locations. These maps are then spatially joined with the Ecological Mapping System of Texas (EMST) developed by the Texas Parks and Wildlife Department [23]. The EMST data contains the classification by vegetation type into 398 distinct classes, out of which 49 classes are present at the network location of interest. The average canopy height for 49 vegetation classes in the network area is then added to the vegetation dataset as a parameter.

C. Spatial Correlation of Data

The purpose of the spatial correlation module is to provide spatial links between different data sets. For example, for every historical outage we want to know the weather conditions at that specific location, the distance between the line and the closest tree, the location of areas that were trimmed, etc.

The spatial correlation module is presented in Fig. 3. We distinguish three parts of the spatial processing module:

- *Weather data processing* encompasses creating the weather data grid that is overlaid on the utility network and has a spatial resolution of 1 km. The weather parameters in each grid cell are calculated from the weather station values using linear interpolation.
- *Vegetation data processing* extracts the vegetation indices, such as distance between the lines and vegetation and growth rate, using spatial links between multiple preprocessed vegetation files. All the calculated parameters are stored as attributes in the final vegetation polyline shapefile.

a) b) c)

Fig. 2. Example of vegetation extraction: a) 40 classes, b) imagery for reference, and c) reclassified (vegetation highlighted)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2868457, IEEE Transactions on Smart Grid

- *Utility data processing* converts the historical tables to the shapefiles identifying the locations of points and polylines based on the line section codes and/or addresses provided in the utility's CSV files. In addition, every reactive tree trimming action is correlated with the outage that lead to it.

To deal with different spatial resolutions of data we used multiple approaches all included in Fig. 3. We used spatial interpolation where weather data was extracted for every location in the network based on the original weather station data with sparse locations. In other instances, data was projected to a nearby location using a spatial join. For example, the distance between the line and vegetation is projected to the line using a spatial join based on distance.

D. Temporal Correlation of Data

The temporal correlation module has five historical input datasets (weather, vegetation, outage, periodic tree trimming, and reactive tree trimming), and real-time weather forecast input. Each dataset contains a variety of parameters (attributes) from Table I, and is stored as a GIS shapefile. Static datasets (network feeders and poles) are assumed not to change over the observed period, and do not require any temporal correlation. Fig. 4 presents an overview of the temporal correlation module

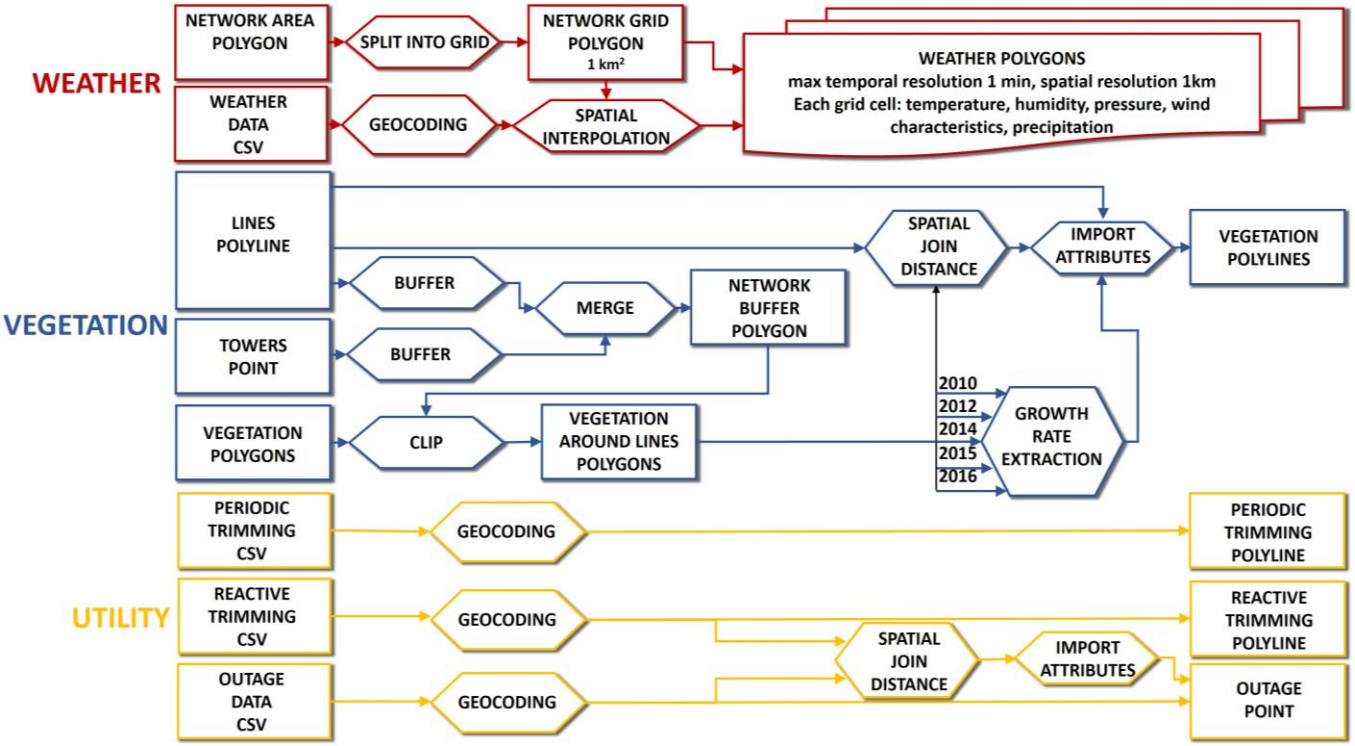

Fig. 3. Spatial correlation of data

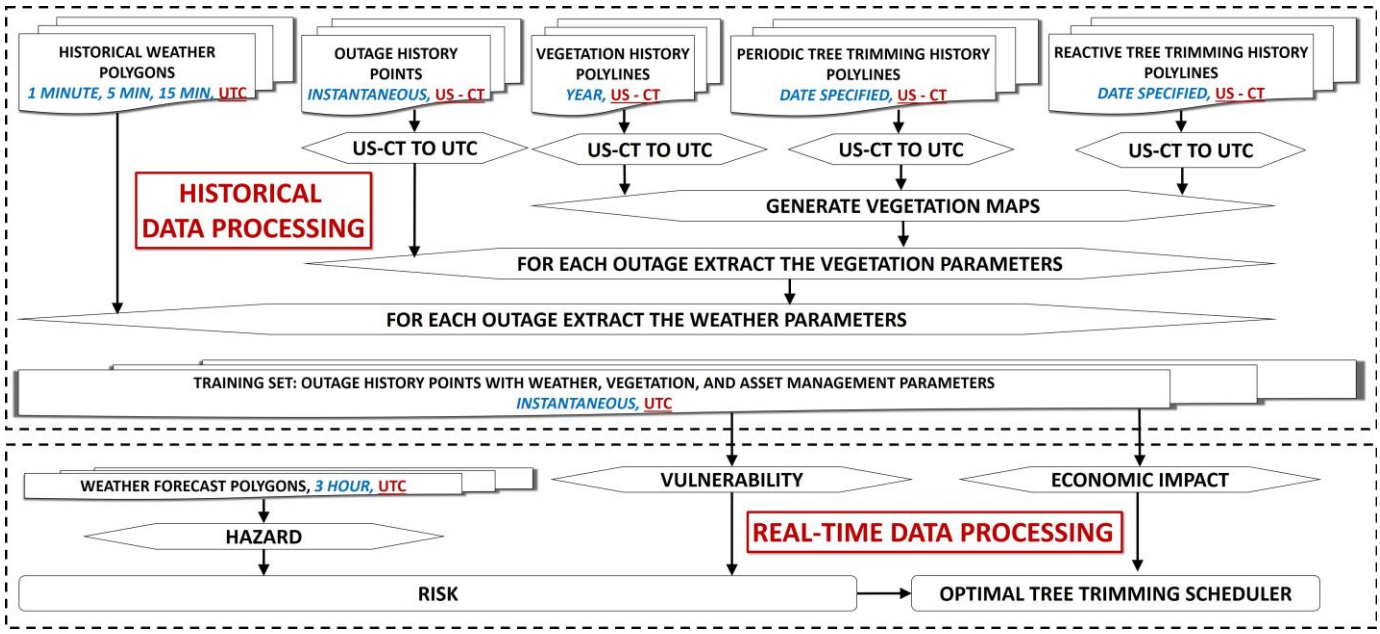

Fig. 4. Temporal correlation of data

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2868457, IEEE Transactions on Smart Grid

containing two major parts: 1) historical data processing, and 2) real-time data processing. The final product of historical data processing is a training list for the prediction algorithm. The real-time data processing generates input data for the real-time risk maps by generating the data for hazard, vulnerability, and economic impact that feeds the dynamic tree trimming scheduler, which will be described in the following sections.

The temporal resolution is guided by the occurrence of outages. For every outage we want to extract, all the relevant information is included as presented in Fig. 4. For each outage, the data points that are closest in time are chosen from each set individually. For example, in the case of historical weather data, the closest data points were within one minute of outage. On the other hand, the closest vegetation maps could be up to several weeks apart.

IV. Vegetation Risk Management

Fig. 5 presents an overview of the predictive spatiotemporal risk model. For every moment of time, each network component is assigned a state of risk value. To enable spatiotemporal analysis, the state of risk R is defined as follows [24]:

$$R(G,t) = P[T(G,t)] \cdot P[C(G,t)|T(G,t)] \quad (1)$$

where G represents the longitude and latitude of a single element, and t represents the moment in time for which the observation is made. A unique state of risk value is assigned to each distribution feeder section. $T(G,t)$ represents the threat intensity. Threat intensity is defined as a qualitative metric of the weather condition severity. The first term in (1) marked $P[T(G,t)]$ is a hazard probability. Hazard represents the probability of occurrence of a severe weather condition with the selected threat intensity. The details on how the Hazard is calculated are provided in Sec. IV A. The second term marked $P[C(G,t)|T(G,t)]$ is network vulnerability, where $C(G,t)$ is an occurrence of a consequence. Vulnerability is a conditional probability of the consequence (vegetation-caused outage) in the distribution network if and when severe weather is present. The details on how the Vulnerability is calculated are provided in Sec. IV B. The risk definition presented here is an adaptation of definition in [25] where the last part of the risk-economic impact is not included. In this paper, the economic impact is calculated separately and included in the optimal tree trimming scheduler as one of the optimization constrains. The details of how the economic impact is combined with the risk framework are described in Sec. V.

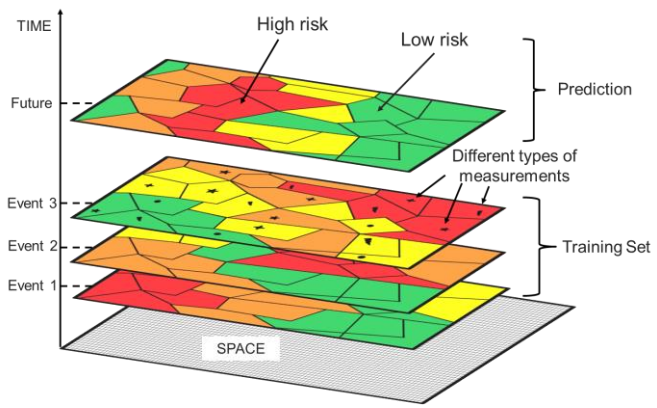

Fig. 5. Spatiotemporal Prediction Model, [15]

A. Hazard

In eq. 1, $P[T(G,t)]$ is a hazard, calculated based on the weather forecast data for a specific time and location. The data from the National Digital Forecast Database (NDFD) [26] is used. The database contains the forecast up to 7 days in the future with time resolution of 3 hours. The updated forecast is provided every 3 hours. The spatial resolution of the weather forecast data is 5 km. Because the weather forecast data is updated every 3 hours with maximum resolution of 3 hours, the risk maps are generated with the same 3 hours resolution.

The following parameters are observed: wind speed, direction, and gust, temperature, relative humidity, convective hazard outlook, probability of critical fire, probability of dry lightning, hail probability, tornado probability, probability of severe thunderstorms, damaging thunderstorm wind probability, extreme hail probability. Based on the values of the observed parameters, the threat level is classified into 6 groups from 0 to 5, where 0 represents normal weather conditions without any potentially severe elements, and 5 represents extremely severe weather conditions. The k-means clustering [27] was used for classification into 6 groups. The k-means clustering enables the construction of hazard consequence levels from the individual weather parameters. This way, multiple different parameters are combined into a single parameter *Threat Intensity* with 6 different states. The clustering is done using historical weather data, where different configurations of weather parameters are associated with their measured impact on the outage occurrence. Then the Hazard is constructed as a heat map in Table II, where each threat level has an assigned probability of occurrence determined based on weather forecast. The construction of heat map is based on the reference [28], where heat maps are constructed following two steps: 1) constructing the probability matrix as in Table III, and 2) constructing the threat intensity matrix as in Table IV. The Hazard value ranges from extremely low marked as the green color in Table II to extremely high marked as the red color in Table II.

B. Vulnerability

A GCRF is used for the prediction of network vulnerability [29]. The GCRF model uses a weighted graph as a data structure, which enables the exploitation of spatial similarities between the nodes for the improved prediction capability. The data are processed in sequential order created during the temporal correlation of data. The algorithm is capable of processing partially observed data [30], which is of benefit since within the collected data, several historical outage instances are missing some of the weather parameters.

The GCRF predicts the state of vegetation impact, denoted y, based on historical measurements in the input vector x. The GCRF expresses the conditional distribution as:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2868457, IEEE Transactions on Smart Grid

TABLE II
Hazard Classification

Probability [%]	Threat Intensity					
	0	1	2	3	4	5
0-20						
20-40						
40-60						
60-80						
80-100						

TABLE III
Probability of Threat Level Occurrence

Probability Range [%]	Description
0-20	Extremely Unlikely
20-40	Highly Unlikely
40-60	Doubtful
60-80	Somewhat Likely
80-100	Very Likely

TABLE IV
Threat Intensity Levels

Category	Description	Example
0	None	No impact on the network
1	Minor	Minor service interruptions, no restauration needed
2	Moderate	Some outages in the network, some restauration needed
3	Low Severe	Moderate number of outages in the network, restauration delays may occur, e.g. rainy weather
4	High Severe	Multiple outages in the network with longer restauration duration, e.g. thunderstorm
5	Catastrophic	The whole network or very large parts of the network under the disconnected – large blackouts, e.g. Hurricane

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z}\exp\left(-\sum_{i=1}^{N}\sum_{k=1}^{K}\alpha_k\left(y_i - R_k(\mathbf{x})\right)^2 - \right.$$
$$\left. -\sum_{i,j}\sum_{l=1}^{L}\beta_l e_{ij}^{(l)} S_{ij}^{(l)}(\mathbf{x})\left(y_i - y_j\right)^2\right) \quad (2)$$

where Z is a normalization constant, x is a set of input variables coming from historical measurements, y is a set of output variables, N is a total number of nodes (line sections) in the network graph, R_k are unstructured models where k is the number of predictors, S_{ij} represent similarities between outputs at nodes i and j determined based on their geographical distance, L is a number of branches, α are parameters of the association, and β are the interaction potentials.

The following historical measurements are stored in the input vector x: wind speed, wind direction, wind gust, precipitation, temperature, humidity, pressure, vegetation distance to the line section, vegetation spread, vegetation growth rate, vegetation health index, pole height, tree trimming period, time since last tree trimming, outage duration, number of customers affected. The output y of the algorithm is the predicted state of vegetation impact on the feeder section.

The parameters α and β from the Eq. 2 can be estimated by maximizing the conditional log-likelihood from our training set, (3) and (4), and applying the gradient descent optimization algorithm:

$$L(\alpha,\beta) = \sum \log P(\mathbf{y}|\mathbf{x}) \quad (3)$$

$$(\boldsymbol{\alpha},\boldsymbol{\beta}) = \arg\max_{\alpha,\beta}\left(L(\boldsymbol{\alpha},\boldsymbol{\beta})\right) \quad (4)$$

The historical outages are an integral part of the Vulnerability. The prediction of future vulnerability is done based on the knowledge collected from the previous outages. As listed in Table I, the historical outage data contains information about the duration of the outage and the number of customers affected by it. This information guides the prediction algorithm to generate higher vulnerability levels in the cases where more customers were affected by the outage and for the greater duration.

V. OPTIMAL TREE TRIMMING SCHEDULER

There are two types of costs associated with the tree trimming:

- Periodic tree trimming has a preset cost since it follows a predetermined schedule.
- Reactive tree trimming includes two types of actions: a) only the faulted area is trimmed, and b) an entire circuit is trimmed. Reactive tree trimming cost varies depending on the events in the network.

The goal of the optimization model is to minimize the overall risk of the system while maintaining the budget allocated for the periodic tree trimming. To achieve that, the quarterly periodic tree trimming schedule is designed based on the risk prediction for the next 3 months. The time instances when the risk map is created are every three hours during a three-month period. A total of T time instances is created each quarter. The risk is calculated for each of the N feeder sections. An optimized tree trimming schedule is determined by solving the following optimization problem:

$$max\left\{R = \sum_{t=1}^{T}\frac{1}{N}\sum_{n=1}^{N}\Delta R_{n,t}\cdot F_{n,t}\right\} \quad (5)$$

$$F_{n,t} = \begin{cases}0, & \text{section } n \text{ not trimmed at time } t\\ 1, & \text{section } n \text{ is trimmed at time } t\end{cases}$$

where $\Delta R_{n,\theta} = R_{n,(\theta-1)} - R_{n,\theta}$ is the difference in risk value for feeder n before and after the tree trimming is performed. The following constraints are enforced:

$$\sum_{t=1}^{T}\sum_{n=1}^{N} F_{n,t}\cdot PC_{n,t} \le PA \quad (6)$$

$$\text{For } t=1,\ldots,T,\ \sum_{n=1}^{N} F_{n,t} \le 1 \quad (7)$$

where R is a total reduction in risk, $PC_{n,t}$ is the cost of tree trimming of section n in the time instance t; and PA is a total budget allocated for the periodic tree trimming during the observed quarter. The optimization problem is nonlinear, and it is solved using the enhanced linear programming relaxation with the Lagrangean relaxation plus heuristic method described in detail in [31].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2868457, IEEE Transactions on Smart Grid

While a reduction in reactive tree trimming cost is not an explicit goal of the optimization problem, it is still calculated to check the impact of risk reduction on the reduction in reactive tree trimming cost. To do that, the reactive tree trimming orders are iterated and for each one it was checked if the developed tree trimming scheduler recommended trimming of the area prior to the outage. If an area is part of the recommended tree trimming schedule in a time frame before the reactive tree trimming was performed, the reactive tree trimming cost is deducted from the total.

VI. Results

The observed utility distribution network has an area of ~2,000 km^2. It contains ~200,000 poles, and ~120,000 lines. The historical outage and weather data were collected for the period from January of 2011 up to the end of April of 2016. Over this period, 505 weather-related outages have been observed in the area, where a total of 331 outages were vegetation-caused (Fig 6). The training set for a prediction algorithm consists of the first 300 historical outages in temporal order. The remaining 31 outages that occurred at the end of 2015 and beginning of 2016 are used as testing set.

A. Risk Maps

The example of the predicted Hazard and Vulnerability map for an outage event that occurred on February 23, 2016 is presented in Fig. 7 and Fig. 8 respectively. The weather hazard is presented as a grid covering the area of the network, while the vulnerability is assigned to each line section individually. The resulting predicted risk map for the observed date is presented in Fig. 9. As it can be seen in the upper right corner the predicted risk value on the faulted section for the outage in the Fig. 9 that occurred on 02/23/2016 was 84%.

The predicted risk values for all 31 test outages are presented in Fig. 10. The minimum risk value during an outage is 64%. There are 6 instances for which the risk probability was less than 75%, all of which occur during the days with a low weather hazard. The authors would like to speculate that in the absence of weather hazard information, when the algorithm is limited to predicting the risk based only on vegetation indices, performance is limited. Further investigation could be conducted with the larger dataset to test the hypothesis.

B. Tree Trimming Scheduler

An example of the developed tree trimming schedule for one quarter is presented in Fig. 11. The zones with different colors (not black) represent the areas of the network that need to be trimmed in the selected quarter. These zones change every quarter. The areas that need to be trimmed sooner are represented with red while the areas that need to be trimmed later are represented with green color.

Overall outage risk for the selected quarter is calculated as follows:

$$R = \sum_{t=1}^{T} \frac{1}{N} \sum_{n=1}^{N} R_{n,t} \qquad (8)$$

The optimal tree trimming schedule reduced the overall outage risk of the network for the period of three months by 32.85%. In addition, the reactive tree trimming total cost described in Sec. V was predicted to be decreased by 27.2%.

VII. Conclusions

The presented research differentiates itself by the use of an extensive set of data. We correlated different datasets and developed a predictive risk model that utilizes spatiotemporal data to produce real-time risk maps for the distribution network. The prediction algorithm, based on a GCRF model, leverages the spatial similarities between different feeder sections to ensure better prediction performance and compensate for missing data. The resulting risk model allows the implementation of a dynamically changing trimming scheduler that optimizes the tree trimming process. It is shown that the

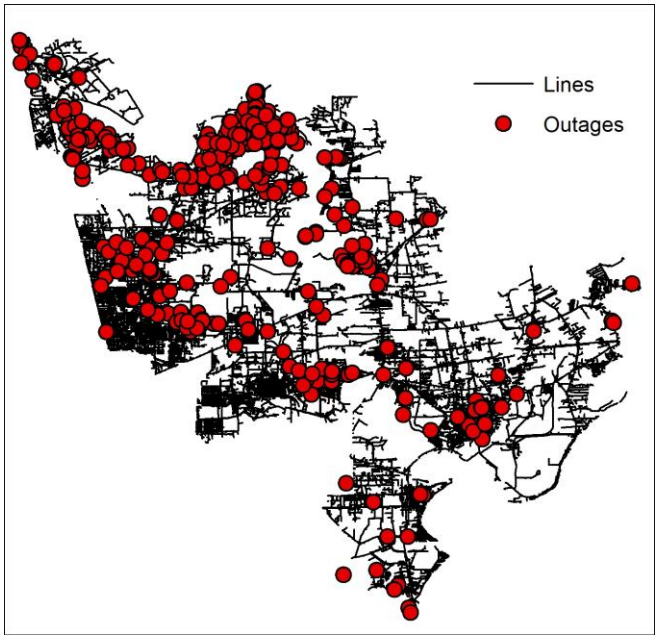

Fig. 6. Distribution of historical vegetation caused outages

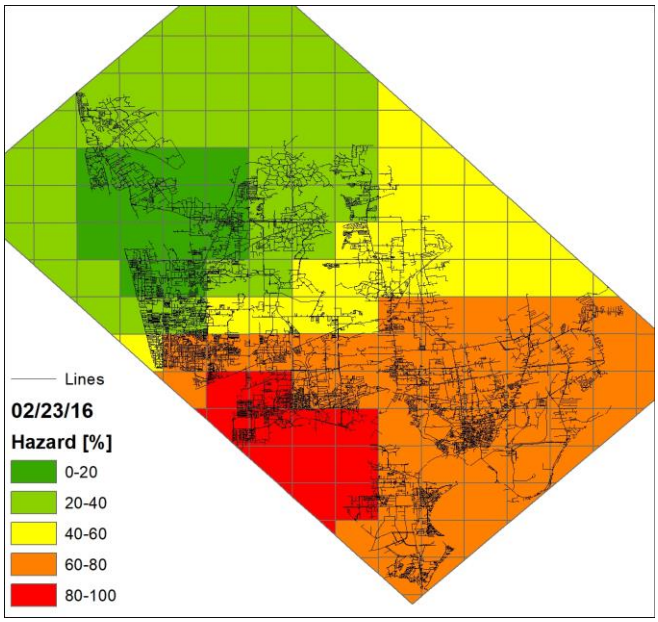

Fig 7. Hazard Map for 02/23/2016

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2018.2868457, IEEE Transactions on Smart Grid

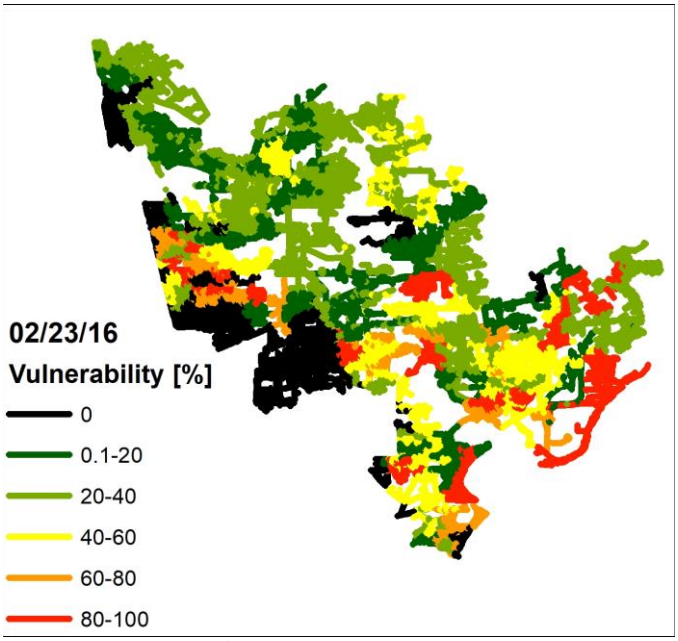

Fig 8. Vulnerability Map for 02/23/2016

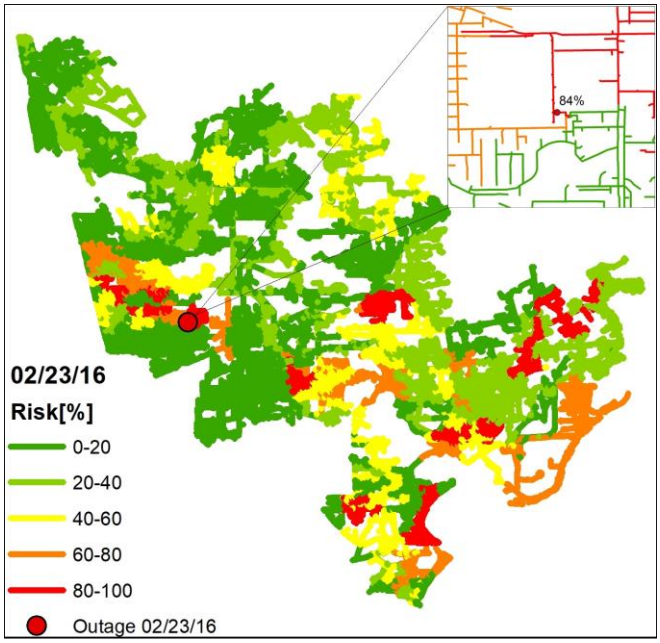

Fig. 9. Risk Map for 02/23/2016

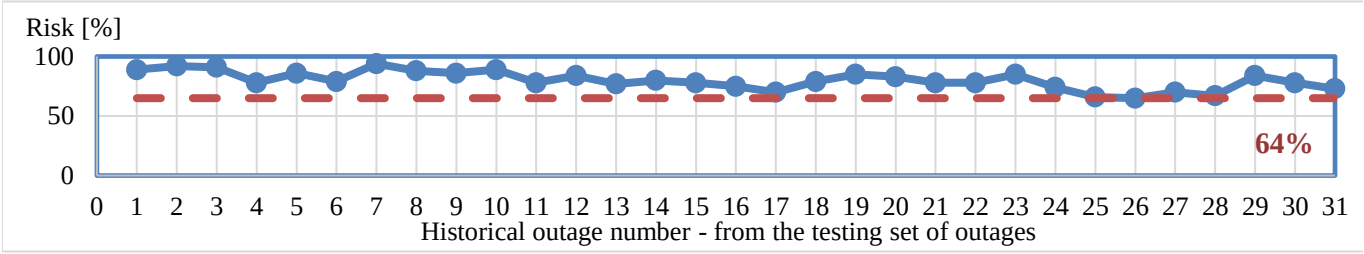

Fig. 10. Calculated risk at the end of training for the outages that occurred at the end of 2015 and beginning of 2016.

achieved reduction in risk has the potential of reducing the cost of reactive tree trimming. The method was applied to a real distribution network and utility data. The testing confirms that the outages occurred in the zones with risk predicted to be greater than 64%, which suggests a new predictive paradigm for vegetation management strategies.

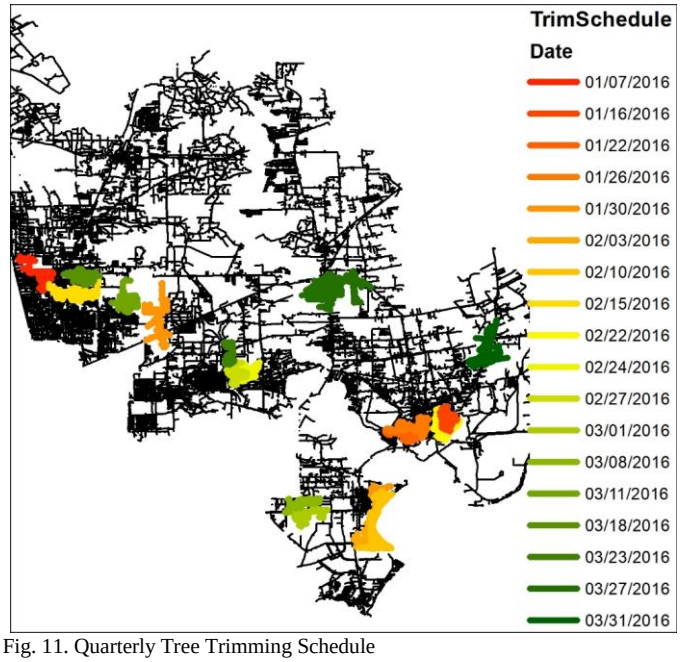

Fig. 11. Quarterly Tree Trimming Schedule

VIII. Acknowledgment

The authors would like to acknowledge Mr. P-C. Chen with Texas A&M University for his comments on hazard model development and data sources. The authors would also like to thank Messrs. R. Greenwood, C. Anderson, G. Sonde, J. George, with CenterPoint Energy, Houston, for sharing their knowledge and comments on the topic of distribution system vegetation management. Special thanks are due to Dr. Obradovic and his Team at Temple University in Philadelphia for introducing us to the GCRF model.

IX. References

[1] Eaton, "Blackout Tracker United States Annual Report 2014," [Online] Available: http://electricalsector.eaton.com/forms/BlackoutTracker AnnualReport

[2] Carroll Electric Cooperative, "Cost Analysis for Integrated Vegetation Management Plan," 20 April 2010. [Online] Available: http://www.carrollecc.com/files/pdf/cecc_finley_cost_study.pdf

[3] 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 Transactions on Power Delivery, Vol. 17, No. 4, pp. 1170-1174, October 2002.

[4] B. D. Russell, C. L. Benner, J. Wischkaemper, W. Jewell, J. McCalley, "Reliability Based Vegetation Management Through Intelligent System monitoring," PSerc Report 07-31, September 2007, [Online] Available: https://pserc.wisc.edu/documents/publications/reports/2007_reports/T-27_Final-Report_Sept-2007.pdf.

[5] Federal Energy Regulatory Commission United States Government, "Utility Vegetation Management – Final Report," Commissioned to

Support the Federal Investigation of the August 14, 2003 Northeast Blackout, March 2004.
[6] P. A. Kuntz, R. D. Christie, S. S. Venkata, "A Reliability Centered Optimal Visual Inspection Model for Distribution Feeders," IEEE Transactions on Power Delivery, Vol. 16, No. 4, pp. 718-723, October 2001.
[7] P. A. Kuntz, R. D. Christie, S. S. Venkata, "Optimal Vegetation Maintenance Scheduling of Overhead Electric Power Distribution Systems," IEEE Transactions on Power Delivery, Vol. 17, No. 4, pp. 1164-1169, October 2002.
[8] S. Guikema, R. A. Davidson, H. Liu, "Statistical Models of the Effects of Tree Trimming on Power System Outages," IEEE Transactions on Power Delivery, Vol. 21, No. 3, pp. 1549-1557, July 2006.
[9] Y. Kobayashi, G. G. Karady, G. T. Heydt, R. G. Olsen, "The Utilization of Satellite Images to Identify Trees Endangering Transmission Lines," IEEE Transactions on Power Delivery, Vol. 24, No. 3, pp. 1703-1709, July 2009.
[10] Y. Kobayashi, G. G. Karady, G. T. Heydt, M. Moeller, R. G. Olsen, "Satellite Imagery for the Identification of Interference with Overhead Power Lines," PSerc Report 08-02, January 2008, [Online] Available: https://pserc.wisc.edu/documents/publications/reports/2008_reports/T-28_Final-Report_Jan-2008.pdf.
[11] J. A. Wischkaemper, C. L. Benner, B. D. Russell, "Electrical Characterization of Vegetation Contacts with Distribution Conductors – Investigation of Progressive Fault Behavior," Transmission and Distribution Conference and Exposition, Chicago, IL, April 2008.
[12] P.-C. Chen, T. Dokic, N. Stokes, D. W. Goldberg, M. Kezunovic, "Predicting Weather-Associated Impacts in Outage Management Utilizing the GIS Framework," IEEE/PES Innovative Smart Grid Technologies Latin America (ISGT-LA), Montevideo, Uruguay, October 2015.
[13] R. Sievanen and T. E. Burk, "Adjusting a process-base growth model for varying site conditions through parameter estimation," Can. J. Forest Res., vol. 23, no. 9, pp. 1837–1851, Sept. 1993.
[14] D. C. Hamlin and R. A. Leary, "An integro—Differential equation model of tree height growth," in Proc. IUFRO Conf., vol. 2, Minneapolis, MN, Aug. 1987, pp. 683–690.
[15] T. Dokic, M. Kezunovic, "Real-Time Weather Hazard Assessment for Power System Emergency Risk Management," CIGRE Grid of the Future, Cleveland, OH, October 2017.
[16] J. Wingfield, "New York Power Authority Develops Vegetation Management Solution for High-Voltage Transmission Lines with GIS," ArcNews, Vol. 27, No. 1, pp. 35-, ESRI Spring 2005.
[17] ArcGIS, Esri. [Online] Available: https://www.arcgis.com
[18] Python, [Online] Available: https://www.python.org/
[19] Python, "Basic date and time types," [Online] Available: https://docs.python.org/2/library/datetime.html
[20] National Centers for Environmental Information, "Automated Surface Observing System (ASOS)," [Online] Available: https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/automated-surface-observing-system-asos
[21] TNRIS, "Maps & Data," [Online] Available: www.tnris.org/get-data
[22] ArcGIS, "Iso Cluster Unsupervised Classification," [Online] Available: http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/iso-cluster-unsupervised-classification.htm
[23] Elliott, Lee F., David D. Diamond, C. Diane True, Clayton F. Blodgett, Dyan Pursell, Duane German, and Amie Treuer-Kuehn. 2014. Ecological Mapping Systems of Texas: Summary Report. Texas Parks & Wildlife Department, Austin, Texas.
[24] M. Kezunovic, Z. Obradovic, T. Dokic, B. Zhang, J. Stojanovic, P. Dehghanian, and P. -C. Chen, "Predicating Spatiotemporal Impacts of Weather on Power Systems using Big Data Science," Springer Verlag, Data Science and Big Data: An Environment of Computational Intelligence, Pedrycz, Witold, Chen, Shyi-Ming (Eds.), ISBN 978-3-319-53474-9, 2017.
[25] Z. Medina-Cetina and F. Nadim F, "Stochastic Design of an Early Warning System", Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, vol. 2, no. 4, pp. 223 – 236, 2008.
[26] National Digital Forecast Database (NDFD) Tkdegrib and GRIB2 DataDownload and ImgGen Tool Tutorial, NWS, NOAA. [Online] Available: http://www.nws.noaa.gov/ndfd /gis/ndfd_tutorial.pdf
[27] W. Tian, Y. Zheng, R. Yang, S. Ji, and J. Wang, "A Survey on Clustering based Meteorological Data Mining," International Journal of Grid Distribution Computing, Vol. 7, No. 6, pp. 229-240, 2014.
[28] B. M. Ayyub, "Risk Analysis in Engineering and Economics," CRC Press, 2003, ISBN: 0203497694, 9780203497692, pp. 57-60.
[29] V. Radosavljevic, S. Vucetic, Z. Obradovic, "Neural Gaussian Conditional Random Fields," Proc. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Nancy, France, September, 2014.
[30] J. Stojanovic, M. Jovanovic, Dj. Gligorijevic, Z. Obradovic, "Semi-supervised learning for structured regression on partially observed attributed graphs" Proceedings of the 2015 SIAM International Conference on Data Mining (SDM 2015) Vancouver, Canada, April 30 - May 02, 2015.
[31] W. Jewell, J. Warner, J. McCalley, Y. Li, S. R. K. Yeddanapudi, "Risk-Based Resource Allocation for Distribution System Maintenance," PSerc Report 06-26, August 2006, [Online] Available: https://pserc.wisc.edu/documents/publications/reports/2006_reports/T-24_Final-Report_Aug-2006.pdf.

X. BIOGRAPHIES

Tatjana Dokic (S'10) received the B.Sc. and M.Sc. degrees in electrical and computer engineering from the University of Novi Sad, Novi Sad, Serbia, in 2012. Currently, she is a Graduate Student with Texas A&M University, College Station, TX, USA. Her main research interests are power system asset and outage management, weather impacts on power systems, big data for power system applications, vegetation management, insulation coordination, fault location.

Mladen Kezunovic (S'77-M'80–SM'85–F'99, LF'17), received the Dipl. Ing., M.S., and Ph.D. degrees in electrical engineering in 1974, 1977, and 1980, respectively. He has been with Texas A&M University for 31 years.

Currently, he is Regents Professor and Eugene E. Webb Professor, Director of the Smart Grid Center, and Site Director of "Power Engineering Research Center, PSerc" consortium. His expertise is in Protective Relaying, Automated Power System Disturbance Analysis, Computational Intelligence, Data Analytics, and Smart Grids. He has published over 550 papers, given over 120 seminars, invited lectures and short courses, and consulted for over 50 companies worldwide. He is the Principal of XpertPower™ Associates, a consulting firm specializing in power systems data analytics.

Dr. Kezunovic is an IEEE Life Fellow, a CIGRE Fellow and Honorary Member, and a Registered Professional Engineer in Texas.