Improved Transmission Line Fault Location Using Automated Correlation of Big Data from Lightning Strikes and Fault-induced Traveling Waves

Mladen Kezunovic	Tatjana Dokic	Po-Chen Chen	Vuk Malbasa
Texas A&M University	Texas A&M University	Texas A&M University	University of Novi Sad
kezunov@ece.tamu.edu	tatjana.djokic@tamu.edu	pchen01@tamu.edu	vmalbasa@gmail.com

Abstract

This paper investigates how correlating cross-domain big data from the lightning surge and the traveling wave measurements in time and space can be used to improve fault location accuracy. The integration and correlation of big data in time and space using Global Positioning System and Geographic Information System respectively improves knowledge about faults on transmission lines caused by lightning. The benefits of proposed method are: a) the decision process can be accelerated through automation, and b) better accuracy of fault location result can be provided due to the data correlation. The benefit is a more efficient outage management procedure.

1. Introduction

Since electricity has been interwoven with humans' daily lives, the reliability of the supply has become increasingly critical in many aspects. Although the utilities are investing paramount efforts to improve the reliability of the operation, severe weather conditions continue to impact electrical networks, particularly since the transmission grid is still largely an overhead infrastructure and hence highly exposed. Storm-related outages are estimated to have cost between $20 billion and $55 billion annually, in cases where the damage to the grid provoked serious consequences to the customers [1]. Providing fast system restoration is essential for reducing this loss. The time needed for system restoration can be significantly reduced by providing more accurate estimate of fault location in real-time (as soon as the fault is detected).

Severe weather conditions are one of the major reasons for electrical outages. In fact, the number of weather-related outages has been increasing annually [2, 3]. According to [2], from 1992 to 2011, 78% of 1333 outages in USA were weather-related where 178 million customers were affected. A study [3] demonstrates that in the US approximately 44% of the blackouts were weather-related during the period of 1984 to 2006, where 11% of these blackouts were caused by lightning activity.

Weather can have different impacts on utility operations [4, 5]. The weather impacts may be caused by variety of circumstances [6]. Transmission systems operators need to identify approximate locations of outages so that crews can be efficiently dispatched [7]. While the line outage is in effect, and crews are inspecting and repairing the damage, the operator may implement adequate switching actions to reduce the impact of the outage on the overall supply [8].

Traveling wave fault location has been explored in literature [9-16] and claimed to be extremely accurate, which requires data sampling in the kilohertz range. Many utilities have either deployed such fault locators or are in the process of evaluating them. The GPS synchronization between two traveling wave recorders on two sides of the transmission line was discussed in [14-16]. However, none of the papers examine how additional data coming from the lightning detection network can be automatically used to improve accuracy of the method, which is done in this paper.

In [17], the real time monitoring of transmission line transients under lightning strikes was presented. Real time electromagnetic transients were measured and correlated with lightning data recorded at the outage location to evaluate the impact on insulation coordination. Such measurements are very intensive exhibiting sampling rates of several megahertz.

This paper demonstrates how the use of traveling wave and lightning surge measurements, correlated with data from Geographic Information System (GIS) and Global Positioning System (GPS) may bring major improvements in the outage management. The utilities use such additional data types today, but data is processed manually and poorly correlated leading to delayed decisions and inaccuracies. The automated method has to address the big data problem due to heterogeneity of the data sets, as well as the high volume and velocity of data.

The paper is organized as follows. Section 2 provides background for the topics of Big Data and individual data sets used in fault location application. In Section 3, the automated improved transmission line fault location is described. Results are presented and discussed in Section 4, while Section 5 lists the conclusions. References are provided at the end.

2. Background

Quick and accurate decision-making is the most essential part of outage management and service restoration. The problem space involves information that is naturally associated with the temporal and spatial scales as well as with other power system properties such as event (contingency) type and time/location of occurrence, power system loading patterns, network switching alternatives, etc. The big data approach can be used to improve the decision-making process in this scenario. The challenge lies in leveraging the large amount of data in real-time to improve the decision-making process, without adding to the complexity. The task is therefore multi-dimensional, multi-scale, and requires cross-domain integration and correlation of data. In this paper the term correlation is used describe spatio-temporal cross-referencing of two data sets (traveling wave recorder data and lightning data), where spatial and temporal referencing is done using GIS and GPS respectively.

2.1. Big data

As denoted in [18], the big data in the electric power industry exhibits following characteristics:

- Large volume,
- High velocity,
- Increasing variety.

The volume of data refers to the quantity of generated and transmitted data described in terms of gigabyte or even terabyte. Velocity refers to the temporal constraints on collecting, processing and analyzing data and it is described in terms of number of samples per second or frequency of data generation and transmission and subsequent recording. Variety refers to data coming from many different sources that are not necessarily part of the traditional electric utility data.

The electric utility measurement data is typically obtained from various types of Intelligent Electronic Devices (IEDs) such as the phasor measurement unit (PMU), automated smart meter (ASM), digital fault recorder (DFR), digital protective relay (DPR), and Supervisory Control and Data Acquisition (SCADA) system. Utility measurements collected from substation IEDs alone can compound to dozens of GB of data per day in large systems due to each device reaching the sampling in the range of kilohertz or even megahertz [19]. The additional sources of big data include the weather reporting system, National Lightning Detection Network, GIS, GPS, and others [20].

Since data is gathered from a variety of sources that follow different standards for data collection and description the database is heterogeneous. Hence, additional steps need to be taken in order to extract the useful information in the format that is compatible with the network model used for simulation as well as specific data analytics used for outage management. Automating this process is of a paramount importance if the decisions are to be made efficiently and accurately. To cope with data variety and the necessity for spatial and temporal correlation of cross-domain sources, operators are also expecting to benefit from visual analytics which can offer an interactive view of multiple data layers.

The big data processing methodology consists of following steps [21]:

- Searching: Implementing advance search process for fast access to data of interest,
- Learning: Identifying important information using machine learning techniques such as classification and clustering.
- Knowledge Extraction: Extracting knowledge from information using different machine learning techniques,
- Correlation: Correlating different pieces of knowledge to form further conclusions using reasoning,
- Prediction: Identifying rules and trends in analyzed data that can be used to predict future behavior,
- Optimization: Based on collected knowledge optimize existing solutions in order to minimize risk and maximize profit.

2.1.1. Lightning data. The faults are usually caused by cloud-to-ground lighting hitting the poles. Depending on the area, the lightning may be very important in influencing electric power network faults. For instance, the research in UK [22] shows that lightning strikes are the second most common factor in weather-related distribution system faults. Due to predicted changes in operating conditions caused by weather and the change of power system infrastructure the percentage of faults induced by lightning is estimated to increase by 40% by 2080s [23]. In this case to minimize the effects of lightning proper protection of network structure (i.e. ground wires) and equipment (i.e. surge protectors) must be implemented by utilities [24].

The lightning detection network data can be used to correlate information about lightning characteristics with other event data gathered from the substation measurements. This provides better situational awareness during the critical events affecting the system and has the potential to improve automated fault location techniques.

Lightning data is gathered by the sensors that are typically located sparsely over the area of interest. There are three common types of lightning sensors:

- Ground-based systems that use multiple antennas to determine distance to the lightning by performing triangulation.
- Mobile systems that use direction and a sensing antenna to calculate distance to the lightning by analyzing surge signal frequency and attenuation.
- Space-based systems installed on artificial satellites that use direct observation to locate the faults.

Typical detection efficiency for a ground-based system is 70-90%, with a accuracy of location within 0.7-1 km, while space-based systems have resolution of 5 to 10 km, [25].

For example, The National Lightning Detection Network (NLDN) [26] uses ground-based system to detect lightning strikes across the United States. After detection data received from sensors in raw form is transmitted via satellite-based communication to the Network Control Center operated by Vaisala Inc. [27].

When it comes to the way data is received by the utility we can distinguish two cases: (i) the lightning sensors are property of the utility, and (ii) lightning data is received from external source. In the first case raw data are received from the sensors, while in second case external sources provide information in the format that is specific to the organization involved. No matter which source is used the lightning data typically includes the following information: a GPS time stamp, latitude and longitude of the strike, peak current, lightning strike polarity, and type of lightning strike (cloud-to-cloud or cloud-to-ground).

2.1.2. Traveling wave fault locator data. Traveling wave recorder data is considered as one of the sources of information for proposed method. Traveling wave fault locator calculates distance to fault automatically based on recorded samples of traveling waves at one or both sides of the line. Mostly used method in modern devices is double ended Type D method with GPS synchronization. The locator calculates arrival time of the fault-induced waves using GPS as a reference. Then, these time tags are sent to the central station where fault location algorithm is used to determine distance to the fault from line terminals. In addition, samples of the recorded signal are transmitted.

The accuracy of traveling wave method is highly dependent on the sampling rate. Modern devices use sampling frequency of 0.1 to 20 MHz.

In case of Type D traveling wave method, GPS is primarily used for synchronization between signals received at two ends of the line. Conveniently, this information can be used for time correlation with lightning detection data that also uses GPS.

2.1.3. GIS and GPS. The main advantage of the Geographic Information System (GIS) is incorporation of spatial data of different forms together in a systematic fashion [32]. Different types of data can be layered jointly, making data management easier. Framework for GIS project is presented in Fig. 1, [33]. Data collected by lightning detection network and traveling wave recorders is mapped and stored in Geodatabase. Geospatial Analysis Tools are used for manipulation of maps. Framework contains one layer for each type of data. Layers are classes or categories of data that can be organized in separate and distinct data structures, but integrated into a single file. These layers can be updated as the new information arrives to the system.

The Global Positioning System (GPS) is a space-based satellite navigation system that provides location and time information for specified targets on the Earth. The latest equipment has a GPS time accuracy of 100 ns with a resolution of 10 ns [35].

3. Spatio-temporal correlation of lightning and fault location data in transmission networks

As a demonstration of utilization of big data for improved outage management, lightning data is correlated with traveling wave fault location results. Both data sets are correlated in time using GPS and in

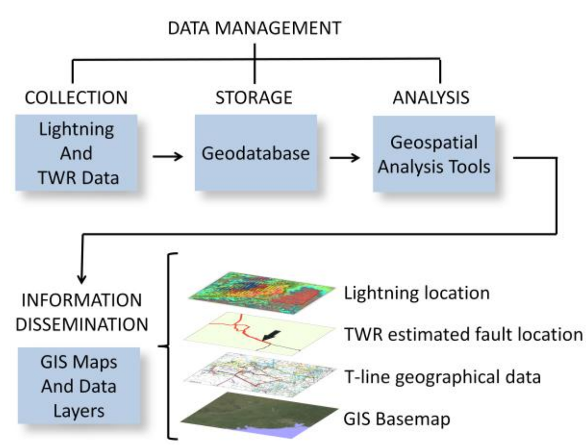

Figure 1: GIS Data Framework

space using GIS. The traveling wave method is assumed to be the main source of information while lightning data is used to enhance the situational awareness and provide better accuracy of fault location result.

Table 1 lists the big data properties of the presented application. The problem falls in the group of big data problems for the following reasons:

- Variety: The database includes sampled waveform data combined with reports from traveling wave fault locator units, lightning detection network, and geographical data. The data files come in different formats that are not compatible and information needs to be extracted so that they match the application. For example, lightning detection network provides the location of lightning strikes in terms of coordinates (longitude and latitude), while traveling wave recorder provides the information in terms of distance to the fault from line terminals.
- Volume: The implementation requires analysis of the extensive set of historical data in order to determine tradeoff between accuracy of traveling wave method and enhancement using the lightning data to determine the confidence of the data gathered during the event analyzed in real time. The lightning data is required for the period of time that covers all events from historical data and each lightning report will generate new map. This is just one level at which the volume of data can be overwhelming. In addition, during the fault, extensive set of data is received and not all of it is used for automated fault location. First, the important information needs to be extracted in an automated way. Typically, this process is done manually by utilities today. With methods used in big data analysis such as indexing for faster search and machine learning for extracting knowledge from data this process can be automated.
- Velocity: The velocity refers to the speed at which data is arriving to the central computing facility. During the fault, multiple sources will send a large amount of data that needs to be stored and ready for analysis. The examples are samples of traveling wave waveforms and coordinates of lightning strikes.

3.1. Traveling wave fault location

The GPS synchronized traveling wave method is used as one source of information for fault location. In order to implement traveling wave fault location the following steps are taken:

- Modeling of the power system: It is done according to the method given in ref, [36]. Transmission line modeling is done using J. Marti model, [37]. This is a frequency dependent line model that uses analog filtering technique for identification of line parameters and can be simulated with ATP EMTP software, [38].
- Simulating the fault transients: Faults are introduced in various locations over the selected transmission line.
- Determining modal transformation for a three-phase system: Signals are transformed into modal components using Clark’s transformation, [39]. After modal transformation a three-phase system is represented by an earth and two aerial modes. The aerial mode 1 is used for fault distance estimation.
- Computing the traveling wave velocity: Method that uses maximum of the first two consecutive peaks of the power delay profile (MPD method, [12]) is used.
- Calculating the arrival time: Wavelet transformation is used to determine the arrival time of the transient peak. The “mother” wavelet that is used is Daubechies wavelet, [13]. Wavelet Toolbox in MATLAB is used [40].
- Calculating fault location: The arrival times of the transient peaks at two TWRs that are located on two line terminals (TA, TB), line length between two TWRs (l) and calculated velocity of wave propagation (ν) are used to calculate the distance θ to fault as

$$\theta = \frac{l + (T_A - T_B)\nu}{2} \tag{3}$$

- Performing time synchronization: Arrival times of two wave fronts are synchronized using GPS, [14-16].

Factors affecting accuracy of the traveling wave fault location methods are:

- Estimation of line length is a major cause of error. As it is presented in [9] not knowing exact line

Table 1: The big data properties of data sources

		VOLUME	VELOCITY
VARIETY	Traveling wave data	4 GB for storage of 2100 records from 8 line modules per substation device	Baud rate of 115200 bits per second
	Lightning data	40 MB of data per day	Sensor baud rate 4800 bits per second, event timing precision of 1μs
	GIS	Additional GIS layer for every type of data, each layer is few MB large	Up to 1000 maps per day can be generated for lightning data

length and line topology can lead to the error close to 500 foot (150 m).

- The traveling waveform is assumed to travel at the speed of light, [35]. When it comes to the overhead transmission lines, velocity of the propagated wave is close to that of the light but not quite the same.
- Time stamping must be very precise to make the system work. As it is stated earlier the latest traveling wave fault locators have GPS time tag accuracy of 100ns, [35].
- Wave-detection error due to interpretation of the transient is a major source of error. This error results from misinterpretation of multiple transients and/or reflected transients. This is a significant concern in the case of lightning strikes. Lightning storms with multiple rapid strikes can cause confusion in terms of which transient was associated with which fault, [41]. In [15], the issue of multiple lightning strikes was investigated and it was reported that travelling wave recorders can produce incorrect results in such cases.
- Current transformers (CT) and capacitive voltage transformers (CVT) can affect the accuracy as well. In [42, 43] modeling techniques for transient response of CTs and CVTs are discussed. It has been pointed out that the differences in the length of the cabling from protection CT to the relay room at each end of the transmission line can affect accuracy [35]. Traveling wave fault location method used in this paper extracts the traveling wave from the current signals collected on the secondary of CTs. The CTs have enough bandwidth to pass the transients, however they do affect accuracy of the method.
- Accuracy of the method is greatly affected in case of the faults with small inception angles (<5°). For the cases of fault inception at zero crossing, theoretically, no traveling wave from the fault location is generated [44].

Because information coming from the lightning detection network is not a part of the conventional traveling wave fault location system it is not affected by all of the described errors. The only parameter that affects both methods is GPS time tag accuracy. Thus, lightning detection network data may complement the fault location method and improves the accuracy of a complete system.

The model of a 400 kV transmission line presented in Fig. 2 is used for simulation in the experimental section. The sampling frequency was 1 MHz. The line length was 120 miles (~193 km). The faults were generated in the range from 10 to 110 miles from the terminal A.

3.2. Dataset description

Lightning detection network collects following:

- Date and time of lightning strike using GPS,
- Location of a strike (latitude and longitude)
- Peak current and lightning strike polarity,
- Type of lightning strike (cloud to cloud or cloud to ground).

Traveling wave fault locator provides following:

- Date and time when event was recorded using GPS,
- Distance to the fault from the line terminals,
- Transient signals recorded at the line terminals.

Additional data that needs to be known for the application is:

- Location of line terminals (latitude and longitude), as well as geographical representation of the line,
- Transmission line characteristics needed for the transient simulation.

The following data is used for modeling:

- Transmission line parameters,
- Physical characteristic of a transmission line and towers,
- Line length,
- Lightning surge peak current

3.3. Automated Correlation of data

In order to automatically correlate traveling wave data with lightning data, for the purpose of more accurate fault detection, it is necessary to identify which faults are likely caused by lightning and to do so with minimal human intervention. This task is accomplished in the following way.

When the traveling wave recorders detect transients indicating that a fault has occurred on a transmission line, they send data with a GPS time stamp to the *Local Control Building*. This data is then transmitted to the *Central Station* where real-time data from the lightning detection system is queried for the lightning activity in the area around the line (in the past 10 minutes, in a 5 km radius). For this step it is necessary to have fast and accurate querying of the lighting detection data.

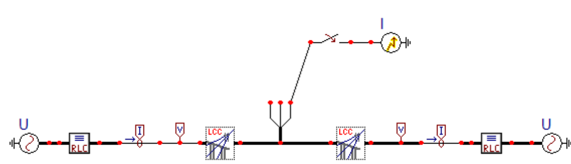

Figure 2: ATPDraw model of the tested line

3.3.1. Identifying faults caused by lightning strikes. By comparing time stamps of events detected by traveling wave recorders and those obtained from querying the lightning detection system it may then be determined whether the disturbance is likely to be caused by lightning activity, as indicated by their closeness in time and space. The flow of information is illustrated in Fig. 3. If it is determined that the disturbance is likely generated by lightning then the complete set of data about the event is gathered at the *Central Station* where correlation of data is leveraged together with analytics to improve fault location. In the *Central Station* the transient simulation of event is run and analysis of data is performed as described next.

3.3.2. Spatio-temporal correlation of diverse data. The data management for the correlation process is shown in Fig. 3. Traveling wave fault recorders are located at both ends of the transmission line. On the other hand lightning sensors are typically not a part of the utility infrastructure and are located sparsely across a wide area. The traveling wave fault location system provides the estimate termed *Automatic Fault Location Result* in Fig. 4. This result is implicitly allocated to the transmission line. Lightning sensors provide an estimated *Location of a Lightning Strike*. This result is presented in terms of longitude and latitude and it is not necessarily located on the transmission line, but rather somewhere in the vicinity of the line.

The location of the lightning strike is projected to the closest point on the transmission line using a "snap" feature. The snap editing in GIS will move the point within the specific distance (tolerance) of the line to the closest point on the line. This snapped point is considered as the lightning detection network estimate of fault location so that the fault location can be described in terms of distance from the line terminals. For the tolerance 1 km distance from the line is chosen. Only lightning strikes that are within 1 km from the line will be used in the correlation process. Then, two fault location results one coming from the traveling wave fault locator and the other coming from the lightning detection network are combined using a Bayesian framework in order to improve the accuracy of the prediction.

Before the beginning of spatio-temporal correlation process lightning data set is reduced to set containing only cloud-to-ground surges, where all instances of cloud-to-cloud surges are removed from data set. Then, the temporal correlation is done. After fault is detected, the time window that contains 2 seconds around the time stamp for fault beginning received from the traveling wave recorder (*FaultStart*) is created. The data received from lightning detection network is searched and only lightning strikes that satisfy following rule are collected inside the *Database A*:

$$|FaultStart - LightningTimeStamp| < 1s \qquad (4)$$

After that the spatial correlation is done. Based on location of line terminals and geographical representation of the line the zone around the line is created that covers area going 300m on both sides of the line. This area has a shape of a polygon. It is to be noted once this area is created it can be used in all future analysis of the observed line and it can be created in advance. The data in *Database A* is searched and only those lightning instances that are inside the area are collected in *Database B*. This problem is called "Point in Polygon" problem, [45]. Created area around the line will typically form a concave polygon. The Grid algorithm is used to determine whether the lightning strike is inside the polygon, [46]. The polygon is divided into grid cells and then the coordinates of lightning are compared with the coordinates of every grid cell.

In the next step lightning instances from *Database B* are searched and the closest one to the traveling wave recorder result is chosen to be correlated as a *LightningDetectionResult*.

3.3.3. Data Analytics. We consider the traveling wave fault location to be the main source of information about the fault event. It processes the recorded data *x*, and makes the maximum likelihood estimate of the

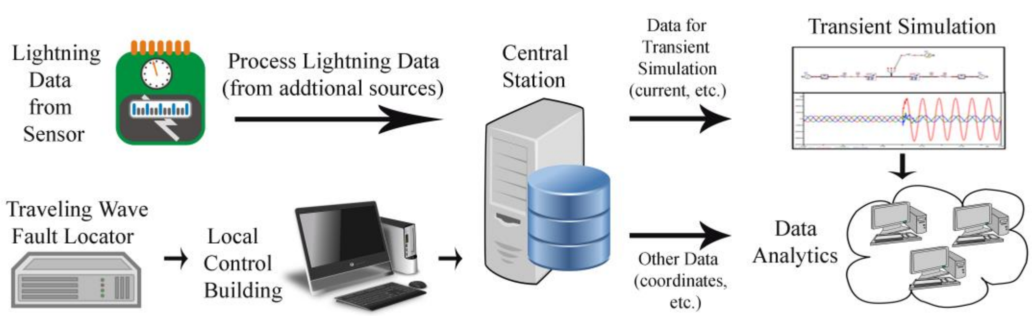

Figure 3: Dataflow during the fault

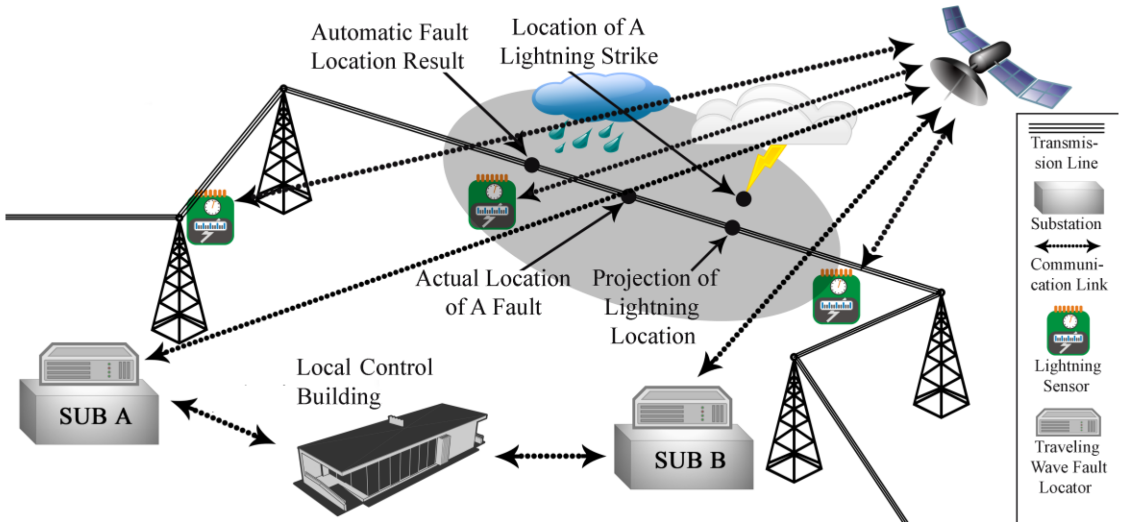

Figure 4: Spatio-temporal correlation of traveling wave fault locator data with lightning, GIS and GPS data

fault location based on this data. The precise value from (3) may be described as the following,

$$FaultLocationResult \approx \arg\max_{\theta} p(x|\theta) \tag{5}$$

It is possible to discern the variance of θ either from historical records or through other means, but these methods may be unreliable and are beyond consideration in this study.

The lightning detection data is considered the prior probability, in this case coming from indirect, side-information and independent of the measurements x,

$$LightningDetectionResult \approx \arg\max_{\theta} p(\theta) \tag{6}$$

The posterior probability of the fault location can then be expressed using Bayes Theorem as,

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \tag{7}$$

In order to compute the necessary maximum a posteriori estimate of fault location,

$$ImprovedPrediction = \arg\max_{\theta} p(\theta|x) \tag{8}$$

it is not necessary to compute the normalization constant $p(x)$ because the same fault recorder data x is considered under all fault location positions θ.

By considering the posterior instead of only the likelihood better predictions are made because cross-domain data is integrated.

Taking the logarithm of the (7) and disregarding the normalization constant,

$$\log p(\theta|x) \sim \log p(x|\theta) + \log p(\theta) \tag{9}$$

Under the normal assumption for both distributions prior and likelihood in (8), the explicit computation of variance is not necessary. Instead it is computationally favorable to compute the optimal trade-off parameter *nu* from the interval [0, 1]. This parameter then controls the trade-off between a bigger or smaller variance in $p(x|\theta)$ and $p(\theta)$, but only in direct proportion to each other and irrespective of $p(x)$. At *nu* = 1 we completely trust the lightning detection network data, and then as *nu* is transitioned towards 0 more and more certainty is placed in the traveling wave fault location.

This approach is computationally favorable to fully Bayesian approaches such as Markov Chain Monte Carlo sampling, which would make the application infeasible for power systems.

Considering the monotonicity of the logarithmic function we may express the improved fault location as the linear combination of

$$\begin{aligned}&\mathrm{argmax}_{\theta} \log p(\theta|x) = \\ &\mathrm{argmax}_{nu}\left[\mathrm{argmax}_{\theta}[nu \cdot p(x|\theta)] + \mathrm{argmax}_{\theta}[(1-nu)\cdot p(\theta)]\right]\end{aligned} \tag{10}$$

The task becomes that of obtaining the precise *nu* to use. In order to compute *nu* a binary search along a line can be used to find optimal values since the problem is one dimensional. This process requires only $O(\log n)$ time to find the optimal nu among n given values. A simple linear combination like this has the advantage of high bias and low variance in machine learning terms, meaning that its predictions are not likely to be very imprecise in addition to having good generalization power across unseen examples. Because of the low computational complexity this kind of algorithm is directly applicable to big data scenarios.

4. Results

In order to assess the performance of the proposed fault location method it was necessary to evaluate its performance on a number of different fault scenarios. Using the model in Fig. 2. 1000 fault scenarios were simulated. First, all fault scenarios were solved using only the traveling wave method for fault location. After simulation the error of this method was calculated as the relative mean absolute error,

$$e(\%) = \frac{|CalcDist - ActualDist|}{LineLength} \times 100 \tag{11}$$

Second, the results from the lightning detection network were calculated as explained in section 3.3 and (10) was used to quantify the error.

After correlation of the two methods, as it was explained in Section 3.3, error of the improved result was calculated using (11). When dealing with a linear combination of predictors it is necessary to assess the generalization performance. Good generalization is indicated by the ability of a fault location method to locate faults accurately even for unforeseen fault locations. In order to quantify the generalization performance of the proposed fault location method it was necessary to compute the generalization error.

In order to estimate the generalization error of the improved fault location method it is necessary to split the data from many different scenarios into a training set and a testing set of data. Determining the optimal *nu* on the training set gives point estimates of the generalization error on the testing set when comparing the improved fault location to the true fault location, and therefore the procedure needs to be repeated for precise estimates, a process often calls for 2-fold cross-validation. The results in Fig. 5 and 6 are average results, computed per scenario, from 100 replications of cross-validation.

A histogram of all three results is presented in Fig. 5, where the *x*-axis represents the error and the *y*-axis represents the frequency of that error, where errors from different scenarios are binned according to a regular grid. From Fig. 5, the proposed approach outperforms the traveling wave fault location, having the largest number of test cases with error that is closer to zero. For every test case our approach shoved better accuracy than the individual methods. Mean Square Error of distance to fault for the lightning data was 0.0076±3.1e-04 miles, for traveling wave it was 0.0012±4.3e-05 miles and the improved method showed 0.0011±4e-05 miles, both the variance and the mean of the error were smaller using the improved method on unseen fault scenarios, when compared to the traveling wave method.

It is significant that the traveling wave method result has much higher accuracy than the one obtained from lightning data. The lightning detection data may only be useful in enhancing the traveling wave fault detection method. Lightning data has very high variance as well compared to other two methods.

Additionally, the proposed method shows no bias in the predictions in Fig. 6, indicating that the fault prediction location neither systematically over- or under-estimated.

Because traveling waves are recorded on both sides of the transmission line, the error does not depend on the distance from the fault, indicating homoscedasticity and confirmed by Engle's ARCH test [47].

As it can be seen in Fig. 5 the tradeoff parameter *nu* between accuracy of traveling wave method and lightning data is estimated to be optimal at 0.871±0.0133 on unseen examples. This can be interpreted as placing 87.1% trust in the result of the traveling wave method, 12.9% in the estimate from lightning. The low variance of *nu* is indicative of the low variance predictor used for improved fault location.

5. Conclusion

This paper demonstrates that correlating automatically multiple sources of data may help enhance fault location calculation. More specifically:

- A method using correlation of cross-domain data for identifying which faults are likely to be caused by lightning strikes is presented.
- A method for locating faults using lightning detection data is presented and its precision quantified.
- A method of integrating lightning detection sensor data with traveling wave fault location measurements is presented and its precision quantified.
- The results indicate that integrating lightning detection sensor data with traveling wave fault detection data improves fault location accuracy.
- Proposed method that correlates traveling wave fault locator data and lightning data exhibits better performances than any of the methods alone.

Acknowledgement

This work has been funded by NSF I/UCRC PSerc under the project titled "Systematic Integration of Large Data Sets for Improved Decision-Making".

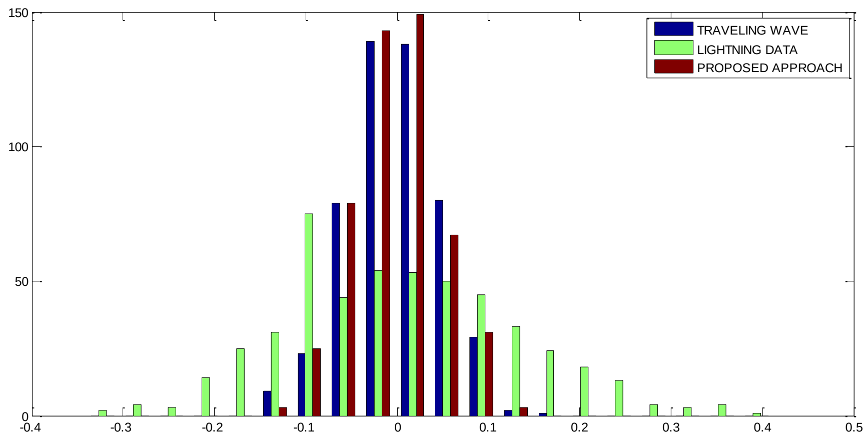

Figure 5: Histogram of an error distribution for individual traveling wave and lightning data; and our approach that combines two methods

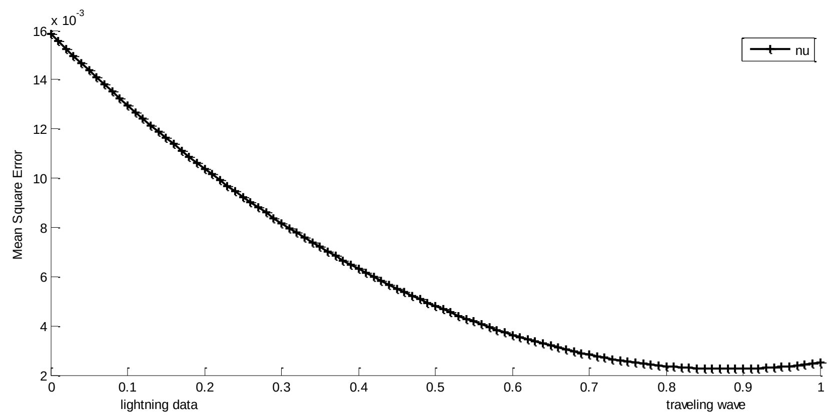

Figure 6: Comparison between traveling wave and lightning data using *nu*

References

[1] Executive Office of the President, "Economic Benefits of Increasing Electric Grid Resilience to Weather Outages," Aug. 2013.

[2] E. Mills, "Electric Grid Disruptions and Extreme Weather", *U.S. Disaster Reanalysis Workshop*, 2013.

[3] P. Hines, *et al.*, "Trends in the History of Large Blackouts in the United States" *IEEE/PES General Meeting*, Jul. 2008.

[4] N. C. Abi-Samra and W. P. Malcolm, "Extreme Weather Effects on Power System," *IEEE/PES General Meeting*, Jul. 2011.

[5] M. Kezunovic, *et al*., "Impact of extreme weather on power system blackouts and forced outages: New challenges," *7th Balkan Power Conf.*, Sep. 2008.

[6] T. Overbye, *et al.*, "The Electric Power Industry and Climate Change: Power Systems Research Possibilities," PSerc Report #07-16, Jun. 2007.

[7] P.-C. Chen, *et al.*, "The Use of Big Data for Outage Management in Distribution Systems," *Int. Conf. on Electricity Distrib.* (*CIRED*) *Workshop*, 2014, in press.

[8] M. Kezunovic, *et al.*, "Reliable Implementation of Robust Adaptive Topology Control," *47th Hawaii Int. Conf. on Syst. Sciences* (*HICCS*), Jan. 2014.

[9] E. O. Schweitzer, III, *et al.*, "Locating Faults by the Traveling Waves They Launch", *Texas A&M Conf. for Protective Relay Engineers*, 2014, in press.

[10] P. F. Gale, *et al*., "Fault location based on travelling waves." *5th Int Conf. on Developments in Power Syst. Protection* (*DPSP*), pp. 54–59, Apr. 1993.

[11] Y. J. Xia, *et al.*, "A novel fault location scheme using voltage traveling-wave of CVTs," *39th Int. Uni. Power Eng. Conf.* (*UPEC*) (*Vol. 2*), vol. 1, pp. 768-772, 2004.

[12] A. Elhaffar and M. Lehtonen, "Multi-end Traveling Wave Fault Location Based on Current Traveling Waves," *16th Power Syst. Computation Conf.* (*PSCC*), Jul. 2008.

[13] F. H. Magnago and A. Abur, "Fault Location Using Wavelets," *IEEE Trans. Power Del.*, vol. 13, no. 4, Oct. 1998.

[14] A. Tabatabaei, *et al*., "Fault location techniques in power system based on traveling wave using wavelet analysis and GPS timing," *Electrical Review*, vol. 88, no. 6, pp. 347-350, 2012.

[15] H. Lee and A. M. Mousa, "GPS travelling wave fault locator systems: investigation into the anomalous measurements related to lightning strikes," *IEEE Trans. Power Del.*, vol. 11, no. 3, pp. 1214-1223, 1996.

[16] W. Zhao, *et al.*, "Improved GPS travelling wave fault locator for power cables by using wavelet analysis," *Int. J. of Electrical Power & Energy Syst.*, vol. 23, no. 5, pp. 403–411, Jun. 2001.

[17] T. Sadovic, *et al*., "Expert System for Transmission Line Lightning Performance Determination", *CIGRE Int. Colloq. on Power Quality and Lightning*, Jun. 2012.

[18] IBM Software, "Managing big data for smart grids and smart meters," [Online] Available: http://www.smartgridnews.com /artman/uploads/1/IBM_anaytics_paper.pdf

[19] M. Patel, *et al.* "Real-Time Application of Synchrophasors for Improving Reliability," *North Amer. Electric Rel. Corporation* (*NERC*), 2010.

[20] M. Kezunovic, *et al.*, "The role of big data in improving power system operation and protection," *Int. Inst. for Research and Education in Power Syst. Dynamics* (*IREP*) *Symp.*, 2013.

[21] G Shroff, "The Intelligent Web: Search, Smart Algorithms and Big Data," Oxford University Press, UK, 2013.

[22] L. McColl, *et al*., "A Climate Change Risk Assessment for the UK Electricity Networks," presented *at Int. Conf. Energy & Meteorology* (*ICEM*), 2011.

[23] L. McColl, *et al.*, "Assessing the potential impact of climate change on the UK's electricity network: electronic supplementary material 3." [Online] Available: http://link.springer.com/content/esm/ art:10.1007/s10584-012-0469-6/file/MediaObjects/ 10584_2012_469_MOESM3_ESM.pdf

[24] Power Guide 2009/Book 07, "Protection against lightning effects – Legrand," Legrand 2009.

[25] U. Finke, *et al.*, "Lightning Detection and Location from Geostationary Satellite Observations," Institut fur Meteorologie und Klimatologie, University Hannover. [Online] Available: http://www.eumetsat .int/website/wcm/idc/idcplg?IdcService=GET_FILE&dDocName=pdf_mtg_em_rep26&RevisionSelectionMethod=LatestReleased&Rendition=Web

[26] K. L. Cummins, *et al.*, "The US National Lightning Detection NetworkTM and applications of cloud-to-ground lightning data by electric power utilities," *IEEE Trans. Electromagn. Compat*., vol. 40 , no. 4, pp. 465-480, Nov. 1998.

[27] Vaisala Inc., "Thunderstorm and Lightning Detection Systems," [Online] Available: http://www.vaisala. com/en/products/thunderstormandlightningdetectionsystems/Pages/default.aspx

[28] *IEEE Guide for Direct Lightning Stroke Shielding of Substations*, IEEE Standard 998-2012, Apr. 2013.

[29] R. B. Anderson, "Lightning performance criteria for electric power systems," *IEE Proc. Gener., Trans. and Distrib.*, vol. 132, no. 6, pp. 298-306, Nov. 1985.

[30] A. M. Mousa, "A study of the engineering model of lightning strokes and its application to unshielded transmission lines," PhD dissertation, The University of British Columbia, Aug. 1986.

[31] A. M. Mousa and K. D. Srivastava, "Effect of shielding by trees on the frequency of lightning strokes to power lines," *IEEE Trans. Power Del.*, vol. 3, no. 2, pp. 724-732, Apr. 1988.

[32] Esri, "What is GIS?," [Online]. Available: http://www. esri.com/what-is-gis/overview#overview_panel

[33] Program and Organizational Performance Division et al., "Best Practices in Geographic Information Systems-Based Transportation Asset Management," Jan. 2012.

[34] D. J. Buckey, Bgis Introduction to GIS. [Online]. Available: http://bgis.sanbi.org/gis-primer/

[35] D. Cole, *et al.*, "Fault Location & System Restoration," Jul. 2013. [Online]. Available: http://www.qualitrolcorp .com/uploadedFiles/Siteroot/Products/Fault_Location_and_System_Restoration.pdf

[36] T. Jinrui, *et al.*, "Modeling Technology in Traveling-Wave Fault Location," TELEKOMNIKA, vol. 11, no. 6, pp. 3333-3340, Jun. 2013.

[37] J. R. Marti "Accurate modeling of frequency-dependent transmission lines in electromagnetic transient simulations," *IEEE Trans. Power Apparatus and Syst.*, vol. PAS-101, no. 1, pp. 147-157, Jan. 1982.

[38] Alternative Transients Program, ATP-EMTP, 2010. [Online]. Available: http://www.emtp.org

[39] E. Clarke, "Circuit Analysis of AC Power Systems, Symmetrical and Related Components," *Wiley*, New York, 1943.

[40] Mathworks Inc. MATLAB R2012b User's Guide. [Online]. Available: http://www.mathworks.com

[41] M. M. Saha, *et al*., "Fault location – Basic concept and characteristic of methods," Fault Location on Power Networks, Springer London, pp. 1-26, 2010.

[42] M. Kezunovic, *et al.*, "Digital models of coupling capacitor voltage transformers for protective relay transient studies," *IEEE Trans. Power Del*., vol. 7, no. 4, pp. 1927-1935, Oct. 1992.

[43] Y. Tang, *et al*., "Study on Effect of Current Transformer and Its Secondary Cable to Travelling Wave Propagation Characteristic of Electric Power Lines," *2nd Int. Conf. on Intell. Syst. Des. and Eng. Appl.*, pp.1495-1498, Jan. 2012.

[44] V. S. Kale, *et al*., "Faulted phase selection based on wavelet analysis of traveling waves', *Int. J. Computer and Elect. Eng.*, vol. 3, no. 3, pp. 421–425, Jun. 2011.

[45] M. J. De Smith, *et al.*, "Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools," *Troubador Publishing Ltd*, 2007.

[46] C.-W. Huang, T.-Y. Shih, "On the Complexity of Point-in-Polygon Algorithm," *Computers & Geosciences, Elsevier Science Ltd*, vol. 23, no. 1, pp. 109-118, 1997.

[47] R. F. Engle, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, vol. 50, pp. 987–1007, 1982.