A Data Fusion Probabilistic Model for Hurricane-Induced Outages in Electric Power Grids

Akwasi F. Mensah
PhD Candidate, Dept. of Civil and Environmental Engineering, Rice University, Houston TX, USA

Leonardo Dueñas-Osorio
Associate Professor, Dept. of Civil and Environmental Engineering, Rice University, Houston TX, USA

ABSTRACT: Prediction of outages in electric power systems before a hurricane can be enhanced by exploiting data not typically used for such purposes, including system contingencies during normal operation. This data-based enhancement is necessary to inform disaster planning and preparedness, as well as to speed up the restoration of the system while capturing local trends. This paper presents a framework that integrates hurricane-induced outage predictions based on component fragilities, physics of power flows, and network responses, along with increasingly available spatio-temporal information of daily outages, so as to better assess outage risks in the electric system. The study adapts Linearly Constrained Least Squares (LCLS) by making space the dependent variable instead of time (as traditionally used), and then determine an optimal linear fusion of the predicted outages with information from the daily outage trackers for enhanced outage distribution assessment. Using the electric power system as an illustrative example, the study shows how data fusion improves the prediction accuracy by including daily outage information, which implicitly contains the spatial structure of the network as well as the current physical state of its components and interactions, including ageing, fatigue and recent hardening or embedding of smart grid technologies. The fusion-based framework predicts an overall outage of 84% in the county under Hurricane Ike winds, which strongly agrees with the reported outage of 86% by the utility provider in the aftermath of the event as opposed to overall outage of 90% predicted without data fusion.

1. INTRODUCTION

Electric power infrastructure is considered the backbone for public health, safety, the economy, security and community life across modern societies. Costly power outages caused by recent natural events such as Hurricane Ike in 2008 and Super Storm Sandy in 2012 highlighted the role of electric power systems in the affected cities, and the need to model, identify and strengthen weaknesses in the systems. Existing outage prediction models can be grouped under three main approaches: statistical (Han et al., 2009), topology-based (Pagani and Aiello, 2011) and fragility-based models (Winkler et al., 2010). Statistical models suffer from paucity of reliable historic data on hurricane events. Purely topology-based models do not consider the physics of power flow in the system. And component fragility-based methods usually do not consider system level interactions. With the emergence of smart-grids and outage trackers that provide volumes of daily outage data, advanced techniques for fusing information from multiple sources or analysis tools can lead to predictive models with improved accuracies and practical appeal given their spatial resolution.

This paper presents a probabilistic model that integrates topological information and components fragilities with spatio-temporal information on daily outages to assess outage risks in an electric power grid under hurricane winds. Taking the electric power grid of Harris County, Texas, USA as an illustrative example, the model predicts outages in distributed 1 km^2 customer blocks across Harris County. The model consists of three main sub-models: a topology-fragility-based model, a spatio-temporal analytical model of outages and a data fusion methodology. The topology-fragility-based model relies on component fragilities based on a given hurricane scenario and predetermined influence networks that inform the electrical long-range effects of failures, to then propagate failures in the transmission system and distribution system along with associated outages. The influence networks are constructed via a DC power flow analysis, which takes into account flow parameters and constraints in the real transmission system of the power grid. Distribution networks, unlike transmission systems, have topologies usually not available to the public. Thus, they are represented by minimum spanning trees in this work, where each tree connects a transmission substation and distributed customer blocks within a service area in a typical radial fashion.

The second sub-model analyzes large records of daily customer outages across the county to generate spatio-temporal data ensembles. Even though the outage data is not hurricane-related, it provides useful insights of weak spots in the system in its current state, while capturing evolving failure trends in the system. Spatio-temporal data is used to evaluate the relative likelihood of outages in the customer blocks. The third component, the data fusion methodology, uses Linearly Constrained least squares (LCLS) (Zhou and Leung, 1997), which provides an optimal fusion with minimum variance and less computational requirements , to integrate predictions from the first and second sub-models to produce enhanced hurricane-induced outage risks in the disrupted square blocks. Thus, the proposed model provides predictions with better accuracy that are consistent with the current operational and geographical state of the electric power system. It can inform decision-making for pre- or post-hurricane intervention and restoration strategies, and also support or exploit emerging smart electric power system technologies.

This paper is organized into five sections. The next section presents the probabilistic outage model for assessing hurricane-induced outages in the electric system. It also contains a description of the Harris County, TX electric grid, which is used as illustrative example in this study as well as discussion of predicted outages. Section 3 discusses daily outage data collected from online trackers. Section 4 introduces the LCLS approach which is used to determine an optimal fusion of the two sets of outage assessments. It also discusses fused outage risks obtained in the study. Finally, Section 5 presents a summary and key conclusions of the research, along with future research needs.

2. PROBABILITISTIC OUTAGE ASSESSMENT MODEL

This paper improves a probabilistic model developed by Mensah and Duenas-Osorio (2014) for assessing outages in an electric power grid. The model is made up of four modules – a hurricane demand module, component fragility module, transmission, and distribution system outage modules. The model uses a Bayesian Networks approach and DC flow analysis to efficiently evaluate hurricane-induced outage probabilities. This section provides only a general overview of the assessment model with emphasis on improvements over the previous model.

2.1.1. Harris County's power grid

The paper illustrates the noted outage assessment models using the electric power system of Harris County, Texas, which is owned and serviced by CenterPoint Energy Inc. This system supplies electricity to about 1.7 million customers via high voltage transmission and low voltage distribution networks. The topology and information on the transmission grid is obtained from Platts (2009). It has 23 power plants and 394 transmission substations connected by 551 transmission lines.

Like in most places in the United States, the structure and information of the distribution network of Harris County is not publicly available. Since distribution networks operate in radial configuration to minimize energy losses

(Cavellucci and Lyra, 1997; Montoya and Ramirez, 2012), this study constructs minimum spanning trees (MSTs) to represent the distribution system. There are a total of 3,330 1 km^2 blocks in the Harris County Key Map, whose midpoints are idealized as distribution load points along main feeders. The county is divided into substation service areas based on the substation tributaries. Each substation is considered to supply electric power to all the distribution load points within its service area. An MST, which originates from the substation to the distribution nodes, is built to represent main feeders and laterals of the distribution network within the service area. Multiple distribution networks can be constructed to originate from a single substation if the substation supplies power to more than 20 distribution nodes to maintain the length and structure of radial networks realistic.

2.1.2. Component performance module

The component performance module consists of fragility curves and failure regression models of the components in the transmission and distribution systems that give the probability of the components failing at a given wind speed. Fragility curves with parameters obtained from HAZUS-MH 4 internal files (2008) of transmission substations are used. The failure probabilities of transmission lines, distribution poles and conductors are approximated based on regression models of their failure rates, which are developed based on historical storm-event data by Quanta (Quanta, 2009). The failure probability of the service drop from a distribution pole to a customer is determined using the probability of wind throw or tree fall.

2.1.3. Outage assessment module

The outage module is made up of two system response models. The first model is a Bayesian-Network (BN)-based model that uses the failure probabilities of substations and transmission lines to evaluate the probabilities of outage at individual substations. By structuring the response model as a BN, the study provides a novel and practical approach for representing the probabilistic dependencies among nodes, and propagating mechanical and electrical damages throughout the system while retaining computational tractability. The study performs *n*-1 node removal vulnerability assessment based on a DC flow analysis to determine the influence of component failures on substations' ability to meet power demands throughout the system. Information obtained is used in the form of influence networks to construct the dependence (edges) between nodes (substations) and their parents (which could be transmission lines and other substations) in the BN. This influence network strategy helps reducing the complexity of conditional probability tables for otherwise naïve BN models, and reflects industry practices to evaluate contingencies. Please refer to Mensah and Duenas-Osorio (2014) for details on futher details and equations for BN-based model.

The second system response model assesses the outage probabilities of individual keymap blocks within the entire system. For each substation service area within the Harris County, the model constructs an adjacency matrix that contains the connections among the substation and the distribution load points based on the minimum spanning trees (distribution networks) built in Section 2.1.1. The model employs the recursive decomposition algorithm (RDA) to evaluate the exact connectivity reliability of each node (distribution load point) within the network using the outage probability of the substation and the failure probabilities of the distribution conductors and poles. RDA is a non-simulation-based algorithm that identifies critical cut and link sets based on shortest paths from the source (substation in this case) and a node, recursively using a graph decomposition scheme (Li and He, 2002; Lim and Song, 2012). Using RDA significantly improves the efficiency of assessing the network reliability over Monte Carlo Simulations. The outage probability within a Keymap is computed as a series system of its disconnection probability computed by RDA and the failure probability of its customer service drop determined in Section 2.1.2.

2.1.4. Predicted Outages

A hurricane scenario with similar characteristics of Hurricane Ike which struck the Houston region on September 13, 2008 is generated using the HAZUS_MH software to illustrate the application of the outage prediction model. The hurricane, which was classified as a large category 2 event resulted in the largest power outage in Texas history. Figure 1 shows outage probabilities within the Keymaps across Harris County as predicted by the power outage model. Widespread outages are seen to be predicted across the entire county with the model predicting an overall outage probability or percentage of customers without power to be 90% owning to structural damage, mainly of components in the electric power distribution system. Centerpoint Energy reported that 86% of its customers within the Harris County were without power in the immediate aftermath of Hurricane Ike 2008 (Centerpoint, 2009).

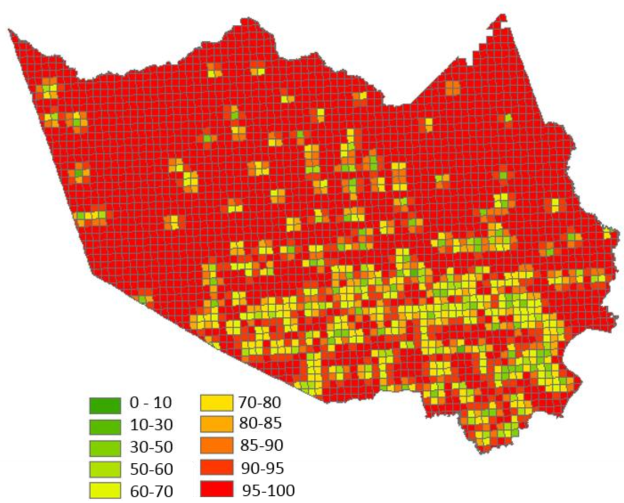

Figure 1: Outage probability (%) in the KeyMap squares of Harris County, TX due to Hurricane Ike winds

Furthermore, comparing the predicted outage percentages and actually reported outages after Hurricane Ike at granular scales show an average of 16% difference between values for the two quantities of individual ZIP code areas. Despite the good performance of the prediction model as compared to the actual figures, simplification of the distribution network and approximation failure information in the modeling of the electric power system are embedded in the predictions. Data from outage tracker, which implicitly captures the structure of the network as well as the current physical state including component ageing, fatigue, and recent hardening or smart grid embedding, may be useful in improving prediction accuracy, reducing uncertainty, and capturing local outage effects. The next section describes daily outage data collected from a recently deployed online outage tracker, which are fused with the physics-based predicted outages.

3. COLLECTED OUTAGE DATA AND ANALYSIS

Centerpoint Energy Inc., the utility provider that owns, operates and maintains the electric transmission and distribution systems of Harris County and surrounding areas, recently deployed an online outage tracker covering the greater Houston area. The outage tracker which is based on geographic information systems (GIS), shows the outage locations and their respective number of customers without power, while also providing estimated restoration times (Centerpoint, 2014). The information on the tracker is updated every 15 minutes and the system runs continuously.

This study collected the outage data from the online site every 15 minutes for a 12-month period, resulting in sizable dataset of temporal-spatial outage information. In order to avoid multiple counting of the outages in the same day, the study extracts the maximum instantaneous number of customers without power as the noteworthy daily outage record per a KeyMap from the recorded data. As an illustration, Figure 2 shows the average daily number of customers without power in Harris County in a pseudo-3D representation for the months of August, September and October of 2013. Incidentally, these months are considered to be critical within the Hurricane Season in the Houston region, and therefore they capture intense convective-induced storm outages in the region. Even though the region did not suffer any hurricane in these months, substantial outages are recorded. For example, if the average values show up to 85 customers experiencing daily outages in a particular Keymap block, it means that there were as many as 2635 customer outages within the 31 days. It is observed in all three months that whereas areas such as the west of the county suffered no to little outages, customers in the central part of the county as well as in specific spots to the south or south west see significant outages.

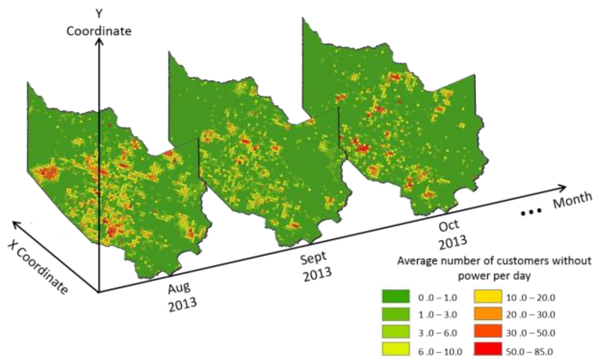

Figure 2: Daily average number of customers without power for three months in the fall of 2013

The daily maximum outages are accumulated over the entire 12-month period and normalized by the total number of service customers within each KeyMap block. Figure 3 shows how many times a customer within a KeyMap block is expected to have experienced an outage for the 12 months of data collected. From the data, the majority of customers, 68% of the county, under no extreme weather conditions are expected to have an annual outage rate of less than 1. Customers living in 292 out of the total of 3,329 Keymap blocks suffer at least 10 outages on average over 12 months. And up to 891 customers are prone to 50 or more outages within a year. The locations with high frequency of outages per year are weak functional spots in the electric power grid and are expected to be highly susceptible to hurricane-induced outages. As such, a method that captures outage-prone spots within the power system along with fragility-based outage distribution assessments, will enhance the accuracy and practicality of outage prediction models.

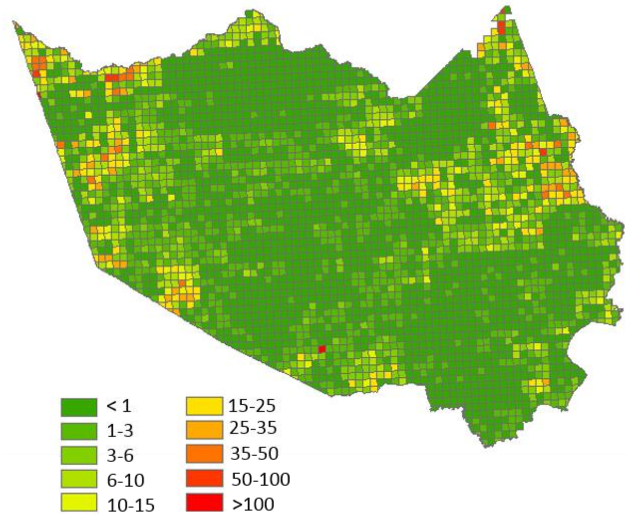

Figure 3: Outage rate per customer within 12-month period

4. DATA FUSION OF PROBABILITY-BASED AND DATA-DRIVEN OUTAGE ASSESSMENTS

Data fusion is the process of integrating data or information from several sources to obtain a unified, consistent and reliable set of information. Data fusion techniques are widely applied in various engineering areas such as remote sensing, signal, image and video processing, and robotics among others (Xia and Leung, 2014). The main purpose of data fusion is to reduce uncertainty by utilizing redundant and complementary information from multiple sources.

There are various data fusion methods depending on the nature of the data. Techniques are here grouped as probabilistic methods such as the Bayesian approach (Khaleghi et al., 2013), non-probabilistic models such as fuzzy logic and interval algebra (Zonta, 2014), and the statistical fusion methods such as linearly constrained least squares (LCLS), covariance intersection (CI), linearly constrained least absolute deviation (CLAD) and constrained least square (CLS) (Xia and Leung, 2014). Statistical fusion methods are not only efficient and simple, but they are also practical while achieving satisfactory performance.

The present study builds upon a linearly constrained least squares (LCLS) approach proposed by Zhou and Leung (1997) for sensor networks, but adapted here to fuse the outage predictions and outage information from simulated and collected outage data. The fusion is formulated as a linearly constrained optimization problem, with less computational requirement, in which the energy (variance) of the linearly fused information is minimized.

4.1.1. Linearly constrained least squares (LCLS) in space

The original formulation of Linearly Constrained Least Squares (LCLS) (Zhou and Leung, 1997) handles multi-source temporal data **x***(t)*. Even though temporal-spatial outage data is collected from the CenterPoint outage tracker, outage prediction models only produce spatial data at an instant of time. Therefore, this study modifies its formulation so that **x** varies in space i.e. **x***(s)* for a given instant of time

Consider $\mathbf{x}(s)=\left[x_1(s),x_2(s)\right]^{\mathrm{T}}$ as the vector containing the predicted outage information $x_1(s)$ and outage information from collected data $x_2(s)$, where *s* denotes a spatial location. Then the fused information is given as a linear combination of the two sets of information,

$$u(s)=\sum_{i=1}^{2}w_i x_i(s)=\mathbf{w}^{\mathrm{T}}\mathbf{x}(s) \tag{1}$$

where $\mathbf{w}=\left[w_1,w_2\right]^{\mathrm{T}}$ denotes the weighting vector. The variance of *u(s)* is given by

$$J=\mathbf{w}^{\mathrm{T}}\mathbf{R}\mathbf{w} \tag{2}$$

where **R** is the covariance matrix, which estimated as

$$\mathbf{R}=\frac{1}{N}\sum_{s=1}^{N}\mathbf{x}(s)\mathbf{x}^{\mathrm{T}}(s) \tag{3}$$

The aim is to find an optimal fusion which produces a set of weights **w** so that the variance of *u(s)*, which is a measure of uncertainty of the fused information, is minimized (Zhou and Leung, 1997). Therefore the LCLS fusion solution is given by

$$\mathbf{w}_{LS}=\arg\min_{\mathbf{w}} J=\arg\min_{\mathbf{w}}\mathbf{w}^{\mathrm{T}}\mathbf{R}\mathbf{w} \tag{4}$$

subject to $\mathbf{a}^{\mathrm{T}}\mathbf{w}=1$, where $\mathbf{a}=[1,1]^{\mathrm{T}}$. Using the method of Lagrange multipliers,

$$\mathbf{w}_{LS}=\mathbf{R}^{-1}\mathbf{a}\Big/\left[\mathbf{a}^{T}\mathbf{R}^{-1}\mathbf{a}\right]. \tag{5}$$

The optimal data fusion is then given by

$$u_{LS}(s)=\left(\mathbf{w}_{LS}\right)^{\mathrm{T}}\mathbf{x}^{*}(s) \tag{6}$$

Note that in this study while the outages in terms of the number of customers are used as $\mathbf{x}(s)$ to estimate $\mathbf{w}_{LS}$, $\mathbf{x}^{*}(s)$ contains the outage probability from the prediction model and the average daily outage rate computed from the tracker data.

4.1.2. Discussion of Predictions

Figure 4 shows outage risks that are produced by the linear combination or fusion of the predicted outages and daily outage rate per customer within each keymap. Hurricane-induced outages predicted for a Hurricane Ike scenario are significantly higher than the outage rates resulting from normal conditions. Therefore, the fused outages show a trend that is dominated by the predicted outages (Figure 1). However, the magnitudes of the fused outages differ from the predicted outages. The overall average outage is 84% as compared to the 90% predicted initially. The new average outage is much closer to the 86%

observed in the immediate aftermath of Hurricane Ike in 2008.

About 2,707 out of the 3,330 (81%) square blocks have at least 80% of their customers without power for this event. In the event of hurricane Ike-like storms, only 6 distribution load points in the entire electric grid have less than 30% of their customers experiencing outages.

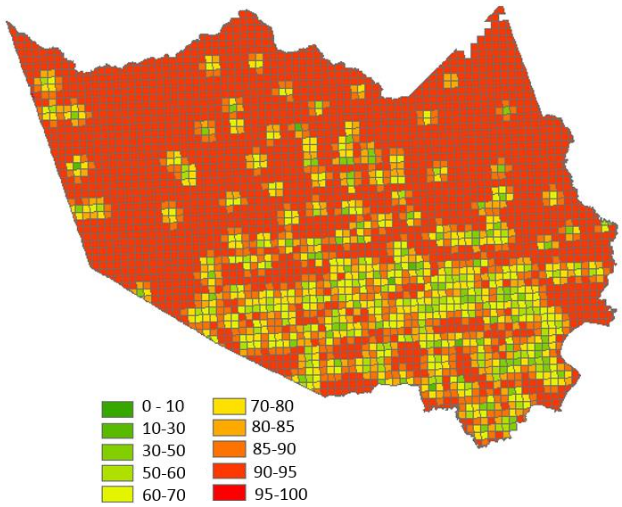

Figure 4: Outage risks in the KeyMap squares produced by data fusion of predicted and observed outages.

5. CONCLUSIONS

Proper prediction of outages in electric power systems before a hurricane can benefit from local sources of information about weakness in the system which can support disaster planning and preparedness with local relevance, as well as speed up the recovery of the system. This paper presents a framework that integrates hurricane-induced outage predictions based on component fragilities, physics of the power flow, and network responses, along with spatio-temporal information on daily outages across Harris County, so as to assess the outage risks in the electric system at various levels of resolution.

The framework consists of two main models. The first is a probabilistic model to predict hurricane-induced outages in an electric power grid. The model is made up of four modules – a hurricane demand module, a component fragility module, and transmission and distribution system outage modules. The model uses a Bayesian Networks approach and DC flow analysis to efficiently evaluate hurricane-induced outage probabilities. The model predicts wide-spread outages for 2008 Hurricane-Ike winds with about 90% customers expected to be without power in the immediate aftermath. The prediction is about 92% in agreement with actual outages reported by the utility provider after the event.

Despite the adequate performance of the prediction model as compared to the actual figures, approximations in the physical layout and failure information in the modeling of the electric power system affect the nature of the predictions. Therefore, a second model collects and analyzes data from an online GIS outage tracker to derive spatial-temporal information that reflects the structure of the network as well as the current physical state including ageing, fatigue, hardening, smart grid technology usage, and other local features of the system and its components. The 12-month daily outage data collected reveals that there are locations within the county, which are prone to outages regularly. As such, these weak spots are more likely to experience outages during severe storm or hurricane events. Hence, this study adapts a Linearly Constrained Least Squares (LCLS) methodology by making space the dependent variable instead of time, to determine an optimal linear fusion of the predicted outages and the information from daily outages. The fusion resulted in outages which generally showed the same trend as the predicted outages because of the significant high hurricane-induced outages over the daily outages. However, the fusion framework reduces the uncertainty across weak and strong spots given its locally-enriched formulation, and thus exploits the best of physics-based outage prediction models with local-level outage details unavailable in most physical models.

This study also forms a good basis for future research which aims at improving the assessment accuracy of the framework and making it easily adaptable for other power systems, and other lifeline systems, in different regions of the country. In particular, the fusion-based framework can be used to provide annual outage risks and support planning decisions. Furthermore, spatio-temporal spectral analyses of the daily outage data, if undertaken, could yield insightful information about restoration trends to support resilience. Also, future studies should determine correlations between daily outages, weather-related events, and other non-weather events.

6. ACKNOWLEDGMENT

This research has been partly funded by the Department of Civil and Environmental Engineering at Rice University, the Office of Public Safety and Homeland Security of the City of Houston under Grant 2008-CP-T8-0023, and the U.S. Department of Defense and its Army Research Office through the MURI grant W911NF-13-1-0340. The authors appreciate the support from these institutions towards resiliency research.

7. REFERENCES

FEMA, 2008, Hazards U.S. Multi-Hazard (HAZUS_MH) Assessment Toool vs 1.4, www.fema.gov/plan/prevent/hazus/index.shtm.

Cavellucci, C., and C. Lyra, 1997, Minimization of energy losses in electric power distribution systems by intelligent search strategies: International Transactions in Operational Research, v. 4, p. 23-33.

Centerpoint, 2009, Hurricane Ike, http://www.centerpointenergy.com/newsroom/stormcenter/ike/, Centerpoint Energy.

Centerpoint, 2014, Outage tracker, http://gis.centerpointenergy.com/outagetracker/.

Han, S.-R., S. D. Guikema, and S. M. Quiring, 2009, Improving the Predictive Accuracy of Hurricane Power Outage Forecasts Using Generalized Additive Models: Risk Analysis, v. 29, p. 1443-1453.

Khaleghi, B., A. Khamis, F. O. Karray, and S. N. Razavi, 2013, Multisensor data fusion: A review of the state-of-the-art: Information Fusion, v. 14, p. 28-44.

Li, J., and J. He, 2002, A recursive decomposition algorithm for network seismic reliability evaluation: Earthquake Engineering & Structural Dynamics, v. 31, p. 1525-1539.

Lim, H.-W., and J. Song, 2012, Efficient risk assessment of lifeline networks under spatially correlated ground motions using selective recursive decomposition algorithm: Earthquake Engineering & Structural Dynamics, v. 41, p. 1861-1882.

Mensah, A. F., and L. Duenas-Osorio, 2014, Outage predictions of electric power systems under Hurricane winds by Bayesian networks: Probabilistic Methods Applied to Power Systems (PMAPS), 2014 International Conference on, p. 1-6.

Montoya, D. P., and J. M. Ramirez, 2012, A minimal spanning tree algorithm for distribution networks configuration: Power and Energy Society General Meeting, 2012 IEEE, p. 1-7.

Pagani, G. A., and M. Aiello, 2011, The Power Grid as a Complex Network: a Survey, http://arxiv.org/abs/1105.3338, **Physics and Society**.

Platts, 2009, Topology of the State of Texas power transmission network, http://www.platts.com/.

Quanta, 2009, Cost benefit analysis of the deployment utility infrastructure upgrades and storm hardening programs, Quanta Technology for Public Utility Commission of Texas.

Winkler, J., L. Dueñas-Osorio, R. Stein, and D. Subramanian, 2010, Performance assessment of topologically diverse power systems subjected to hurricane events: Reliability Engineering & System Safety, v. 95, p. 323-336.

Xia, Y., and H. Leung, 2014, Performance analysis of statistical optimal data fusion algorithms: Information Sciences, v. 277, p. 808-824.

Zhou, Y., and H. Leung, 1997, Linearly constrained least squares approach for multisensor data fusion, p. 118-129.

Zonta, D., 2014, Sensor data analysis, reduction and fusion for assessing and monitoring civil infrastructures: Sensor Technologies for Civil Infrastructures: Applications in Structural Health Monitoring, v. 2, p. 33.